Lectures Notes on Data Structures and Algorithms

Florian Rabe

2017

Contents

These notes were originally prepared for my 2nd semester CS course at Jacobs University in Spring 2017. In that course, the following chapters were skipped or treated only very superficially: 6, 7, 8, 13, 15, 20, 23, 24, 25. In the other chapters, almost all material was covered; only a few subsections were skipped.

Part I

Introduction and Foundations

Chapter 1

Meta-Remarks

Important stuff that you should read carefully!

State of these notes I constantly work on my lecture notes. Therefore, keep in mind that:

- I am developing these notes in parallel with the lecture—they can grow or change throughout the semester.
- These notes are neither a subset nor a superset of the material discussed in the lecture.
- Unless mentioned otherwise, all material in these notes is exam-relevant (in addition to all material discussed in the lectures).

Collaboration on these notes I am writing these notes using LaTeX and storing them in a git repository on GitHub at https://github.com/florian-rabe/Teaching. Familiarity with LaTeX as well as Git and GitHub is not part of this lecture. But it is essential skill for you. Ask in the lecture if you have difficulty figuring it out on your own.

As an experiment in teaching, I am inviting all of you to collaborate on these lecture notes with me.

By forking and by submitting pull requests for this repository, you can suggest changes to these notes. For example, you are encouraged to:

- Fix typos and other errors.
- Add examples and diagrams that I develop on the board during lectures.
- Add solutions for the homeworks if I did not provide any (of course, I will only integrate solutions after the deadline).
- Add additional examples, exercises, or explanations that you came up or found in other sources. If you use material from other sources (e.g., by copying an diagram from some website), make sure that you have the license to use it and that you acknowledge sources appropriately!

The TAs and I will review and approve or reject the changes. If you make substantial contributions, I will list you as a contributor (i.e., something you can put in your CV).

Any improvement you make will not only help your fellow students, it will also increase your own understanding of the material. Therefore, I can give you up to 10% bonus credit for such contributions. (Make sure your git commits carry a user name that I can connect to you.) Because this is an experiment, I will have to figure out the details along the way.

Other Advice I maintain a list of useful advice for students at https://svn.kwarc.info/repos/frabe/Teaching/general/advice_for_students.pdf. It is mostly targeted at older students who work in individual projects with me (e.g., students who work on their BSc thesis). But much of it is useful for you already now or will become useful soon. So have a look.

Chapter 2

Basic Concepts

These lecture notes do not follow a particular textbook.
Students interested in additional literature may safely use [CLR10] (available online), one of the most widely used textbooks. Knuth's book series on the Art of Computer Programming [Knu73], although not usually used as a modern textbook, is also interesting as the most famous and historically significant book on the topic.

2.1 What are Data Structures and Algorithms?

Data structures and algorithms are among the most fundamental concepts in computer science.

2.1.1 Static vs. Dynamic

In all areas of life and science, we often find a pair of concept such that one concept captures static and the other one dynamic aspects. This is best understood by example:

area	static	dynamic
in life		
existence	be	become
events	situation	development
food	ingredients	cooking
in science		
mathematics	sets	functions
physics	space	time
chemistry	molecules	reactions
engineering	materials	construction
in computer science		
hardware	memory	processing
abstract machines	states	transitions
programming	types	functions
software design	**data structures**	**algorithms**

The static aspects describes things as they are at one point in time. The dynamic aspects describes how they change over time.
Data structures and algorithms have this role in software design. Data structures are sets of objects (the data) that describe the domain that our software is meant to be used for. Algorithms are operations that describe how the objects in that domain change.

2.1.2 Basic Definition and Examples

Definition 2.1 (Data Structure)**.** Assume some set of effective objects.
A data structure defines a subset of these objects by providing effective methods for determining

- whether an object is in the data structure or not,
- whether two objects are equal.

In practice, a data structure is often bundled with several algorithms for it.

Definition 2.2 (Algorithm)**.** An algorithm consists of

- a data structure that defines the possible input objects
- a data structure that defines the possible output objects
- an effective method for transforming an input object into an output object

These definitions are not very helpful—they define the words "data structure" and "algorithm" by using other not-defined words, namely "effective object" and "effective method". Let us look at some examples before discussing effective objects and methods in Sect. 2.1.3.

Example 2.3 (Natural Numbers)*.* The most important data structure are the natural numbers.
It is defined as follows:

- The string 0 is a natural number.
- If n is a natural number, then the string $s(n)$ is a natural number.
- All natural numbers are obtained by applying the previous step finitely many times, and these are all different.

We immediately define the usual abbreviations $1, 2, \ldots,$. It is also straightforward to define algorithms for the basic functions on natural numbers such as $m + n$, $m - n$, $m * n$, etc.

Example 2.4 (Euclidean Algorithm)*.* The Euclidean algorithm (see also Sect. 2.1.4) computes the greatest common divisor $\gcd(m, n)$ of two natural numbers $m, n \in \mathbb{N}$. It consists of the following components:

- input: $\mathbb{N} \times \mathbb{N}$
- output: $\mathbb{N}$
- effective method:

```
fun gcd(m : ℕ, n : ℕ) : ℕ =
    x := m                                   introduce variables, initialize with input data
    y := n
    while x ≠ y                              repeat as long as gcd(x, y) ≠ x
        if x < y          subtract the smaller number from the bigger one, which does not affect gcd(x, y)
            y := y − x
        else
            x := x − y
    return x                                 now trivially gcd(x, y) = x
```

The algorithm starts by introducing variables x and y and initializes them with the input data m and n. Then it repeatedly subtracts the smaller number from the greater one until both are equal. This works because $\gcd(x, y) = \gcd(x - y, y)$. If x and y are equal, we can return the output because $\gcd(x, x) = x$.

This algorithm has a subtle bug (Can you see it?) that we will fix in Ex. 3.14.

For a simpler example, consider the definition of the factorial $n! = 1 \cdot \ldots n$ for $n \in \mathbb{N}$.

Example 2.5 (Factorial)*.* The factorial can be defined as follows:

- input: $\mathbb{N}$
- output: $\mathbb{N}$

- effective method:

```
fun fact(n : ℕ) : ℕ =
    product := 1
    factor := 1
    while factor ≤ n
        product := product · factor
        factor := factor + 1
    return product
```

  Here the variable *factor* runs through all values from 1 to n and the variable *product* collects the product of those values.

Notation 2.6. It is convenient to give the effective method of an algorithm as a function definition using pseudo-code. That way the input and output do not have to be spelled out separately because they are clear from the data structures used in the header of the function definition.

2.1.3 Effective Objects and Methods

It is now a central task in computer science to define data structures and algorithms that correspond to given sets and functions. This question that was first asked by David Hilbert in 1920, one of the most influential mathematicians at the same time. In modern terminology, he wanted to define data structures for all sets and algorithms for all functions and then machines to mechanize all mathematics.

In the 1930s, several scientist worked on this problem and eventually realized that it cannot be done. These scientists included Alonzo Church, Kurt Gödel, John von Neumann, and Alan Turing. Their work provided partial solutions and theoretical limits to the problem. In retrospect, this was the birth of computer science.

Not every set and not every function can be represented by a data structure or an algorithm (see Sect. 2.1.5 for the reason why not). That limitations bring us back to the question of effective objects and methods:

Definition 2.7 (Effective Object)**.** An effective object is any object that can be stored, manipulated, and communicated by a physical machine.

Here, *physical* means any machine that we can build in the physical world.[1]

Thus, every physical machine defines its own kind of effective objects. All digital machines (which includes all modern computers) use the same effective objects: lists of bits. These are stored in memory or on hard drives, which provide essentially one very, very long list of bits.

Data structures use fragments of these lists to represents sets. For example, the set $\mathbb{Z}_{2^{32}}$ of 32-bit-integers is represented by a list of 32 bits.

Definition 2.8 (Effective Method)**.** An effective method consists of a sequence of instructions such that

- any reasonably intelligent human can carry out the instructions
- and all such humans will carry out the instructions in exactly the same way (in particular reaching the same result).

The first condition makes sure that any prior knowledge needed to understand the instructions is be explicitly stated or referenced. The second condition makes sure that an effective method has a well-defined result: There may be no ambiguity, randomness, or unspecified choice.

Example 2.9. The second condition excludes for example the following instructions

- "Let x be the factorial of 5.": Different humans could compute the factorial differently because it is not clear which algorithm to use for the factorial.
- "Let x be a random integer.": Randomness is not allowed.
- "Let x be an element of the list l.": It is not specified which element should be chosen.

[1] Sometimes we use hypothetical machines. For example, quantum computers are physical machines that we think we can build but have not been able to build in practice at useful scales yet.

2.1.4 History

One of the earliest and most famous (arguably *the* earliest) algorithms is Euclid's algorithm for computing the greatest common divisor (see Ex. 2.4). It is given around 300 BC in Euclid's Elements [EucBC, Book VII, Proposition 2], maybe the most influential textbook of all time.

The word *algorithm* is much younger. It is derived from the name of the 9th century scientist al-Khwarizmi. He was one of the most important scientists of his millennium but is relatively unknown in the Western world because he was an Arab and wrote in Arabic. Translations of his work on arithmetic in the 12th century spread several new results to the Western world.
This included the use of numbers as abstract objects as opposed to geometric distances that had dominated Europe since the work of the Greek mathematicians (such as Euclid). It also included the positional number system and the base-10 digits that are still in use today. The corresponding arithmetical operations on numbers were named *algorismus* after him in Latin, which developed into the modern word. He also worked on algorithms for solving linear and quadratic equations, and one of his basic operations called *al-jabr* gave rise to the word *algebra*.

The modern *meaning* of the word *algorithm* is even younger: Its formalization was effected by a major development in the 1920s and 1930s that eventually gave to modern computer science itself. Hilbert was the most influential mathematician in the early 20th century. One of his legacies was to call for solutions to certain fundamental problems [Hil00]. Another legacy was his program [Hil26], a call for the formalization of mathematics that (among other things) should yield an algorithm for determining whether any given mathematical formula is a theorem.
Hilbert's program inspired seminal work by (among others) Alonzo Church, Kurt Gödel, and Alan Turing. This led to several concrete definitions of *algorithm*, including Turing-machines and the λ-calculus, from which all modern programming languages are derived. It also led to an understanding of the limits of what algorithms can do (see Sect. 2.1.5), which has led to the modern theory of computation.

2.1.5 The Limits of Data Structures and Algorithms

Countability of Data Structures and Algorithms

We can now see immediately why not all mathematical objects are effective in digital machines: There are only countably many lists of bits. Therefore, there can only be countably many effective objects.
Similarly, any data structure we define must be defined as a list of characters in some language. But there are only countably many such lists. Therefore, there can only be countably many data structures. For the same reason, there can only be countably many algorithms.

Inspecting the sizes of the constructed sets from Sect. A.5, we can observe that

- If all arguments are finite, so is the constructed set—except for lists.
- If all arguments are at most countable, so is the constructed set—except for function and power sets.

Because of these exceptions, we cannot restrict attention to finite or countable sets only—working with them invariably leads to uncountable sets.

Computability

At best, we can hope to give data structures for all countable sets. But not even that is possible. Because countable sets have uncountably many subsets, we cannot give data structures for every subset of every countable set.
Therefore, we give the sets that have data structures a special name:

Definition 2.10 (Decidable)**.** A set is called **decidable** if we can give a data structure for it.

Similarly, at best we can hope to give algorithms for all functions between decidable sets. Again that is not possible. Because countable sets have uncountably many functions between them, we cannot give algorithms for all functions between decidable sets.
Therefore, we give the functions that have algorithms a special name:

Definition 2.11 (Computable)**.** A function between decidable sets is called **computable** if we can give an algorithm for it.

Decidability and computability are discussed in detail in—typically—a second year course in theoretical computer science.

The Role of Programming Languages

Vagueness of the Definitions It is not possible to precisely define effective objects and methods—every definition eventually uses not-defined concepts like "machine" or "instruction". Thus, it impossible to precisely define what data structures and algorithms are. Instead, we must assume those concepts to exist a priori.

That may seem flawed—but it is actually very normal. We can compare the situation to physics where there is also no precise definition of *space* and *time*. In fact, the question what space and time are is among the difficult of all of physics.[2]

Similarly, the question of what data structures and algorithms are is among the most fundamental of computer science. Every machine and evey programming language give their own answers to the question.

Data Description and Programming Languages The only way to have a precise definition of *data structure* and *algorithm* is to choose a concrete formal language.

Definition 2.12 (Languages)**.** A **data description language** is a formal language for writing objects and data structures.
A **programming language** is a formal language for writing algorithms.

Because algorithms require data structures, every programming language includes a data description language. And because all data structures usually come with specific algorithms, we are mostly interested in programming languages.

But there are some languages that are pure data description languages. These are useful when storing data on hard drives or when exchanging data between programs and computers (e.g., on the internet). Examples of pure data description languages are JSON, XML, HTML, and UML.

Types of Programming Languages Programming languages can vary widely in how they represent data structures.

We can roughly distinguish the following groups:

- Untyped languages avoid explicit definitions of data structures. Instead, they use algorithms such as *isNat* to check, e.g., if an object is a natural number.
  Examples are Javascript and Python.
- Functional languages primarily use inductive data types to write data structures and recursive functions to write algorithms.
  Examples are SML and Haskell.
- Object-oriented languages primarily use classes to write data structures and imperative loops to write algorithms.
  Examples are Java and C++.
- Multi-paradigm languages combine functional and object-oriented features.
  Examples are Scala and F#.

Independence of the Choice of Language Above we have seen that the concrete meaning of *data structure* and *algorithm* seems to depend on the choice of programming language. Thus, it seems that whether a set is decidable or a function computable also depends on the choice of programming language.

One of the most amazing and deepest results of theoretical computer science is that this is not the case:

[2] For example, even today physicists have no agreed-upon answer to the question why time moves forwards but not backwards.

Theorem 2.13 (Church-Turing Thesis)**.** *All known machines and programming languages (including theoretical ones such as Turing machines)*

- *can define data structures for exactly the same sets,*
- *can define algorithms for exactly the same functions.*

Thus, it does not depend on the chosen machine or programming language

- *whether a set is decidable,*
- *whether a function is computable.*

Proof. The proof is very complex. For every program of every language, we must provide an equivalent program in every other language.
However, this can be done (and has been done) for all languages. □

A related (stronger) theorem is that every programming language P allows defining for every programming language Q a program that executes Q-programs.
We can only prove these theorems for all *known* languages. It is generally believed to be true also for all *possible* languages, but that is impossible to prove.

2.2 Specification vs. Design vs. Implementation

Above we have seen sets and functions as well as data structures and algorithms. Moreover, we have already mentioned programs consisting of types and functions.
The following table gives a overview of the relation between these concepts:

Specification	Design/Architecture	Implementation
sets functions	data structures algorithms	types functions

Software development consists of 3 steps:

1. The **specification** describes the intended behavior in terms of mathematical sets and functions.
   It does not prescribe in any way how these sets and functions are realized. The same specification can have multiple different correct realizations differing among others in size, maintainability, or efficiency.
   A good specification should be:
   - adequate: actually describe the problem that needs solving
   - simple: easy to understand
   - unambiguous: impossible to misunderstand
   - consistent: possible to realize
   - (optionally) complete: no freedom in what it means (An incomplete specification is not necessarily a flaw. For example, one might specify a function on integers without saying what should happen for negative input.)
2. The **architecture** makes concrete choices for the data structures and algorithms that realize the needed sets and functions.
   It usually defines many auxiliary data structures and algorithms that are not part of the specification.
   The architecture does not prescribe a programming language. It can be correctly realized in any programming language.
3. The **implementation** chooses a programming languages and then writes a **program** in it that realizes the architecture. The program includes concrete choices for the type and function definitions that realize the needed data structures and algorithms.
   It usually defines many auxiliary types and functions that are not part of the architecture.

Terminology 2.14. *Design* and *architecture* can usually be used synonymously.
The words *specification*, *design*, and *implementation* can refer to both the process and the result. For example, we can say that the result of implementation is one implementation.

It is critical to distinguish the three steps in software development:

- Specification changes are much more expensive than design changes. Changing the specification may completely change, which design is appropriate. Therefore, every single design decision must be revisited and checked for appropriateness.
- Design changes are much more expensive than implementation changes. Changing the design may completely change which components of the implementation are needed and how they interact.
  Therefore, every part of the program must potentially be revisited.
  In particular, whenever the design of component X is changed, we have to revisit every place of the program that uses X. This often introduces bugs.

Typically any specification change entails bigger design changes, and any design change entails bigger implementation changes. Moreover, specification changes require

- re-verification (i.e., checking that the implementation still correctly implements the specification)
- re-certification by regulatory agencies (if applicable to the specific software)
- changes to documentation, manuals, and tutorials, re-training of users, etc.
- distribution of software updates, which confuses and disrupts their workflows
- need for other software projects to adapt to the updated software

An ideal programmer proceeds in the order specification-design-implementation. However, it is often necessary to loop back: The design phase may reveal problems in the specification, and the implementation phase may reveal problems in the design. Therefore, we usually have to work on all 3 parts in parallel—but with a strong preference against changing specification and design.

Many self-taught or not-well-taught programmer do not understand the difference between the 3 steps or do not systematically apply it. There are many such programmers, who never studied CS or got a degree without taking a rigorous foundations course. Their programs are typically awful because:

- They begin programming without writing down the specification. Consequently, they do not realize that they have not actually understood the specification. This results in programs that do not meet the specification, which then leads to retroactive changes to the design. Over time the program becomes (sometimes called "spaghetti code") that is unmaintainable and cannot be understood by other programmers, often not even by the programmer herself.
- They begin programming without consciously choosing a design. Consequently, they end up with a random design that may or may not be appropriate for the task. Over time they change the design multiple times (without being aware that they are changing the design). Each change introduces new bugs and more mess.

Example 2.15 (Greatest Common Divider). The specification of the greatest common divider function gcd is as follows: Given natural numbers m and n, return a natural number g such that

- $g|m$ and $g|n$
- for every number h such that $h|m$ and $h|n$ we have that $h|g$

Before we design an algorithm, we should check whether gcd is indeed a function:

- Consistency: Does such a $g = \gcd(m, n)$ always exists?
- Uniqueness: Could there be more than one such g?

Using mathematics, we can prove that g indeed exists uniquely.

Now we design an algorithm. Let us assume that we have already designed data structures for the natural numbers with the usual operations. There are many reasonable algorithms, among them the one from Ex. 2.4. For the sake of example, we use a different one here:

```
fun gcdRec(m : ℕ, n : ℕ) : ℕ =
    if n == 0
        m
    else
        gcd(n, m mod n)
```

This is a recursive algorithm: The instructions may recursive call the algorithm itself with new input.

Finally, we implement the algorithm. We choose SML as the programming language. First we implement the data structure for natural numbers and the function $mod : nat * nat \to nat$ that were assumed by the specification. Note that this requires some auxiliary functions that were not part of the algorithm:

```
datatype nat = zero | succ of nat

fun leq(m: nat, n: nat): bool = case (m,n) of
  (zero,     zero)     => true
| (zero,     succ(y)) => true
| (succ(x), zero)     => false
| (succ(x), succ(y)) => leq(x,y)

fun minus(m: nat, n: nat): nat = case (m,n) of
  (zero,     zero)     => zero
| (zero,     succ(y)) => zero (* error case, should not happen *)
| (succ(x), zero)     => succ(x)
| (succ(x), succ(y)) => minus(x,y)

fun mod(m:nat, n:nat):nat =
  if m = n then zero
  else if leq(m,n) then m
  else mod(minus(m,n), n)
```

Then we define

```
fun gcdRec(m:nat, n: nat): nat = if n = zero then m else gcdRec(n,mod(m,n))
```

2.3 Stateful Aspects

2.3.1 Immutable vs. Mutable Data Structures

Consider a data structure for the set $\mathbb{N}^*$ of lists of natural number and assume we have a variable $x : \mathbb{N}^*$.

Immutable Data Structures and Call-by-Value

We can always assign a new value to x as a whole. For example, after executing $x := [1, 3, 5]$, we have the following data stored in memory:

variable	type	value
x	$\mathbb{N}^*$	$[1, 3, 5]$

location	value
P	$[1, 3, 5]$

Here the left part shows the variables as seen by the programmer. The right part shows the objects as they are maintained in memory by the programming language. P is some name for the memory location holding the value of x. Importantly, the programmer is completely unaware of the organization of the data in memory and only sees the value of x.

In particular, x is just an abbreviation for the value $[1, 3, 5]$. If we pass x to a function f, there is no difference between saying $f(x)$ and $f([1, 3, 5])$. That is called **call-by-value**.

For example, if we execute the instruction $y = delete(x, 2)$, we obtain:

variable	type	value
x	$\mathbb{N}^*$	$[1, 3, 5]$
y	$\mathbb{N}^*$	$[1, 3]$

location	value
P	$[1, 3, 5]$
Q	$[1, 3]$

All old data is as before. For the new variable y, a new memory location Q has been allocated and filled with the result of the operation. This has the drawback that the entire list was duplicated, and we now use twice as much memory as before.

Immutable data structures and call-by-value are the usual way how functions work in mathematics. Such data structures are closely related to their specification and make writing, understanding, and analyzing algorithms very easy.

Mutable Data Structures and Call-by-Name

If our data structure is mutable, the value of a variable x is just a reference to the memory location where the value is stored.
For example, after executing $x := [1, 3, 5]$, we have the following data stored in memory:

variable	type	value
x	$\mathbb{N}^*$	P

location	value
P	$[1, 3, 5]$

The value of x is now the reference to the memory location. The programmer still cannot see P directly.[3]
But there are two carefully-designed ways how P can be accessed indirectly. Firstly, we can assign new values to each component of x. For example, after $x.1 := 4$, the memory looks like

variable	type	value
x	$\mathbb{N}^*$	P

location	value
P	$[1, 4, 5]$

The old value at location P is gone and has been replaced by the new value.
Secondly, when we pass x to a function f, we pass the reference to the value, not the value itself. This is called **call-by-name** or **call-by-reference**.
For example, after executing $delete(x, 2)$, we have

variable	type	value
x	$\mathbb{N}^*$	P

location	value
P	$[1, 4]$

No additional memory location has been allocated for the result, and no copying took place. That makes the operation much more time- and memory-efficient. But from a mathematical perspective, this is very odd: The function call $delete(x, 2)$ *changed* the value of x under the hood.

In many programming languages (in particular object-oriented ones), mutable data structures are called *classes*. Some functions involving a mutable data structure will make use of mutability, some will not. This must be part of the specification of each function.

2.3.2 Environments and Side Effects

So far we have said that algorithms realize mathematical functions. That makes algorithms very close to the specification and makes writing, understanding, and analyzing them very easy. But it is not the whole picture in computer science—computer science needs a generalization:

Definition 2.16 (Stateful Functions)**.** Let E be the set of environments. An **effectful function** from A to B is a function $A \times E \to B \times E$.

Again this is a vague definition because the word "environment" is not defined. That is normal—there is no universally recognized definition for it. Intuitively, an object $e \in E$ represents the state of the environment. e contains all information that is visible from the outside of our algorithms and that can be acted on by the algorithm. These usually include the global variables, all kinds of input/output, threads, and exceptions.
An effectful function f from A to B can do two things besides returning a result of type B:

- It can use the environment (because E occurs in its input type). Thus, calling f twice on the same $a \in A$ may return different results if the environment has changed in between.
  Formally, if $f(a, e_1) = (b_1, e_1')$ and $f(a, e_2) = (b_2, e_2')$ always implies $b_1 = b_2$, we say that f is **environment-independent**.

[3]Some programming languages allow explicitly creating and manipulating these references. The most notable example is *C* (where the references are called *pointers*). With a few caveats (most importantly that it can allow for maximal optimization), that can be considered a design flaw in the programming language.

- It can change the environment (because E occurs in its output type). Thus, programmers must be careful when to call f and how often to call f because every call may have an effect that can be observed by the user. Formally, if $f(a, e) = (b, e')$ always implies $e = e'$, we say that f is **side-effect-free**.

If f is both environment-independent and side-effect-free, f is called **pure**. In that case, we always have $f(a, e) = (g(a), e)$ for some function $g : A \to B$, i.e., we can ignore environments entirely. Thus, pure functions are essentially the same as the usual mathematical functions.

An environment $e \in E$ is usually a big tuple containing among others

- the current values of all accessible mutable variables
- console input/output:
  - the list of characters to be printed out to the user
  - the list of characters typed by the user that are available for reading
- file and peripheral network input/output: for every open file, network connection or similar
  - the list of data to be written to the connection
  - the list of data that is are available for reading
- information about exceptions
  - by depending on this aspect of the environment, effectful functions can handle exceptions
  - by effecting this aspect of the environment, effectful functions can raise exceptions
- the set of currently active threads
- additional components depending on the features of the respective programming language

Environment-dependency and side effects are important. Without input/output side effect, the user could never provide input for algorithms and could never find out what the output is. Moreover, computers could not be used to read sensor data or control peripheral devices.

But they also present major challenges to algorithm design. Because the precise definition of E depends on the details of the programming language, it is very difficult to precisely specify effectful functions. And without a precise specification, the programmer never knows whether an algorithm is designed and implemented correctly. Therefore, some programming languages such as Haskell try to systematically restrict environment-dependency and side-effects as much as possible.

2.4 Parametric Polymorphism

Many important data structures and algorithms are polymorphic in the following sense:

Definition 2.17 (Polymorphism)**.** A **polymorphic data structure** D is an operator that maps data structures $D_1, \ldots, D_n$ to a data structure $D[D_1, \ldots, D_n]$.
A **polymorphic algorithm** F is an operator that maps data structures $D_1, \ldots, D_n$ to an algorithm $F[D_1, \ldots, D_n]$.
The D_i are called the **type parameters** or **type arguments** of the data structure/algorithm.

This is best understood by example:

Example 2.18 (Lists)*.* Lists are a polymorphic data structure. A^* is the set of lists whose elements have type A. Any data structure for A^* should take A as a type parameter.
For example, $List[A]$ may be a data structure such that $List[int]$ is the type of lists of integers.
Most algorithms about lists are polymorphic as well. For example, reversing a list can be realized using an algorithm

```
fun reverse[A](x : List[A]) : List[A] =
    ...
```

Terminology 2.19. There are many different concepts of polymorphism that are (correctly, confusingly, or even wrongly) called *polymorphism*. The special kind described here is usually called *parametric polymorphism*.
Both terminology and notations vary across programming languages, communities, and textbooks.

A more difficult example arises if we want to sort a list: To sort a list over A, we need a comparison function $\leq (x : A, y : A) : bool$. Moreover, $\leq$ has to be a total order. We can handle that using abstract classes:

Example 2.20. Consider the following polymorphic abstract class for total orders:

abstract class $TotOrd[A]()$
  fun $lessOrEqual(x : A, y : A) : bool$

It requires a function *lessOrEqual* that provides the comparison $\leq$. The axioms for being a total order can usually not be programmed—they can only be added as part of the documentation.
Then a polymorphic sorting algorithm could look like

fun $sort[A](ord : TotOrd[A], x : List[A]) : List[A] =$
  ...

Notation 2.21 (Omitting Type Parameters). Most of the time it is possible to omit the type parameters when calling a polymorphic function without ambiguity. For example, if $l : List[int]$, we can simply say $revert(l)$ instead of $revert[int](l)$—both human readers and compiler can infer the type argument.
Most programming languages that allow polymorphism also allow omitting parameters if they can be inferred uniquely. It is also allowed to do so in examples and pseudo-code.

In Programming Languages

Even though polymorphism is relatively simple mathematically, not all programming languages do a good job of implementing it. Therefore, we will often gloss over issues of polymorphism when giving algorithms.
But we give a few examples of polymorphism in a few typed programming languages.

Scala Scala's syntax is very similar to te pseudo-code used in these notes:

```
abstract class TotOrd[A] {
  def lessOrEqual(x:A, y:A): Boolean
}

object IntSmaller extends TotOrd[Int] {
  def lessOrEqual(x:Int, y:Int): Boolean = x <= y
}

object Sort {
  def sort[A](ord: TotOrd[A], x: List[A]): List[A] = {
    ...
  }
}

object Test {
  def main(args: Array[String]) {
    sort[Int](IntSmaller, List(4,3,5))
  }
}
```

Java In Java, polymorphic data structures are called *generics*. It uses angular instead of square brackets and puts the parameter types of a polymorphic algorithm before the return type instead of after the name:

```
interface TotOrd<A> {
  public Boolean lessOrEqual(A x, A y);
}

class Sort {
```

```
  static <A> List<A> sort(TotOrd<A> ord, List<A> x) {
    ...
  }
}

class IntSmaller implements TotOrd<Integer> {
  public Boolean lessOrEqual(Integer x, Integer y) {
    return x <= y;
  }
  public static IntSmaller it = new IntSmaller();
}

class Test {
  public static void main (String[] args) {
    Sort.sort(IntSmaller.it, Arrays.asList(3,5,4));
  }
}
```

C++ In C++, we can use templates to implement polymorphism. C++ also uses angular brackets, and the parameter types of classes and functions must be declared using the template keyword.

```
using namespace std;
#include <list>

template <class A>
class TotOrd {
  bool lessOrEqual(A x, A y);
};

class IntSmaller: public TotOrd<int> {
  bool lessOrEqual(int x, int y) {return x <= y;}
};
IntSmaller* is = new IntSmaller();

template <class A>
list<A> sort(TotOrd<A> ord, list<A> x) {
  ...
};

int test() {
    sort<int>(*is, {3,5,4});
}
```

SML In SML, we do not have abstract classes, but we can use a datatype instead. The type parameters of polymorphic types and functions are not declared explicitly. Instead, they are implicit given as variables like 'a.

```
datatype 'a TotOrder = TotOrder of 'a * 'a -> bool
fun lessOrEqual(ord: 'a TotOrder): 'a * 'a -> bool = case ord of TotOrder(f) => f

val IntSmaller: int TotOrder = TotOrder(fn (x,y) => x <= y)

fun sort(ord: 'a TotOrder, x: 'a list) = ...

fun test() = sort(IntSmaller, [3,5,4])
```

Chapter 3

Design Goals

3.1 Correctness

3.1.1 General Definition

The most important goal of design is *correctness*:

Definition 3.1. We say that:
- A data structure D is correct for a set S if the objects of D correspond exactly to the elements of S.
- An algorithm A is correct for a function F if for every possible input x the result of running A on x has output $F(x)$.

Data Structures

Obviously, an incorrect algorithm is simply a bug.[1]
However, incorrect data structures are often used.

Example 3.2. The data structure *int* is not correct for the sets $\mathbb{N}$ or the $\mathbb{Z}$. In both cases, *int* has not enough objects. *int* even has objects that are not in $\mathbb{N}$ at all (namely negative numbers).
However, *int* is routinely used in practice as if it were a correct data structure for $\mathbb{N}$ and $\mathbb{Z}$. If *int* uses 32 bits, it only covers the numbers between -2^{31} and $2^{31}-1$. As long as all involved numbers are between -2^{31} and 2^{31}, this is no problem.
It is possible to define correct data structure for $\mathbb{N}$ and $\mathbb{Z}$. But that can be disadvantageous because
- operations on *int* are much faster,
- interfacing with other program components may be difficult if they use different data structures.

Example 3.3. There is no data structure that is correct for $\mathbb{R}$.
Therefore, the data structure *float* is used in practice as if it were a correct data structure for $\mathbb{R}$. This always leads to rounding errors so that all results about *float* are only approximate.
float is often also used as if it were a correct data structure for $\mathbb{Q}$. That is a bad habit because computations on *float* are only approximate even if the inputs are exact. For example, there is no guarantee that 1.0/2.0 returns 0.5 and not 0.49999999999.

Example 3.4. Object-oriented languages use class types. Because of the *null* pointer, a class A that implements a set S actually implements the set $S^?$—a value of type A can be *null* or an instance of A.
Therefore, many good programmers systematically avoid ever using *null*. Still, the use of *null* is wide-spread in practice.

[1] However, there are advanced areas of computer science that study approximation algorithms. For example, we may want to use a fast algorithm that is almost correct for a function for which no fast algorithm exists.

Example 3.5. Assume we have a correct data structure for A.

Then we can give a correct data structure for $\{x \in A | P(x)\}$ if $P \in A \to \mathbb{B}$ is computable. However, because the set of computable functions is itself not decidable, programming languages usually do not allow defining correct data structures for $\{x \in A | P(x)\}$.

More severely, we cannot in general give a correct data structure for $\{F(x) : x \in A\}$ at all. Even if F is computable, we cannot give an algorithm that determines whether a given object is in that set.

Neither can we give a correct data structure for A/r for $r \in A \times A \to \mathbb{B}$. Even if r is computable, we cannot give an algorithm for equality of elements of A/r.

Algorithms

The process of making sure that an algorithm is correct is called *verification*. Verification is very difficult. In particular, the function that determines whether a data structure or algorithm is correct is itself not computable. Therefore, we have to prove the correctness of each data structure or algorithm individually.

Good programmers design algorithms that are close to the specification. That makes it easier to verify the design.

To make verification more systematic, we usually split the specification into two parts: precondition and postcondition. Independently, we split the verification arguments into two independent steps: termination and partial correctness. The definitions are as follows:

Definition 3.6. Consider an algorithm A for a function $f(x_1 \in I_1, \ldots, x_n \in I_n) \in O$.

We define:

- A **precondition** for A is a formula $Pre(x_1, \ldots, x_n)$ about the inputs.
- A **postcondition** for A is a formula $Post(x_1, \ldots, x_n, r)$ about the inputs and the output.
- A **terminates for** $v_1, \ldots, v_n$ if running A with these inputs finishes in finitely many steps.
- A **terminates** if it terminates whenever $Pre(v_1, \ldots, v_n)$.
- A is **partially correct** if for all $v_1, \ldots, v_n$
  - if $Pre(v_1, \ldots, v_n)$ and
  - A terminates for $v_1, \ldots, v_n$ with return value r, then
  - $Post(v_1, \ldots, v_n, r)$
- A is **totally correct** it is partially correct and terminates.

Finally we can recover Def. 3.1 by saying that A is a correct algorithm for a function f it is totally correct with

- precondition: nothing (always true)
- postcondition: $r == f(x_1, \ldots, x_n)$

The reason for splitting correctness up is that partial correctness and termination are often proved separately in very different ways. So it is good to have separate definitions for them. Sect. 3.1.2 and 3.1.3 describe the most important techniques.

The reason for splitting the specification into pre- and postcondition is to make fine-granular statements about what input an algorithm expects and what output it provides. They can be seen as a trade between the programmer W who writes function F and the programmer C who calls F. The precondition is the price that C has to pay (by making sure the precondition holds before calling F). And the postcondition in the service that W provides in exchange (by returning a value that satisfies the postcondition).

In particular, W may assume that the precondition holds—she does not have to check it. Instead, C has to check it. Vice versa, C may assume that the postcondition holds afterwards.

Example 3.7 (Pre/Postcondition). Consider a variant $gcd32(x : int, y : int) : int$ of the Euclidean algorithm that uses 32-bit integers. This can never be correct because it cannot handle arbitrarily large natural numbers. Moreover, the input and output type now allow negative values, which we want to exclude.

So we could use the following:

- precondition: $Pre(x, y) = 0 \leq x \leq MaxInt \wedge 0 \leq y \leq MaxInt$
- postcondition: $Post(x, y, r) = 0 \leq r \leq MaxInt \wedge r == \gcd(x, y)$

where $MaxInt$ is the maximal value of the type int.

Note that this specification makes the strong requirement that there will be no overflows. That works out for the Euclidean algorithm because all its intermediate results are smaller than the input.

For other algorithms, like a 32-bit algorithm $fib32(n : int) : int$ for Fibonacci numbers, the input has to be much smaller than $MaxInt$ to make sure the output fits into a 32-bit integer. So we might use:

- precondition: $Pre(n) = 0 \leq x \leq 46$
- postcondition: $Post(n, r) = r == fib(n)$

3.1.2 Partial Correctness

Loop Invariants for while-Loops

Many algorithms use while-loops. Verifying the correctness of while-loops is notoriously difficult.

Therefore, many good programmers try to avoid while-loops altogether. Instead, they prefer operations on lists (like *map*, *fold*, and *foreach*) or recursive algorithms.

The central method for verifying the partial correctness of a while-loop is the *loop invariant*:

Definition 3.8 (Loop Invariant)**.** Consider a loop of the form **while** $C(\vec{x})$ $\{code\}$. Here $\vec{x} = (x_1, \ldots, x_n)$ are the names that are in scope before entering the loop (i.e., excluding any names declared only in *code*).

A formula $F(\vec{x})$ is a **loop invariant** for this loop if F is preserved by the loop: if F holds before executing *code*, it also holds afterwards. Specifically, for all $\vec{v}$, the following must hold

$$C(\vec{v}) \text{ and } F(\vec{v}) \qquad \text{implies} \qquad F(code(\vec{v}))$$

where $code(v) = (v'_1, \ldots, v'_n)$ contains the values of the x_i after executing $x_1 := v_1; \ldots; x_n := v_n; code$.

If we have a loop invariant, we can use it as follows:

Theorem 3.9. *Consider a loop* **while** $C(\vec{x})$ $\{code\}$ *with a loop invariant* $F(\vec{x})$*.*
Assume that $F(\vec{v})$ *holds where* v_i *is the value of* x_i *before executing the while-loop.*

Then $\neg C(\vec{x}) \wedge F(\vec{x})$ *holds if and when the while-loop has been executed.*[2]

Proof. After the while-loop $C(\vec{x})$ cannot hold—otherwise, the while-loop would continue. Because $F(\vec{x})$ held before executing the loop and is preserved by every iteration of *code*, it also holds after executing the loop. □

Note that Thm. 3.9 says *if and when* the while-loop has been executed. That is because it is not guaranteed that the while-loop terminates. We still have to prove termination separately.

Example 3.10 (Euclidean Algorithm). We prove partial correctness of the algorithm from Ex. 2.4. We proceed statement-by-statement.

The first two statements are easy to handle: Their effect is that $x == m$ and $y == n$.

But now we reach a while-loop. We have $\vec{x} = (m, n, x, y)$ and $C(m, n, x, y) = x \neq y$. A loop invariant is given by $F(m, n, x, y) = \gcd(m, n) == \gcd(x, y)$. The intuition of this loop-invariant is that we only apply operations to x and y that do not change their gcd.

To work with the while-loop, we prove that F is a loop invariant:

- We show that F holds before the loop.
  Before reaching the loop, we have $x == m$ and $y == n$. Thus, immediately $gcd(m, n) == gcd(x, y)$.
- We show that F is preserved by the loop.
  Let us assume that $C(m, n, x, y)$ holds, i.e., $x \neq y$ (i).
  Moreover, let us assume that $F(m, n, x, y)$ holds, i.e., $\gcd(m, n) == \gcd(x, y)$ (ii).
  Let $code(m, n, x, y) = (m', n', x', y')$.

[2] We assume here that the evaluation of $C(\vec{x})$ has no side-effects and thus may not change the values of the x_i. In most programming languages, that would be allowed, but is a very bad practice precisely because it makes loop-invariant arguments more complicated.

We have to prove $F(m', n', x', y')$, i.e., $\gcd(m, n) = \gcd(x', y')$.
To do that, we have to distinguish two cases according to the if-statement:
- $x < y$: Then $(m', n', x', y') = (m, n, x, y - x)$. Thus we have to prove that $\gcd(m, n) = \gcd(x, y - x)$. Because of (ii), it is sufficient to prove $\gcd(x, y) = \gcd(x, y - x)$. That follows from the mathematical properties of gcd.
- $y < x$: Then $(m', n', x', y') = (m, n, x - y, y)$. We have to prove that $\gcd(m, n) = \gcd(x - y, x)$. That follows in the same way as in the first case.
- We do not need a case for $x == y$ because that is excluded by (i).

Now we can continue. The next statement is **return** x. Using Thm. 3.9, we obtain that $\neg C(m, n, x, y) \wedge F(m, n, x, y)$ holds, i.e., $\neg x \neq y \wedge \gcd(m, n) == \gcd(x, y)$. That yields $x == y$ and therefore $\gcd(m, n) == \gcd(x, x) == x$. Thus, the returned value is indeed $\gcd(m, n)$.
To prove total correctness, we still have to show that the while-loop terminates, which we do in Ex. 3.14

Induction for Recursive Functions

Proving partial correctness of recursive functions is very easy because we can simply use the postcondition about the recursive call. Formally, this means we do an induction proof on the number of recursive calls.

Example 3.11 (Recursive Euclidean Algorithm). We prove partial correctness for the algorithm $gcdRec(m : \mathbb{N}, n : \mathbb{N})$ from Ex. 2.15.
We have to prove the postcondition $gcdRec(m, n) == \gcd(m, n)$ where r is the return value. We proceed by induction, i.e., we assume that the property holds for all recursive calls. Then we have to handle two cases for the two branches of the if-statement:
- $n == 0$: Then $gcdRec(m, n) = m$, and the postcondition follows from $\gcd(m, 0) == m$.
- $n \neq 0$: Then, by using the induction hypothesis, $gcdRec(n, m \bmod n) == \gcd(n, m \bmod n)$. Then the postcondition follows from $\gcd(m, n) == \gcd(n, m \bmod n)$.

To prove total correctness, we still have to show that the recursion terminates which we in Ex. 3.18.

3.1.3 Termination

Verifying the termination of an algorithm is also very hard. The halting function is the function that takes as input an algorithm A and an object I and returns as output the following boolean: *true* if A terminates with input I and *false* otherwise. One of the most important results of theoretical computer science is that the halting function is not computable, i.e., there is no algorithm for it.
Thus, even if do not care what our algorithm actually does and only want to know if it terminates at all, all we can do is prove it manually for each input.
Termination is trivial for assignment, for-loop[3], if-statement, and the return-statement. Only while-loops and recursion are tricky. The most important technique to prove termination is to use a termination ordering.

Termination Orderings for While-Loops

Definition 3.12 (Termination Ordering)**.** Consider a while-loop of the form **while** $C(\vec{x})$ $\{code\}$.
A **termination ordering** for it is a function $T(\vec{x}) \in \mathbb{N}$ such that for all $\vec{v}$ we have that

$$C(\vec{v}) \quad \text{implies} \quad T(\vec{v}) > T(code(\vec{v})).$$

The intuition behind a termination ordering is that $T(\vec{x})$ strictly decreases in every iteration of the loop. Because it cannot decrease indefinitely, there can only be finitely many iterations, i.e., the loop must terminate. The following theorem makes that precise:

[3] In some programming languages, it is possible to write non-terminating for-loops by explicitly assigning to the counter variable in the body of the loop. That is a very bad practice precisely because it endangers termination.

Theorem 3.13 (Termination Ordering)**.** *Consider a the loop* **while** $C(\vec{x})$ $\{code\}$ *and a termination ordering* $T(\vec{x})$ *for it.*
Then the while-loop terminates for all initial values $\vec{v}$ *of* $\vec{x}$*.*

Proof. We define a sequence $\vec{v}^0, \vec{v}^1, \ldots$ such that $\vec{v}^i$ contains the values of $\vec{x}$ after i iterations of executing *code*:

$$\vec{v}^0 = \vec{v}$$

$$\vec{v}^{i+1} = code(\vec{v}^i) \qquad \text{for } i > 0$$

We use an indirect proof: We assume the while-loop does not terminate and show a contradiction.
If the loop does not terminate, the condition must always be true, i.e., $C(\vec{v}^i)$ for all $i \in \mathbb{N}$.
Then the termination ordering yields $T(\vec{v}^i) > T(\vec{v}^{i+1})$ for all $i \in \mathbb{N}$.
That yields an infinite sequence $T(\vec{v}^0) > T(\vec{v}^1) > \ldots$ of natural numbers.
But such a sequence cannot exist, which yields the needed contradiction. □

Example 3.14 (Euclidean Algorithm). We prove that the algorithm from Ex. 2.4 terminates for all inputs. Only the while-loop presents a problem.
A termination ordering for the while-loop is given by $T(m, n, x, y) = x + y$. The intuition of this termination ordering is that the loop makes either x or y smaller. Therefore, it must make their sum smaller.

We show that T is indeed a termination ordering.
As when proving the loop-invariant, we put $(m', n', x', y') = code(m, n, x, y)$.
We have to show that $T(m, n, x, y) > T(m', n', x', y')$, i.e., $x + y > x' + y'$.
We again distinguish two cases according to the if-statement:

- $x < y$ and thus $(m', n', x', y') = (m, n, x, y - x)$: We have to show $x + y > x + y - x$.
- $x > y$ and thus $(m', n', x', y') = (m, n, x - y, y)$: We have to show $x + y > x - y + y$.

Both cases are trivially true for all $x, y \in \mathbb{N} \setminus \{0\}$.
But what happens if $x == 0$ or $y == 0$? Indeed, the proof of the termination ordering property does not go through.
Inspecting the algorithm again, we realize that we have found a bug: If exactly one of the two inputs is 0, the algorithm never terminates.
We can fix the algorithm in two ways:

- We change the specification to match the behavior of the algorithm. That means to change the input data structure such that $m, n \in \mathbb{N} \setminus \{0\}$.
- We change the algorithm to match the specification. We can do that by adding the lines

  if $(x == 0)$ {**return** y}
  if $(y == 0)$ {**return** x}

  Now the loop can be analyzed with the assumption that $x \neq 0$ and $y \neq 0$.

Termination Orderings for Recursion

Termination orderings for recursion work in essentially the same way. But the precise definition is a little bit trickier.

Definition 3.15 (Termination Ordering for Recursion)**.** Consider a recursive function $f(\vec{x})$.
A **termination ordering** for f is a function $T(\vec{x}) \in \mathbb{N}$ such that: whenever f is called with arguments $\vec{v}$ and recursively calls itself with arguments $\vec{v'}$, then $T(\vec{v}) > T(\vec{v'})$.

Then we can prove the corresponding theorem:

Definition 3.16 (Relative Termination)**.** Consider a recursive function $f(\vec{x})$.
We say that f **terminates relatively** if the following holds: f terminates for all arguments under the assumption that all recursive calls terminate.

Theorem 3.17 (Termination Ordering for Recursion)**.** *Consider a recursive function $f(\vec{x})$ with a termination ordering T for it.*
If f terminates relatively, then it terminates for all arguments.

Proof. This is proved in the same way as for while-loops. □

Example 3.18 (Recursive Euclidean Algorithm). Consider the recursive algorithm from Ex. 2.15.
It is easy to see that the arguments never get bigger during the recursion. So we might try $T(m,n) = m+n$ as a termination ordering. But that does not work because if $m < n$, the recursive call is to $gcd(n,m)$, which just flips the arguments. In that case, $T(m,n) = m+n$ does not become strictly smaller.
It becomes easier to show termination if we expand the recursive call once. That yields the equivalent function:

```
fun gcd(m : ℕ, n : ℕ) : ℕ =
    if n == 0
        m
    else
        if m mod n == 0
            n
        else
            gcd(m mod n, n mod(m mod n))
```

Relative termination is trivial either way: Under the assumption that the recursive call returns, the function consists only of if-statements and therefore terminates.
And for the expanded function, $T(m,n) = m+n$ is a termination ordering. We have to prove $m+n > (m \bmod n)+(n \bmod(m \bmod n))$, which is easy to see.

3.1.4 Implementing Loop Invariants and Termination Orderings

Loop invariants and termination orderings an be tricky to understand for beginners. Therefore, the following gives a more concrete explanation.
Consider an arbitrary algorithm that uses a while-loop, e.g.,

```
fun fact(n : ℕ) : ℕ =
    product := 1
    factor := 1
    while factor ≤ n
        product := product · factor
        factor := factor + 1
    return product
```

To exemplify the role of a termination ordering, we modify it as follows:

```
fun T(n : ℕ, product : ℕ, factor : ℕ) : ℕ =
    ???

fun fact(n : ℕ) : ℕ =
    product := 1
    factor := 1
    print T(n, product, factor)
```

```
    while factor ≤ n
        product := product · factor
        factor := factor + 1
        print T(n, product, factor)
    return product
```

Our goal is to implement T such that running the algorithm prints strictly decreasing natural numbers. Any such implementation of T is a termination ordering and proves that the while loop terminates.

To exemplify the role of a loop invariant, we modify the algorithm in a very similar way:

```
fun F(n : ℕ, product : ℕ, factor : ℕ) : bool =
    ???

fun fact(n : ℕ) : ℕ =
    product := 1
    factor := 1
    print F(n, product, factor)
    while factor ≤ n
        product := product · factor
        factor := factor + 1
        print F(n, product, factor)
    return product
```

Our goal is to implement F such that running the algorithm prints only *true*. In that case,

- F is true before the loop
- F is a loop invariant, i.e., if it is true before, it is also true after executing the body of the loop.

Thus, if the while-loop should terminate, afterwards F must be true and the condition of the loop must be false.

There are many possible ways to implement F—already **return** *true* trivially satisfies the requirements. A practically useful implementation of F should tell us something that helps establish the postcondition (which in this case is $fact(n) == n!$).

3.2 Efficiency

An algorithm is efficient if it can be run with low cost. *Complexity* measures that cost.[4] Thus, an efficient algorithm has low complexity and vice versa.

There are two kinds of complexity: *time* and *space* complexity. Time complexity measures how long it takes for an algorithm to terminate. Space complexity measures how much temporary memory is needed along the way. Without qualification, the word *complexity* usually but not always means *time complexity*

In this section, we focus on time complexity. While termination describes whether an algorithm A terminates at all, its time complexity describes how long it takes to terminate. The time complexity of A is a function $C : \mathbb{N} \to \mathbb{N}$ such that $C(n)$ is the number of steps needed until A terminates for input of size n.

3.2.1 Exact Complexity

Exact complexity is tricky because the number of steps and the sizes of inputs depend on the programming language and the physical machine that is used. For example, we might try to use the following definitions for a simple programming language:

Example 3.19 (Counting Steps Exactly). For a typical programming language implemented on a digital machine, the following definition is roughly right:

[4] At Jacobs University, complexity is discussed in detail in a special course in the 2*nd* year.

For the execution of a statement:

- $\text{Steps}(C; D) = \text{Steps}(C) + \text{Steps}(D)$
- $\text{Steps}(x := E) = \text{Steps}(E) + 1$
  - $\text{Steps}(E)$ steps to evaluate the expression E
  - 1 step to make the assignment
- $\text{Steps}(\textbf{return } E) = \text{Steps}(E) + 1$
  - $\text{Steps}(E)$ steps to evaluate the expression E
  - 1 step to return
- $\text{Steps}(\textbf{if } (C) \{T\} \textbf{ else } \{E\}) = \text{Steps}(C) + 1 + \begin{cases} \text{Steps}(T) & \text{if } C == \textit{true} \\ \text{Steps}(E) & \text{if } C == \textit{false} \end{cases}$
  - $\text{Steps}(C)$ steps to evaluate the condition
  - 1 step to branch
  - $\text{Steps}(T)$ or $\text{Steps}(E)$ steps depending on the branch
- $\text{Steps}(\textbf{while } C \{B\}) = (n+1) \cdot \text{Steps}(C) + n \cdot \text{Steps}(B)$ where n is the number of times that the loop is repeated
  - $\text{Steps}(C)$ steps to evaluate the condition $n+1$ times
  - 1 step to branch after each evaluation of the condition
  - $\text{Steps}(B)$ steps to execute the body

For the evaluation of an expression:

- Retrieving a variable: $\text{Steps}(x) = 1$
- Applying built-in operators O such as + or &&: $\text{Steps}(O(E_1, \ldots, E_n) = \text{Steps}(E_1) + \ldots + \text{Steps}(E_n) + 1$
  - $\text{Steps}(E_i)$ steps to evaluate the arguments
  - 1 step to apply the operator
- Calling a function: $\text{Steps}(f(E_1, \ldots, E_n)) = \text{Steps}(E_1) + \ldots + \text{Steps}(E_n) + 1 + n$
  - $\text{Steps}(E_i)$ steps to evaluate the arguments
  - 1 step to create jump into the definition of f
  - 1 step each to pass the arguments to f

The size of an object depends on the data structure:

- For *int*, *float*, *char*, and $\mathbb{B}$, the size is 1.
- For *string*, the size is the length of the string.
- For lists, the size is the sum of the sizes of the elements plus 1 more for each element. The "1 more" is needed because each element needs a pointer to the next element of the list.

In actuality however, a number of subtleties about the implementation of the programming language, its compiler, and the physical machine can affect the run-time of a program. For example:

- We usually assume that all arithmetic operations take 1 step. But actually, that only applies to arithmetic operations on the type *int* of 32 or 64-bit integers.
  - Any arithmetic operation that can handle arbitrarily large numbers takes longer for larger numbers. Most such arithmetic operations have complexity closely related to the number of digits needed to represent the arguments. That number is logarithmic in the size of the arguments.
  - Multiplication and related operations usually take longer than addition and related operations. Similarly, exponentiation usually takes longer than multiplication.
  - Any operation not built into the hardware must be implemented using software, which makes it take longer. Operations on larger numbers may take longer even if they are of type *int*.
- Floating point operations may take more than 1 step.
- The programming language may provide built-in operations that are actually just abbreviations for non-trivial functions. For example, concatenation of strings usually require copying one or both of the strings, which takes at least 1 step for each character. In that case, concatenating longer strings takes longer.
- The programming language's compiler may perform arbitrary optimizations in order to make execution faster.

For example, we may have Steps(**if** (*false*) $\{E\}$) $= 0$ because the compiler removes the statement entirely. On the other hand, optimization may occasionally use a bad trade-off and make execution slower.
- A smart compiler may generate code that is optimize for multi-core machines, such that, e.g., 2 steps are executed in 1 step.
- Calling a function may take much more than 1 step to jump to the function. Usually, it requires memory allocation, which can be a complex operation.
- For advanced operations, like instantiating a class, it is essentially unpredictable how many steps are required.
- From a complexity perspective, IO-operations (printing, networking, file access, etc.) take as many steps as the size of the sent data. But they take much more time than anything else.

The dependency of exact complexity on programming language, implementation, and physical machine is awkward because it precludes analyzing an algorithm independent of its implementation. Therefore, it is common to work with asymptotic complexity instead.

The idea is that dependencies are usually harmless in the sense that they can be "rounded away". For example, it does not matter much whether $\text{Steps}(x := E) = \text{Steps}(E) + 1$ or $\text{Steps}(x := E) = \text{Steps}(E) + 2$. It just means that every program takes a little longer. It would matter more if $\text{Steps}(x := E) = 2 \cdot \text{Steps}(E) + 1$, which is unlikely.

We introduce the formal definitions in Sect. 3.2.2 and apply them in Sect. 3.2.3.

3.2.2 Asymptotic Notation

The field of complexity theory usually works with with Bachmann-Landau notations.[5] The basic idea is to focus on the rough shape of the function $C(n)$ instead of its details. For example, $C(n) = an + b$ is linear, and $C(n) = 2^{an+b}$ is exponential. The distinction linear vs. exponential is often much more important than the distinction $an + b$ vs. $a'n + b'$.

Therefore, we define classes of functions like linear, exponential, etc.:

Definition 3.20 (O-Notation)**.** Let $\mathbb{R}^+$ be the set of positive-or-zero real numbers.
We define a relation on functions $f, g : \mathbb{N} \to \mathbb{R}^+$ by

$$f \lessapprox g \quad \text{iff} \quad \exists N \in \mathbb{N}.\, \exists k > 0.\, \forall n > N.\, f(n) \leq k \cdot g(n)$$

If $f \lessapprox g$, we say that f is **asymptotically smaller** than g.
We write $f \eqslantless g$ if $f \lessapprox g$ and $g \lessapprox f$.
Moreover, for a function $g : \mathbb{N} \to \mathbb{R}^+$, we define the following sets of functions

$$O(g) = \{f : \mathbb{N} \to \mathbb{R}^+ \mid f \lessapprox g\}$$

$$\Omega(g) = \{h : \mathbb{N} \to \mathbb{R}^+ \mid g \lessapprox h\}$$

$$\Theta(g) = \{f : \mathbb{N} \to \mathbb{R}^+ \mid f \eqslantless g\} = O(g) \cap \Omega(g)$$

Intuitively, $f \lessapprox g$ means that f is essentially smaller than g. More precisely, f is smaller than g *for sufficiently large arguments* and *up to a constant factor*. The other definitions are straightforward: $O(g)$ is the set of everything smaller than g, $\Omega(g)$ is the set of everything larger than g, and $\Theta(g)$ is the set of everything essentially as great as g (i.e., both smaller and larger).

Remark 3.21 (A Slightly Simpler Definition). The following statement is not true in general. However, it is easier to remember and true for all functions that come up when analyzing algorithms: $f \lessapprox g$ iff $\exists a > 0. \exists b > 0. \forall n. f(n) \leq a \cdot g(n) + b$.
We can verbalize that condition as "f is smaller than g except for a constant factor and a constant summand". Those are the two aspects of run time that we can typically make up for by building faster machines.

[5] In the definition below, only O, Ω, and Θ are the standard BachmannLandau notations. The symbols $\lessapprox$ and $\eqslantless$ are specific to these lecture notes.

Example 3.22 (Complexity Classes). Now we can easily define some important classes of functions grouped by their rough shape:

- $\Theta(1)$ is the set of $(*)$ constant functions
- $\Theta(n)$ is the set of $(*)$ linear functions
- $\Theta(n^2)$ is the set of $(*)$ quadratic functions
- and so on

Technically, we should always insert "asymptotically" at $(*)$. For example, $\Theta(n)$ contains not only the linear functions but also all functions whose shape is similar to linear when we go to infinity. But that word is often omitted for brevity.

If we use O instead of Θ, we obtain the sets of *at most* constant/linear/quadratic/etc. functions. For example, $O(n)$ includes the constant functions whereas $\Theta(n)$ does not.

Similarly, if we use Ω instead of Θ, we obtain the sets of *at least* constant/linear/quadratic/etc. functions. For example, $\Omega(n)$ includes the quadratic functions whereas $\Theta(n)$ does not.

Of particular importance in complexity analysis is the set of polynomial functions: It includes all functions whose shape is similar to a polynomial.

The following table introduces a few more classes and arranges them by increasing size:

$O(1)$	constant
$O(\log_c \log_c n)$	doubly logarithmic
$O(\log_c n)$	logarithmic
$O(n)$	linear
$O(n \log_c n)$	quasi-linear
$O(n^2)$	quadratic
$O(n^3)$	cubic
$\vdots$	$\vdots$
$Poly = \bigcup_{k \in \mathbb{N}} O(n^k)$	polynomial
$Exp = \bigcup_{f \in Poly} O(c^{f(n)})$	exponential
$\bigcup_{f \in Exp} O(c^{f(n)})$	doubly exponential

Here $c > 1$ is arbitrary—all choices yield the same classes of functions.

We also say sub-X for strictly lower and super-X for strictly greater complexity than X. For example $\log_c n$ is sub-linear, and n^2 is super-linear.

The following theorem collects the basic properties of asymptotic notation:

Theorem 3.23 (Asymptotic Notation)**.** *We have the following properties for all f, g, h, f', g':*

- $\lesssim$ *is*
  - *reflexive:* $f \lesssim f$
  - *transitive: if* $f \lesssim g$ *and* $g \lesssim h$*, then* $f \lesssim h$

  Thus, it is a preorder.
- *If* $f \lesssim f'$ *and* $g \lesssim g'$*, then* $\lesssim$ *is preserved by*
  - *addition:* $f + g \lesssim f' + g'$
  - *multiplication:* $f \cdot g \lesssim f' \cdot g'$
- $\eqsim$ *is*
  - *reflexive:* $f \eqsim f$
  - *transitive: if* $f \eqsim g$ *and* $g \eqsim h$*, then* $f \eqsim h$
  - *symmetric: if* $f \eqsim g$*, then* $g \eqsim f$

  Thus, it is an equivalence relation.
- *The following are equivalent:*
  - $f \lesssim g$
  - $O(f) \subseteq O(g)$
  - $\Omega(f) \supseteq \Omega(g)$
  - $f \in O(g)$
  - $g \in \Omega(f)$

All statements express that f is essentially smaller than g.
- *The following are equivalent:*
  - $f \in \Theta(g)$
  - $g \in \Theta(f)$
  - $\Theta(f) = \Theta(g)$

  All statements express that f is essentially as great as g.

Proof. Exercise. □

Notation 3.24. The community has gotten used to using $O(f(n))$ as if it were a function. If $f(n) - g(n) \in O(r(n))$, it is common to write $f(n) = g(n) + O(r(n))$. The intuition is that f arises by adding some function in $O(r(n))$ to g. This is usually done when r is smaller than g, i.e., r is a rest that can be discarded.
Similarly, people often write $f = O(r(n))$ instead of $f \in O(r(n))$ to express that f is equal to some function in $O(r(n))$.

These notations are not technically correct and should generally be avoided. But they are often convenient.

Example 3.25. Using Not. 3.24, we can write $2^n + 5n^2 + 3 = 2^n + O(n^2)$. This expresses that 2^n is the dominating term and the polynomial rest can be rounded away.
Or we can write $6n^3 + 5n^2 + \log n = O(n^3)$.

Remark 3.26 (Other Notations). There are a few more notations like O, Ω, and Θ. They include o and ω. They are less important and are omitted here to avoid confusion.

3.2.3 Asymptotic Complexity

Equipped with asymptotic notations, we can now compute the run time of algorithms in a way that is mostly independent of the implementation and the machine.

Example 3.27. Consider the algorithm from Ex. 2.5. Let $C(n)$ be the number of steps it takes with input n.
Because we are only interested in the complexity class of C, this is quite simple:
1. The while-loop must be repeated n-times. So the algorithm is at least linear.
2. Each iteration of the while-loop requires one comparison, one multiplication, and two assignments. These operations take a constant number c of steps.[6]
   So the entire loop takes $c \cdot n$ steps. The value of c does not matter because we can ignore all constant factors. Thus, the entire loop takes $\Theta(n)$ steps.
3. The assignments in the first two lines and the return statement take constant time each. Because $C(n)$ is at least linear, we can ignore them entirely.
4. Thus, we obtain $C(n) \in \Theta(n)$ as the complexity class of the algorithm.

Note how all the subtleties described in Sect. 3.2.1 are rounded away by looking at Θ-classes.

In many cases, the run time primarily depends on the *size* of the input, i.e., the exact choice of input does not matter as long as we know its size. Therefore, complexity of an algorithm A is usually measured as a function $C(n)$ that returns the number that A will taken when called on input of size n.
- If the input is an integer $x \in \mathbb{Z}$, its size is $n = \log |x|$, which is the number of bits needed to represent x.
- If the input is a list, its size is usually the length of the list. But sometimes it may matter how big the elements of the list are.
- If there are multiple inputs, the size is often the sum of the sizes. But we may also use a function $C(m, n)$ that takes two sizes as arguments.

But sometimes different inputs of the same size make lead to very different run times. For example, Ex. 2.4 happens to terminate immediately if the inputs are equal (no matter what the size they have) but takes very long in other cases (specifically when they are consecutive Fibonacci numbers).
Thus, we have to distinguish between:

- worst-case complexity $C_w(n)$: This is the maximal possible number of steps for input of size n. If there is no additional information, this is usually what the author means.
- average-case complexity $C_a(n)$: This is the average number of steps for input of size n. This is more useful in practice, but it is more difficult because we need a probabilistic analysis to compute the average.
- best-case complexity: This is the minimal possible numbeer of steps for input of size n. This is rarely useful but occasionally helps put a lower bound on the complexity.

There is no universal convention how these details are formalized. Instead, we often have to consider the context to understand what the author means.

Example 3.28 (Euclidean Algorithm). Consider the algorithm from Ex. 2.4. Let $n = \max(a, b)$ and let $C(n)$ be the worst-case number of steps the algorithm takes for input a, b (i.e., we use the maximum value of the inputs as the argument of the complexity function).

It is not that easy to see what the worst case actually is. But we can immediately see that the loop is repeated at most n times. Each iteration requires one comparison, one subtraction, and one assignment, which we can sum up to a constant factor.[7] Thus, the critical question is how often the loop can be repeated.

We can answer that question by going backwards. Because x and y are constantly decreased but stay positive, the worst case must arise if they are both decreased all the way down to 1. Then computing through the loop backwards, we obtain $1, 1, 2, 3, 5, 8, 13$ as the previous values, i.e., the Fibonacci numbers.

Indeed, the worst-case of the Euclidean algorithm arises if m and n are consecutive Fibonacci numbers. By applying some general math (see Sect. 4.2), we obtain that $Fib(k) \in \Theta(2^k)$. Thus, if n is a Fibonacci number, the number of repetitions of the loop is in $\Theta(\log n)$.

Thus, $C(n) \in \Theta(\log n)$.

3.2.4 Discussion

Asymptotic Analysis

Asymptotic analysis is the dominant form of assessing the complexity of algorithms. It has the huge advantages that it

- is mostly largely independent of the implementation and the physical machine,
- abstract away from minor details that do not significantly affect the quality of the algorithms.

But it has some disadvantages. Most importantly, the terms that it ignores can be huge. For example, $n+2^{(2^{10000})} \in O(n)$ is linear. But the constant term is so huge that an algorithm with that complexity will never terminate in practice.

More formally, $f \lneqq g$ only means that f is smaller than g for *sufficiently large* input. Thus, $f \lneqq g$ does not mean that f is better than g. It only means that f is better than g if we need the results for sufficiently large inputs.

Judging Complexity

Θ-classes for complexity are usually a very reliable indicator of the performance of an algorithm. If two algorithms were designed naturally without extreme focus on complexity, we can usually assume that:

- For small inputs, they are both fast, and it does not matter which one we use.
- For large inputs, the one in the smaller complexity class will outperform the other.

Note that large inputs are usually not encountered by the programmer: the programmer often only tests his programs with small test cases and examples. Instead, large input is encountered by users. Therefore, complexity analysis is an important tool for the programmer to judge algorithms. Most of the time this boils down to relatively simple rules of thumb:

- Avoid doing something linearly if you can do it logarithmically or in constant time.
- Avoid doing something quadratically if you do it quasi-linearly or linearly.
- Avoid doing something exponentially if you can do it polynomially.

The distinction between exponential and polynomial has received particularly much attention in complexity theory. For example, in cryptography, as a rule of thumb, polynomial is considered easy in the sense that anything that

[7] Again we assume that all arithmetic operations take constant time.

takes only polynomial amount of time to hack is considered insecure. Exponential on the other hand is considered hard and therefore secure. For example, the time needed to break a password through brute force is exponential in the length of the password. So increasing the length and variety of characters from time to time is enough to stay ahead of brute force attacks.

Algorithm Complexity vs. Specification Complexity

Note that we have only considered the complexity of *algorithms* here.

We can also define the **complexity of a specification**: Given a mathematical function f, its complexity is that of the most efficient correct algorithm A for it. In this context, f is usually called the problem and A a solution.

It is generally much harder to analyze the complexity of a problem than that of an algorithm. It is easy to establish an upper bound for the complexity of a problem: Every algorithm for f defines an upper bound for the complexity of f. But to give a lower bound, we have to prove that there is no better algorithm for f (on any physical machine we might be able to build). Proving the absence of something is generally quite difficult.

An example is the $P \neq NP$ conjecture, which is the most famous open question in computer science. P is the class of all problems that have polynomial complexity, and NP is a related class that contains P. It is generally assumed that NP is strictly larger than P. But to prove that, one has to show that there is no polynomial algorithm for some problem in NP.

Algorithm Complexity vs. Implementation Complexity

The **complexity of an implementation** is its actual run-time. It is usually assumed that this corresponds to the complexity of an algorithm.

But occasionally, the subtleties discussed in see Sect. 3.2.1 have to be considered because they do not get rounded away. These subtleties can usually not make the implementation less complex than the algorithm, but they may make it more complex. Most importantly, when analyzing the complexity of algorithms, we often assume that arithmetic operations can be performed in $O(1)$. In practice, that is only true for numbers within the limits of underlying CPU, e.g., 64-bit numbers. If we implement the data structures for numbers correctly (i.e., for arbitrarily large numbers), the complexity of the arithmetic operations will be greater.

More generally, when analyzing algorithm complexity, we must make assumptions about the complexity of the primitive operations used in the algorithm. Then the complexity of the implementation is equal to complexity of the algorithm only if the implementation of the primitive operations satisfies these assumptions.

> *Example* 3.29 (Euclidean Algorithm). The implementation in Ex. 2.15 uses a very inefficient implementation for the data structure $\mathbb{N}$. It does not satisfy the assumption that arithmetic operations are done in $O(1)$. In fact, already the function implementing $\leq$ is in $\Theta(n)$. Consequently, the complexity of this particular implementation of gcd is higher than $\Theta(n)$.
>
> But there are efficient correct implementations of $\mathbb{N}$, which we could use instead. For example, if we use base-2 representation, we can implement natural numbers as lists of bits. Because the number of bits of n is $\Theta(\log_2 n)$, most arithmetic operations end up being $O(p(\log_2 n))$ for a polynomial p. For example, addition and subtraction take time linear in the number of bits. Multiplication and related operations such as mod are super-linear. That is more than $O(1)$ but still small enough to often be inessential.
>
> With an efficient implementation of $\mathbb{N}$ and its arithmetic operations, the implementation of gcd, which uses $\Theta(\log_2 n)$ steps and applies mod at every step, has a complexity somewhat bigger than $O((\log_2 n)^2)$. The details depend on how we implement mod.

3.3 Simplicity

An important and often under-estimated design goal is simplicity.

An algorithm should be elegant in the sense that it is very close to its mathematical specification. That makes it easy to understand, verify, document, and maintain.

Often simplicity is much more important than efficiency. The enemy of simplicity is optimization: Optimization increases efficiency usually at the cost of simplicity.

In practice, programmers must balance these two conflicting goals carefully.

Example 3.30 (Building a List). A frequent problem is to read a bunch of values and store them in a list. This usually requires appending every value to the end of the list as in:

```
data := []
while moreData
    d := getData
    data := append(data, d)
return data
```

But appending to *data* may take linear time in the length of the list. This is because *data* points to the beginning of the list, and the append operation must traverse the entire list to reach the end. Thus, traversal takes 1 step for the first element that is appended, 2 for the second, and so on. The total time for appending n elements in a row is $1 + 2 + \ldots + n = n(n+1)/2 \in \Theta(n^2)$. Thus, we implement a linear problem with a quadratic algorithm.

A common solution is the following:

```
data := []
while moreData
    d := getData
    data := prepend(d, data)
return reverse(data)
```

This *prepends* all elements to the list. Because no traversal is required, each prepend operation takes $O(1)$. So the whole loop takes $\Theta(n)$ steps.

But we build the list in the wrong order. Therefore, we revert it before returning it. Reversal must traverse and copy the entire list once, which takes linear time again.

Thus, the second algorithm runs in $\Theta(n)$ overall.

But it requires an additional function call, i.e., it is less simple. In a very large program, it is possible that the calls to *prepend* and *reverse* occur in two different program locations that are far away from each other. A programmer who joins the project may not realize that these two calls are related and may introduce a bug.

It is non-obvious which algorithm should be preferred. The decision has to be made on a case-by-case basis keeping all goals in mind. For example, if the data is ultimately read from or written to a hard drive, that will be linear. But it will usually be much slower than building the list in memory, no matter whether the list is built in linear or quadratic time.

3.4 Advanced Goals

There are a number of additional properties that algorithms should have. These can be formally part of the specification, in which case they are subsumed by the correctness properties. But often they are deliberately or accidentally ignored when writing the specification.

Reliability An algorithm is **reliable** if it minimizes the damage that can be caused by external factors. For example, power outages, network failures, user error, available memory and CPU, communication with peripherals (printers, hard drive, etc.) can all introduce problems even if all data structures and algorithms are correct.

Safety A system is safe if it cannot cause any harm to property or humans. For example, an algorithm governing a self-driving car must make sure not to hit a human.

Often safety involves interpreting signals received from and sending signals to external devices that operate in the real world, e.g., the cameras and the engine of the car. This introduces additional uncertainty (not to mention the other cars and pedestrians) that can be difficult to anticipate in the specification.

Security A system is secure if it cannot be maliciously influenced from the outside. This includes all defenses against hacking.

Security is often not part of the specification. In fact, attacking a system often requires intentionally violating the specification in order to call algorithms with input that the programmer did not anticipate.

Secure algorithms must catch all such invalid input.

Privacy Privacy is the requirement that only the output of an algorithm is visible to the user. Perfect privacy is impossible to realize because all computation leaks some information other than the output: This reaches from runtime and resource use to obscure effects like the development of heat due to CPU activity.

More critically, badly designed systems may expose intermediate data that occurred during execution but is not specified to be part of the output. For example, when choosing a password, the output should only the cryptographic hash of the password, not the password itself.

Additionally, a system may behave according to its specification, but the user may be unaware of it. For example, a user may not be aware that her word document stored its previous revision, thus accidentally exposing an early draft.

Maintainability An often-underestimated goal being able to maintain a program. Software usually lives for years, often decades, and programmers will come and go during its life time. One of the biggest sources of problems can be unclear or undocumented code—even if it is well-designed, correct, and efficient.

Simple data structures and elegant algorithms that are derived systematically from the specification help here. It leads to implementations that are easier to understand, which allows new programmers to take over seamlessly.

Minor optimizations should generally be avoided because they make the implementation less maintainable. Even major optimizations (e.g., linear instead of quadratic) must be weighed against the danger of introducing bugs in the long run.

Chapter 4

Arithmetic Examples

4.1 Exponentiation

4.1.1 Specification

The function $power(x \in \mathbb{Z}, n \in \mathbb{N}) \in \mathbb{N}$ (also written as x^n) returns the n-the power of x defined by

$$x^0 = 1$$

$$x^n = x \cdot x^{n-1} \qquad \text{if } n > 0$$

By induction on n, we show this indeed specifies a unique function.

4.1.2 Naive Algorithm

It is straightforward to give an algorithm for exponentiation. For example,

```
fun power(x : ℤ, n : ℕ) : ℕ =
    if n == 0
        1
    else
        x · power(x, n − 1)
```

Correctness The correctness of this algorithm is immediate because it follows the specification literally. For example, $T(x, n) = n$ is already a termination ordering.

Complexity Assuming that all multiplications take $O(1)$ no matter how big x is, the complexity of this algorithm is $\Theta(n)$ because we need n multiplications and recursive calls.

4.1.3 Square-and-Multiply Algorithm

It is easy to think that $\Theta(n)$ is also the complexity of the specification, i.e., that there is no sub-linear algorithm for it. But that is not true.

Consider the square-and-multiply algorithm:

```
fun sqmult(x : ℤ, n : ℕ) : ℕ =
    if n == 0
        1
    else
        r := sqmult(x, n div 2)
        sq := r · r
        if (n mod 2 == 0) {sq} else {x · sq}
```

Correctness To prove the correctness of this algorithm, we note that

$$x^{2i+0} = (x^i)^2$$

$$x^{2i+1} = x \cdot (x^i)^2$$

Moreover, we know that $n = 2(n \operatorname{div} 2) + (n \bmod 2)$. Partial correctness of *sqmult* follows immediately.
To prove termination, we observe that $T(x, n) = n$ is a termination ordering: $n \operatorname{div} 2$ always decreases (because $n \neq 0$) and remains positive.

Complexity Computing the run time of a recursive function often leads to a recurrence relation: The function occurs on both sides with different arguments. In this case, we get:

$$C(n) = C(n \operatorname{div} 2) + c$$

where $c \in O(1)$ is the constant-time effort needed in each iteration. We systematically expand this further

$$C(n) = C(n \operatorname{div} 2) + c = C(n \operatorname{div} 2 \operatorname{div} 2) + 2 \cdot c = \ldots = C(n \overbrace{\operatorname{div} 2 \ldots \operatorname{div} 2}^{k+1 \text{ times}}) + (k+1) \cdot c$$

Now let $n = (b_k \ \ldots \ b_0)_2$ be the binary representation of the exponent. We know that $k = \lfloor \log_2 n \rfloor$ and $n \overbrace{\operatorname{div} 2 \ldots \operatorname{div} 2}^{k+1 \text{ times}} = 0$. Moreover, we know from the base case that $C(0) = 1$.
Substituting these above yield

$$C(n) \in O(1) + \Theta(\log_2 n) \cdot O(1) = \Theta(\log_2 n)$$

Thus, we can compute *power* in logarithmic time.

4.2 Fibonacci Numbers

4.2.1 Specification

The Fibonacci numbers $Fib(n \in \mathbb{N}) \in \mathbb{N}$ are defined by

$$fib(0) = 0$$

$$fib(1) = 1$$

$$fib(n) = fib(n-1) + fib(n-2) \qquad \text{if } n > 1$$

By induction on n, we prove that this indeed specifies a unique function.
Moreover, we can prove the non-obvious result that

$$fib(n) = \frac{\varphi^n - (1-\varphi)^n}{\sqrt{5}} \qquad \text{for } \varphi = \frac{1+\sqrt{5}}{2}$$

(φ is also called the golden ratio.) That can be further simplified to

$$fib(n) = round\left(\frac{\varphi^n}{\sqrt{5}}\right)$$

where we round to the nearest integer.

4.2.2 Naive Algorithm

It is straightforward to give an algorithm for computing Fibonacci numbers. For example:

```
fun fib(n : ℕ) : ℕ =
    if n ≤ 1
        n
    else
        fib(n − 1) + fib(n − 2)
```

Correctness The correctness of this algorithm is immediate because it follows the specification literally. For example, $T(n) = n$ is a termination ordering.

Complexity We obtain the recurrence relation $C(n) = C(n-1)+C(n-2)+c$ where $c \in O(1)$ is the constant-time effort of the recursion. That is the same recurrence as for the definition of the Fibonacci numbers themselves, thus $C(n) \in O(fib(n)) = Exp$.
This naive approach is exponential because every functions spawns 2 further calls. Each time n is reduced only by 1 or 2, so we have to double the number of calls about n times to $\Theta(2^n)$ calls.

4.2.3 Linear Algorithm

It is straightforward to improve on the naive algorithm, turning an exponential into a linear solution. For example:

```
fun fib(n : ℕ) : ℕ =
    if n ≤ 1
        n
    else
        prev := 0
        current := 1
        i = 1
        while i < n
            next := current + prev
            prev := current
            current := next
            i := i + 1
        return current
```

Correctness As a loop invariant, we can use

$$F(n, prev, current, i) = prev == fib(i-1) \wedge current == fib(i)$$

which is straightforward to verify. After the loop, we have $i == n$ and thus $current = fib(n)$, which yields partial correctness.
As a termination ordering, we can use $T(n, prev, current, i) = n - i$. Again this is straightforward to verify.

Complexity Both the code before and inside the loop take $O(1)$, and the loop is repeated $n-1$ times. Thus, the complexity is $O(n)$.

4.2.4 Inexact Algorithm

It is tempting to compute $fib(n)$ directly using $fib(n) = round(\varphi^n/\sqrt{5})$. Because we can precompute $1/\sqrt{5}$, that requires $n+1$ floating point multiplications, i.e., also $O(n)$.
However, it is next to impossible to verify the correctness of the algorithm. While termination is trivial, partial correctness does not hold. We know that the formula $fib(n) = round(\varphi^n/\sqrt{5})$ is true, but that has no immediate use for floating point arithmetic. Rounding errors will accumulate over time and may eventually lead to a false result.

4.2.5 Sublinear Algorithm

Maybe surprisingly, we can still do better. Inspecting the body of the while loop in the linear algorithm, we see that we can rewrite the assignments as

$$(current, prev) := (current + prev, current)$$

which we can write in matrix form as

$$(current, prev) := (current, prev) \cdot \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$$

Thus, we obtain

$$(fib(n), fib(n-1)) = (1,0) \cdot \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}^n \qquad \text{for } n > 0$$

We can now pick any algorithm for computing the n-power of a matrix, e.g., by using square-and-multiply from Sect. 4.1.3 for matrices.

Correctness Correctness follows from the correctness of square-and-multiply.

Complexity Square-and-multiply has complexity $O(\log n)$. Thus, we can compute $fib(n)$ with logarithmic complexity.

4.3 Matrices

4.3.1 Specification

We write $\mathbb{Z}^{mn}$ for the set $(\mathbb{Z}^n)^m$ of vectors over vectors (i.e., matrices) over integers.
We define two operations on matrices:

- Addition: For of $x, y \in \mathbb{Z}^{mn}$, we define $x + y \in \mathbb{Z}^{mn}$ by

$$(x + y)_{ij} = x_{ij} + y_{ij}$$

- Multiplication: For $x \in \mathbb{Z}^{lm}$ and $y \in \mathbb{Z}^{mn}$, we define $x \cdot y \in \mathbb{Z}^{ln}$ by

$$(x \cdot y)_{ij} = x_{i1} \cdot y_{1j} + \ldots + x_{im} \cdot y_{mj}$$

4.3.2 Naive Algorithms

Vectors and matrices are best stored using arrays. We assume that

- Mat is the data structure of two-dimensional arrays of integers (i.e., arrays of arrays of the same length),
- if x is an object of Mat, then $x.rows$ is the length of the array and $x.columns$ is the length of the inner arrays,
- **new** $Mat(m, n)$ produces a new array of length m of arrays of length n in which all fields are initialized as 0.

Then we have the straightforward algorithms

```
fun add(x : Mat, y : Mat) : Mat =
    r = new Mat(x.rows, x.columns)
    for i from 1 to x.rows
        for j from 1 to x.columns
            r.i.j := x.i.j + y.i.j
    return r

fun mult(x : Mat, y : Mat) : Mat =
    r = new Mat(x.rows, y.columns)
    for i from 1 to x.rows
```

```
    for j from 1 to y.columns
        for k from 1 to x.columns
            r.i.j := r.i.j + x.i.k · y.k.j
return r
```

Correctness The algorithms directly implement the definitions. Thus, correctness—seemingly—obvious.

But there is one subtlety: The functions take two arbitrary matrices—there is no way to force the user to pass matrices of the correct dimensions. Therefore, we have to state correctness a bit more carefully:

- for $z := add(x, y)$
  precondition: $x.rows == y.rows$ and $x.columns == y.columns$,
  postcondition: $z == x + y$ and $z.rows == x.rows$ and $z.columns == x.columns$.
- for $z := mult(x, y)$
  precondition: $x.columns == y.rows$
  postcondition: $z := mult(x, y)$ is $x \cdot y$ and $z.rows == x.rows$ and $z.columns == y.columns$

Then we can easily show that *add* and *mult* are correct in the sense that the precondition implies the postcondition.

Complexity Assuming that all additions and multiplications take constant time, the complexity is easy to analyze. For addition it is $\Theta(mn)$ and for multiplication $\Theta(lmn)$ where l, m, and n are the dimensions of the respective matrices.

For addition, we can immediately see that we cannot improve on $\Theta(mn)$: Just creating the new array and returning it already takes $\Theta(mn)$ steps. Thus, $\Theta(mn)$ is the complexity of the specification, and the naive algorithm is optimal.

This is not obvious for multiplication. Using the same argument, we can say that the complexity of multiplication is $\Omega(ln)$. But there cannot be an $\Theta(ln)$-algorithm because m must matter—if m increases, it must take longer.

4.3.3 Strassen's Multiplication Algorithm

Inspecting the definition of matrix multiplication, we see that we can split up matrices into rectangular areas of submatrices, for example, like so:

$$\begin{pmatrix} x_{11} & x_{12} & x_{13} & x_{14} \\ x_{21} & x_{22} & x_{23} & x_{24} \\ x_{31} & x_{32} & x_{33} & x_{34} \\ x_{41} & x_{42} & x_{43} & x_{44} \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix} & \begin{pmatrix} x_{13} & x_{14} \\ x_{23} & x_{24} \end{pmatrix} \\ \begin{pmatrix} x_{31} & x_{32} \\ x_{41} & x_{42} \end{pmatrix} & \begin{pmatrix} x_{33} & x_{34} \\ x_{43} & x_{44} \end{pmatrix} \end{pmatrix}$$

Moreover, if matrices are split up like that, we can still obtain their product in the same way using recursive matrix multiplication:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \begin{pmatrix} e & f \\ g & h \end{pmatrix} = \begin{pmatrix} ae + bg & af + bh \\ ce + dg & cf + dh \end{pmatrix} = \begin{pmatrix} p & q \\ r & s \end{pmatrix}$$

Strassen's algorithm works in the general. But for simplicity, we only consider the case $l = m = n$, i.e., we are multiplying square matrices. Then the naive algorithm has complexity $\Theta(n^3)$, and we know the specification has complexity $\Omega(n^2)$. The question is to find a solution in between.

We further simplify to $n = 2^k$, i.e., we can recursively subdivide our 2^k-matrices to 4 2^{k-1}-matrices. Then we can design a recursive algorithm that only needs k nested recursions.

The complexity depends on the details of the implementation. Naively, computing p, q, r, s requires 8 recursive calls to multiplications and 4 additions of 2^{k-1}-matrices. That yields

$$C(n) = 8 \cdot C(n/2) + \Theta(n^2) = \ldots = 8^k \cdot C(1) + \Theta(n^2)$$

Because $k = \log_2 n$ and $C(1) \in O(1)$, that yields $C(n) \in \Theta(n^{\log_2 8}) = \Theta(n^3)$.

However, Strassen observed that we can do better. With some fiddling around, we can replace the 8 multiplications and 4 additions with 7 multiplications and 18 additions:

$$M_1 = a(f - h)$$

$$M_2 = (a+b)h$$
$$M_3 = (c+d)e$$
$$M_4 = d(g-e)$$
$$M_5 = (a+d)(e+h)$$
$$M_6 = (b-d)(g+h)$$
$$M_7 = (a-c)(e+f)$$
$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \begin{pmatrix} e & f \\ g & h \end{pmatrix} = \begin{pmatrix} M_5 + M_4 + M_2 + M_6 & M_1 + M_2 \\ M_3 + M_4 & M_1 + M_5 - M_3 - M_7 \end{pmatrix}$$

The extra additions do not harm because they are $\Theta(n^2)$. But turning the 8 into a 7 yields $C(n) = \Theta(n^{\log_2 7})$. Thus, Strassen's algorithm reduces n^3 to $n^{2.81\cdots}$, which can yield practically relevant improvements for relatively small n, e.g., $n \approx 30$.

Even more efficient algorithms are found regularly. The current record is $\Theta(n^{2.37\cdots})$. However, the sufficiently large n for which these algorithms are actually faster than Strassen's algorithm is so large that they have no practical relevance at the moment.

Chapter 5

Example: Lists and Sorting

5.1 Specification

Lists are the most important non-primitive data structure in computer science, and sorting is not only the most important problem about lists but also one of the historically most important algorithmic problems of computer science.

5.1.1 Lists

For a set A, the set A^* contains all lists $[a_0, \ldots, a_{l-1}]$ with elements $a_i \in A$ for some $l \in \mathbb{N}$. l is called the length of the list.
Because A^* is a set for an arbitrary set A, data structures for lists must be polymorphic with a type parameter A.

Immutable Lists The following table specifies the most important functions involving lists:

function	returns	abbreviation
$nil[A] \in A^*$	$[]$	
$range(m \in \mathbb{N}, n \in \mathbb{N}) \in \mathbb{N}^*$	$[m, \ldots, n-1]$ or $[]$ if $m \geq n$	
below, let $l \in A^*$ be of the form $[a_0, \ldots, a_{l-1}]$ and assume $n < l$		
$length[A](x \in A^*) \in \mathbb{N}$	l	
$get[A](x \in A^*, n \in \mathbb{N}) \in A$	a_n	x_n or $x[n]$
$concat[A](x \in A^*, y \in A^*) \in A^*$	$[a_0, \ldots, a_{l-1}, b_0, \ldots, b_{k-1}]$ if $y = [b_0, \ldots, b_{k-1}]$	$x + y$
$map[A, B](x \in A^*, f \in A \to B) \in B^*$	$[f(a_0), \ldots, f(a_{l-1})]$	$l\ map\ f$
$fold[A, B](x \in A^*, b \in B, f \in A \times B \to B) \in B$	$f(a_1, f(a_2, \ldots, f(a_n, b)) \ldots)$	
$prepend[A](a \in A, x \in A^*) \in A^*$	$[a, a_0, \ldots, a_{l-1}]$	
$append[A](x \in A^*, a \in A) \in A^*$	$[a_0, \ldots, a_{l-1}, a]$	
$revert[A](x \in A^*) \in A^*$	$[a_{l-1}, \ldots, a_0]$	
$delete[A](x \in A^*, n \in \mathbb{N}) \in A^*$	$[a_0, \ldots, a_{n-1}, a_{n+1}, \ldots, a_{l-1}]$	
$insert[A](x \in A^*, a \in A, n \in \mathbb{N}) \in A^*$	$[a_0, \ldots, a_{n-1}, a, a_n, a_{n+1}, \ldots, a_{l-1}]$	
$update[A](x \in A^*, a \in A, n \in \mathbb{N}) \in A^*$	$[a_0, \ldots, a_{n-1}, a, a_{n+1}, \ldots, a_{l-1}]$	

Most of them are polymorphic. *map* and $fold$ even take a second type parameter for the return type of the function.
These are split into three groups:

- The first group contains functions to create new lists. These are important to have any lists.
- The second group contains functions that take a list $x \in A^*$ and return data about l or use l to build new data.
- The third group also takes a list $l \in A^*$ but also returns an element of A^*. This distinction is irrelevant in mathematics but critical in computer science: These functions may be implemented using in-place-updates. With in-place update, the list l is changed to become the intended result. The original value of l is lost in the process. If this is the case, we speak of **mutable** lists.

Mutable Lists The following table specifies the most important functions on mutable lists that differ from immutable lists. Instead of returning a new list, they have the effect of assigning a new value to the first argument.

function	returns	effect	abbreviation
below, let $l \in A^*$ be of the form $[a_0, \dots, a_{l-1}]$ and assume $n < l$			
$delete[A](x \in A^*, n \in \mathbb{N})$	nothing	$x := [a_0, \dots, a_{n-1}, a_{n+1}, \dots, a_{l-1}]$	
$insert[A](x \in A^*, a \in A, n \in \mathbb{N})$	nothing	$x := [a_0, \dots, a_{n-1}, a, a_n, a_{n+1}, \dots, a_{l-1}]$	
$update[A](x \in A^*, a \in A, n \in \mathbb{N})$	nothing	$x := [a_0, \dots, a_{n-1}, a, a_{n+1}, \dots, a_{l-1}]$	$x_n := a$ or $x[n] := a$

The other functions such as *length* and *get* are not affected.

5.1.2 Sorting

Sorting a list is intuitively straightforward. We need a function that takes a list and returns a list with the same elements in a different order, namely such that all elements occur according to their size.

Example 5.1. Consider $x = [4, 6, 5, 3, 5, 0] \in \mathbb{N}^*$. Then $sort(x)$ must yield $[0, 3, 4, 5, 5, 6]$.
Here we made the implicit assumption that we want to sort with respect to the $\leq$-order on $\mathbb{N}$. We could also use the $\geq$-order. Then $sort(x)$ should return $[6, 5, 5, 4, 3, 0]$.
Thus, sorting always depends on the chosen order.

Definition 5.2 (Sorting)**.** Fix a set A and a total order $\leq$ on A.
A list $x = [a_0, \dots, a_l] \in A^*$ is called $\leq$-**sorted** if $a_0 \leq a_1 \leq \dots \leq a_{l-1} \leq a_l$.
Let $count(x \in A^*, a \in A) \in \mathbb{N}$ be the number of times that a occurs in x. Two list $x, y \in A^*$ are a **permutation** of each other if $count(x, a) = count(y, a)$ for all $a \in A$.
$sort : A^* \to A^*$ is called a $\leq$-**sorting** function if for all $x \in A^*$, the list $sort(x)$ is a $\leq$-sorted permutation of x.

As usual we check that the specification indeed defines a function:

Theorem 5.3 (Uniqueness)**.** *The function sort from Def. 5.2 exists uniquely.*

Proof. Because $\leq$ is assumed to be total, every list x has a unique least element, which must occur first in $sort(x)$. By induction on the length of x, we show that all elements of $sort(x)$ are determined. □

For immutable lists, the above definition is all the specification we need. For mutable lists, we specify an alternative sorting function that does not create a new list:

Definition 5.4 (In-place Sorting)**.** An effectful function *sort* that takes an argument $x \in A^*$ and has the side-effect of modifying the value v of x to v' is called an **in-place** $\leq$-**sorting** function if $v' = s(v)$ for a $\leq$-sorting function s.

5.1.3 Sorting by a Property

Often we do not have a total order on A, and we want to sort according to a certain property. The property must be given by a function $p : A \to P$ such that we have a total order $\leq$ on P.
For example, we may want to sort a list of students by age. Then $A = Student$, $P = \mathbb{N}$, and $p : (s \in Student) \mapsto age(s)$.
However, there may be ties: A list may contain multiple different elements that agree in the value of p. To break, we require that the order in the original list should be preserved. Formally:

Definition 5.5 (Sorting by Property)**.** Fix sets A and P, a function $p : A \to P$, and a total order $\leq$ on P.
Given a list $x \in A^*$, we define a total order $\leq^p$ on the elements of x as follows:

$$x_i \leq^p x_j \quad \text{iff} \quad p(x_i) < p(x_j) \quad \text{or} \quad p(x_i) = p(x_j) \text{ and } i \leq j$$

$sort : A^* \to A^*$ is called a **stable sorting** function for p and $\leq$ if it is a sort function for $\leq^p$.

Note that normal sorting becomes a special case of sorting by property using $P = A$ and $p(a) = a$.

5.1.4 Why Do We Care About Sorting?

Nowadays, sorting is a solved problem. Computer scientists almost never need to implement sorting because all programming languages come with sophisticated ready-to-use solutions.

This is captured in the following exchange where a good, modern programmer is quizzed on sorting:

1. How do you implement sorting a list? — I call the *sort* function of my programming language's basic library.
2. OK, but what if there is no *sort* function? — I import a library that provides it.
3. OK, but what if there is no such library? — I use a different programming language.
4. OK, but what if circumstances beyond your control prevent you from using third-party libraries? — I copy-paste a definition from the internet.[1]

Thus, for most people the only realistic situations in which to implement sorting algorithms is in exams, job interviews, or similar situations. Then the question is never actually about sorting—it just uses sorting as an example to see whether the programmer understands how to design algorithms, analyze their complexity, and verify their correctness.

In any case, sorting is an extremely good subject for an introductory computer science class because it

- is an elementary problem that is easy to understand for students,
- is complex enough to exhibit many important general principles in interesting ways,
- is simple enough for all analysis to be doable manually,
- has multiple, very different solutions, none of which is better than all the others,
- is extremely well-studied,
- is widely taught so that the internet is full of good tutorials, examples, and visualizations that help learners.

5.2 Design: Data Structures for Lists

Besides natural numbers, the most important examples of a data structure are lists. There are many different data structures for lists that differ subtly in how simply and/or efficiently the various functions can be implemented. We will write $List[A]$ whenever we mean an arbitrary data structure for lists.

5.2.1 Immutable Lists

For immutable lists, functions like *delete*, *insert*, and *update* (see Sect. A.5.4) always return new lists. That requires copying (parts of) the old list, which takes more time and memory.

Without further qualification, this is usually what $List[A]$ refers to.

Functional Style: Lists as an Inductive Type

Functional languages usually implement lists an in inductive data type:

$$\textbf{data}\ IndList[A] = nil \mid cons(head : A,\ tail : IndList[A])$$

Now the list $[1, 2, 3]$ is built as $cons(1, cons(2, cons(3, nil)))$.

Then functions on lists are implemented using recursion and pattern-matching. For example:

$$\begin{array}{l}\textbf{fun}\ map(x : IndList[A],\ f : A \to B) : IndList[B] = \\ \quad \textbf{match}\ x \\ \qquad nil \mapsto nil \\ \qquad cons(h, t) \mapsto cons(f(h), map(t, f))\end{array}$$

[1] Nowadays an internet search for elementary problems almost always finds a solution for every programming language, usually on http://www.stackexchange.org.

Object-Oriented Style: Linked Lists

Every inductive data type can also be systematically realized in an object-oriented language. The correspondence is as follows:

inductive type	class	example: lists
name of the type	abstract class	$IndList$
parameters of the type	parameters of the class	A
constructor	concrete subclass	e.g., *cons*
constructor arguments	constructor arguments	$head : A, tail : IndList[A]$

A basic realization looks as follows:

abstract class $IndList[A]()$

class $nil[A]()$ **extends** $IndList[A]()$

class $cons[A](head : A, tail : List[A])$ **extends** $IndList[A]()$

Now the list $[1, 2, 3]$ is built as **new** $cons(1,$ **new** $cons(2,$ **new** $cons(3,$ **new** $nil())))$.
Instead of pattern-matching, we have to use instance-checking to split cases. For example:

```
fun map(x : IndList[A], f : A → B) : IndList[B] =
    if x isInstanceOf nil
        new nil()
    else
        xc := x asInstanceOf cons
        new cons(f(xc.head), map(x.tail, f))
```

Moreover, we have to override equality so that, e.g., two instances of *cons* are equal iff they used equal constructor arguments.

Complexity

Complexity of lists is measured in the lenght n of the list.

Most operations on lists are linear because the algorithm must traverse the whole list. For example, the straightforward implementation of $length$ takes $\Theta(n)$.

Similarly, $get(x, i)$ takes i steps to find the element. This is n in the worst case and $n/2$ on average. So it also takes $\Theta(n)$.

In general, immutable lists require copying the list whenever we insert, delete, or update elements. These algorithms must traverse the list. Therefore, they usually take $\Theta(n)$ time for the traversal *and* $\Theta(n)$ space to store the result list.

In the case of $map(x, f)$ and $fold(x, a, f)$, the complexity depends on the passed function f. However, in the typical case where the run time of f does not depend on the length of the list, we can assume it takes constant time c. Thus, the overall run time is $\Theta(cn) = \Theta(n)$.

However, there is one important exception: $prepend$ takes $\Theta(1)$. This is because we can implement $prepend(a, x)$ simply by calling $cons(a, x)$. Correspondingly, removing the first element takes $\Theta(1)$.

5.2.2 Mutable Lists

Mutable lists allow assignments to the individual elements of the list. This allows updating an element without copying the list, thus allowing for many operations with $\Theta(1)$ time or space complexity.

Because we can update the list in place, it becomes critical for efficiency how exactly the list is stored in memory. Several cases are of great importance, all with advantages and disadvantages:

data structure	memory layout	remark
array	all in a row	easy to find elements but difficult to insert/delete
(singly-)linked list	every element points to next one	easy to insert/delete but traversal needed
doubly-linked list	every element points to next and previous one	traversal in both directions possible, more overhead
growable array	linked list of arrays	compromise between the above

Arrays

The data structure $Array[A]$ stores all elements in a row in memory. Arrays must be a primitive feature of the programming language and are so in most languages.
For example, the list $x = [1, 2, 5]$ is stored in 3 consecutive memory locations:

variable	type	value
x	$\mathbb{N}^*$	P

location	value
P	1
$P+1$	3
$P+2$	5

That allows implementing *get* and *update* in $\Theta(1)$. $get(x, n)$ is evaluated by retrieving the element in memory location $P + n$. That takes one step to retrieve x, one step for the addition, and one step to retrieve the element at $P + n$. $update(x, a, n)$ works accordingly.
Inserting and deleting elements still takes $\Theta(n)$. For example, we can implement deleting by:

```
fun delete(x : Array[A], n : ℕ) =
    for i from n to length(x) − 1
        x[i] := x[i + 1]
```

Inserting an element into an array is difficult though: The memory location behind the array may not be available because it may have already been used for something else. Therefore, arrays are often realized in such a way that the programmer chooses in advance the maximal length of the array. Thus, technically this data structure does not realize the set A^* but the set A^n for some length n. This may waste memory if n is chosen too large. But arrays are unbeatable in the common situation where we know that we will never call *insert* anyway.

Linked Lists

Mutable linked lists consist of a reference to the first element. Each element consists of a value and a reference to its successor. We can implement that using classes (or similar primitives like structs in C):

```
class LinkedList[A](head : Elem[A])

class Elem[A](value : A, next : Elem[A])
```

Technically, *head* and *next* should have the type $Elem(A)^?$ to allow for empty lists and the end of the list, respectively. However, object-oriented programmers usually use a trick where the built-in value *null* is used:

- If *head* is null, we have the empty list.
- If *next* is null, we have the last element of the list.

Now the list $[1, 2, 5]$ is built as $x :=$ **new** $LinkedList($**new** $Elem(1,$ **new** $Elem(2,$ **new** $Elem(5, null))))$. It is stored in memory as

variable	type	value
x	$\mathbb{N}^*$	P

location	value
$P.head$	Q
$Q.value$	1
$Q.next$	R
$R.value$	2
$R.next$	S
$S.value$	5
$S.next$	$null$

Deletion can now be realized in-place as follows:

```
fun delete(x : LinkedList[A], n : ℕ) =
    if n == 0
        x.head := x.head.next
    else
        previous := x.head
        current := x.head.next
        for i from 1 to n − 1
            previous := current
            current := current.next
            previous.next := current.next
```

Like immutable lists, linked lists take $\Theta(n)$ time for most operations. However, they still perform better because changes can be done in-place. Moreover, many operations can be done in $\Theta(1)$ memory whereas immutable lists often require $\Theta(n)$ memory.

An interesting exception is the following variant of *insert*: Instead of taking the position n at which to insert (which takes linear time to find), it takes the element after which to insert:

```
fun insert(x : LinkedList[A], after : Elem[A], a : A) =
    after.next := new Elem(a, after.next)
```

A similar trick for deleting does not work so well: We can implement $delete(x : LinkedList[A], after : Elem[A])$ in $\Theta(1)$ if we know after which element to delete. But a function $delete(x : LinkedList[A], e : Elem[A])$ where e is to be deleted still requires $\Theta(n)$ to find e in the linked list.

Doubly-Linked Lists

Doubly-linked linked list are the same as linked lists except that each element also knows its predecessor (*null* for the first element). Moreover, the list knows its first and last element.

```
class DoubleLinkedList[A](head : Elem[A], last : Elem[A])

class Elem[A](value : A, previous : Elem[A], next : Elem[A])
```

Now the list $x = [1, 2, 5]$ is stored in memory as

variable	type	value
x	$\mathbb{N}^*$	P

location	value
$P.head$	Q
$P.last$	S
$Q.value$	1
$Q.previous$	$null$
$Q.next$	R
$R.value$	2
$R.previous$	Q
$R.next$	S
$S.value$	5
$S.previous$	R
$S.next$	$null$

Operations on doubly-linked lists are usually in the same complexity class as the corresponding ones for singly-linked lists.

A doubly-linked list has more memory overhead and thus copying and update operations have more time overhead. But doubly-linked lists can be traversed efficiently in *both* directions. For example, processing the elements of a singly-linked list in reverse order requires two traversals: one to find the last element, one to process. The same operation on a doubly-linked list requires only one traversal. Both are $\Theta(n)$, but the latter may be twice as fast.

In a double-linked list, we can also define nice constant-time variants for both *insert* and *delete*. For example:

```
fun delete(x : DoubleLinkedList[A], e : Elem[A]) =
    if e.previous == null
        x.head := e.next
    else
        e.previous.next := e.next
    if e.next == null
        x.last := e.previous
    else
        e.next.previous := e.previous
```

The following table summarizes the complexity of some operations on arrays, linked lists and doubly-linked lists in terms of the length l:

	$length[A]$	$get[A]$	$update[A]$	$insert[A]$	$delete[A]$	$prepend[A]$	$append[A]$	$reverse[A]$
		at position n						
Array	$\Theta(1)$	$\Theta(1)$		$\Theta(l-n)$		$\Theta(l)$	$\Theta(1)$	$\Theta(l)$
Linked list	$\Theta(l)$	$\Theta(n)$		$\Theta(n)$		$\Theta(1)$	$\Theta(l)$	$\Theta(l)$
Doubly-linked List	$\Theta(l)$	$\Theta(n)$		$\Theta(n)$		$\Theta(1)$	$\Theta(1)$	$\Theta(l)$

Growable Arrays

Growable arrays are a compromise between arrays and linked lists. Initially, they behave like an array with a fixed length l. However, when inserting an element increases the length beyond l, we create a second array of length l (elsewhere in memory) and remember their connection by storing a list containing the two elements. Thus, a growable array is a linked list of fixed-length arrays. The choice of l is up to the data structure designer, who may allow the programmer to tweak it.

Retrieval and update technically are linear now. To access the element in position n, we have to make n/l retrievals to jump to the needed array. Because l is constant, that yields $\Theta(n)$ retrievals. However, l is usually large so that element access is only a little slower than for an array and much faster than for a linked list.

5.3 Design: Algorithms for Sorting

We assume a fixed set A and a fixed comparison function $\leq: A \times A \to \mathbb{B}$. For $x \in A^*$, we write $Sorted(x)$ if x is $\leq$-sorted.

Auxiliary Functions Many in-place sorting algorithms have to swap two elements in a mutable list at some point. Therefore, we define an auxiliary function

```
fun swap(x : MutableList[A], i : ℕ, j : ℕ) =
    h := x[i]
    x[i] := x[j]
    x[j] := h
```

Here $MutableList$ is any of the mutable data structures from above.
It is easy to see that this function indeed has the effect of swapping two elements in x. For arrays, the time complexity of *swap* is $\Theta(1)$. For linked lists, it is $\Theta(n)$.

5.3.1 Bubblesort

Bubblesort is a stable in-place sorting algorithm that closely follows the natural way how a human would sort. The idea is to find two elements that are not in order and swap them. If no such elements exist, the list is sorted.

```
fun bubblesort(x : Array[A]) =
    sorted := false
    while !sorted
        sorted := true
        for i from 0 to length(x) − 2
            if !x[i] ≤ x[i + 1]
                sorted := false
                swap(x, i, i + 1)
```

Correctness The for-loop compares all $length(x) - 1$ pairs of neighboring elements. It sets *sorted* to *false* if the list is not sorted. Thus, we obtain the loop invariant $F(x, sorted) = sorted == Sorted(x)$, which immediately yields partial correctness.
Total correctness follows from the termination ordering

$$T(x, sorted) = \text{number of pairs } i, j \text{ such that } !x_i \leq x_j + \begin{cases} 1 & \text{if } sorted == false \\ 0 & \text{if } sorted == true \end{cases}$$

Indeed, this number decreases in every iteration of the loop in which x is not sorted. The second summand is necessary to make sure $T(x, sorted)$ also decreases when x is already sorted (which happens exactly once in the last iteration).

Complexity If n is the length of x, each iteration of the while-loop has complexity $\Theta(n)$. Moreover, the while-loop iterates at most n times. That happens in the worst-case: when x is reversely sorted initially. Thus, the complexity is $\Theta(n^2)$.
In the best-case, when x is already sorted initially, the complexity is $\Theta(n)$. That is already optimal because it requires $n - 1$ comparisons to determine that a list is sorted.

5.3.2 Insertionsort

Insertion is also a stable in-place algorithm.
The idea is to sort increasingly large prefixes of a list x. If $[x_0, \ldots, x_{i-1}]$ is sorted already, the element x_i is inserted among them.

```
fun insertionsort(x : Array[A]) =
    for i from 1 to length(x) − 1
        current := x[i]
        pos := i
```

```
    while pos > 0 && current < x[pos − 1]          shift elements to the right to make space for current
        x[pos] := x[pos − 1]
        pos := pos − 1
    x[pos] := current
```

Correctness We use a loop-invariant for the for-loop: $F(x, i) = Sorted([x_0, \ldots, x_{i-1}])$. The preservation of the loop-invariant is non-obvious but straightforward to verify. It holds initially because the empty list is trivially sorted. That yields partial correctness.

Termination is easy to show using the termination ordering $T(x, i, current, pos) = pos$ for the while-loop.

Complexity If n is the length of x, the for-loop runs n times with $i = 0, \ldots, n-1$ Inside, the while-loop runs i times in the worst-case: if x is reversely sorted, all i elements before *current* must be shifted to the right. That sums up to $0 + 1 + \ldots + n - 1 \in \Theta(n^2)$.

Everything else is $O(n)$. Thus, the worst-case complexity is $\Theta(n^2)$.

In the best-case, if x is already sorted, the while-loop never runs, and the complexity is $\Theta(n)$.

5.3.3 Mergesort

Mergesort is based on the observation that

- sorting smaller lists is much easier than sorting larger lists (because the number of pairs that have to be compared in $\Theta(n^2)$),
- merging two sorted lists is easy (linear time).

Thus, we can divide a list into two halves, sort them recursively, then merge the results. This is similar to the idea of square-and-multiply (Sect. 4.1.3) and an example of the family of divide-and-conquer algorithms.

Because it needs auxiliary memory to do the merging of two half lists into one, it is easiest to implement as non-in-place algorithm. Then the input data structure does not matter and can be assumed to be immutable. The following is a straightforward realization:

```
fun mergesort(x : List[A]) : List[A] =
    n := length(x)
    if n < 2
        x
    else
        k := n div 2
        l := mergesort([x_0, ..., x_{k−1}])
        r := mergesort([x_k, ..., x_{n−1}])
        return merge(l, r)

fun merge(x : List[A], y : List[A]) : List[A] =
    xRest := x
    yRest := y
    res = []
    while nonempty(xRest) || nonempty(yRest)
        takefromX := empty(yRest) || (nonempty(xRest) && xRest.head ≤ yRest.head)
        if takefromX
            res := cons(xRest.head, res)
            xRest := xRest.tail
        else
            res := cons(yRest.head, res)
            yRest := yRest.tail
    return reverse(res)
```

Correctness Because the function *merge* is not part of the specification, we have to first specify which property we want to prove about it. The needed property for $z := merge(x, y)$ is:

- precondition: $Sorted(x)$ and $Sorted(y)$
- postcondition: $Sorted(z)$ and z is a permutation of $x + y$

Now we can prove each function correct.

First we consider *mergesort*. Partial correctness means to prove $Sorted(mergesort(x))$. That is very easy:

- If $n < 2$, x is trivially sorted.
- Otherwise:
  - $Sorted(a)$ and $Sorted(b)$ follow from the postcondition of the recursive call.
  - Then the postcondition of *merge* yields $Sorted(merge(a, b))$.

Relative termination is immediate (assuming that *merge* always terminates, which we prove below). A termination ordering is given by $T(x) = length(x)$. Indeed, *mergesort* recurses only into strictly shorted lists.

Second we consider *merge*. We use a loop invariant $F(x, y, xRest, yRest, res)$ that states that

- $Sorted(reverse(res))$ and $Sorted(xRest)$ and $Sorted(yRest)$
- All elements in res are in $\leq$-relation to all elements in $xRest + yRest$.
- $res + xRest + yRest$ is a permutation of $x + y$

It is non-obvious but it is straightforward to see that this is indeed a loop invariant:

- $reverse(res)$ remains sorted because we always take the smallest element in $yRest + xRight$ and prepend it to res. In particular, because $xRest$ and $yRest$ are sorted, the smallest element must be $xRest.head$ or $yRest.head$.
- For the same reason, all elements of res remain smaller than the ones of $xRest$ and $yRest$.
- Because we only remove elements from $xRest$ and $yRest$, they remain sorted.
- Because every element that is removed from $xRest$ or $yRest$ is immediately added to res, they remain a permutation.

To show partial correctness, we see that

- The loop invariant holds initially, which is obvious.
- After completing the loop, $xRest$ and $yRest$ are empty.
- Then, using the loop invariant, it is easy to show that $reverse(res)$ is sorted and a permutation of $x + y$.

To show termination, we use $T(x, y, xRest, yRest, res) = length(xRest) + length(yRest)$. It is easy to see that T is a the termination ordering for the while-loop.

Complexity We have to analyze the complexity of both functions.

First we consider *merge*. Let $n = length(x) + length(y)$.

- The three assignments in the beginning are $O(1)$.
- The while-loop is repeated once for every element of x and y, which requires $\Theta(n)$ steps. The body of the loop takes $O(1)$. So $\Theta(n)$ in total.
- The last step requires reverting res, which has n elements at this point. Reverting a list requires building a new list by traversing the old one. That is $\Theta(n)$ as well.

Thus, the total complexity of *merge* is $\Theta(n) = \Theta(length(x) + length(y))$.

Second we consider *mergesort*. Let $n = length(x)$. We compute the time complexity $C(n)$:

- The assignments and the if-statement are in $O(1)$.
- The recursive calls to *mergesort* take $C(n/2)$ each.
- The call to *merge* takes $\Theta(length(a) + length(b)) = \Theta(n)$.

That yields

$$C(n) = 2 \cdot C(n/2) + \Theta(n) = \ldots = 2^k \cdot C(n/2^k) + k \cdot \Theta(n)$$

By choosing $k = \log_2 n$ and $C(1) = C(0) \in O(1)$, we obtain

$$C(n) = n \cdot O(1) + \log_2 n \cdot \Theta(n) = \Theta(n \log_2 n)$$

Thus, mergesort is quasilinear and thus strictly more efficient than bubblesort and insertionsort.

Contrary to bubblesort and insertionsort, mergesort takes the same amount of time no matter how sorted the input already is. The recursion and the merging happen in essentially the same way independent of the input list. Thus, its best-case complexity is also $\Theta(n \log_2 n)$.

Remark 5.6 (Building the list reversely in *merge*). *merge* could be simplified by always adding the element $xLeft.head$ or $yLeft.head$ to the *end* of *res* instead of the beginning. However, as discussed in Sect. 5.2, adding an element to the beginning of an immutable list takes constant time whereas adding to the end takes linear time. Therefore, if we added elements to the end of *res* would become quadratic instead of linear. Then mergesort as a whole would also be quadratic.

5.3.4 Quicksort

Quicksort is similar to mergesort in that two sublists are sorted recursively. The main differences are:

- It does not divide the list x in half. Instead it picks some element a from the list (called the *pivot*). Then it divides x into sublists a and b containing the elements smaller and greater than x respectively.
  No merging is necessary because all elements in a are smaller than all elements in b. Thus the sorted list is $quicksort(a) + x + quicksort(b)$.
- To divide the list, quicksort has to traverse and reorder the list anyway. Therefore, it can easily be implemented in-place avoiding the use of auxiliary memory.

When implemented as an in-place sorting algorithm, the recursive call takes two additional arguments: two numbers $first$ and $last$ that describe the sublist that should be sorted.

Remark 5.7 (Additional Arguments in a Recursion). Carrying along auxiliary information is very typical for recursive algorithms. Therefore, we often find pairs of function:

- A recursive function that takes additional arguments.
  That is *quicksortSublist* below, which takes the entire list and the information about which sublist to sort.
- A non-recursive function that does nothing but call the other function with the initial arguments.
  That is *quicksort* below, which calls *quicksortSublist* on the entire list (e.g., on the sublist from 0 to the end of x).

```
fun quicksort(x : Array[A]) =
    quicksortSublist(x, 0, length(x) − 1)

fun quicksortSublist(x : Array[A], first : ℕ, last : ℕ) =
    if first ≥ last
        return
    else
        pivot := x[last]
        pivotPos := first
                loop invariant: x[k] ≤ pivot for k = first, ..., pivotPos − 1 and pivot ≤ x[k] for k = pivotPos, ..., j − 1
        for j from first to last − 1
            if x[j] ≤ pivot
                swap(x, pivotPos, j)
                pivotPos := pivotPos + 1
        swap(x, pivotPos, last)

        quicksortSubList(x, first, pivotPos − 1)
        quicksortSubList(x, pivotPos + 1, last)
```

Correctness Before proving correctness we have to specify the behavior of the auxiliary function *quicksortSublist*:

- precondition: none
- postcondition: $Sorted([x_{first}, \ldots, x_{last}])$

Then the correctness of *quicksort* follows immediately from that of *quicksortSublist*.

Now we prove the partial correctness of *quicksortSublist*. First, the base case is trivially correct: It does nothing for lists of length 0 or 1. For the recursive case, we prove that the following two properties hold just before the two recursive calls:

- The sublist $[x_{first}, \ldots, x_{last}]$ is a permutation of its original value, and no other elements of x has changed. That is easy to see because we only change x by calling *swap* on positions between $first$ and $last$.
- All values x_k are
  - smaller than *pivot* for $k = first, \ldots, pivotPos - 1$,
  - equal to *pivot* for $k = pivotPos$,
  - greater than *pivot* for $k = pivotPos + 1, \ldots, last$.

  We prove that by using the indicated loop invariant for the for-loop. It is trivially true before the for-loop because $first = pivotPos$ and $pivotPos = j$. It is straightforward to check that it is preserved by the for-loop. Therefore, it holds after the for-loop for the value $j = last - 1$. The last call to *swap* moves the pivot element into $x_{pivotPos}$ so that the loop invariant is now also true for $j = last$. Then the needed properties can be seen easily.

To prove the termination of *quicksortSublist*, we use the termination ordering $T(x, first, last) = last - first + 1$ (which is the length of the sublist). That value always decreases because the pivot element is never part of the recursive call.

Complexity Let $n = last - first - 1$ be the length of the sublist. It is easy to see that, apart from the recursion, *quicksortSublist* takes $\Theta(n)$ steps because the for-loop traverses the sublist. Thus, the complexity of quicksort depends entirely on the lengths of the sublists in the recursive calls. However, the pivot position and therefore those lengths are hard to predict.

The best-case complexity arises if the pivot always happens to be in the middle. Then the same reasoning as for mergesort yields best-case complexity $\Theta(n \log_2 n)$. The worst-case arises if the list is already sorted: then the pivot position will always be the last one, and the two sublists have sizes $n - 1$ and 0. That results in n recursive calls on sublists of length n, $n - 1$, ..., 1 as well as n calls on empty sublists. Consequently, the worst-case complexity is $\Theta(n^2)$.

However, the worst-case complexity does not do quicksort justice because it is much higher than its average-case complexity. Because there are only finitely many permutations for a list of fixed length, the average-case complexity can be worked out systematically. The result is $\Theta(n \log_2 n)$.

It may seem that quicksort is less attractive than mergesort because of its higher worst-case complexity. However, that is a minor effect because the algorithms have the same best-case and average-case complexity. Instead, the constant factors, which are rounded away by using Θ-classes, become important to compare two algorithms with such similar complexity.

Here quicksort is superior to mergesort. Moreover, quicksort can be optimized in many ways. In particular, the choice of the pivot can be tuned in order to increase the likelihood that the two sublists end up having the same size. For example, we can randomly pick 3 elements of the sublist and use the middle-size one as the pivot. With such optimizations, quicksort can become substantially faster than mergesort.

5.3.5 Other Algorithms

There is a number of other sorting algorithms that we will not go into here. Examples include counting sort, radix sort, and bucket sort.

One particularly important sorting algorithm is heap sort, which we discuss in Sect. 9.6.4 (after introducing heaps).

5.3.6 In Programming Languages

Most programming languages come with a standard library that includes efficient sorting algorithms. Moreover, other libraries for other algorithms may be around. In some cases, languages only specify the interface and leave the implementation (and thus the choice of algorithm) to individual implementations of compilers/interpreters.

The following gives some examples.

Python uses Timsort (named after the programmer), which is a hybrid of mergesort and insertionsort with various optimizations. It is written directly in C.

Java used to use just quicksort. Java 7 uses either Timsort (ported to Java) or a variant of quicksort that uses two pivot elements.

Scala defers to Java's implementation.

C++'s std library specification does not prescribe a sorting algorithm but requires $O(n \log_2 n)$ worst-case complexity (average-case in earlier versions). Implementations vary in their choice of algorithm, e.g., using hybrid algorithms that perform some iterations of quicksort before switching to insertionsort for the resulting small lists.

For Javascript, the choice is up to the browser (because every browser is a separate implementation of Javascript).

Part II

Important Data Structures

Chapter 6

Finite Data Structures

6.1 Void

The set *void* contains no elements.

Not surprisingly, it is rarely used. However, it is nice to have when dealing with operations that do not return. For example, we say that throwing an exception or terminating the program returns an element of *void*.

Most programs do not need the type void. And most programming language either do not have it or only have it under the hood.

6.2 Unit

The set *unit* contains exactly one element, which we write ().

It is rarely used because if we know that $x \in unit$, we already know the value of x. Thus, having a value of type *unit* gives us no information.

However, *unit* is nice to have when dealing with operations that do return, but do not return a value. In that case, we say that the operation returns type *unit*.

For example, assignments, loops, and print statements return *unit*. Many methods of mutable data structures also return *unit*. For example, using *unit*, we can specify *insert* for a mutable list from Sect. 5.1.1 as $insert(x \in A^*, a \in A, n \in \mathbb{N}) \in unit$.

Functional programming languages usually have a built-in type *unit*. That way, in a functional programming langauge, every operation has a return type.

6.3 Booleans

The set *bool* contains exactly two elements, which we call *true* and *false*.

Most programming languages have a built-in type *bool*, which is the result type of the equality operator.

6.4 Integers Modulo

For $m > 0$, the set $\mathbb{Z}_m$ consists of the elements $\{0, \ldots, m-1\}$.

Most programming languages do not offer $\mathbb{Z}_m$ for every m. Usually, they offer at most $\mathbb{Z}_{2^k}$ for $k = 8$ (usally called *byte*), $k = 16$ (*word*), $k = 32$ (*integer*), and/or $k = 64$ (*long*).

Note that, depending on the programming language, the built-in type *int* may refer to one of those (usually for $k = 32$) or to $\mathbb{Z}$.

If we need $\mathbb{Z}_m$ for a specific m, we usually work with *int* and use the mod operation to ensure we remain inside $\mathbb{Z}_m$. [1]

[1] Note that some programming languages implement div and mod in unexpected ways for negative arguments.

6.5 Enumerations

For fresh names $l_1, \ldots, l_n$, the set $enum\{l_1, \ldots, l_n\}$ has exactly n elements, which are called $l_1, \ldots, l_n$.

The names l_i must be fresh. That means they may not have been defined previously. This is similar to how the name of a new function or class must be fresh. This is because defining an enumeration set introduces new values, namely the l_i.

Most programming languages allow defining enumeration types in some way. For example, in SML:

```
datatype answer = yes | no | maybe
```

Or in C:

```
enum {yes, no, maybe} answer;
```

Chapter 7

Number-Based Data-Structures

7.1 Countable Sets

The sets $\mathbb{N}$, $\mathbb{Z}$, and $\mathbb{Q}$ are well-known from mathematics.

Working with $\mathbb{Z}$ (as opposed to $\mathbb{Z}_m$ for some m) is called *arbitrary precision arithmetic*. $\mathbb{Z}$ may or may not be the built-in type *int*—that depends on the programming language. If not, *int* is $\mathbb{Z}_m$ for some m—in those languages, there is usually a library that defines $\mathbb{Z}$.

A data structure for $\mathbb{Q}$ can be defined by using pairs of integers.

We usually do not use a special data structure for $\mathbb{N}$ and instead just use the positive values of $\mathbb{Z}$. Alternatively, we can give a (very inefficient) definition of $\mathbb{N}$ as an inductive type as in Ex. 2.15.

7.2 Uncountable Sets

We cannot implement data structures for $\mathbb{R}$ and $\mathbb{C}$ because they are uncountable.

There are some approximate solutions to work with $\mathbb{R}$. For example, we can simply represent a real number r as a function $\mathbb{N} \to \{0, \ldots, 9\}$ that provides the infinite decimal expansion of r. Because we can only represent countably many functions as effective objects, not all real numbers can be represented like. However, all practically useful ones can. A major drawback of this representation is that we cannot give an algorithm for equality (because we would have to check that two functions are equal for infinitely many arguments), thus crippling the data structure.

For the $\mathbb{C}$, it is often sufficient to work with the countable set $\mathbb{Q} + \mathbb{Q}i$, which is the set of complex numbers whose real and imaginary parts are rational.

Chapter 8

Option-Like Data Structures

8.1 Specification

$A^?$ is a set containing all the elements of A and one additional element $\perp$.
$A^?$ is used to represent an optional value of A. The element $\perp$ is used to represent an undefined/absent value.
Options are usually immutable. The main operations on $A^?$ are

function	returns	effect
$getOrElse(x \in A^?, default : A) \in A$	get the optional value or a default value	none
$get(x \in A^?) \in A$	get the optional value	error if absent
$map(x \in A^?, f \in A \to B) \in B^?$	apply f to the optional value	none

8.2 Data Structures

8.2.1 Using Inductive Types

In functional programming languages, a data structure for optional values can be realized as an inductive type:

data $Option[A] = Some(value : A) \mid None$

Such a definition (except for possibly using different names) is usually part of the standard library of the language.

8.2.2 Using Pointers

In languages that use pointers, we can represent $Option[A]$ as the type $Pointer[A]$ (written $A*$ in C) of pointers to elements of A. In that case, the *null* pointer represents $\perp$.
Object-oriented languages do not necessarily expose pointers to the programmers (e.g., C++ does, but Java does not). However, even then they use pointers internally, and any class-type provides the value *null*. In this situation it is impossible to represent the set A correctly as a class—any class for A is automatically a data structure for $A^?$. This often causes ambiguous specifications and subtle errors. Therefore, it is good practice to never use *null* even when possible.

Chapter 9

List-Like Data Structures

The specification and several data structures for mutable and immutable lists are already discussed in Sect. 5.1. Here we only discuss some additional data structures for the set A^*.

9.1 Stacks

$Stack[A]$ is a data structure for the set A^*.

$Stack[A]$ is very similar to $List[A]$. The difference is that $Stack[A]$ provides *less* functionality. While $List[A]$ is a general-purpose list, $Stack[A]$ is custom-fitted to one specific, very common use case. By requiring fewer operations, they allow more optimized implementations.

Stacks can be mutable or (less commonly) immutable. Here we will use the mutable variant. The functions for mutable stacks are:

function	returns	effect
$push(x \in A^*, a \in A) \in unit$	nothing	prepend a to x
$pop(x \in A^*) \in A^?$	the first element of x (if any)	remove the first element of x
$top(x \in A^*) \in A^?$	the first element of x (if any)	none

The intuition behind stacks is that they provide a LIFO store of data. LIFO means last-in-first-out because every *pop* returns the most recently pushed value. This is exactly the behavior of a literal stack of items: We can put an item on top of a stack (*push*), remove an item from the stack (*pop*), or check what item is on top (*top*). We cannot easily see or remove the other items.

Very often, the LIFO behavior is exactly what is needed. For example, when we solve a maze, we can push every decision we make. When we hit a dead end, we trace back our steps—for that, we have to pop the most recent decision, and so on.

9.2 Queues

Queues are very similar to stacks. Everything about stacks also applies to queues except for the following.

The functions for mutable queues are:

function	returns	effect
$enqueue(x \in A^*, a \in A) \in unit$	nothing	append x to A
$dequeue(x \in A^*) \in A^?$	the first element of x	remove the first element of x
$empty(x \in A^*) \in bool$	true if x is empty	none

The intuition behind queues is that they provide a FIFO store of data. FIFO means first-in-first-out because every *dequeue* returns the least recently enqueued value. This is exactly the behavior of a literal queue of people: Every newcomer has to queue up at the end of the queue (*enqueue*), and every time a server is ready the first in line gets served (*dequeue*). Newcomers cannot cut in line, and the server cannot easily see who else is waiting.

Very often, the FIFO behavior is exactly what is needed. For example, when we have a list of tasks that need to be done. Every time we create a new task, we enqueue it, and whenever we have time we dequeue the next task.

Queues are often used when components exchange messages or commands. In that case, some components—called the producers—only call enqueue, and other components—called the consumers—only call dequeue. For example, the producers can be different programs, A is the type of print jobs, and the consumers are different printers.

More complex queue data structures may also for dequeueing based on priority (see also Sect. 9.6.3).

9.3 Buffers

Buffers are conceptually very similar to queues. But $Buffer[A]$ is usually optimized for enqueueing and dequeueing many elements of A at once. Therefore, while stacks and queues can be implemented well using linked lists, buffers usually use arrays to be faster.

A typical $Buffer[A]$ consists of three components:

- an $Array[A]$ b
- two integers $begin$ and end indicating the first and last valid entry in the array.

Enqueueing writes to $b[end + 1]$ and increments end. Dequeueing reads from $b[begin]$ and increments $begin$.

A buffer overflow occurs when incrementing $begin$

For example, when a browser receives a web page, its network component loads the page into a $Buffer[char]$. In parallel, its HTML parser component starts processing the partially received page. That way the HTML page can be displayed partially already before it is fully loaded.

Buffers are almost always used automatically when a program is writing to a file. In that case, a $Buffer[int]$ or $Buffer[char]$ is used that holds the data written to the file. The write command does not actually write data to the file directly—it only enqueues it in the buffer. That is advantageous because enqueueing to a buffer in memory is much faster than writing to the hard drive. While the program is already moving on, the programming language libraries or the operating system work in the background to periodically dequeue and write all characters to the file.

When working with files, an important operation is *flushing* the buffer. This forces the immediate processing of all data in the buffer. Flushing happens automatically at the latest when the program terminates. However, occasionally manual flushing is necessary:

- When a program terminates with an error, buffers have to be flushed to avoid losing data.
- When a program writes log data to a file that the programmer wants to read immediately, it is important to flush regularly to make sure the programmer reads updated information.

9.4 Iterators

9.4.1 Specification

$Iterator[A]$ is a data structure for the set A^*.

Iterators are usually mutable. Their functionality is even more restricted than the one of stacks and queues:

function	returns	effect
$getNext(x \in A^*) \in A$	the first element of x	remove the first element of x
$hasNext(x \in A^*) \in bool$	$true$ if x is not empty	none

The typical way to use an iterator $i \in Iterator[A]$ is the following:

```
while hasNext(i)
    a := getNext(i)
    do something with a here
```

This is called **traversing** the iterator. Afterwards the iterator is traversed and cannot be used again.

$Iterator[A]$ may look somewhat boring. In order to understand the value of iterators, we have to make one definition: A data structure $D[A]$ is called **iterable** if there is a function

$$iterator(x \in D[A]) \in Iterator[A]$$

Now the imoprtance of iterators follows from two facts:

- Many data structures D are iterable (see Sect. 9.4.3).
- Many important operations for D can be realized using only the functionality of iterators (see Sect. 9.4.2).

Thus, iterators provide a sweet-spot in the trade-off between simplicity and expressivity—they are very simple, but we can do a lot with them.

> *Remark* 9.1 (Simplicity vs. Expressivity). The trade-offs between simplicity and expressivity comes up again and again in computer science. The best data structures combine both properties, but usually they are mutually exclusive.
> All the important data structures presented in Part II have become important because they do well in this way.

9.4.2 Working with Iterable Data Structures

Let us assume an iterable data structure $D[A]$. Our goal is to define functions on $x \in D[A]$ that use only $iterator(x)$. There are indeed many of those. Some important ones are:

function			returns
		below, let $X = iterator(x)$	
$length$	$(x \in D[A])$	$\in \mathbb{N}$	numbers of elements in X
$contains$	$(x \in D[A],\ a \in A)$	$\in bool$	*true* if a occurs in X
$index$	$(x \in D[A],\ a \in A)$	$\in \mathbb{N}^?$	the position of the first occurrence of a in X (if any)
$find$	$(x \in D[A],\ p \in A \to bool)$	$\in A^?$	the first element a in X (if any) such that $p(a)$ is *true*
$count$	$(x \in D[A],\ p \in A \to bool)$	$\in \mathbb{N}$	the number of elements a in X for which $p(a)$ is *true*
$forall$	$(x \in D[A],\ p \in A \to bool)$	$\in bool$	*true* if $p(a)$ is *true* for every element a in X
$exists$	$(x \in D[A],\ p \in A \to bool)$	$\in bool$	*true* if $p(a)$ is *true* for some element a in X
map	$(x \in D[A], f \in A \to B)$	$\in Iterator[B]$	an iterator for $[f(a_1), \ldots, f(a_n)]$ where $x = [a_1, \ldots, a_n]$
$filter$	$(x \in D[A], p \in A \to bool)$	$\in Iterator[B]$	like x but skips elements that do not satisfy p
$results$	$(x \in D[A],\ f \in A \to B)$	$\in List[B]$	the list of results from applying f to all a in X
$fold$	$(x \in D[A],\ b \in B, f \in A \times B \to B)$	$\in B$	$f(a_1, f(a_2, \ldots, f(a_n, b))\ldots)$ with $X = [a_1, \ldots, a_n]$

All of the above functions should not have a side-effect. However, some of them take other functions as arguments. It is usually a bad to do so, but it is technically possible that these functions have side-effects. There is only one exception where we explicitly allow f to have a side-effect:

function	returns	effect
$foreach(x \in D[A], f \in A \to unit) \in unit$	nothing	apply f to all a in X

The trick behind *map* (and the difference to *results*) is that x is not traversed right away. Instead, we create a new iterator that, when traversed, applies f. That way we ensure that f is applied only as often as necessary.

9.4.3 Making Data Structures Iterable

We can give a data structure for iterators as an abstract class:

```
abstract class Iterator[A]()
  fun hasNext() : bool                          precondition for getNext is hasNext == true
  fun getNext() : A
```

Then we can define, e.g., *map* as follows:

```
class Map[A, B](x : Iterator[A], f : A → B) extends Iterator[B]
  fun hasNext() : bool = {x.hasNext}
  fun getNext() : B = {f(x.getNext)}

fun map(x : D[A], f : A → B) = {new Map[A, B](iterator(x), f)}
```

Many important data structures are naturally iterable, and that can be realized by implementing the abstract class. That includes in particular all data structures for lists:

```
class ListIterator[A](l : List[A]) extends Iterator[A]
    index := 0
    fun hasNext() : bool = {index < length(l)}
    fun getNext() : A =
        a := get(l, index)
        index := index + 1
        a

fun iterator(l : List[A]) : Iterator[A] = {new ListIterator(l)}
```

9.5 Streams

$Stream[A]$ is not a data structure for the set A^*. Instead, it is a data structure for the set $A^{\mathbb{N}}$.

The set $A^{\mathbb{N}}$ contains functions $f : \mathbb{N} \to A$, which we can think of as infinite lists $[f(0), f(1), \ldots]$. Because they are so similar to lists, they are usually treated together with lists, even though they do not realize the same set.

The set $A^{\mathbb{N}}$ is uncountable. Therefore, not all possible streams are effective objects that can be represented in a physical machine. However, for many practical purposes, it is fine to treat $Stream[A]$ as if it were the type of all possible streams.

$Stream[A]$ is usually implemented in the same way as $Iterator[A]$ with the understanding that $hasNext$ is always $true$, i.e., the stream is never over.

Consequently, the functions on $Iterator[A]$ behave slightly differently when used for $Stream[A]$. For exapmle:

- We cannot call $length$, $count$, $results$, $fold$, and $foreach$ on streams.
- We can call $contains$ on a stream. However, the function may run forever if the searched-for element is not in the stream. The same caveat applies to $index$, $find$, $forall$, and $exists$.
- We can call map (because map returns a new iterator without actually applying the map-function to all elements right away).

9.6 Heaps

Heaps are formally defined in Sect. 10.1.3.

$Heap[A, O]$ is not a data structure for the set A^*. Instead, it is a data structure for the subset of A^* containing only lists sorted according to O. Therefore, heaps are very useful for sorting and prioritizing. We discuss applications of heaps to lists in Sect. 9.6.3 and 9.6.4.

First we introduce some basic operations on heaps in Sect. 9.6.1.

9.6.1 Operations on Heaps

Because heaps are mostly used for efficiency, they are usually mutable. The main operations on a heap are similar to those on a stack:

function	returns	effect
$insert(x \in Heap[A, O], a \in A) \in unit$	nothing	add a to x in any position
$extract(x \in Heap[A, O]) \in A^?$	the O-smallest element of x (if any)	remove that element from x
$find(x \in Heap[A, O]) \in A^?$	the O-smallest element of x (if any)	none

$insert$, $extract$, and $find$ for heaps correspond exactly to $push$, pop, and top for stacks. The crucial different is that $insert(x, a)$ does not prepend a to x—instead, it is unspecified where and how x is added. $extract$ and $find$ do not return the most recently added element—instead, they return the smallest element with respect to O.

It is unspecified what exactly a heap looks like and where and how $insert$ actually performs the insertion. That way heaps have a lot of freedom to organize the data in an efficient way. That freedom is exploited to make the operations $extract$ and $find$ fast.

Because $Heap[A, O]$ is underspecified, there are many different options how to implement heaps. In practice, there are dozens of competing variants using different efficiency trade-offs. A critical property is that all operations take only $O(\log n)$ where n is the number of elements in the heap.

9.6.2 Implementations

The most important case of $Heap[A, O]$ are binary heaps H, i.e., binary trees over A that are also heaps. There is a wide variety of optimized implementations of heaps.

Using Trees

For a straightforward implementation, we use a tree.

Let n be the number of nodes in H and h be the height of H. All operations are such that H remains almost-perfect: for every depth $d < h$ there are maximally many nodes, i.e., 2^d nodes. At depth h, we have to allow for fewer than 2^h nodes because not every n there is a perfect heap. We use the convention that the nodes at level h are as far to the left as possible. That way, we always have $h \leq \log_2 n$, and all branches have length h or $h-1$, i.e., $O(\log_2 n)$.

$find$ is trivial: We return the root of H. That takes $O(1)$.

$insert(H, x)$ inserts x into one of the branches with minimal length. If the heap is perfect, we extend it to a new level $h+1$ and insert x all the way to the left. Otherwise, we add it in the left-most free slot at level h. The insertion occurs at the position that keeps the branch sorted. Because it was sorted already, that requires $O(l)$ operations, where l is the length of the branch, i.e., $O(\log_2 n)$. It is easy to check that the resulting tree is again a heap.

$extract$ removes the root of H and returns it. That takes $O(1)$. Additionally, we have to repair the heap property. To do that, we take some leaf l of H and put it at the root. Now we have an almost-perfect binary tree again, but it is not a heap yet: l may be too big to be the root. Therefore, we push l down by iteratively swapping it with its smallest child until we have a heap. Finding a leaf and pushing along some branch takes $O(\log_2 n)$.

Using Arrays

For efficiency, it is often preferable to store the nodes of the heap as an array.

This requires a bijection that translates between positions in the heap to positions in the array. Let h be the height of the heap. Then a position in the heap is a list $p \in \{left, right\}^*$ of length up to h that describes the path from the root to a node. A position in the array is an integer $i \in \{0, \ldots, 2^{h+1} - 1\}$.

It is straightforward to give such a bijection. For example, we can number the nodes of the heap in BFS order. Then the heap-position p corresponds to the array position $2^{length(p)} + j$ where j is the number obtained by treating p as a binary number (with $left$ and $right$ corresponding to 0 and 1).

9.6.3 Priority Queues

A $PriorityQueue[A]$ behaves like a $Queue[A]$ except that dequeueing returns the element with the highest priority. This is achieved by using a data structure for $Heap[A, O]$ where O orders elements by decreasing priority. Then $insert$ and $extract$ correspond to $enqueue$ and $dequeue$.

9.6.4 Heapsort Algorithm

Heapsort is a sorting algorithm that runs in $\Theta(n \log n)$.

If $\leq$ is the total order for sorting, a simple heapsort is given by

```
fun heapsort(x : A*) : A* =
    h := new Heap[A, ≥]()

    left := x
    for i from 0 to length(x) − 1
        next := left.head
        insert(h, next)
```

```
    left := left.tail

res := Nil
for i from 0 to length(x) − 1
    next := extract(h)
    res := prepend(next, res)
```

This uses two loops using $length(n)$ iterations each. The first loop throws all elements of x into the heap; the second loop pulls them out again and builds the list res to be returned. Because *extract* always returns the greatest element, the result is automatically sorted. Any other implementation of a priority queue yields a corresponding sorting algorithm.

If n is the length of the list, each *insert* and *extract* operation takes at most $\Theta(\log n)$. Thus, heapsort runs in $\Theta(n \log n)$.

There are much more optimized implementations of heapsort than the above example, possibly using optimized implementations of heaps. In particular, there are encodings of the heap structure in an array, which allow using heapsort as an in-place sorting algorithm. With those optimizations, heapsort is among the fastest sorting algorithms (but still takes $\Theta(n \log n)$).

Chapter 10

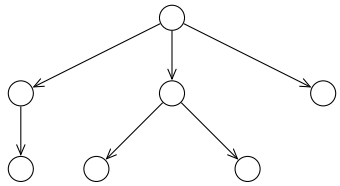
Tree-Like Data Structures

After lists, trees are the next most important data structure in computer science. They can be seen as a generalization of lists where the elements are not arranged in a row, but branching is allowed.

10.1 Specification

10.1.1 General Trees

There are many equivalent definitions. The easiest is by graphical example: A tree is something that looks like



A more formal definition is this:

Definition 10.1 (Tree)**.** A **tree** is a connected directed graph in which

- there is exactly one node (called the **root**) with in-degree 0,
- all other nodes have in-degree 1.

Here we already used the more general concept of graphs, which we define formally in Sect. 12.

Talking about the shape and parts of a tree can be confusing. Therefore, we introduce some vocabulary that helps us:

Definition 10.2 (Parts of a Tree)**.** For every edge from p to c, we call p the **parent** of c and n a **child** of p. Thus, the root has no parent; every non-root node has exactly one parent. A node may have any number of children. A node with 0 children is called a **leaf**. A node that is neither the root nor a child is called an **inner node**.

For every path from a to d, we call a an **ancestor** of d and d a **descendant** of a. Thus, all nodes are descendants of the root Every node is an ancestor/descendant of itself; a **proper** ancestor/descendant of n is an ancestor/descendant that is not n itself.

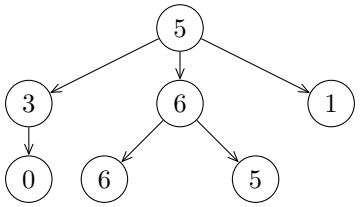
The number of proper ancestors of n is called the **depth** of n. Thus, the root has depth 0.

For a node n, the descendants of n form a tree again, which has root n. It is called the **subtree** at n.

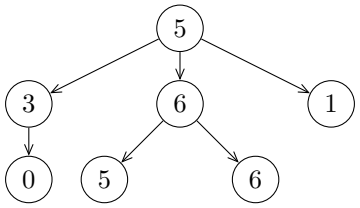
A path from the root to a leaf is called a **branch**. Thus, every leaf l is part of exactly one branch, whose length is the depth of l. The length of the longest branch(es) is called the **height** of the tree.

Remark 10.3. Contrary to all these tree metaphors, computer scientists prefer drawing trees with the root at the top and the leafs at the bottom.

Def. 10.1 only defines the abstract shape of trees. But trees are only useful if we can store some data in each node. For example, the following is a tree of integers:



Once we store data in a tree, we have to be a bit more careful: the order of children matters now. For example, the above tree of integers is different from the tree of integers below even both are based on the same tree.



Keeping track of the order makes the definition more complicated. The following definition is one out of several equivalent formal definitions:

Definition 10.4 (Trees over a Set)**.** The set $Tree[A]$ contains the **trees over the set** A. Such a tree over A consists of

- a set N (whose elements we call the **nodes**),
- a function $label : N \to A$ that maps nodes to elements of A ($label(n)$ is called the **label** of n, it is the data stored in each node),
- a function $children : N \to N^*$ that maps every node to its list of children,

such that N and $children$ form a tree.

10.1.2 Binary Trees

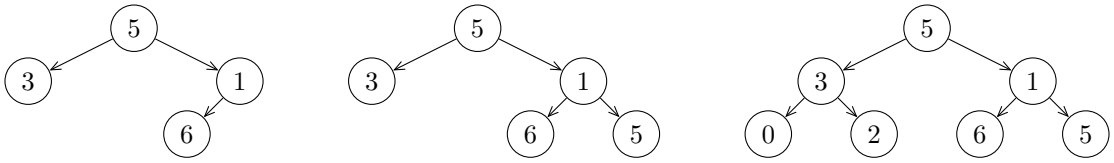
Binary trees are an important special case:

Definition 10.5 (Binary Tree)**.** A **binary tree** is a tree in which all nodes have at most 2 children. If a node has 2 children, the first and second child are called the **left** and **right** child, respectively.

Binary trees over a set are defined accordingly.

A binary tree is called **full** if all non-leaf nodes have exactly two children. A full binary tree is called **perfect** all leafs have the same depth. It is called **almost-perfect** if it is perfect except for missing some nodes at the deepest level as far to the right as possible.

For example, the following are, from left to right, a non-full, a full but not perfect, and a perfect binary tree of integers:



The middle tree would be almost-perfect if 6 and 5 were children of 3 instead of 1.
It is important to know the number of nodes in a binary tree:

Theorem 10.6. *A binary tree of height h has at most 2^n nodes at depth n. It has at most $2^{h+1}-1$ nodes in total. If it is perfect, it has exactly 2^n nodes at depth n and exactly $2^{h+1}-1$ nodes in total.*

Proof. Exercise. □

In particular, the number of nodes grows exponentially with the depth. Vice versa, we can organize n nodes as a binary tree of height $\log_2 n$. The latter property is often useful to obtain logarithmic implementations: if we organize n elements in a (nearly) perfect binary tree, we can reach any element in $\log_2 n$ steps.

10.1.3 Trees for Ordered Sets

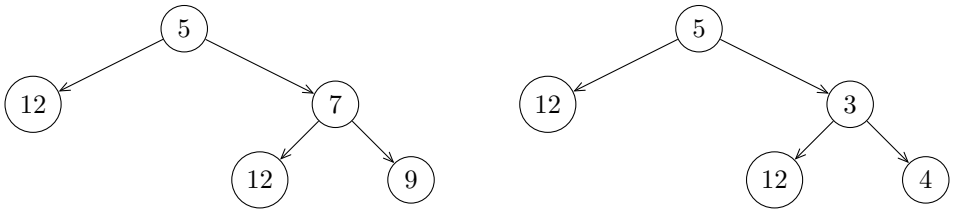
Special cases of trees are sometimes used to store sets of values for which we have a total order O.

Heaps

Assume a fixed total order O on A.

Definition 10.7 (Heap)**.** $Heap[A, O]$ is the subset of $Tree[A]$ containing only trees in which all branches are sorted with respect to O.

The elements of $Heap[\mathbb{Z}, \leq]$ are also called **min-heaps**. The elements of $Heap[\mathbb{Z}, \geq]$ are also called **max-heaps**.
The left tree below is a (binary) min-heap, the right one is neither a min-heap nor a max-heap:



In a heap, every node is smaller (with respect to O) than all its descendants. The root is always the smallest element in the heap. That makes heaps practical for sorting. Applications are presented in Sect. 9.6.

Binary Search Trees

Binary search trees are similar to heaps but the order property is different. In a heap every node is smaller than both its children. In a binary search tree, every node is greater than all its left descendants and smaller than all its right descendants.
They are discussed in Sect. 11.2.4.

10.1.4 Variants

Trees are simple enough to come up everywhere. But they are difficult enough to defy standardization. Contrary to, e.g., lists, the definition of tree can vary subtly across textbooks, programming language libraries, and computer scientists.
The following lists some details to watch out for when interacting with what someone else calls *trees*.

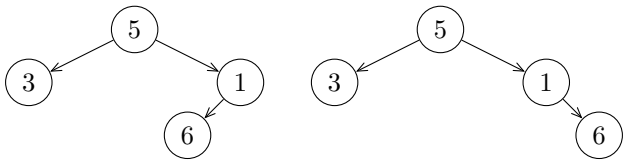
Rooted Trees Some definitions speak of *rooted trees*. That is usually redundant because there are no trees without a root.
But some definitions (unlike ours) allow for trees where the root is undetermined and multiple nodes could be the root. Then rooted trees are trees with a distinguished root node.

Trees vs. Labeled Trees We distinguish between trees, which just define the shape, and trees over a set, where the nodes are labeled with data. Others may or may not make that distinction and may use the word *tree* to refer to either concept.

Order of Children Some definition may make the nodes in a tree a *set* of children instead of (as in our definition) a *list*.

Leaf-Labeled Trees Our $Tree[A]$ data structure contains trees in which *every* node stores data from A. Occasionally, we are also interested in trees where only the *leafs* are labeled. And sometimes we need trees where inner nodes are labeled with elements of A and leafs with elements of B.

Single Children in Binary Trees Some people will speak of binary trees if every node has 0 or 2 nodes (but not 1).
When nodes with 1 child are allowed (like in our definition), definitions may or may not distinguish whether that one child is the left or the right child. Thus, they may consider the following trees to be the same (like in our definition) or different (which would make the definition of binary search tree in Sect. 11.2.4 simpler):

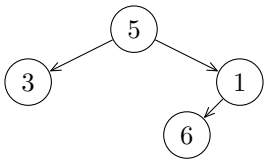


Properties of Binary Trees The properties *complete*, *full*, *balanced*, and *perfect* are all similar. They all relate to the goal of arranging a fixed number of nodes into a tree of small height.
But their definitions vary slightly.

Heaps Some people say *heap* to refer exclusively to heaps of integers.
Some people will assume that heaps are always binary trees.

10.2 Data Structures

Trees can be mutable or immutable. However, trees are mostly used to store data. Many algorithms work with a single mutable tree and insert data into it or delete data from it over time.
We consider two different data structures and use the following as an example tree



10.2.1 Using Lists

The simplest data structure for trees uses lists:

class $Tree[A](data : A,\ children : List[Tree[A]])\{\}$

The example tree is represented as

$$\textbf{new}\ Tree[\mathbb{Z}](5,\ [\textbf{new}\ Tree[\mathbb{Z}](3, Nil),\ \textbf{new}\ Tree[\mathbb{Z}](1, [\textbf{new}\ Tree[\mathbb{Z}](6, Nil)])])$$

10.2.2 Using Sibling Pointers

Some programmers or programming languages prefer a more awkward (but less memory-intensive) data structure that does not use lists.

Here every node has two pointers: one to its first child and one to its next sibling:

class $Node[A](data : A,\ firstChild : Node[A], nextSibling : Node[A])\{\}$

For leafs, the field $firstChild$ is $null$; for the last child of a node, the field $nextSibling$ is $null$. It would be better not to use $null$. But programmers who use this data structure usually do not mind.

The example tree is represented as

$$\textbf{new}\ Node[\mathbb{Z}](5,\ \textbf{new}\ Tree[\mathbb{Z}](3, null,\ \textbf{new}\ Tree[A](1, \textbf{new}\ Tree[\mathbb{Z}](6, null, null)), null), null)$$

10.3 Applications

The usefulness of lists and sets is self-evident because set- and list-structured data pervades all sciences. The usefulness of trees is more questionable by comparison. Therefore, we explain the most important applications.

10.3.1 Tree-Structured Data

Tree-structured data is not as pervasive as set- or list-structured data. But if tree structures come up, they are often critical as a structure-giving concept. That is because sets and lists provide only comparatively trivial structure.

Hierarchic Grouping

The most convincing application of trees comes up when recursively grouping data into subsets. We give two examples.

Sectioning structure of documents Any document (book, paper, etc.) naturally has a sectioning structure given by its **sections**. Here *section* is the generic term; in practice, sections can be called, e.g., chapters, (sub)section, paragraphs, sentences, etc.

The sectioning structure forms a tree: the overall document is the root, and every subsection is a child of its containing section. The root and each inner node are labeled with the their title. They may also be labeled with other data, e.g., each element in a collection of articles can be labeled with its author(s). The leafs (typically paragraphs or sentences) are labeled with strings that contain the actual text. The table of contents shows the entire tree.

Almost every document description language provides an explicit means for sectioning. For example, this includes the `hn` and `div` elements in HTML[1], the `\section` etc. commands in LaTeX, and the `text:h` and `text:p` elements in the OASIS OpenDocument standard[2] for text documents.

In many documents, the sections are numbered, and the numbers provide unique identifiers for the children of a section. Examples are software specifications and legal documents. Often the path from the root to a node n is used to identify the section at n. These paths are written as identifiers like 3.5.2 for the second child of the fifth child of the third child of the root.

Structure of programs Programs can be seen as a special case of documents. The structure into packages, classes, functions, and similar grouping declarations corresponds very closely to the sectioning structure of documents.

[1] https://www.w3.org/TR/html5/
[2] http://docs.oasis-open.org/office/v1.2/os/OpenDocument-v1.2-os-part1.html

File Systems File systems usually employ a tree structure consisting of directories for inner nodes and files for leafs. However, in a file system, the order of children usually does not matter. Instead, each node is labeled with a *name* that all nodes with the same parent have different names.

The root of the file system tree is written `/` in Unix systems. Paths like `/a/b` are used to describe the child named `b` of the child named `a` of the root.

Institutions Most institutions like bureaucracies or companies structure their members in a tree structure. The root corresponds to the, e.g., president, and the path from the root to a leaf is the *chain of command*.

Decision Trees

Many processes or developments can be described naturally as a sequence of decision steps. At each step, an agent has to select one out of several options. The agents include both software and hardware systems as well as humans.

Each node corresponds to a point in time. The root corresponds to the start. For every node, the children are the options among which to choose. After making a choice, time progresses to the corresponding child, and a new decision must be made.

This yields a discrete model of time where the past is linear (the path from the root to the current node) and the future is branching (the subtree at the current node).

10.3.2 Structure-Sharing for Lists

A prefix of the list $[x_0, \ldots, x_i]$ is any list $[x_0, \ldots, x_i]$ for $i \leq n$. A suffix is any list $[x_i, \ldots, x_n]$. Very often we have to store a large set of lists that share common prefixes or suffixes.

If we represent all lists such that the suffixes are shared, using immutable lists automatically leads to a tree structure. For example, consider the following lists:

$a := []$

$b := cons(1, a)$
$c := cons(2, a)$

$d := cons(3, b)$
$e := cons(4, b)$
$f := cons(5, b)$
$g := cons(6, d)$

Because *cons* for immutable lists does not trigger copying the argument list, the seven lists form a tree in memory. Here the root is the empty list, and each list corresponds to the path from a node to the root. Note that, contrary to the usual data structure for trees, in this tree each node points to its parent instead of the other way around.

Such lists with shared suffixes occur a lot in hierarchic data structures and recursive functions for them.

For example, when compiling or interpreting a program, we must maintain the function Var, which maps places p in the program to the set of variables $Var(p)$ that are in scope at p.

Maintaining the lists $Var(p)$ for all p is done most efficiently using a $T : Tree[List[V]]$ where V is the type of variable declarations. Every p corresponds to a node of T labeled with the variables declared locally in p, and the list $Var(p)$ is given by the concatenation of all lists in the path from p to the root.

If, e.g., p is a line inside a method f inside a class c, then $Var(p) = Decl(f) + Decl(c) + global$ where $Decl(f)$ are the local variables declared in f so far, $Decl(c)$ are the declarations of c, and *global* is the list of global variables.

10.3.3 Efficient Storage of Sets or Lists

Sometimes the tree structures is an artifact of storing a set of objects efficiently. Concretely, we store a set or a list C of objects as a tree T. The correspondence between C and T is such that every node or every leaf of T is labeled with an element of C.

This can allow for more efficient access to an object of C. For example, if C has size n and T is a perfect binary tree, we access every object of C using $O(\log n)$ steps.

Examples are heaps (see Sect. 9.6) and binary search trees (see 11.2.4).

10.3.4 Generic Data Description Languages

Generic data description languages usually provide at two primitive constructors: lists and trees. Intuitively, trees are used for hierarchically grouping and lists are used for aggregating objects.

Examples of such languages are XML and JSON.

This indicates that, in practice, all data structures can be encoded relatively conveniently by combining only lists and trees.

10.4 Important Algorithms

10.4.1 Search

Trees are often used to represent a problem.

> *Example* 10.8. Consider a labyrinth in which some treasure is hidden. We represent it as a tree. The entrance is the root. Every fork in the path is a node with multiple children—one child per direction we can go in. Every dead end is a leaf. One node in the tree is special because it has the treasure.
>
> To find the treasure, we have to explore the labyrinth. That means we have to visit every node of the tree until we find the treasure.

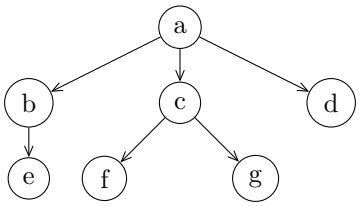
Many problems in real life can be seen as labyrinths in the sense that we have to make a series of decisions, each time choose between multiple options.

Therefore, many problems can naturally be represented as trees. Moreover, if we do not have any special knowledge (e.g., a map leading to the treasure), the only thing we can do is systematically explore all nodes of the tree.

That is straightforward in principle, but we have to decide in which order we explore the nodes. Two strategies are important:

- In Breadth-First Search (BFS), we explore nodes in increasing order of depth: first the root, then the children, then the grandchildren of the root, and so on. We can visualize this as searching top-to-bottom (if the tree is drawn in the usual way with the root at the top). Thus, we search the entire breadth before moving on to deeper nodes.
- In Depth-First Search (DFS), we first explore all descendants of a node n before moving on the siblings of n. We can visualize this as searching left-to-right. Thus, we search as deep as we can before moving on to the siblings.

Consider the tree below. BFS yields abcdefg. DFS yields abecfgd.



BFS has the drawback of back-and-forth movement. For the tree above, we have to go from a to b, back up to a and down to c, back up and down to d, then all the way back to b, so that we can go e, back up all the way to a, down to c again, and so on. DFS is much simpler.

However, it is much more common to have a very high tree (i.e., long branches) than a very wide tree (i.e., lots of branches). This is because we often have many decisions to make, but each decision only has a few options. For example, many games consist of an unlimited number of turns where at each turn we have to choose from a limited number of moves. In those situations, if DFS picks the wrong cild of the root early on, it may have to explore a huge subtree before coming back to pick the right child.

BFS is more balanced and predictable. If we know that the probability of finding a solution becomes smaller at greater depths, BFS makes sure that we explore the most promising nodes first.

Depth-First Search

DFS can be realized quite easily with a recursive function, especially if we use the data structure from Sect. 10.2.1. We use an arbitrary function f as the payload, i.e., a function that is to be called at every node n. For example, f can check if n is the needed solution or do some other work on n.

```
fun DFS[A](n : Tree[A], f : Tree[A] → unit) =
    f(n)
    foreach(n.children, x ↦ DFS[A](x, f))
```

In this variant of DFS, f acts on every node n before it recurses into the children. It is also possible to switch those two, i.e., first recurse into the children, then call $f(n)$.

Breadth-First Search

BFS is a bit more complicated. One way to do it is to use a queue that stores all nodes that we have already seen but not acted on yet. That way we can avoid the back-and-forth movement.

```
fun BFS[A](n : Tree[A], f : Tree[A] → unit) =
    needToVisit := new Queue[Tree[A]]()
    enqueue(needToVisit, n)
    while !empty(needToVisit)
        n := dequeue(needToVisit)
        f(n)
        foreach(n.children, x ↦ enqueue(needToVisit, x))
```

Here in every iteration of the loop, we process the next node n (dequeue) and then put its children at the end of the queue. That way all children of n are guaranteed to be processed before any grandchildren of n.

The above BFS-algorithm is interesting because we can easily turn it into a DFS-algorithm: all we have to do is use a stack instead of a queue. That way all descendants of n are processed before anything else.

10.4.2 Min-Max Algorithm

Many games can be represented as trees. Consider a 2-player game in which the players alternate taking turns. At every turn, a player has to choose among multple moves. We assume there is no luck (e.g., no dice-rolling) and no hidden information (e.g., no bluffing).

We can represent all possible courses of the games in a single tree as follows:

- Every node represents a turn.
  - root: initial state
  - nodes of even depth (including root): turn of player 1
  - nodes of odd depth: turn of player 2
  - leafs: terminal states (when the game is over)
- For every node n, the children of n are the possible moves in that turn.
- Every branch represents a possible course of the game.

For leafs l, let $score(l) \in \mathbb{Z}^\infty$ represent the outcome:

- ∞: player 1 wins
- positive values: player 1 is ahead
- 0: draw
- negative values: player 2 is ahead
- $-\infty$: player 2 wins

Thus, player 1 wants to maximize the result, player 2 wants to minimize it.

The min-max algorithm builds the entire tree by exploring all possible courses of the game. Let *State* be the type of game states. We assume some basic functions $isTerminal : State \to bool$ and (for terminal states) $result : State \to \mathbb{Z}^\infty$ that represent the rules of the game.

Let us assume we have built the tree $game : Tree[State]$. Then we can call the minmax algorithm with $minxmax(game, 0)$ to aggregate the results of the terminal states:

```
fun minmax(current : Tree[State], depth : ℕ) : ℤ^∞ =
    state := current.data
    if isTerminal(state)
        result(state)
    else
        childResults := map(current.children, n ↦ minmax(n, depth + 1))
        if even(depth)
            max(childResults)
        else
            min(childResults)
```

If $minxmax(game, 0) = \infty$, then player 1 has a perfect strategy to win every game. Correspondingly for player 2. If $minxmax(game, 0) = 0$, then both players have a perfect strategies to hold a draw.

In practice, the tree is usually far too big to build. Therefore, instead of obtaining the result at terminal states, we must estimate the result at cut-off. For example, at depth 6, we estimate the current score using heuristic function $State \to \mathbb{Z}^\infty$.

This is a basic design used in artificially intelligent computer players for many games. Many optimizations are needed to obtain strong players.

Chapter 11

Set-Like Data Structures

11.1 Specification

The set $Set[A]$ contains the finite subsets of A. It is countable if A is countable.
Sets can be mutable or immutable. The main operations for immutable sets are:

function	returns	effect
$contains(x \in Set[A], a \in A) \in bool$	*true* iff $a \in x$	none
$insert(x \in Set[A], a \in A) \in Set[A]$	$x \cup \{a\}$	none
$delete(x \in Set[A], a \in A) \in Set[A]$	$x \setminus \{a\}$	none

The main operations for mutable sets are:

function	returns	effect
$contains(x \in Set[A], a \in A) \in bool$	*true* iff $a \in x$	none
$insert(x \in Set[A], a \in A) \in unit$	nothing	$x := x \cup \{a\}$
$delete(x \in Set[A], a \in A) \in unit$	nothing	$x := x \setminus \{a\}$

In both cases, we often need operations for combining and comparing sets. We only consider the immutable case:

function	returns	effect
$equal(x \in Set[A], y \in Set[A]) \in bool$	*true* iff $x = y$	none
$union(x \in Set[A], y \in Set[A]) \in Set[A]$	$x \cup y$	none
$inter(x \in Set[A], y \in Set[A]) \in Set[A]$	$x \cap y$	none
$diff(x \in Set[A], y \in Set[A]) \in Set[A]$	$x \setminus y$	none

Equality is listed explicitly here because it can be very complex. For most data structures such as the ones for lists and trees, equality is straightforward. This may or may not be the case for data structures for sets.

11.2 Data Structures

The complexity of data structures for sets is usually measured in terms of the size n of the set.

11.2.1 Bit Vectors

Design

If A is finite with $|A| = m$, an easy data structure for $Set[A]$ are bit vectors of length m such as $Array[bool](m)$. Given such a vector a, we put $a[i] = true$ to represent that i is in the set.

Complexity

We can implement *insert* and *delete* easily in $\Theta(1)$. We can also implement *equal*, *union*, *inter*, and *diff* easily in $\Theta(m)$.
A major drawback is the memory requirement: We need $\Theta(m)$ for each $x : Set[A]$, which is only feasible for small m.

11.2.2 List Sets

Design

If A is much larger than the sets to be represented, a better data structure for $Set[A]$ is $ListSet[A]$. It represents the set $\{a_1, \ldots, a_n\}$ as the list $[a_1, \ldots, a_n]$. Thus, it represents sets as lists without repetition.
The operations on $ListSet[A]$ are defined in the same way as for $List[A]$ with one exception: the $insert(x, a)$ operation does nothing if x already contains a.

Complexity

If n is the size of the $ListSet$, the operations *contains*, *insert*, and *delete* takes $\Theta(n)$. However, higher-level operations like building a set with n elements step-by-step by calling *insert* n times requires n insertions and thus costs $\Theta(n^2)$.
Moreover, these operations require calls to the equality on A. For example, to implement $contains(x, a)$, we have to compare a to every element of x. That may be easy, e.g., if $A = int$. But it can be arbitrarily costly if A is more complex data structure itself.
For equality, union, intersection, and difference of x and y, we may have to compare every element of x with every element of y. So it may take $O(|x| \cdot |y|)$.
These operations quickly become too costly for large subsets of A.

11.2.3 Hash Sets

Design

Hash sets try to combine the advantages of bit vector and list sets. The key parameter is a function $hash : A \to \mathbb{Z}_m$. This is called the hash function.
hash has two purposes:

- The set A is supported by a finite, managably small set $\mathbb{Z}_m$. That makes it feasible to use arrays of length m.
- The equality operation on A is supported by the $O(1)$ equality on $\mathbb{Z}_m$. To check $a = a'$, we first check $hash(a) = hash(a')$. If false, we know $a \neq a'$; otherwise, we call the usual equality on A. That minimizes the number of times equality on A is called.

Of course, the function *hash* will usually not be injective. A **collision** is a pair $x, y \in A$ such that $hash(x) = hash(y)$.
A good hash function should be fast and rarely have collisions. An (unrealistically) ideal hash function runs in $O(1)$ and the probability of $hash(x) = hash(y)$ is $1/m$. Those two properties work against each other: For example, it is easy to be fast by always returning 0, but that has maximally many collisions. Vice versa, it is easy to minimize collisions by choosing *hash* carefully, but then *hash* may be very expensive to compute. Thus, hash functions must make a trade-off.
For a fixed hash function $hash : A \to \mathbb{Z}_m$, the data structure $HashSet[A]$ is given by

type $HashSet[A] = Array[ListSet[A]](m)$
fun $insert(h : HashSet[A], a : A) = \{insert(h[hash(a)], a)\}$
fun $delete(h : HashSet[A], a : A) = \{delete(h[hash(a)], a)\}$

The elements of the array are called hash **buckets**. Thus, the bucket for i contains all elements of the set whose hash value is i.

Complexity

If n is the size of the subset of A, the sets $h[0], \ldots, h[m-1]$ have average size n/m. Thus, *contains*, *insert*, and *delete* take n/m on average. *equal*, *union*, *inter*, and $diff$ are similarly sped up.

Asymptotically, hash sets do not beat list sets because they only spped up by a constant factor. But that constant factor is a critical improvement.

The speed up is bigger if m is bigger. However, the memory requirement increases linearly with m: Even the empty subset requires $\Theta(m)$ space and $\Theta(m)$ time to initialize that space.

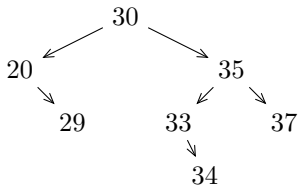
Optimized data structures for hash sets can dynamically choose m in order to find a good trade-off. Often users of the *HashSet* data structure can choose the value of m. That can help if they know in advance how big the subset is going to get and what kind of operations will be called.

11.2.4 Binary Search Trees

Design

If we have a total order $\leq$ on A, we can use binary trees to realize $Set[A]$. The idea is that the nodes of the tree hold the elements of the set, and every node n splits a range of values into two subranges: all left descendants of n have smaller and all right descendants have greater values than n.

Example 11.1. We represent the set $\{20, 29, 30, 33, 34, 37\} \subseteq \mathbb{N}$ as the following binary tree over $\mathbb{N}$:



Now we can locate an element efficiently by traversing just one branch of the tree.

Note how a binary search tree has to distinguish between the left and the right subtree even if there is only one subtree (e.g., for the nodes for 20 and 33). In those cases, we have to use a dummy node as the omitted left child. These are usually labeled with *null* or $\perp$ and omitted when drawing the tree.

The formal definition is as follows:

Definition 11.2 (Binary Search Trees)**.** $BST[A, O]$ is the subset of $Tree[A^?]$ containing only full binary trees satisfying the following properties:

- All leafs are labeled with the value $\perp$; all other nodes are labeled with elements of A.
- For every non-leaf node n:
  - All values in the left subtree of n are strictly smaller than the one at n.
  - All values in the right subtree of n are strictly greater than the one at n.

The leafs hold dummy values as placeholders for elements that we may want to insert later. When drawing binary search trees, we always omit them. The reason for labeling all leafs with $\perp$ is a technicality: we already need dummy leafs for omitted left children; then it can be simpler to just define that all leafs are dummy nodes.

Complexity

Ideally, the binary tree is perfect or nearly perfect. That happens when every node is labeled with the median of all the values among its descendants.

In that case, the height of the tree is logarithmic in the size n of the set. Then *contains*, *insert*, and *delete* can be implemented in $O(\log n)$.

But a series of random insertions and deletions may make the tree arbitrarily imperfect. In the worst case, the tree degenerates to a list. Therefore, a binary search tree must be rearranged from time to time.

This can be done from scratch in one big operation. For example, we can do a depth-first traversal of the tree to obtain a sorted list of all elements. Then we can create a perfect binary tree of height $\log n$ and put the elements into it. Overall that takes $\Theta(n)$.

Alternatively, we can rearrange the tree incrementally. Here we try to make minor changes to the tree after every insertion or deletion. To keep the tree near perfect. One way to do that is to use red-black trees (see Sect. 11.2.5).

11.2.5 Red-Black Trees

Red-black trees are a sophisticated variant of binary search trees. They guarantee $\Theta(\log n)$ cost for insertion and deletion.

11.2.6 Characteristic Functions

[1]

For every set $x : Set[A]$, we can define its characteristic function $\texttt{Char}_x : A \to bool$ by

$$\texttt{Char}_x(a) = contains(a, x)$$

In mathematics, the inverse is true as well: for every function $f : A \to bool$, we can define the set $\{x : A | f(x)\}$. However, we can usually not implement that set in programming languages. That is easy to see: bit vectors, list sets, hash sets, or binary search trees can only ever represent *finite* sets. But functions $f : A \to bool$ can easily represent *infinite* sets as well. For example,

$$(x : string) \mapsto true \quad : \quad string \to bool$$

is the characteristic function of the infinite set of all strings.

Therefore, characteristic functions are a representation of sets that is more expressive than all of the above. If we want to represent infinite sets, characteristic functions are the only option.

All set operations on characteristic functions can be implemented in $O(1)$ with the following exceptions:

- *contains* must call the characteristic function. There is no way to predict how long that will take.
- It is impossible to implement *equal* correctly if A is infinite.
- Operations that require finiteness such as computing the size of a set cannot be implemented.

[1] This section was not part of the course.

Chapter 12

Graph-Like Data Structures

After lists, and trees, graphs are the most important data structure in computer science. In fact, just like lists are a special of trees, trees are a special of graphs.

Data structures for lists and trees are of course used to represent lists and trees. But they are also used to represent other data. For example, we can represent a set as a list (Sect. 11.2.2) or as a tree (Sect. 11.2.5) or a list as a tree (Sect. 9.6). That is because choosing the more complex data structure (i.e., a tree instead of a list) can allow for more efficient algorithms.

Data structures for graphs on the other hand are almost exclusively used to represent graphs. That is because they are rather difficult to work with. But they are needed because graph-structured data occurs very frequently.

Practical examples of graph-structured data include

- Social networks: nodes are given by people and edges by the social connection relation. These relations may be asymmetric (e.g., the follows-relation of twitter), which leads to directed graphs, or symmetric (e.g., the are-friends-with-each-other relation of facebook), which leads to undirected graphs.
- Roads: nodes are given by cities and intersections and edges by roads between them. Flight routes, subway systems, etc. work accordingly.
- Utilities supply networks for water, electricity, internet, etc.: nodes are given by power plants/switchboards/hubs/etc. and households and edges by pipes/cables/etc.
- Mazes and other explorable territories: nodes are given by intersections and edges by paths.
- Games: generalizing the tree intuition from Sect. 10.4.2, nodes are given by states and edges by steps/moves/developments.
- Neighborhood relations: nodes are given by countries or other territories and edges by the share-a-border-with-each-other relation.

In fact, binary function and binary relations are the most ubiquitous complex mathematical objects. Algorithms are our primary representation method for the former and graphs for the latter.

12.1 Specification

Definition 12.1 (Graph)**.** A **graph** is a pair $G = (N, E)$ such that E is a binary relation on N.
If E is symmetric, G is called **undirected**, otherwise **directed**.

The set N is usually but not necessarily finite.

Like for trees, there are many definitions to talk about the parts of a graph:

Definition 12.2 (Parts of a Graph)**.** Consider a graph $G = (N, E)$.
An element $n \in N$ is called a **node** or a **vertex**. An element $(m, n) \in E$ is called an **edge** from m to n. It is also called an **incoming edge** of n and an **outgoing edge** of m.
For every node n, the number of incoming edges is called the **in-degree** of n, and the number of outgoing edges is called the **out-degree** of n. If G is undirected graph, incoming and outgoing edges are the same, and we simply speak of the **degree** of n.

A **path** from a_0 to a_n is a list $[a_0, \ldots, a_n] \in N^*$ such that there is an edge from a_{i-1} to a_i for $i = 1, \ldots, n$.
n is called the **length** of the path. If $n = 0$ (and thus $a_0 = a_n$), the path is called **empty**. If there is a path from a_0 to a_n, then a_n is called **reachable** from a_0.
A **cycle** is a non-empty path from a to itself. If G has (no) cycles, it is called **(a)cyclic**.
Let us write $\overline{G}$ for the undirected graph $(N, E \cup E^{-1})$ in which all edges go both ways. Then G is called **connected** if all nodes in $\overline{G}$ are reachable from each other.
A **clique** is a subset C of N such that there is an edge from every $a \in C$ to every other $b \in C$. G is called **complete** if N is a clique.

Visualization A good intuition to think of graphs is to imagine the nodes as places and the edges as streets between them. In a directed graph, all edges are one-way streets.
All concepts about graphs also have very intuitive visual aspects:

	Visual Intuition	
Concept	undirected	directed
node	point	
edge from a to b	line from a to b	arrow from a to b
incoming edge of a		arrow pointing at a
outgoing edge of a		arrow pointing away from a
b reachable from a	we can walk from a to b along edges	... in arrow direction
path from a to b of length n	a walk from a to b in n steps	... in arrow direction
weight[1] of an edge	cost intuition: length of the line capacity intuition: width of the line	
complete	we can walk everywhere in 1 step	... in arrow direction
cycle	we can walk in a circle	... in arrow direction
connected	graph can be drawn in one stroke	

Reachability Relation Many graph properties are just rephrasings of or closely related to relation properties. Most importantly:

Theorem 12.3 (Reachability)**.** *For every graph G, the relation "b is reachable from a" is*
- *reflexive and transitive*
- *symmetric iff G is undirected*
- *anti-symmetric iff G is acyclic*

Proof. Exercise. □

Labeled Graphs Like for trees, graphs are only useful for computation, if we can store data in them. Contrary to trees, we often need to store data in the nodes *and* the edges.

Definition 12.4 (Labeled Graph)**.** A A-B-**labeled** graph is a triple of
- a graph $G = (N, E)$
- a function $nodeLabel : N \to A$
- a function $edgeLabel : E \to B$

$Graph[A, B]$ is the set of A-B-labeled graphs.

The most important special case arises when the nodes are not labeled (i.e., we put $A = unit$) and the edges are labeled with numbers, i.e., $B = \mathbb{Z}$:

[1] See below for weighted graphs.

Definition 12.5. A **weighted** graph is a *unit-B*-labeled graph where B is any set of numbers. The label of an edge is called its **weight**.

Most often B is $\mathbb{N}$, but $\mathbb{Z}$, $\mathbb{R}$, or $\mathbb{Z}^\infty$, etc. are also common.
There are two important applications of weighted graphs that use different interpretations of the weights:

- Cost intuition: The weight of an edge is the cost of moving along the edge. For example, if the nodes represent cities and the edges flight routes, the weight can be the distance.
- Capacity intuition: The weight of an edge is the capacity for moving objects along the edge. For example, if the nodes represent cities and the edges flight routes, the weight can be the number of flights per day.

Correspondingly, we define:

Definition 12.6. Consider a weighted graph. We write $weight(i, j)$ for the weight of the edge from i to j.
We make $weight : N \times N \to \mathbb{N}^\infty$ a total function by using a default value whenever there is no edge from i to j:
- A **cost-weighted** graph uses the default $weight(i, j) = \infty$.
- A **capacity-weighted** graph uses the default $weight(i, j) = 0$.

In a cost-weighted graph, the **cost of a path** is the sum of the weights of all edges.
In a capacity-weighted graph, the **capacity of a path** is the minimal weight of any edge in it.

The intuition behind the default values is that if there is no edge from i to j,
- cost intuition: it is impossible to go from i to j, i.e., which corresponds to infinite cost,
- capacity intuition: it is impossible to move objects from i to j, which corresponds to empty capacity.

12.2 Data Structures

We consider only graphs $G = (N, E)$ where $|N| = m$, i.e., is finite. We number the elements of N (in any order) so that we can assume $N = \mathbb{Z}_m$.
Moreover, we consider only graphs whose edges are labeled with weights from set B. Note that an unlabeled graph can be seen as the special case where $B = unit$.
Graphs are among the trickiest data structures to design because it is difficult to represent the set E efficiently. There are a number of frequently-needed operations, whose complexity depends on the data structure. We may want to obtain
- $edge(G, i, j)$: for two nodes i, j, the edge from i to j (if any),
- $outgoing(G, i)$: for a node i, the list of outgoing edges,
- $incoming(G, i)$: for a node i, the list of incoming edges,
- $edges(G)$: an iterator over all edges.

If we have one of those operations, we can compute the others—but not necessarily efficiently.

12.2.1 Adjacency Matrix

Design An often useful representation is via a matrix, called the **adjacency matrix** of G. It stores the entire function that maps two nodes to their edge in a single object.

Definition 12.7 (Adjacency Matrix)**.** The adjacency matrix of an unlabeled graph G is the matrix $Adj \in bool^{mm}$ where $Adj_{ij} == true$ iff there is an edge from i to j in G.

Adjacency matrices have the nice property that we can multiply them. Here matrix multiplication is computed using conjunction and disjunction instead of multiplication and addition:

Definition 12.8. For $X, Y \in bool^{mm}$, we define $(X \cdot Y)_{ik} := \bigvee_{j=0,\dots,m-1} X_{ij} \wedge Y_{jk}$.

This is useful because:

Theorem 12.9. *If Adj is the adjacency matrix of G, then $(Adj^n)_{ij}$ iff there is a path of length n from i to j in G.*

This is advantageous because it lets us compute all paths of length n efficiently by computing Adj^n, e.g., using square-and-multiply (Sect. 4.1.3). Moreover, in an acyclic graph, there are only finitely many paths. Thus, we eventually have $Adj^n = Adj^{n+1} = \dots$, at which point we have computed all paths.
If the edges are labeled with weights from B, we can use an adjacency matrix $W \in B^{mm}$ such that W_{ij} is the weight of the edge from i to j. If there is no edge, we use a default value, e.g., ∞, 0, or *null*.

Complexity A drawback of the adjacency matrix is that its size $|N|^2$. Moreover, for an undirected graph, half the space is wasted because we always have $Adj_{ij} = Adj_{ji}$.
$edge(G, i, j)$ takes $\Theta(1)$ (assuming we store the matrix efficiently using arrays).
$outgoing(G, i)$ and $incoming(G, i)$ must pull out a row or column from the adjacency matrix. That takes $\Theta(1)$ or $\Theta(m)$, depending on how we store the arrays.
$edges(G)$ takes $\Theta(1)$ because we can use the iterator of the array.

12.2.2 Adjacency Lists

Design For graphs with many nodes and few edges, it is better to store adjacency lists. Those are the lists of outgoing edges for each node:

Definition 12.10 (Adjacency List)**.** For an unlabeled graph, the adjacency list of a node i is the sorted list $L_i \in List[\mathbb{Z}_m]$ of all j such that there is an edge from i to j in G.
The adjacency list–representation of G consists of an list $[(0, L_0), \dots, (m-1, L_{m-1})]$ pairing every node with its adjacency list.

If the edges are labeled with weights from A, the adjacency list $L_i \in List[\mathbb{Z}_m \times A]$ contains pairs (j, w) where W is the weight of the edge from i to j. We do not need default values because we can simply omit those j for which there is no edge.

Complexity The size of the adjacency list–representation is $|N| + |E|$, which is usually much smaller than $|N|^2$.
$edge(G, i, j)$ takes $\Theta(d)$ where d is the maximal number of outgoing edges of any node. d is usually much smaller than $|E|$.
$outgoing(G, i)$ takes $\Theta(1)$.
$incoming(G, i)$ is inefficient. We have to check each node for an edge into i and collect the results.
$edges(G)$ takes $\Theta(m)$ because we have to concatenate m iterators.

12.2.3 Opposite Adjacency List

Design The adjacency list stores for every node the list of *outgoing* edges.
For directed graphs, we can alternatively store the list of incoming edges. That is the same as storing the adjacency lists for the graph in which all edges are flipped.

Complexity All results are opposite to the previous case.

12.2.4 Redundant Adjacency Lists

Design If we want to be able to do both, it is best to store both adjacency lists for each node.

Complexity The size of the representation is now twice as big as when using a single adjacency list.
$edge(G, i, j)$ takes $\Theta(d)$ as before.
$outgoing(G, i)$ and $incoming(G, i)$ take $\Theta(1)$.
$edges(G)$ takes $\Theta(m)$ because we have to concatenate m iterators.
The main drawback of this data structure is that it is more difficult to implement because we have to keep the adjacency lists in sync. Every time we add or remove an edge from i to j, we have to change two lists, namely $incoming(G, j)$ and $outgoing(G, i)$.

12.3 Important Algorithms

Unless mentioned otherwise, we use a connected finite directed graph $G = (N, E)$, whose edges are labeled with natural numbers. We assume $N = \mathbb{Z}_m$.

12.3.1 Reachability

We work with unlabeled graphs in this section.

Problem Given a start node $n \in N$, our goal is to explore all nodes that are reachable from n.
This problem of exploring all reachable nodes comes up all the time in practice. In a maze, it is a way to find the exit. In a game, it means to analyze a particular situation in the game.

Algorithm We use a data structure for which *outgoing* is efficient. Then we start at n are recursively call *outgoing* until we reach no further nodes.
Like for trees, there are two typical options: DFS and BFS. Contrary to trees, we have to watch for cycles: if G has a cycle and we do not keep track of which nodes we have found already, we would never terminate. Therefore, we use two auxiliary data structures:

- a set $reachable : List[N]$ where we store all nodes in the order in which we found them and which we return eventually
- a queue $horizon : List[N]$ where we store all nodes that we have found already but whose outgoing edges we have not yet looked at

For BFS, the algorithm looks as follows:

```
fun BFS(G : Graph, start : N) : List[N] =
    reachable : List[N] = Nil
    horizon : Queue[N] = new Queue[N]()
    enqueue(horizon, start)
    while !empty(horizon)
        i = dequeue(horizon)
        reachable = prepend(i, reachable)
        foreach(outgoing(G, i), j ↦
            if !contains(reachable, j)∧!contains(horizon, j)        the critical check to avoid cycles
                enqueue(horizon, j)
        )
    reverse(reachable)
```

Like for trees, we obtain a DFS exploration algorithm if we use a stack instead of a queue.

Correctness The postcondition is that BFS return the nodes reachable from *start*. As a loop invariant we use that

- every node reachable from *start* or from a node in *reachable* is
  - an element of *reachable* or
  - reachable from some element of *horizon*,

  and

- all nodes in *reachable* or *horizon* are reachable from *start*.

As a termination ordering we use $|N| - length(reachable)$. Indeed, reachable grows in every iteration of the loop and (because it never contains duplicates) becomes at most $|N|$.

Complexity Let $r < |N|$ be the number of reachable nodes and e be the number of edges between reachable nodes. Then BFS takes $\Theta(r + e)$ because every reachable node is visited once and every edge is traversed once.

12.3.2 Minimal Spanning Tree

We work with undirected graphs in this section.

Problem A **spanning tree** for a graph G is a subgraph of G that is a tree and includes all nodes of G. In other words, it is a subset of edges of G such that G becomes a tree.

Spanning trees are closely related to the exploration of reachable nodes: they provide a minimal set of edges needed to reach all reachable nodes. For example, when planning to roads or laying cables a spanning tree can provide the minimal amount of connections needed to reach all households.

We can capture the minimality condition with the following observations. A spanning tree has $|N|-1$ edges because every node but the root has exactly one incoming edge. It is not possible to use fewer edges because any connected subgraph of G that includes all nodes must have at least $|N|-1$ edges: 1 edge can connect two nodes; any additional edge can connect only one additional node.

The problem becomes more interesting if the edges of G are cost-weighted. Then the **weight of a subgraph** T of G is the sum of the weights of the edges in T. Then our goal is not only to find *some* spanning tree but a minimal one, i.e., a spanning tree with the least possible weight.

For example, when planning to build roads, the cost of an edge could be the distance (more generally: the financial cost of building the road) between two points. A minimal spanning tree minimizes the overall cost of building the roads.

Algorithm The BFS and DFS exploration algorithms immediately yield algorithms to find *some* spanning tree. All we have to do is to store the nodes we find in a $reachable : Tree[N]$ instead of $reachable : List[N]$.

It is less obvious how to find a minimal spanning tree of a cost-weighted graph. One of the most well-known solutions is Kruskal's algorithm:

```
                                                                    G = (N, E)
fun Kruskal(G : Graph) : Set[E] =
    edges := list of edges e ∈ E, sorted by increasing weight
    sot := new Set[E]()
    foreach(edges, e ↦ if (isSetOfTrees(sot ∪ {e})) {insert(sot, e)})
    sot
```

The algorithm returns the set of edges that make up a minimal spanning tree. Here the function $isSetOfTrees(S : Set[E]) : bool$ has the following postcondition: if $S \subseteq E$ is a set of edges of G, then $isSetOfTrees(S)$ is true if the subgraph of G containing only the edges in S is a set of trees.[2]

Correctness and Complexity This is an example of a greedy algorithm. We discuss its correctness and complexity in Ch. 21.

12.3.3 Cheapest Path

In this section, we interpret all edge weights as costs. We think of it as the cost of moving along the edge, where the cost represents the physical distance or any kind of resource (money, gas, effort, amount of material, etc.) that must be expended. Recall that the cost of a path is the sum of the costs of its edges.

[2]Such a graph is called a **forest**.

When comparing paths, the literature often uses the words length/short/long, whereas we will use the words cost/cheap/expensive. We do that to avoid confusing the cost of a path with its number of edges, both of which are often called the *length*. The cheapest path does not necessarily contain the least number of edges: a detour along cheap edges can lead to lower cost than the direct path along expensive edges. For example, direct flights usually cost more money than multi-leg flights, and the cheapest flight route is usually not a direct flight.

Problem We are interested in finding the cheapest paths. This has obvious applications in logistics, navigation, and similar situations. Whenever we move in any kind of network, we prefer taking the cheapest path.

Note there may be multiple different paths with the same cost. So technically, we are looking for *a* cheapest path, not *the* cheapest path.

We write $Cost(p)$ for the cost of a path and $Cost(i, j)$ for the cost of the lowest cost of any path from i to j.

There are multiple variants of the cheapest path problem:

- for fixed nodes i and j, find a cheapest path from i to j,
- for a fixed start node i, find cheapest paths to any node j,
- for a fixed end node j, find cheapest paths from any node i,
- find cheapest paths for all pairs of node i and j.

Clearly, all variants can be reduced to each other. But that might not be efficient: finding the cheapest paths for all pairs at once can be much faster than the total time of doing so individually for each pair. That is because cheapest paths are related, especially:

- if p is a cheapest path, then so is any sub-path of p
- if p and q are cheapest paths from i to j and j to k, then $Cost(i, k) \leq Cost(p) + Cost(q)$.

Algorithm We look at the special case where we have a fixed start node i and want to find the cheapest path to every node. It is easy to see that these paths form a spanning tree of G with root i.

One well-known solution is Dijkstra's algorithm. It is also a greedy algorithm: we start with the root only, and each iteration adds to the tree whichever node is cheapest to reach.

```
                                                                  G = (N, E), m = |N|
fun Dijkstra(G : Graph, start : N) : Array[List[N]] =
    cheapest := new Array[List[N]](m)          cheapest[j] is the cheapest known path from start to j
    foreach(N, n ↦ cheapest[n] := null)                                  no path known yet
    cheapest[start] := [start]                                             empty path

    cost := new Array[ℕ^∞](m)                                  cost[j] is the cost of cheapest[j]
    foreach(N, n ↦ cost[n] := ∞)                                         no path known yet
    cost[start] := 0                                                       cost of empty path

    rest := new Set[N]()
    foreach(N, n ↦ insert(rest, n))                          remaining nodes to add to spanning tree
    while !empty(rest)
        i := a node i ∈ rest with minimal value of cost[i]
        delete(rest, i)
        foreach(outgoing(G, i), j ↦
                                                          check where we can go next after going to i
            c := cost[i] + weight(i, j)                                  cost if we go to j via i
            if c < cost[j]                                          cheaper than best known path
                cheapest[j] := append(cheapest[i], j)                          use the path via i
                cost[j] := c
        )
    cheapest
```

Correctness Termination is straightforward using the termination order $length(rest)$.

Partial correctness is harder. We can easily establish the loop invariant that $cheapest[j]$ (if not $null$) is a path from $start$ to j whose cost is $cost[j]$. That is easy to see.

Moreover, with a little thinking we see that for every node that is reachable from *start*, *cheapest*[j] will not be *null* at the end.

It remains to show that these paths are in fact the cheapest paths. The following additional loop invariant is the key:

- for nodes $j \notin rest$: *cheapest*[j] is a cheapest path from *start* to j
- for nodes $j \in rest$: *cheapest*[j] is cheapest among those paths from *start* to j whose intermediate nodes are not in *rest*.

Because *rest* is empty at the end, this yields the desired result.

Because *rest* contains all nodes initially, the invariant trivially holds before the loop.

It remains to show that it is indeed a loop invariant.

Complexity It is easy to see that everything outside the while-loop takes $\Theta(|N|)$.

The while-loop is run $|N|$ times. But it is not so obvious how long the body of the while-loop takes:

- The algorithm does not clarify how to find i. To find it efficiently, we can use a priority queue, which takes at most $O(\log|rest|) \subseteq O(\log|N|)$.
  So the first part of the while-loop takes $|N|O(\log|N|)$.
- We need a special priority queue where we change the priorities (the values *cost*[i]) from time to time. We know that takes at most $O(\log|rest|) \subseteq O(\log|N|)$ each time.
- The body of the *foreach* command is run once for every outgoing edge of i. It is unclear how many such edges there are. But we can merge the analysis with the while-loop: Overall the body of the *foreach* is run once for every outgoing edge of every node, i.e., once for every edge.
  Every time we have to allow for $O(\log|N|)$ because we may change a priority.
  Thus, the second part of the while-loop takes $|E|O(\log|N|)$ overall.

Adding everything up, we obtain $O((|N|+|E|)\log|N|)$ as an upper bound for the worst case time complexity. The average time complexity is lower because of the if-statement: we do not have to change a priority every time.

The worst-case time complexity becomes lower if we use better-optimized data structures for the priority queue.

Other Algorithms There are also important algorithms for finding all cheapest paths between any nodes at once. For example, the Floyd-Warshall algorithm.

12.3.4 Greatest Flow

In this section, we interpret all edge weights as capacities. We assume that $G = (N, E)$ has two distinguished nodes *source* and *sink*. Such a graph is called a **flow network**.

Intuitively, we think of the edges as pipes, and our goal is to make a liquid flow from the source to the sink.

Flows Intuitively, a flow says how much is flowing along an edge.

A flow from *source* to *sink* maps every $e \in E$ to a number $f(e) \in \{0, \dots, weight(e)\}$ such that

- Nothing flows into the start node: $f(e) = 0$ for incoming edges of *source*.
- Nothing flows out of the end node: $f(e) = 0$ for outgoing edges of *sink*.
- What flows into a node must flow out of it and vice versa: for all nodes n except for *source* and *sink*

$$\Sigma_{e \in incoming(n)} f(e) = \Sigma_{e \in outgoing(n)} f(e)$$

From that we can prove that

$$\Sigma_{e \in outgoing(source)} f(e) = \Sigma_{e \in incoming(sink)} f(e)$$

and that number is called the capacity $Cap(f)$ of f. It describes how much is flowing through the network from the source to the sink.

Problem Now our goal is to find f such that $Cap(f)$ is as big as possible. This is called the greatest flow, maximal flow, or max-flow problem.

For example, the flow network describes goods that have to be shipped from one place to another via a variety of paths, the maximal flow maximizes the amount of goods that are shipped. Similar examples abound in logistics.

Correspondence between Paths and Flows Recall that paths from *start* to *end* are lists of nodes $[p_0, \ldots, p_n]$ such that

- The path begins at *start*: $p_0 = e$.
- The path ends at *end*: $p_n = end$.
- There are edges that connect each node to is successor: for each $i = 0, \ldots, n-1$, there is an edge from p_i to $p_i + 1$.

Then the definitions of paths from *start* to *end* and flows from *start* to *end* are elegantly similar. Moreover, the cost of a path corresponds to the capacity of path, and the cheapest path problem corresponds to the greatest flow problem.

Algorithm A well-known algorithm for the greatest flow is the Ford-Fulkerson algorithm. It is also a greedy algorithm.

The basic idea is to

- start with the empty flow where $f(e) = 0$ for all edges e,
- repeatedly look for some path from *source* to *sink* along which every edge still has free capacity and increase the flow accordingly.

However, the algorithm also uses a subtle trick to allow for reducing the flow along some edge.

That has the effect of adjusting sub-optimal choices from previous iterations.

We omit the details.

Chapter 13

Function-Like Data Structures

13.1 Specification

Functions implement the set $A \to B$, also written B^A. Two fundamentally different ways of treating functions must be distinguished.

Algorithmic functions are defined by algorithms that transform the input of type A into the output of type B. They are always immutable, and the only operation on them is function application:

function	returns
$apply[A, B](f \in A \to B, a \in A) \in B$	$f(a)$

Tabular functions are finite sets of pairs $(a, b) \in A \times B$ indicating that a should be mapped to b. They are usually partial functions (unless A is finite and the function contains a pair for every $a \in A$). Apart from possibly being partial, they are a special case of algorithmic functions: We can define $f(a)$ by searching for the pair (a, b) in f and returning b.

Tabular functions can be mutable or immutable. We only give the mutable case, which is the most useful in practice. The main operations are

function	returns	effect
$isDefined[A, B](f \in A \to B, a \in A) \in \mathbb{B}$	true if f is defined for a	none
$apply[A, B](f \in A \to B, a \in A) \in B^?$	$f(a)$ if defined	none
$keys[A, B](f \in A \to B) \in List[A]$	values for which f is defined	none
$update[A, B](f \in A \to B, a \in A, b \in B) \in unit$	nothing	change f to also map a to b

13.2 Data Structures for Algorithmic Functions

13.2.1 As a Primitive Type

All functional programming languages include a primitive feature to form the function type $A \to B$. In fact, the existence of this type is the defining characteristic of being a functional language. Individual functions that map every $x : A$ to $t(x)$ are built using λ-abstraction $(x : A) \mapsto t(x)$.

Most untyped languages also provide an untyped λ-abstraction as $x \mapsto t(x)$.

13.2.2 As a Class

In non-functional object-oriented languages, we can define the type $A \to B$ as a class:

```
                                                    A → B
abstract class Function[A, B]()
    fun apply(x : A) : B
                                                (x : A) ↦ t(x)
```

```
new Function[A, B]()
    fun apply(x : A) : B =
        t(x)
```

Thus, object-oriented languages are in some sense also functional, except that the syntax for λ-abstractions is very awkward.

13.3 Data Structures for Tabular Functions

Data structures for tabular functions $TableFun[A, B]$ can be easily realized by using data structures for $Set[A \times B]$. If the programming language does not offer product types or in order to optimize better, the corresponding data structures for sets can be modified to handle functions. We give some examples.

13.3.1 List Functions

Design It is straightforward to define the data structure $ListFun[A, B]$ of list-backed functions in analogy to $ListSet[A]$. It represents the function that maps a_i to b_i as the list $[(a_1, b_1), \ldots, (a_n, b_m)]$.

Complexity The complexity properties are essentially the same as for $ListSet[A]$. $isDefined$, *apply*, and *update* are linear in the size of the function because we have to find the corresponding pair in the list.

13.3.2 Hash Tables

Design For a hash function $hash : A \to \mathbb{Z}_m$, the data structure $HashTable[A, B] = Array[ListFun[A, B]](m)$ holds functions backed by a hash set.
For a hash table $h : HashTable[A, B]$, each $h[i]$ holds the list of pairs (a, b) for which $hash(a) = i$.

Complexity The complexity properties are essentially the same as for $HashSet[A]$.

13.3.3 Binary Search Trees

Design Binary search trees can be easily generalized to represent a function f. The basic idea is to make a binary search tree for the set of keys. Each node in the tree then stores not only the key a but also the corresponding value $f(a)$.

Complexity The complexity properties are essentially is the same as when representing sets.

Chapter 14

Product-Like Data Structures

There are three important data structures for the Cartesian product $A \times B$.

14.1 Positional Products

The simplest data structure directly implements the Cartesian product. There are multiple ways to do that.

14.1.1 Primitive Products

Functional programming languages usually include product types as a primitive feature. This includes the types $A * B$ in SML or (A, B) in Scala.

14.1.2 Products as an Inductive Data Type

Products can also easily be defined as an inductive data type:

```
data Product[A, B] = Pair(A, B)

fun ProjectLeft[A, B](p : Product[A, B]) : A =
  match p
    Pair(a, b) ↦ a

fun ProjectRight[A, B](p : Product[A, B]) : B =
  match p
    Pair(a, b) ↦ b
```

14.2 Labeled Products: Structures/Records

Very often, we need the Cartesian product of more than two types. Moreover, it can become tedious to keep track of the individual components. For example, if we work with the product $\mathbb{Z} \times \mathbb{Z} \times \mathit{string} \times \mathbb{N}$, code like $p.3$ to access the third component can become hard to read.
Labeled products introduce a name for every component and use those names instead of numbers to access the components.

14.2.1 Specification

Let $l_1, \ldots, l_n$ be some fresh unique names.
The set $P = \{l_1 : A_1, \ldots, l_n : A_n\}$ is isomorphic to the set $A_1 \times \ldots \times A_n$.

Its elements are $\{l_1 = a_1, \ldots, l_n = a_n\}$ for $a_i \in A_i$ with the convention that the components can be given in any order.

Given $u \in P$, we write $u.l_i$ for the value of the component l_i in u.

14.2.2 Data Structures

Data structures for labeled products are part of almost every programming language. However, they occur with very different names and notations.

For example, they are called

- records in SML,
- structs in C,
- classes in object-oriented languages,
- objects in Javascript,
- dictionaries in Python.

In untyped languages like Javascript and Python, of course the type $\{l_1 : A_1, \ldots, l_n : A_n\}$ does not exist. But the values $\{l_1 = a_1, \ldots, l_n = a_n\}$ and $u.l_i$ work as in typed languages.

In object-oriented languages, there is often no special data structure for labeled products. Instead, labeled products are simply a special case of classes as in

$\{l_1 : A_1, \ldots, l_n : A_n\}$

abstract class $P()$
  val $l_1 : A_1$
  $\vdots$
  val $l_n : A_n$

$\{l_1 = a_1, \ldots, l_n = a_n\}$

new $P()$
  val $l_1 := a_1$
  $\vdots$
  val $l_n := a_n$

14.3 Recursive Products: Classes

Often we want products to be recursive in the sense that the fields of the product refer to the whole product. This is a typical application of classes.

We have seen examples already, e.g., in the data structures for lists. For example, we can understand $List[A]$ as the product of A (for the head) and $List[A]$ (for the tail).

In the C-family of languages, recursive products are usually realized as structs S that contain pointers to elements of type S. For example, we might use

```
struct IntList {
  int head;
  IntList* tail;
}
```

Chapter 15

Union-Like Data Structures

Union types provide data structures for the disjoint union $A + B$ of sets.
They are dual to product types and behave in very much the same way.

15.1 Positional Unions

The simplest data structure directly implements the disjoint union.

15.1.1 Primitive Unions

Functional programming languages could include disjoint union as a primitive feature. However, this is rare because disjoint unions turn out to be much less practically relevant than products.
Instead, languages almost exclusively define them as inductive data types.

15.1.2 Unions as an Inductive Data Type

Disjoint unions can also easily be defined as an inductive data type:

```
data DisjointUnion[A, B] = InjectionLeft(A) | InjectionRight(B)

fun matchUnion[A, B](u : DisjointUnion[A, B], f : A → C, g : B → C) : C =
  match u
    InjectionLeft(a) ↦ f(a)
    InjectionRight(b) ↦ g(b)
```

15.2 Labeled Unions

Labeled unions define the disjoint union of multiple types.

15.2.1 Specification

Let $l_1, \ldots, l_n$ be some fresh unique names.
The set $U = l_1(A_1)| \ldots |l_n(: A_n)$ is isomorphic to the set $A_1 + \ldots + A_n$.
Its elements are of the form $l_1(a_1)$ or $\ldots$ or $l_n(a_n)$ for $a_i \in A_i$.
Given $u \in U$, we use pattern-matching to distinguish between the n possible cases for u.

15.2.2 Data Structures

Data structures for labeled unions are rarely used.

Instead, they can easily be implemented as special cases of inductive data types as in

$l_1(A_1)| \ldots |l_n(A_n)$

data $U = l_1(A_1) \mid \ldots \mid l_n(A_n)$

15.3 Recursive Disjoint Unions: Inductive Data Types

Finally inductive data types arise as the most general case of disjoint unions. Here the components of the union may already refer to the whole disjoint union.

We have seen examples already, e.g., in the inductive data type for lists. For example, we can understand $List[A]$ as the disjoint union of the unit type (which contains a single element corresponding to the empty list) and the product of A (for the head) and $List[A]$ (for the tail).

Inductive data types are part of all functional programming languages and can be mimicked using classes in object-oriented programming languages (as described in Sect. 5.2).

Chapter 16

Algebraic Data Structures

16.1 Specification

Algebraic data types are a not clearly delineated family of record types. Typically one field of the record is a type, and the other fields are operations on that type.

Many of them represent mathematical theories, in which case their elements represent mathematical structures. Therefore, they tend to come up a lot. But their high level of abstraction leads to them often being neglected or not understood.

Three classes are of major importance: these are the data types based on one type U and

- one binary relation $U \times U \to \mathit{bool}$ such as in graphs, preorders, orders, and equivalence relations,
- one binary function $U \times U \to U$ such as in monoids, groups, and semi-lattices,
- two binary functions $U \times U \to U$ such as in rings, fields, and lattices

and several axioms such as transitivity, associativity, or distributivity.

The most important special cases are specified in Sect. A.1 for a binary relation and in Sect. A.2 for one binary function. We omit the case for two binary functions.

16.2 Data Structures

Apart from the axioms, we can implement algebraic data types very well as polymorphic abstract classes that take U as the type parameter. However, the axioms can usually not be implemented elegantly (except in very advanced programming languages) and must be realized as comments.

16.2.1 One Binary Relation

We already implemented some of them when sorting lists.

For example, the type of orders on U can be realized as

```
abstract class Relation[A]()
    fun rel(x : A, y : A) : bool =
        ...
abstract class Preorder[A]() extends Relation[A]
                                                    rel is reflexive and transitive
abstract class Order[A]() extends Preorder[A]
                                                    rel is anti-symmetric
abstract class TotalOrder[A]() extends Order[A]
                                                    rel is total
```

Individual structures can now be implemented as concrete classes that implement the abstract ones. We have already implemented some concrete orders such as $\leq$ on *int*, $|$ on *int*, or the lexicographic ordering on *string*.

16.2.2 One Binary Function

For example, the type of monoids on U can be realized as

abstract class $BinOp[A]()$
  fun $op(x : A, y : A) : A =$
    ...
abstract class $Monoid[A]()$ **extends** $BinOp[A]$
  fun $e() : A =$
    ...
op is associative and has neutral element *e*

Individual structures can now be implemented as concrete classes that implement the abstract ones.
The following implements some example monoids:

class $Addition()$ **extends** $Monoid[int]$
  fun $op(x : int, y : int) = \{x + y\}$
  fun $e() : A = \{0\}$

class $Multiplication()$ **extends** $Monoid[int]$
  fun $op(x : int, y : int) = \{x * y\}$
  fun $e() : A = \{1\}$

class $Concatenation()$ **extends** $Monoid[string]$
  fun $op(x : string, y : string) = \{x + y\}$
  fun $e() : A = \{""\}$

class $Maximum()$ **extends** $Monoid[\mathbb{N}]$
  fun $op(x : \mathbb{N}, y : \mathbb{N}) = \{$**if** $(x \leq y)\ \{y\}$ **else** $\{x\}\}$
  fun $e() : A = \{0\}$

In each case, we have to check that the axioms (associativity and neutrality) actually hold to show the correctness of the implementations.
For a more complex example, consider the monoid of 2×2 matrices under multiplication:

class $Matrix22(a : int, b : int, c : int, d : int)$

class $Matrix22Multiplication()$ **extends** $Monoid[Matrix22]$
  fun $op(x : \mathbb{N}, y : \mathbb{N}) = \{...\}$
  fun $e() : A = \{$**new** $Matrix22(1, 0, 0, 1)\}$

where **new** $Mat(a, b, c, d)$ represents the matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$.
There are many more monoids that come up all the time. Some examples are

- $\wedge$ and *true* yield a $Monoid[bool]$.
- $\vee$ and *false* yield a $Monoid[bool]$.
- Concatenation and empty list yield a $Monoid[List[A]]$ for every A.
- Union and empty set yield a $Monoid[Set[A]]$ for every A.
- If we can implement the set $Full[A] : Set[A]$ containing all elements of A, then intersection and $Full[A]$ yield a $Monoid[Set[A]]$ for every A.
- Minimum and ∞ yield a $Monoid[\mathbb{N}^\infty]$.
- *gcd* and 0 yield a $Monoid[\mathbb{N}]$.
- Function composition and identity function yield a $Monoid[A \to A]$.
- Concatenation of paths and empty path yield a monoid of paths in a graph. (See Ch. 12 for paths in a graph.)

16.2.3 Two Binary Functions

The types of, e.g., rings and fields are implemented accordingly.

16.3 Important Algorithms

The implementation of algebraic data types and algorithms that

- are used in concrete algebraic structures
- compute new algebraic structures from existing ones

are studied in the field of *computer algebra*. This is essential for mathematical computation.

Additionally, many algorithms can be easily generalized to take an arbitrary structure as their input. For example, sorting can be implemented as a function

$$sort(x : List[int]) : List[int].$$

But it is much more general, equally easy, and essentially as fast to implement a function

$$sort[A](ord : TotOrd[A], x : List[A]) : List[A]$$

that sorts with respect to an arbitrary total order.
We give some examples.

16.3.1 Folding Lists over a Monoid

The function

$$fold[A, B](x : List[A], f : A \times B \to B) : B$$

for lists is important but often confusing.
It becomes much simpler if we consider the special case $A = B$:

$$fold[A](x : List[A], e : A, f : A \times A \to A) : A$$

If we additionally write f as infix, i.e., write $x\, f\, y$ instead of $f(x, y)$, we have

$$fold[A]([a_1, \ldots, a_n], e, f) = a_1\, f\, (a_2\ \ldots\ (a_{n-1}\, f\, (a_n\, f\, e)) \ldots)$$

In general, the bracketing matters. But if f is associative, the bracketing becomes irrelevant. If additionally e is a neutral element for f, then f and e form a $Monoid[A]$. Then we can write $fold$ as

$$monoidFold[A](mon : Monoid[A], x : List[A])$$

$$monoidFold[A](mon, [a_1, \ldots, a_n]) = a_1\, mon.op\, a_2\ \ldots\ a_{n-1}\, mon.op\, a_n\, mon.op\, mon.e$$

In particular, we have

$$monoidFold[A](mon, []) = mon.e \qquad \text{and} \qquad monoidFold[A](mon, [a]) = a.$$

For example,

- $monoidFold[int](\textbf{new}\ Addition(), x)$ is the sum of all elements in x,
- $monoidFold[int](\textbf{new}\ Multiplication(), x)$ is the product of all elements in x,
- $monoidFold[\mathbb{N}](\textbf{new}\ Maximum(), x)$ is the greatest element in x,
- $monoidFold[string](\textbf{new}\ Concatenation(), x)$ is the concatenation of all strings in x.

16.3.2 Square-and-Multiply

We can finally give the square-and-multiply algorithm from Sect. 4.1.3 in full generality.
This should be a function

```
fun sqmult[A](mon : Monoid[A], x : A, n : ℕ) : A =
    ...
```

that satisfies the specification

$$sqmult(mon, x, 0) = mon.e$$

$$sqmult(mon, x, n + 1) = mon.op(sqmult(mon, x, n), x)$$

and whose time complexity is $\Theta(\log_2 n)$.

Exercise 16.1. Implement this function.

Part III

Important Families of Algorithms

Chapter 17

Recursion

17.1 Overview

We speak of recursion if a function calls itself. More generally, a set of functions is recursive if they call each other in a cyclic way.

A recursive algorithm is a good design choice if we can reduce the problem $P(x)$ to the problem $P(x')$ such that x' is (in some sense) smaller than x. Usually, x' is just a little smaller than x, e.g., we might have $x \in \mathbb{N}$ and $x' = x - 1$.

Recursive algorithms have been used in multiple places throughout these notes. In the following, we look at some frequent patterns of recursive algorithms.

Here we only consider the case where $P(x)$ is reduced to slightly smaller instances, e.g., $P(x-1)$. There are other patterns, which we discuss in separate chapters, that may or may not be implemented using recursion. The following gives an overview:

- backtracking: $P(x)$ is reduced to some smaller instance, but we have to try multiple instances before finding the right one.
- divide-and-conquer: $P(x)$ is reduced to multiple much smaller instances, e.g., $P(x/2)$.
- dynamic programming: $P(x)$ is computed by first computing all smaller instances, e.g., $P(0), \ldots, P(x-1)$.
- parallelization: $P(x)$ is reduced to multiple smaller instances, whose solutions are computed in parallel.

17.2 Induction

Induction is the special case of recursion that comes up in conjunction with inductive data types such as natural numbers or lists. It is used to define most operations on these types. We consider some examples.

17.2.1 Natural Numbers

Consider the data type of natural numbers:

data $nat = zero \mid succ(pred : nat)$

An inductive algorithm for $P(n : nat)$ employs a case split on n, usually using pattern-matching:

- if $n = zero$: return $P(zero)$ directly without recursion (the base case),
- if $n = succ(n')$: recursively compute $y = P(n')$, then use y to compute $P(succ(n'))$ without further recursion (step case).

As an example, consider the inductive algorithm for the factorial on the left:

```
fun fact(n : nat) =
    match n
        zero ↦ 1
        succ(n′) ↦
            y := fact(n′)
            y · succ(n′)
```

```
fun fact(n : int) =
    if n == 0
        1
    else
        y := fact(n − 1)
        y · n
```

Essentially the same algorithms can also be used if *nat* is not defined as an inductive type. Instead of pattern-matching, this requires an if-statement. For example, if we use *int* to represent natural numbers, we obtain the algorithm on the right above.

Virtually all arithmetic operations can be elegantly defined as recursive algorithms along these lines.

17.2.2 Lists

Consider the inductive type of lists:

```
data IndList[A] = nil | cons(head : A, tail : IndList[A])
```

It is slightly more complicated than *nat*. But an inductive algorithm for $P(x : IndList[A])$ proceeds in essentially the same way:

- if $x = nil$: return $P(nil)$ directly without recursion (the base case),
- if $x = cons(hd, tl)$: recursively compute $y := P(tl)$, then use hd and y to compute $P(cons(hd, tl))$ without further recursion (step case).

Some examples were already given elsewhere in these notes. The simplest example is the *length* function:

```
fun length(x : IndList[A]) : int =
    match x
        nil ↦ 0
        cons(hd, tl) ↦
            y := length(tl)
            y + 1
```

Here the *cons*-case ignores the value *hd*.

Like for natural numbers, the corresponding recursive algorithms can also be used if lists are not defined as an inductive type.

17.2.3 Binary Trees

Full binary trees are often defined as the following inductive type. If all nodes are labeled with values from A, we obtain

```
data BinTree[A] = Leaf(label : A) | Node(label : A, left : BinTree[A], right : BinTree[A])
```

It is more complicated than $IndList[A]$ because it uses two inductive arguments. But an inductive algorithm for $P(x : BinTree[A])$ proceeds in essentially the same way:

- if $x = Leaf(a)$: compute the result from a directly without recursion,
- if $x = Node(a, l, r)$: recursively compute $y := P(l)$ and $y' = P(r)$, then use a, y, and y' without further recursion.

As an example, consider a function that returns the list of labels in DFS-order:

```
fun labels(x : BinTree[A]) : List[A] =
    match x
        Leaf(a) ↦ cons(a, nil)
```

```
    Node(a, l, r) ↦
        y := labels(l)
        y' := labels(r)
        cons(a, concat(y, y'))
```

Here the *cons*-case ignores the value *hd*.
Like for other types, the corresponding recursive algorithms can also be used if trees are not defined as an inductive type.

17.3 Mutual Recursion

We speak of *mutual recursion* if multiple functions call each other.
A very simple non-trivial example is the *even-odd* recursion:

```
fun even(n : ℕ) : bool =
    if n == 0
        true
    else
        odd(n − 1)

fun odd(n : ℕ) : bool =
    if n == 0
        false
    else
        even(n − 1)
```

Of course, that is very inefficient for large n and only makes sense if the $n \bmod 2$ function is not available.
Termination arguments for sets of mutually recursive functions are more difficult than for single recursive functions. But the basic idea is the same: every recursive call should make the argument smaller in some sense.

17.4 Recursion vs. While-Loops

Technically, recursion is redundant if the programming language allows for while-loops. But recursion is such a versatile concept that it is part of every practical programming language.
The following example shows how a while-loop can be systematically replaced with a recursive function:

```
x := x0
y := y0
z := z0
while C(x, y, z)
    (x, y, z) := Body(x, y, z)
```

```
fun f(x, y, z) =
    if C(x, y, z)
        (x', y', z') := Body(x, y, z)
        f(x', y', z')
    else
        (x, y, z)
f(x0, y0, z0)
```

On the left, x, y, and z are the variables, whose values may change during an iteration of the loop, and $C(x, y, z)$ is the loop condition. *Body* is some piece of code that may use the current values of x, y, and z, and assign new values to them. This is indicated in the line $(x', y', z') := Body(x, y, z)$, which treats *Body* as a function that takes the old values, performs arbitrary additional operations, and eventually returns the new values of the variables.
The right side shows the equivalent recursive function.
Above we use three variables x, y, and z. Any other number works accordingly—the only critical aspect is that all variables whose value may be changed by *Body* must become argument of the recursive function.
The example below uses two mutable variables *result* and i. It computes the factorial of n in the variable *result*. $C(result, i)$ is the formula $i > 0$.

```
result := 1
i := n
while i > 0
    (result, i) := (result · i, i − 1)
```

```
fun f(result, i) =
    if i > 0
        (result′, i′) := (result · i, i − 1)
        f(result′, i′)
    else
        (result, i)
f(1, n)
```

The example may look a bit odd because *Body* is written in a way that emphasizes the correspondence between while-loops and recursions. If we rewrite them separately into more familiar styles, we obtain

```
result := 1
i := n
while i > 0
    result := result · i
    i := i − 1
```

```
fun f(result, i) =
    if i > 0
        f(result · i, i − 1)
    else
        result
f(1, n)
```

Note that the recursive functions that correspond to while-loops have some special properties:

- There is exactly one recursive call, and no other code is executed after it.
- The return value (if any) is not used for further computations.
- The base case returns all function arguments. In order to return anything useful, at least one of the arguments must have the role of collecting the result. In the example, the argument *result* has that role: it is not used in f except for building the result; therefore, the base case can simply return *result*.

Thus, while-loops correspond to a very small class of recursive functions. It is also possible to translate *any* recursive function (even any set of mutually recursive functions) into a while-loops. But the translation is a bit more complicated, and the two translations are not inverse to each other.

17.5 Tail-Recursion

17.5.1 Tail-Recursive Functions

In Sect. 17.4, we encountered a class of recursive functions with special properties.

Concretely, a recursive function f is called **tail-recursive**, if all its recursive calls occur in positions where the result of the recursive call directly becomes the result of the f.

For example, the following is a tail-recursive implementation of the factorial of n:

```
fun f(result, i) =
    if i > 0
        f(result · i, i − 1)
    else
        result
f(1, n)
```

17.5.2 Optimization

If f is tail-recursive, a interpreter/compiler may turn f into the corresponding while-loop, which usually yields much more efficient code. Note that this is one of the rare situations, where the complexity of an algorithm depends on the interpreter/compiler. For example, Java does not do it; many C compilers do.

The reason for the efficiency gain is the following. Both the recursive function and the while-loop must pass information from one iteration to the next. The while-loop uses assignments to mutable variables; because the mutable variables reside in the same memory location for the current and the next iteration, no physical passing of

data is needed. But in a recursive function, the recursive function call (like any other function call) allocates new memory locations and then copies the function arguments into them. This overhead does not change the Θ-class but can still be substantial.

Moreover, many interpreters/compilers allocate only a fixed amount of memory for the stack (see Sect. **??**). For large function arguments, a recursive function may create so many nested function calls that it exhausts the available stack space (causing a *stack overflow error*). Tail-recursive functions are a special case where this danger can be averted by optimizing them into while-loops.

The exact way in which tail-recursion optimization happens is up to the interpreter/compiler. Usually no new stack frame is allocated for the recursive call—instead the current stack frame is reused. That works out because—due to the call being tail-recursive—the variable values stored in the current stack frame will never be used again and can therefore be safely overridden.

17.5.3 The Call Stack and Stack Frames

Most interpreters/compilers use a stack to keep track of the nesting of function calls. This data structure is usually called the **call stack** (or just *the stack*), and its elements are called **stack frames**.

When a function call $f(t_1, \ldots, t_n)$ of a function $f(x_1, \ldots, x_n)$ is processed,

- a new stack frame is created containing at least
  - the variable definitions $x_1 := t_1, \ldots, x_n := t_n$
  - the current program counter (i.e., the position of the next statement to be executed)
- the frame is pushed onto the stack
- execution continues with the body of f.

When f returns, the stack frame of f (which is at the top of the call stack now) is popped from the stack, and execution continues at the position where f was called.

Chapter 18

Backtracking

18.1 Overview

Consider a problem that involves multiple successive choices. A backtracking algorithm always chooses one of the options and proceeds. If that leads to a situation where no further progress is possible, the algorithm reverts all steps since the most recent choice and chooses a different option. The reversal is called **backtracking**.

The typical example is finding the exit of a maze: at any given intersection, we have to choose one of the possible paths. If we ever reach a dead end, we backtrack all steps since the most recent intersection and choose a different path.

Backtracking algorithms make sense if

- we have little or no information to predict the best choice right away,
- after making a choice, we can detect quickly that it was the wrong choice.

If the first condition is not met, we can usually do better by finding the best choice instead of trying out a random choice. If the second condition is not met, the backtracking algorithm degenerates into DFS.

For the maze example, we cannot detect early on that a path leads to a dead end. Therefore, we have to walk all the way to the dead end before we can backtrack. Thus, we end up performing a DFS of the maze (seen as a graph whose nodes are the intersections).

18.2 General Structure

Consider a problem whose solution involves a series of choices among finitely many options. We can represent this as finding a branch in a tree:

- the root is the starting point,
- at every node n, we have to choose one out of the children of n,
- at a leaf, we can test whether we have a solution or not.

Thus, we obtain the following general problem: given a tree $T : Tree[A]$ and a property $solution : List[A] \to bool$, find a branch b of T such that $solution(b)$.

Backtracking works well if we additionally have a test $abort : List[A] \to bool$ such that $abort(b)$ implies that there is no solution that starts with b.

To be efficient, we avoid ever building the entire tree T—otherwise, we would waste the cost-saving effect of $abort$. Then we can give a general backtracking algorithm as

```
fun search(state : List[A]) : Option[List[A]] =
    if (abort(state)) {return None}
    if (solution(state)) {return Some(state)}
    foreach(choices(state), c ↦
        x := search(state + [c])
        if (x ≠ None) {return x}
    )
```

return *None*

Here the argument *state* represents the path from the root to the current node, and *choices*(*state*) yields its children. The algorithm is essentially a DFS-algorithm in the tree T that

- never builds T as a whole,
- stops as soon as the first solution has been found,
- uses *abort* to avoid traversing a subtree.

Example 18.1. A standard example is the 8-queens problem: place 8 queens on a chess-board such that none threatens the other.

Clearly, there has to be exactly one queen in each row. Thus, we can represent every solution as a list $[c_1, \dots, c_8]$ such that c_i is the column coordinate of the queen in row i.

The possible choices are defined by

$$choices(state) \begin{cases} [1,2,3,4,5,6,7,8] & \text{if } length(state) < 8 \\ [] & \text{if } length(state) = 8 \end{cases}$$

i.e., in every row we have to choose one out of 8 columns, and we reach possible solutions after making 8 choices.

The function *abort*(*state*) return *true* if any two queens threaten each other. Because this can be tested efficiently already if $length(state) < 8$, we can use it to backtrack early.

18.3 Constraint Satisfaction Problems

A constraint satisfaction problem (CSP) consists of a set $x_1 : A_1, \dots, x_m : A_m$ of variables and a set $C_1(\vec{x}), \dots, C_n(\vec{x})$ of boolean expressions (called the **constraints**) about the variables.

A typical example is a system of n linear equations in m variables. But it is not required that the constraints are equations.

Many practical problems can be captured as CSPs. This includes many games and puzzles:

Example 18.2 (Sudoku). The Sudoku game is a CSP with $m = 81$ variables $x_{ij} \in \{1, \dots, 9\}$ for $i, j = 1, \dots, 9$. It has at least the 27 constraints of the form $\{x_{i1}, \dots, x_{i9}\} = \{1, \dots, 9\}$ for each row i and correspondingly for each column and each of the 9 $3x3$ squares.

Additionally, each instance of Sudoku has a few constraints of the form $x_{ij} = c$, which fix the values of certain variables.

Some CSPs can be solved efficiently using backtracking algorithms. Step i tries to fill in a value for x_i. After i steps *state* is the list $[v_1, \dots, v_i]$ of currently tried values for $x_1, \dots, x_i$. $choices(state) = A_{i+1}$ returns the possible values for the next variable x_{i+1}. *abort*(*state*) checks if there is some C_j such that $C_j(v_1, \dots, v_i, x_{i+1}, \dots, x_n)$ is false no matter which values are used for $x_{i+1}, \dots, x_n$. *solution*(*state*) evaluates the conjunction of all constraints.

Example 18.3 (SAT). The SAT problem is the special case of CSP where $A_i = bool$ for all i and all C_i are boolean expressions (i.e., using only $\neg$, $\wedge$, and $\vee$). Because there are only 2 choices for each variable and boolean expressions can be evaluated very easily, it can be solved relatively quickly using a backtracking algorithm.

However, even with a good *abort* function, the fastest known SAT algorithms are exponential.

Chapter 19

Divide and Conquer

19.1 Overview

A divide-and-conquer algorithm consists of 3 steps:

1. the input is divided into parts,
2. each part is processed recursively,
3. the results of the parts are combined into the result of the whole.

Divide-and-conquer works whenever a problem of size n can be reduced to multiple smaller subproblems of the same kind.

The simplest special case arises when working on a list that is divided into two parts. *mergesort* from Sect. 5.3.3 is a typical example. Here the *merge* function combines the partial results into the result of the whole.

A more difficult example is Strassen's multiplication algorithm from Sect. 4.3.3. It splits the arguments matrices into 4 parts each and recurses 7 times. Both the inputs of the recursions and the combination of the results are obtained by substantial computations.

19.2 General Structure

19.2.1 Design

Consider a problem with a size parameter $n \in \mathbb{Z}$ (e.g., the length of a list or the dimension of a square matrix).

The following is the rough general shape of a divide-and-conquer algorithm using two constants d and r:

1. argument/input: problem of size n
2. if $n < d$, solve the problem directly
3. otherwise:
   (a) divide: create r subproblems of size n/d
   (b) conquer: recursively solve the subproblems
   (c) return/output: combine the solutions of the subproblems into the solution of the overall problem

Here d stands for *division*: every iteration reduces the size of the problem from n to n/d. Technically, the size of the subproblems is $n \operatorname{div} d$. But we gloss over this problem here and assume that n is a power of d, i.e., $n = d^k$. All results of this section apply to the general case as well.

r stands for *recursion*: every iteration recurses r times. The value of r varies between algorithms:

- Very often we have $r = d$, i.e., every subproblem is solved separately. For example, for mergesort $r = d = 2$.
- But sometimes we do not have to recurse for every subproblem. A common case is $r = 1$, i.e., we can identify a specific subproblem that is sufficient to solve. For example, for binary search, we know which sublist has the needed value and recurse into that one.
- In some algorithms, the subproblems are not *parts* of the original problem. For example, in Strassen's algorithm, the smaller matrices that get multiplied recursively are not submatrices of the original matrix. Here we have $d = 2$ (if we consider the dimension of the matrices to be the size of the problem) and $r = 7$.

19.2.2 Correctness

The correctness must be argued separately in each case.

However, this is usually easy by induction:

- check that the base cases are handled correctly
- assuming the recursive calls yield correct result, check that the divide-and-conquer steps are correct.

19.2.3 Complexity

In many cases the time complexity of a divide-and-conquer algorithm can be obtained using a general method. Let $C(n)$ be the time complexity for input of size n.

Substitution Method Let $div(n)$ and $combine(n)$ be the costs of dividing and combining, and let $f(n) = div(n) + combine(n)$. Then we have

$$C(n) = div(n) + r \cdot C(n/d) + combine(n) = r \cdot C(n/d) + f(n)$$

By recursively substituting this formula into itself, we obtain

$$C(n) = r \cdot C(n/d) + f(n) = r(r \cdot C(n/d^2) + f(n/d)) + f(n)$$

$$= \ldots = r^k \cdot C(n/d^k) + r^{k-1} \cdot f(n/d^{k-1}) + \ldots + r \cdot f(n/d) + f(n)$$

Because the cost of the base cases does not depend on n, we have $C(n) \in O(1)$ for $n < d$. Recalling that $n = d^k$, we get

$$C(n) = r^k \cdot O(1) + r^{k-1} \cdot f(n/d^{k-1}) + \ldots + r \cdot f(n/d) + f(n)$$

Tree Intuition We can visualize the execution of a divide-and-conquer algorithm as a tree:

- the root is the original call
- for each node, the children are the recursive calls
- the leafs are the base cases

Then we get a tree of height k in which every non-leaf node has r children. In total there are r^i nodes at depth i and in particular r^k leafs.

The terms in the above complexity formula now have an intuitive interpretation:

- The terms $r^i \cdot f(n/d^i)$ are the cost of dividing and combining in the r^i nodes at depth i, which process problems of size n/d^i.
- In particular, the term $f(n) = r^0 \cdot f(n/d^0)$ is the cost of dividing and combining at the root.
- The term $r^k \cdot O(1)$ is the cost of base cases at the r^k leafs.

Thus, the conquering cost at each node is given by its subtrees. The total cost of each node is obtained by summing the values in its subtree, and the overall cost of executing the algorithm is the sum of the costs at all nodes.

Mathematical Preliminaries Before we continue we recall two general formulas that we will use:

- the geometric series for $q \neq 1$
$$\sum_{i=0}^{k-1} q^i = \frac{1-q^k}{1-q}$$
- the following logarithm swap
$$x^{\log_y z} = z^{\log_y x}$$

Solving the Formula It remains to simplify the clunky formula

$$C(n) = r^k \cdot O(1) + r^{k-1} \cdot f(n/d^{k-1}) + \ldots + r \cdot f(n/d) + f(n)$$

into something nicer.

Clearly that depends on f. Let us assume that f is polynomial, i.e., $f \in \Theta(n^c)$ for some $c > 0$. (If f is super-polynomial, the algorithm is probably not very useful anyway.)

Then we get

$$C(n) \in r^k \cdot O(1) + r^{k-1} \cdot \Theta(n^c/d^{c(k-1)}) + \ldots + r \cdot \Theta(n^c/d^c) + \Theta(n^c)$$
$$= \Theta\left(r^k + n^c \cdot \sum_{i=0}^{k-1} \left(\frac{r}{d^c}\right)^i\right)$$

After abbreviating $q = r/d^c$, assuming $q \neq 1$ (We treat the case $q = 1$ below.) and applying the geometric series, and recalling that $k = \log_d n$, that becomes

$$= \Theta\left(r^{\log_d n} + n^c \cdot \frac{1 - q^{\log_d n}}{1 - q}\right)$$

After applying the logarithm swap twice and using $\log_d q = \log_d r - \log_d d^c$, that becomes

$$= \Theta\left(n^{\log_d r} + \frac{n^c - n^{c+\log_d q}}{1 - q}\right) = \Theta\left(n^{\log_d r} + \frac{n^c - n^{\log_d r}}{1 - q}\right)$$

Now we want to drop the constant factor $1 - q$ from inside Θ. But that is only allowed if $1 - q > 0$.
So we eventually have to distinguish three cases:

- $r = d^c$ and thus $q = 1$ and $\log_d r = c$.
  Then we cannot apply the formula for the geometric series. Instead the Σ-sum reduces to $k - 1 = \log_d n - 1$.
  Overall we get
  $$C(n) \in \Theta\left(n^{\log_d r} + n^c(\log_d n - 1)\right) = \Theta(n^c \log_d n)$$
- $r < d^c$ and thus $0 < q < 1$ and $\log_d r < c$.
  Then we can drop $1 - q > 0$ from inside Θ, and we obtain
  $$C(n) \in \Theta(n^c)$$
- $r > d^c$ and thus $q > 1$ and $\log_d r > c$.
  Then $1 - q$ is negative and we can drop it only if we also flip the sign. That yields
  $$C(n) \in \Theta\left(n^{\log_d r} - n^c + n^{\log_d r}\right) = \Theta\left(n^{\log_d r}\right)$$

The so-called Master theorem collects these three cases into a handy cheat sheet:

Theorem 19.1 (Master theorem)**.** *For the time complexity $C(n)$ of a divide-and-conquer algorithm that*
- *requires $f(n) \in \Theta(n^c)$ time for dividing and combining*
- *recurses into r subproblems of size n/d whenever $n \geq d$*

we have

$$\begin{array}{ll} \text{if } r < d^c\text{:} & C(n) \in \Theta(n^c) \\ \text{if } r = d^c\text{:} & C(n) \in \Theta(n^c \log_d n) \\ \text{if } r > d^c\text{:} & C(n) \in \Theta(n^{\log_d r}) = \Theta(r^{\log_d n}) \end{array}$$

The theorem holds independent of our assumption that even if n is not a power of d.
Not surprisingly, all cases require at least $\Theta(n^c)$, which is already the cost of dividing and combining at the root.
If $c > \log_d r$, that is the entire cost, and the cost of the recursions and the base cases can be neglected.
If $c = \log_d r$, the recursion cost yields a logarithmic factor corresponding to the depth of the recursion.
If $c < \log_d r$, the cost of dividing and combining can be neglected and the cost of the $r^{\log_d n}$ base cases dominates the overall cost.

19.3 Examples

19.3.1 Mergesort

For mergeseort from Sect. 5.3.3, we have $r = d = 2$. Dividing and combining requires linear time, i.e., $c = 1$.
The Master theorem indeed yields $\Theta(n \log_2 n)$ for the time complexity of mergesort.

19.3.2 Binary Search

Binary search checks whether a sorted list x of length n contains the value a. The algorithm uses $d = 2$ and $r = 1$.
The base case for $n < d$ is easy. The divide steps splits x into two parts of length $n/2$. The conquer step uses the property of being sorted to determine whether x is the lower or the upper half and recurses only for that one. No combination of results is needed.
The overall time complexity is $\Theta(\log_2 n)$.

19.3.3 Karatsuba's Multiplication of Polynomials

Problem

We want to multiply two polynomials $p(x) = p_m X^m + \ldots + p_1 X + p_0$ and $q(x) = q_m X^m + \ldots + q_1 X + q_0$ with integer coefficients $p_i, q_i \in \mathbb{Z}$. The size n of the problem is the number of coefficients per polynomial, i.e., $n = m + 1$.
Without loss of generality, we assume that $n = 2^k$, i.e., $m = 2^k - 1$. (If the polynomials have a different degree, we can simply add 0-coefficients to increase m.)

Algorithm

Karatsuba's divide-and-conquer algorithm uses $d = 2$ and $r = 3$.

Data Structure For the implementation, we need a data structure for polynomials. The easiest choice is to use the list of coefficients. So we assume that every polynomial is a list $[p_m, \ldots, p_0] \in List[\mathbb{Z}]$.
Addition/subtraction of two polynomials can be implemented easily as component-wise addition of the elements in the lists.
Note that two polynomials are equal if they only differ in initial 0-coefficients. We have two options:

- forbid lists that start with 0
- make two polynomials equal if they only differ in initial 0s

Neither option is essential for the algorithm. However, it is convenient to pick the second option: that way we can easily add initial 0s to adjust the size of a polynomial.

Base Case If $n = 0$, both polynomials are the 0-polynomial, and their product is again the 0-polynomial. If $n = 1$, both polynomials are integers, and we use plain integer multiplication: $pq = [p_0][q_0] = [p_0 q_0]$.

Idea To understand the key idea, let us first look at the special case $n = 2$. Then we have $pq = [p_1, p_0][q_1, 1_0] = [p_1 q_1, p_1 q_0 + p_0 q_1, p_0 q_0]$.
The naive computation takes 4 multiplications and 1 addition. But we can compute the result cleverly using only 3 multiplications and 4 additions:

$$pq = [a, b - a - c, c] \qquad \text{where } a = p_1 q_1,\ b = (p_1 + p_0)(q_1 + q_0),\ c = p_0 q_0$$

Because multiplication is much more complex than addition (addition is linear, multiplication is not), this is preferable even though we need additional additions.

Divide Let $n' = n/2 = (m + 1)/2$, i.e., $m = 2n' - 1$.
We split p into two lists $p^u = [p_m, \ldots, p_{n'}]$ and $p^l = [p_{n'-1}, \ldots, p_0]$ of length n'. We split q into q^u and q^l accordingly.

Now (as polynomials) $p(X) = p^u X^{n'} + p^l$ and $q(X) = q^u X^{n'} + q^l$.

Conquer We recursively multiply 3 pairs of polynomials of size n':

$$a = p^u q^u \qquad b = (p^u + p^l)(q^u + q^l) \qquad c = p^l q^l$$

Combine We combine the results a, b, and c as follows:

$$pq = aX^{2n'} + (b - a - c)X^{n'} + c$$

Correctness

The correctness follows immediately from the construction of the algorithm.
We just have to check the mathematics of the divide and the combine step.

Complexity

This is left as an exercise.

19.3.4 Associative Folding

Problem

Consider folding over a monoid from Sect. 16.3.1.

Algorithm

We give a divide-and-conquer algorithm for it using $d = r = 2$.

```
fun monoidFold[A](mon : Monoid[A], x : List[A]) : A =
    if empty(x)
        mon.e
    else
        n := length(x)
        i := n div 2
        lower := [x_0, ..., x_{i-1}]
        upper := [x_i, ..., x_{n-1}]
        lowerFold := monoidFold(mon, lower)
        upperFold := monoidFold(mon, upper)
        mon.op(lowerFold, upperFold)
```

Correctness

The correctness is straightforward. The key insight is that associativity allows us to bracket the monoid operations any way we want, e.g.,

$$x_0\ mon.op\ \ldots\ mon.op\ x_{i-1}\ mon.op\ x_i\ mon.op\ \ldots mon.op\ x_{n-1} =$$

$$(x_0\ mon.op\ \ldots\ mon.op\ x_{i-1})\ mon.op\ (x_i\ mon.op\ \ldots mon.op\ x_{n-1})$$

Complexity

Let us assume that we use arrays to split the lists in constant time. Combining only requires one monoid operation. So the complexity of dividing and combining is in $O(1)$ and thus $c = 0$.
Using $r = d = 2$ and $c = 0$, we have $r > d^c$ and $\log_d r = 1$. Thus, the Master theorem yields $C(n) \in \Theta(n)$.
Thus, the divide-and-conquer algorithm has the same complexity as the naive algorithm. This is not surprising because all we do is change the bracketing. The number of occurrences of *mon.op* remains the same, i.e., we have to apply the monoid operation the same number of times.
So not every divide-and-conquer algorithm yields an improvement.

Chapter 20

Parallelization and Distribution

20.1 Overview

Multi-Threading Multi-threading executes multiple parts of a program at the same on the same machine. These parts are called **threads**.
Ideally, this uses a separate CPU for each thread to maximize the speed-up. But there can also be more threads running in parallel than there are CPUs. In that case, the threads must be scheduled.

Speed-Up The speed up is the gain in time complexity if parallel execution is taken into account.
Formally, let $C^k(n)$ be the time complexity for input of size n if k CPUs are available for parallel execution. In particular, $C^1(n) = C(n)$. Then the speedup for k machines is defined by $C^k(n)/C(n)$.

Blocking If multiple threads are present, a thread may block, i.e., it waits until a certain condition is fulfilled. This gives a thread the ability to yield CPU access to other threads.

Fairness Usually there is no guarantee when or in which order threads gain access to a CPU. In extreme cases, a thread may be finished completely before the next thread is started.
Fairness refers to a scheduling that gives each thread some guarantee on getting CPU access eventually.
Parallel algorithms are usually designed independent of fairness: they have to yield correct results no matter how the threads are scheduled.

Distributed Algorithm If the parts of a program are executed on completely different machines, we speak of a distributed algorithm.

20.2 General Structure

20.2.1 Threads

Most mainstream programming languages offer multi-threading. The details vary a lot.
But typically a thread is a class with a method $run() : unit$. When run is called, it returns immediately, and the program continues normally. But from now on, the body of the method run is executed in parallel with the remainder of the program.

20.2.2 Parallel List Operations

List-like data structures are great for parallelization because we often apply the same operation to all elements in a list.
We specify two new operations:

function	returns	effect
below, let $l \in A^*$ be of the form $[a_0, \ldots, a_{l-1}]$		
$parMap[B](l \in A^*, f : A \to B) \in B^*$	nothing	$[f(a_0), \ldots, f(a_{l-1})]$ computed in parallel
$parForeach(l \in A^*, f : A \to unit) \in unit$	$f(a_0), \ldots, f(a_{l-1}$ run in parallel	nothing

20.2.3 Parallel Composition

Let C and D be functions that take no arguments. Then $C|D$ is the command that runs $C()$ and $D()$ in parallel. This can be seen as a special case of a $parForeach$ using the list $[C, D]$.

20.3 Examples

20.3.1 Parallel Depth-First Search

Often we can turn a normal algorithm into a parallel algorithm by replacing map with $parMap$ or $foreach$ with $parForeach$.
For example, we obtain a parallel DFS traversal of a tree as follows:

```
fun DFS[A](n : Tree[A], f : Tree[A] → unit) =
    f(n)
    parForeach(n.children, x ↦ DFS[A](x, f))
```

This executes $f(n)$ in parallel for every node n. Of course, the actual order in which nodes are visited is not entirely predicable anymore. It is still guaranteed that $f(n)$ terminates before $f(d)$ begins for any proper descendant d of n. But it is unpredictable in which order the children of n are processed.

20.3.2 Associative Folding

Consider the associative folding problem from Sect. 19.3.4. The divide-and-conquer algorithm did not lower the complexity.
But we obtain a speedup if we recurse in parallel:

```
fun monoidFold[A](mon : Monoid[A], x : List[A]) : A =
    if empty(x)
        mon.e
    else
        n := length(x)
        i := n div 2
        lower := [x_0, ..., x_{i-1}]
        upper := [x_i, ..., x_{n-1}]
        lowerFold := monoidFold(mon, lower)    |    upperFold := monoidFold(mon, upper)
        mon.op(lowerFold, upperFold)
```

If we have more CPUs than $n = length(x)$, this runs in $\Theta(\log n)$.

Chapter 21

Greedy Algorithms

21.1 Overview

Greedy algorithms are an informal grouping of algorithms characterized by the following property: We have to make a number of choices until we are done, and at each step we choose the most attractive option.

A greedy algorithm emphasizes local optimization over global optimization: at each step it makes the optimal local choice. That may or may not yield a good result overall. For example, eating the cheapest available food every day saves money in the short run (locally optimal) but is not healthy in the long run (globally optimal).

But greedy algorithms are easy to implement and usually very efficient: making a locally optimal choice is usually much easier than making a globally optimal choice. For example, finding the cheapest available food just requires browsing items in the market. But finding out what food is healthy in the long run may require extensive research. Humans have a strong tendency towards making local choices because they require so much fewer mental effort and yield immediate gratification.

21.2 General Structure

Consider a set M with a weight function $w : M \to \mathbb{N}$ (or any other other set of positive numbers). For a set $S \subseteq M$, we define the weight of S by summing the weights of the elements.

$$w(S) = \Sigma_{x \in S} w(x)$$

Moreover, consider a property *Acceptable*, i.e., if $S \subseteq M$, then $Acceptable(S)$ is a boolean.

Our goal is to find an acceptable subset of M with maximal weight.

The generic greedy algorithm proceeds as follows:

```
                                  precondition: elements is the list of all elements of M sorted decreasingly by w
fun greedy[M](elements : List[M], Acceptable : Set[M] → bool) =
    solution := new Set[M]()
    foreach(elements, x ↦ if (Acceptable(solution ∪ {x})) {insert(solution, x)})
    solution
```

Here $solution \cup \{x\}$ is an immutable operation that returns a new set, and $insert(solution, x)$ is a mutable operation that changes *solution*.

We can also use the function *greedy* to find an acceptable subset with *minimal* weight: we simply sort *elements* by *increasing* weight.

21.3 Matroids

To understand when the generic greedy algorithm yields optimal results, we introduce matroids:

Definition 21.1 (Matroid)**.** A **matroid** consists of

- a finite set M
- a property $Acceptable : Set[M] \to \mathbb{B}$ (subsets with this property are called **acceptable**[1])

such that the following holds

- M has at least one acceptable subset.
- *Acceptable* is subset-closed, i.e., subsets of acceptable sets are also acceptable.
- If A and B are acceptable and $|A| > |B|$, then B can be increased to an acceptable set $B \cup \{x\}$ by adding an element $x \in A$ to B. (Thus, we must have $x \in A \setminus B$.)

The third property is the critical one: it guarantees that it does not matter which elements we add to an acceptable set, we always eventually get an acceptable set of maximal size. Thus, local choices (which element to add) can never lead to a dead end. More formally, we can state this as follows:

Theorem 21.2. *We call an acceptable set S that has no acceptable superset $S' \supset S$ a **base**. Then, in a matroid, all bases have the same size.*

Proof. Exercise. □

21.4 General Structure for Matroids

21.4.1 Correctness

Finding a base is always easy: start with the empty set and keep adding elements as long as the resulting set remains acceptable.

Now the matroid property guarantees that all bases have the same size. Thus, it does not matter which elements we add—eventually we get an acceptable set of maximal size.

If we want to find not only an acceptable set of maximal *size* but an acceptable set of maximal *weight*, we simply add the elements in order of weight—that is exactly what the greedy algorithm does. Formally, we have:

Theorem 21.3. *If M and Acceptable form a matroid, then the greedy algorithm finds a base with greatest possible weight.*

The corresponding theorem holds for finding the base with smallest possible weight.

21.4.2 Complexity

The main structure of the greedy algorithm is linear in the number $|M|$ of elements. But we also have to sort the elements once and check *Acceptable* at every step. We know sorting takes $\Theta(|M| \log |M|)$. So if we can check $Acceptable(S)$ in $O(\log |S|)$, the overall run time (including sorting) is in $O(|M| \log |M|)$.

To be efficient, we usually implement the acceptability check in a greedy algorithm slightly smartly:

```
class Solution[M]()
    fun insert(m : M) =
        ...
    fun acceptableWith(m : M) =
        ...

precondition: elements is the list of all elements of M sorted decreasingly by w
fun greedy[M](elements : List[M]) =
    solution := new Solution[M]()
    foreach(elements, x ↦ if (solution.acceptableWith(x)) {solution.insert(x)})
    solution
```

[1] The literature calls them *independent*, but *acceptable* is a more intuitive name for greedy algorithms.

Here $S.acceptableWith(x)$ is specified as follows:

- precondition: $Acceptable(S)$
- postcondition: if $S.acceptableWith(x)$, then $Acceptable(S \cup \{x\})$

This works because the greedy algorithm only needs to check $Acceptable(S \cup \{x\})$, and only if S is already known to be acceptable.

$S.acceptableWith(x)$ can often be implemented much faster than $Acceptable(S \cup \{x\})$ because:

- We do not have to copy the set S to build $S \cup \{x\}$.
- The acceptability check can use the information that S is already acceptable.

21.5 Examples

21.5.1 Kruskal's Algorithm

Kruskal's algorithm from Sect. 12.3.2 is a simple example of a greedy algorithm.

Correctness To show that it is correct, we only have to show that it operates on a matroid.

We use the following matroid:

- The set M is the set of edges of the graph $G = (N, E)$.
- A set $S \subseteq E$ is acceptable if the graph (N, S) is a set of trees.

We have to prove the matroid properties:

- There is an acceptable set. For example, $(N, \varnothing)$ is a graph where every node is a tree by itself.
- Subsets of acceptable sets are acceptable. Taking an edge away from a tree splits it into two trees. Thus, removing edges from a set of trees again yields a set of trees.
- For the critical third property, assume that (N, A) and (N, B) are sets of trees such that A contains more edges than B. We have to find an edge $x \in A \setminus B$ that we can add to B. We pick any $x \in A$ that connects two nodes that are not in the same tree in (N, B). Then $(N, B \cup \{x\})$ is a set of trees again. We only have to check that such an x exists: If there were no such x, the trees in A and B would consist of the same nodes; but then A cannot have more edges then B because the number of edges in a tree is already fixed by the number of nodes.

Thus, we immediately know that Kruskal's algorithm is correct.

Complexity We have $Acceptable = isSetOfTrees$.

To improve efficiency, we implement an appropriate data structure for the class $Solution$. This is indeed possible in $O(\log |S|)$. Then we obtain $\Theta(|E| \log |E|)$ as the overall run time of Kruskal's algorithm.

The idea behind the implementation is that $S.acceptableWith(x)$ only has to check that x connects two nodes from different trees. By cleverly storing the set of trees, we can check that in $O(\log |S|)$.

21.5.2 Dijkstra's Algorithm

Because the term *greedy algorithm* is not defined precisely, not every algorithm that has a greedy flavor is a special case of the matroid algorithm.

A counter-example is Dijkstra's algorithm from Sect. 12.3.3.

21.5.3 Scheduling with Deadlines and Penalties

Problem We are given n tasks. Each task takes the same amount of time (e.g., 1 day), and all tasks have to be done separately without overlap (e.g., it takes n days in total). We want to find the best order in which to do all tasks.

Each task has a deadline: Task i must be completed by time $D(i) \in \mathbb{N}$ for $0 < D(i) \leq n$. Otherwise, we have to pay a penalty $w(i)$ for $w(i) > 0$.

Our goal is to minimize the total penalty we have to pay (i.e., the sum of all $w(i)$ for all i that are done after $D(i)$). Equivalently, we want to maximize the total penalty that we do not have to pay (i.e., the sum of all $w(i)$ for all i that are done by $D(i)$).

Design We define a matroid as follows:

- M is the set $\{1, \ldots, n\}$ representing the n tasks.
- A set $S \subseteq M$ is acceptable if it is possible to do all tasks in S on time.

We instantiate the generic greedy algorithm using this matroid and weight function w. The algorithm returns the optimal set S_{opt}.

We now schedule the tasks as follows: First we schedule all tasks in S_{opt} in some way that they are all done on time. (This is possible because the greedy algorithm returns an acceptable set.) Then we schedule all other tasks (all of which will be late because the greedy algorithm returns a maximally big acceptable set) arbitrarily.

Example 21.4. Consider 5 tasks as follows

task i	1	2	3	4	5
deadline $D(i)$	3	1	3	2	2
penalty $w(i)$	10	20	5	25	15

Ordered by decreasing penalty the tasks are $[4, 2, 5, 1, 3]$. The greedy algorithm proceeds as follows:

considered task	decision	known schedule
4	insert	?4???
2	insert	24???
5	skip	24???
1	insert	241??
3	skip	241??

Here the right column tracks not only the set *solution* but also the best schedule for its elements. That makes it easy to implement $acceptableWith$ by checking whether there is still a slot available before the deadline.

The algorithm return $S_{opt} = \{1, 2, 4\}$. Finally, we insert the remaining tasks at the end obtaining the schedule 24135. We have to pay the penalty $w(3) + w(5) = 20$.

Correctness We only have to show that the above structure is indeed a matroid. This is left as an exercise.

21.5.4 Knapsack Problem

Problem The knapsack problem is one of the most commonly studied algorithmic problems.

Fix a set $M = \{1, \ldots, k\}$ of objects. Each $i \in M$ has size $s(i)$ and value $w(i)$. Now given a knapsack of capacity n, we want to find the subset of M with biggest total value that fits into the knapsack.

Design We call a subset of $S \subseteq M$ acceptable if it fits into the knapsack, i.e., if $\Sigma_{i \in S} s(i) \leq n$.

Then we can instantiate the generic greedy algorithm. However, the greedy algorithm does not yield the optimal result.

For example, consider $n = 10$ and $k = 3$ objects

object i	1	2	3
size $s(i)$	8	6	4
value $w(i)$	8	6	4

The greedy algorithm takes object 1 first even though the best fit is achieved by using objects 2 and 3.

Indeed, the above structure does not satisfy the matroid properties: Consider the sets $A = \{2, 3\}$ and $B = \{1\}$. Both are acceptable and $|A| > |B|$. But there is no $x \in A$ such that $B \cup \{x\}$ is acceptable.

Chapter 22

Dynamic Programming

22.1 Overview

As indicated in the introduction to Ch. 17, there is a certain duality between induction and dynamic programming. For example,

- An inductive algorithm for $P(n)$ recurses into $P(n-1), \ldots, P(0)$ (in that order), then propagates the results back. All $n+1$ function calls are open at the same time.
- A dynamic programming algorithm computes the results in the order $P(0),\ldots, P(n-1), P(n)$ and stores them in a table. This is usually done with a for-loop or similar construct.

Dynamic programming can be superior to induction in multiple situations:

- $P(n)$ needs multiple previous results. For example, for the Fibonacci numbers we need $P(n-1)$ and $P(n-2)$. An inductive algorithm would be exponential because we compute the same values multiple times. A dynamic programming algorithm remains linear because all results for smaller inputs are available in a table.
- Recursion causes overhead. Especially for large inputs and non-tail-recursive functions, induction requires the creation and removal of many stack frames. A dynamic programming algorithm in a for-loop can be much faster even if both solutions are linear.
- For functions that are called a lot on many different inputs, it may make sense anyway to store the entire table of results.

But dynamic programming has the disadvantage of storing all results for smaller inputs in a table. That increases the space complexity, whereas induction usually only requires constant space.[1] This can be substantial overhead if only $P(n)$ is needed or if n is large.

22.2 General Structure

22.2.1 Design

In the simplest case a dynamic programming algorithm looks like this:

```
fun dynamic(n : ℕ) : B =
    results := new Array[B](n)
    for i in 0, ..., n
        r := P(i)                          may use results[0], ..., results[i − 1]
        results[i] := r
    results[n]
```

Here r is the result of computing $P(i)$ using all results for lower values. In this variant of dynamic programming, the table *results* is thrown away after each call to *dynamic*. Alternatively, the table could be kept for later function calls.

[1] This comparison is not quite fair though: recursion (unless tail-recursive) also requires linear space in practice, namely for the frames on the stack. But that space is hidden by the programming language.

However, this is not particularly powerful. A much more powerful class of algorithms arises if the table-variable is not part of the input. Instead of solving the problem $P(n)$ with a table variable $i = 0, \dots, n$, we use an input $a \in A$ and generalize the problem $P(a)$ to a problem $Q(a, i)$ for an auxiliary variable $i = 0, \dots, k$. The dynamic program then iterates over i and eventually returns $P(a) = Q(a, k)$.

Very often this design allows for elegant iteration formulas that compute $Q(x, i)$ from the values $Q(y, j)$ for all $y \in A$ and $j = 0, \dots, i-1$. Consequently, this requires a two-dimensional table storing $Q(x, i)$ for all possible inputs $x \in A$ and all $i = 0, \dots, k$.

The general algorithm becomes

```
fun dynamic(a : A) : B =
    results := new Array[B](|A|, k)
    for i in 0, ..., k
        for x in A
            r := Q(x, i)                    may use results[y, 0], ..., results[y, i − 1] for any y ∈ A
            results[x, i] := r
    results[a, k]
```

Remark 22.1. The technique of changing a problem $P(a)$ to a problem $Q(a, i)$ by

- adding an auxiliary variable i that makes the problem more general
- using a fixed k to recover the original problem as the special case $P(a) = Q(a, k)$

is not specific to dynamic programming.

We find it in many recursive algorithms as well. Examples are quicksort from Sect. 5.3.4 and tail-recursion from Sect. 17.5.

In all cases, the more general problem is — counter-intuitively — easier to solve because the extra generality makes deep properties visible that can be used to reduce problems to smaller subproblems and then design efficient algorithms.

22.2.2 Correctness

The main condition for correctness is that $P(x) = Q(x, k)$ for all $x \in A$. Then we only have to verify that we correctly compute $Q(x, i)$ from all values $Q(y, j)$ with $j < i$.

22.2.3 Complexity

The time complexity is $\Theta(|A| \cdot k \cdot f(|A|, k))$ where $f(|A|, k)$ is the complexity of computing $Q(x, i)$ for $x \in A$ and $i \le k$.

The space complexity is $\Theta(|A| \cdot k)$.

22.3 Examples

22.3.1 Coin Change

Problem Consider a currency whose coins have the denominations $D = \{d_1, \dots, d_n\}$. For example, for Euros, we have $D = \{1, 2, 5, 10, 20, 50, 100, 200\}$ (giving all denominations in cents). Our goal is to find the minimum number $P(n)$ of coins whose denominations add up to n. For example, $P(0) = 0$ (using no coins), $P(1) = P(2) = 1$, $P(3) = 2$, and so on.

Design We do not need an auxiliary variable. We compute $P(i)$ directly from all $P(j)$ for $j < i$ as follows:

$$P(0) = 0 \qquad \text{and} \qquad P(i) = 1 + \min\{P(i - d) \mid d \in D,\, d < i\}$$

Here $1 + P(i - d)$ is the minimum number of coins adding up to n if one of the coins has denomination d.

Plugging this into the general algorithm, we obtain

```
fun coinChange(n : ℕ) : ℕ =
    results := new Array(ℕ)(n)
    for i in 0, ..., n
        r := if (i == 0) {0} else {1 + min{results[i − d] | d ∈ D, d < i}}
        results[i] := r
    results[n]
```

Correctness The correctness follows immediately from the correctness of the formula for $P(i)$.

Complexity The complexity for the minimum operation is $\Theta(|D|)$. So the overall complexity is $\Theta(n \cdot |D|)$.

22.3.2 Cheapest Path

Problem We revisit the problem of finding the cheapest path in a directed, cost-weighted graph $G = (N, E)$. As before, we write $weight(x, y)$ for the weight of the edge from x to y, using 0 if there is no edge. We assume $weight(x, y) > 0$ so that the cheapest path can have at most length $k = |E|$.

Design We use $A = N \times N$, and for $x = (x_1, x_2)$, we want to find the cost of the cheapest path $P(x)$ from x_1 to x_2. Using dynamic programming, we will compute $P(x)$ for all $x \in N \times N$ in one go.
We use the following generalized problem: for $x = (x_1, x_2)$ and $i = 0, \ldots, k$, we say that $Q(x, i)$ is the cost of the cheapest path from x_1 to x_2 that has length at most i. Thus, the length of the path is our auxiliary variable.
This lets us use the following formula for $Q(x, i)$:

$$Q((x_1, x_2), 0) = E_{x_1 x_2} := \begin{cases} \infty & \text{if } x_1 \neq x_2 \\ 0 & \text{if } x_1 = x_2 \end{cases}$$

$$Q((x_1, x_2), i) = \min\{Q((x_1, y), i - 1) + weight(y, x_2) \mid y \in N\} \qquad \text{for } i > 0$$

Here $Q((x_1, y), i - 1) + weight(y, x_2)$ is the cost of

- a path p from x_1 to y of length at most $i - 1$
- the edge from y to x_2.

Because p is chosen optimally each time, the cheapest path from x_1 to x_2 of length at most i is the minimum of over all possible choices for y.
The general dynamic programming algorithm becomes:

```
fun cheapestPath(a_1 : N, a_2 : N) : B =
    results := new Array[ℕ∞](|N|, |N|, k)
    for i in 0, ..., k
        for (x_1, x_2) in N × N
            r := if (i == 0) {E_{x_1 x_2}} else {min{results[x_1, y, i − 1] + weight(y, x_2) | y ∈ N}}
            results[x_1, x_2, i] := r
    results[a_1, a_2, k]
```

Correctness The correctness follows immediately from the correctness of the formula for $Q((x_1, x_2), i)$.

Complexity The complexity of the minimum operation is $\Theta(|N|)$. Thus, the overall complexity is $\Theta(|A| \cdot k \cdot |N|) = \Theta(|N|^3 \cdot |E|)$.

Matrix Formulation We can simplify the dynamic program further by using a smarter technique for computing Q.
Let $N = \{1, \ldots, n\}$ and let $Adj \in (\mathbb{N}^\infty)^{n \times n}$ be the adjacency matrix of G, i.e., $weight(x, y) = Adj_{xy}$.
For $A, B \in (\mathbb{N}^\infty)^{n \times n}$, we write $A \odot B \in (\mathbb{N}^\infty)^{n \times n}$ for the matrix product computed with

- min instead of +
- + instead of ·

Formally, we have

$$(A \odot B)_{ik} = \min\{A_{ij} + B_{jk} \mid j = 1, \dots, n\}$$

Like normal matrix multiplication, $\odot$ is associative. It also has a neutral element, namely the matrix E from above with 0s along the diagonal and ∞ everywhere else.

We can now see that the dynamic program becomes

```
fun cheapestPath(a1 : N, a2 : N) : B =
    results := new Array[(ℕ^∞)^{n×n}](k)
    for i in 0, ..., k
        results[i] := if (i == 0) {E} else {results[i − 1] ⊙ Adj}
    results[k]_{a1,a2}
```

which we can further simplify to

```
fun cheapestPath(a1 : N, a2 : N) : B =
    (Adj^k)_{a1,a2}
```

Because $((\mathbb{N}^\infty)^{n\times n}, \odot, E)$ is a monoid, we can apply square-and-multiply to compute Adj^k in $\log_2 k$ steps. $\odot$ takes $O(|N|^3)$. So the overall complexity becomes $\Theta(|N|^3 \cdot \log_2 |E|)$.

Similar to Strassen's algorithm, we can further improve the algorithm by computing $\odot$ in less than $\Theta(|N|^3)$.

22.3.3 Floyd Warshall Algorithm for the Cheapest Path

An alternative dynamic programming solution is the Floyd-Warshall algorithm.

Given a graph $G = (N, E)$ with $N = \{1, \dots, k\}$, it uses the following related problem: $Q((x_1, x_2), i)$ is the cheapest path from x_1 to x_2 whose intermediate nodes are chosen from $\{1, \dots, k\}$ only.

Dijkstra's algorithm is closely related to this.

22.3.4 Knapsack Problem

Problem We revisit the knapsack problem from Sect. 21.5.4.

For simplicity, we only want to compute the total value of the optimal set of objects, not the set itself. Given a set $S \subseteq O = \{1, \dots, k\}$, we write $s(S)$ for its total size and $v(S)$ for its total value. We want to compute the maximal total value of any set of objects whose total size is below or equal to n, i.e.,

$$K(n) = \max\{v(S) \mid S \subseteq O,\ s(S) \leq n\}$$

Design We simplify the problem even further: Let

$$P(n) = \max\{v(S) \mid S \subseteq O,\ s(S) = n\}$$

It is sufficient to solve $P(n)$ because we can compute

$$K(n) = \max\{P(i) \mid i = 0, \dots, n\}$$

To solve $P(n)$, we use one auxiliary variable $i = 0, \dots, k$ and define $Q(x, i)$ to be the maximum total value of objects of total size x that use only objects $1, \dots, i$. Formally, we have

$$Q(x, i) = \max\{v(S) \mid S \subseteq \{1, \dots, i\},\ s(S) = x\}$$

We obtain the following formula for Q:

$$Q(x, 0) = 0 \qquad \text{and} \qquad Q(x, i) = \max\{Q(x, i-1), Q(x - s(i), i-1) + v(i)\}$$

Here $Q(x, i-1)$ is the value if we do not use object i, and $Q(x - s(i)) + v(i)$ is the value if we do.

As usual, we have $P(n) = Q(n, k)$.

Correctness The correctness follows immediately from the formula for Q.

Complexity The formula for Q takes constant time. So the overall complexity to compute $P(n)$ is $\Theta(n \cdot k)$. The complexity for $K(n)$ is the same because we can compute it from $P(n)$ in linear time.

Chapter 23

Protocols

A protocol is an algorithm that involves multiple parties that exchange messages. The programs implementing the various parties usually run on different machines, and messages are exchanged over a network. Notably, all parties must implement the protocol specification independently, possibly in different programming languages.

A typical example is the HTTP protocol between two parties called **client** and **server**. The client sends an HTTP request to the server, and then the server sends an HTTP response back to the client.

For example, retrieving an HTML page usually uses the HTTP protocol (or its encrypted variant HTTPS). Here the client is the web browser, and the server is the web server.

Chapter 24

Randomization

A randomized algorithm is an algorithm that systematically use randomness as part of their implementation. Typically, the randomness makes the result (or at least the run time) unpredictable.

Randomized algorithms present two theoretical problems:

1. Because the result is not determined by the input, they are technically not algorithms at all.
2. They may not always return a correct result.

Often randomized algorithm are used when it would be too expensive to find a correct solution. Ideally, but not necessarily, if an incorrect result is returned, it is at least approximately correct with respect to some measure.

For example, we could give a randomized variant of quicksort that chooses the pivot element randomly. That algorithm has a high probability of running in $\Theta(n \log n)$.

Randomized algorithms are often used in problems in statistical analysis or scientific simulation where correct solutions are infeasible anyway. Often a correct algorithm is run on a random sample of the input, from which the overall result is estimated.

Chapter 25

Quantum Algorithms

Quantum algorithms use a different physical machine, namely a quantum computer that exploits quantum effects. In the most important idea, the superposition of quantum states is used: that allows performing a computation on multiple inputs at the same time, from which a single result value is obtained through measurement.

Quantum algorithms do not allow computing more, i.e., the definition of computable and decidable is not affected. But they can solve certain problems faster, e.g., we can give polynomial quantum algorithms for some classically super-polynomial problems. That makes them interesting for many application, in particular in cryptography where the non-existence of polynomial solutions is often used as a criterion for security.

For example, a quantum algorithm can be used to test $f(0) = f(1)$ in such a way that f is executed only once. We call f on the superposition of the 0 and 1 states of a quantum bit to compute the superposition of $f(0)$ and $f(1)$. Then we perform a certain quantum transformation that maps superpositions of equal values t0 0 and superpositions of different values to 1. Eventually, we measure the result.

Quantum algorithms were hypothetical until researchers succeeded in building quantum computers in recent years. Existing quantum computers are still small (e.g., 16 bits), and it remains open if and when they can be used cost-effectively for large problems.

Part IV

Appendix

Appendix A

Mathematical Preliminaries

A.1 Binary Relations

A binary relation on a set A is a subset $\# \subseteq A \times A$. We usually write $(x, y) \in \#$ as $x\#y$.

A.1.1 Classification

Definition A.1 (Properties of Binary Relations)**.** We say that $\#$ is ... if the following holds:

- reflexive: for all x, $x\#x$
- irreflexive: for no x, $x\#x$
- transitive: for all x, y, z, if $x\#y$ and $y\#z$, then $x\#z$
- a strict order: irreflexive and transitive
- a preorder: reflexive and transitive
- anti-symmetric: for all x, y, if $x\#y$ and $y\#x$, then $x = y$
- symmetric: for all x, y, if $x\#y$, then $y\#x$
- an order[1]: preorder and anti-symmetric (= reflexive, transitive, and anti-symmetric)
- an equivalence: preorder and symmetric (= reflexive, transitive, and symmetric)
- a partial equivalence: transitive and symmetric
- a total order: order and for all x, y, $x\#y$ or $y\#x$

An element $a \in A$ is called ... of $\#$ if the following holds:

- least element: for all x, $a\#x$
- greatest element: for all x, $x\#a$
- least upper bound for x, y: $x\#a$ and $y\#a$ and for all z, if $x\#z$ and $y\#z$, then $a\#z$
- greatest lower bound for x, y: $a\#x$ and $a\#y$ and for all z, if $z\#x$ and $z\#y$, then $z\#a$

Definition A.2 (Dual Relation)**.** For every relation $\#$, the relation $\#^{-1}$ is defined by $x \#^{-1} y$ iff $y \# x$.
$\#^{-1}$ is called the **dual** of $\#$.

Theorem A.3 (Dual Relation)**.** *If a relation is reflexive/irreflexive/transitive/symmetric/antisymmetric/total, then so is its dual.*

A.1.2 Equivalence Relations

Equivalence relations are usually written using infix symbols whose shape is reminiscent of horizontal lines, such as $=$, $\sim$, or $\equiv$. Often vertically symmetric symbols are used to emphasize the symmetry property.

[1] Orders are also called *partial order*, *poset* (for partially ordered set), or *ordering*.

Definition A.4 (Quotient)**.** Consider a relation $\equiv$ on A. Then

- For $x \in A$, the set $\{y \in A \mid x \equiv y\}$ is called the (equivalence) **class** of x. It is often written as $[x]_{\equiv}$.
- $A/\equiv$ is the set of all classes. It is called the **quotient** of A by $\equiv$.

Theorem A.5. *For a relation $\equiv$ on A, the following are equivalent[2]:*

- *$\equiv$ is an equivalence.*
- *There is a set B and a function $f : A \to B$ such that $x \equiv y$ iff $f(x) = f(y)$.*
- *Every element of A is in exactly one class in $A/\equiv$.*

In particular, the elements of $A/\equiv$

- *are pairwise disjoint,*
- *have A as their overall union.*

Consider a partial equivalence relation $\equiv$ on A. $\equiv$ is not an equivalence because it is not reflexive. However, we can easily prove: if $x \equiv y$, then $x \equiv x$ and $y \equiv y$.
Thus, we have:

Theorem A.6. *A partial equivalence relation $\equiv$ on A is the same as an equivalence relation on a subset of A.*

A.1.3 Orders

Theorem A.7 (Strict Order vs. Order)**.** *For every strict order $<$ on A, the relation "$x < y$ or $x = y$" is an order.*
For every order $\leq$ on A, the relation "$x \leq y$ and $x \neq y$" is a strict order.

Thus, strict orders and orders come in pairs that carry the same information.
Strict orders are usually written using infix symbols whose shape is reminiscent of a semi-circle that is open to the right, such as $<$, $\subset$, or $\prec$. This emphasizes the anti-symmetry ($x < y$ is very different from $y < x$.) and the transitivity ($< \ldots <$ is still $<$.) The corresponding order is written with an additional horizontal bar at the bottom, i.e., $\leq$, $\subseteq$, or $\preceq$. In both cases, the mirrored symbol is used for the dual relation, i.e., $>$, $\supset$, or $\succ$, and $\geq$, $\supseteq$, and $\succeq$.

Theorem A.8. *If $\leq$ is an order, then least element, greatest element, least upper bound of x, y, and greatest lower bound of x, y are unique whenever they exist.*

Theorem A.9 (Preorder vs. Order)**.** *For every preorder $\leq$ on A, the relation "$x \leq y$ and $y \leq x$" is an equivalence.*
For equivalence classes X and Y of the resulting quotient, $x \leq y$ holds for either all pairs or no pairs $(x, y) \in X \times Y$. If it holds for all pairs, we write $X \leq Y$.
The relation $\leq$ on the quotient is an order.

Remark A.10 (Order vs. Total Order). If $\leq$ is a preorder, then for all elements x, y, there are four mutually exclusive options:

	$x \leq y$	$x \geq y$	$x = y$
x strictly smaller than y, i.e., $x > y$	true	false	false
x strictly greater than y, i.e., $x < y$	false	true	false
x and y incomparable	false	false	false
x and y similar	true	true	maybe

Now anti-symmetry excludes the option of similarity (except when $x = y$ in which case trivially $x \leq y$ and $x \geq y$). And totality excludes the option of incomparability.
Combining the two exclusions, a total order only allows for $x > y$, $y < x$, and $x = y$.

A.2 Binary Functions

A binary function on A is a function $\circ : A \times A \to A$. We usually write $\circ(x, y)$ as $x \circ y$.

Definition A.11 (Properties of Binary Functions)**.** We say that $\circ$ is ... if the following holds:

- associative: for all x, y, z, $x \circ (y \circ z) = (x \circ y) \circ z$
- commutative: for all x, y, $x \circ y = y \circ x$
- idempotent: for all x, $x \circ x = x$

An element $a \in A$ is called a ... element of $\circ$ if the following holds:

- left-neutral: for all x, $a \circ x = x$
- right-neutral: for all x, and $x \circ a = x$
- neutral: left-neutral and right-neutral
- left-absorbing: for all x, $a \circ x = a$
- right-absorbing: for all x, $x \circ a = a$
- absorbing: left-absorbing and right-absorbing
- if e is a neutral element:
  - left-inverse of x: $a \circ x = e$
  - right-inverse of x: $x \circ a = e$
  - inverse of x: left-neutral and right-neutral of x

Moreover, we say that $\circ$ is a ... if it is/has:

- semigroup: associative
- monoid: associative and neutral element
- group: monoid and inverse elements for all x
- semilattice: associative, commutative, and idempotent
- bounded semilattice: semilattice and neutral element

Terminology A.12. The terminology for *absorbing* is not well-standardized. *Attractive* is an alternative word sometimes used instead.

Theorem A.13. *Neutral and absorbing element of $\circ$ are unique whenever they exist.*
If $\circ$ is a monoid, then the inverse of x is unique whenever it exists.

A.3 The Integer Numbers

A.3.1 Divisibility

Definition A.14 (Divisibility)**.** For $x, y \in \mathbb{Z}$, we write $x|y$ iff there is a $k \in \mathbb{Z}$ such that $x * k = y$.
We say that y is divisible by x or that x divides y.

Remark A.15 (Divisible by 0 and 1). Even though division by 0 is forbidden, the case $x = 0$ is perfectly fine. But it is boring: $0|x$ iff $x = 0$.
Similarly, the case $x = 1$ is trivial: $1|x$ for all x.

Theorem A.16 (Divisibility)**.** *Divisibility has the following properties for all $x, y, z \in Z$*

- *reflexive: $x|x$*
- *transitive: if $x|y$ and $y|z$ then $x|z$*
- *anti-symmetric for natural numbers $x, y \in \mathbb{N}$: if $x|y$ and $y|x$, then $x = y$*
- *1 is a least element: $1|x$*
- *0 is a greatest element: $x|0$*
- *$\gcd(x, y)$ is a greatest lower bound of x, y*
- *$\mathrm{lcm}(x, y)$ is a least upper bound of x, y*

Thus, $|$ is a preorder on $\mathbb{Z}$ and an order on $\mathbb{N}$.

Divisibility is preserved by arithmetic operations: If $x|m$ and $y|m$, then

- *preserved by addition: $x + y|m$*
- *preserved by subtraction: $x - y|m$*
- *preserved by multiplication: $x * y|m$*
- *preserved by division if $x/y \in Z$: $x/y|m$*
- *preserved by negation of any argument: $-x|m$ and $x| - m$*

gcd *has the following properties for all $x, y \in \mathbb{N}$:*

- *associative:* $\gcd(\gcd(x, y), z) = \gcd(x, \gcd(y, z))$
- *commutative:* $\gcd(x, y) = \gcd(y, x)$
- *idempotent:* $\gcd(x, x) = x$
- 0 *is a neutral element:* $\gcd(0, x) = x$
- 1 *is an absorbing element:* $\gcd(1, x) = 1$

lcm *has the same properties as* gcd *except that* 1 *is neutral and* 0 *is absorbing.*

Theorem A.17. *For all $x, y \in \mathbb{Z}$, there are numbers $a, b \in \mathbb{Z}$ such that $ax + by = \gcd(x, y)$.*
a and b can be computed using the extended Euclidean algorithm.

Definition A.18. If $\gcd(x, y) = 1$, we call x and y **coprime**.
For $x \in \mathbb{N}$, the number of coprime $y \in \{0, \ldots, x - 1\}$ is called $\varphi(x)$. φ is called Euler's **totient function**.

Example A.19. We have $\varphi(0) = 0$, $\varphi(1) = \varphi(2) = 1$, $\varphi(3) = \varphi(4) = 2$, and so on. Because $\gcd(x, 0) = x$, we have $\varphi(x) \leq x - 1$. x is prime iff $\varphi(x) = x - 1$.

A.3.2 Equivalence Modulo

Definition A.20 (Equivalence Modulo)**.** For $x, y, m \in \mathbb{Z}$, we write $x \equiv_m y$ iff $m|x - y$.

Theorem A.21 (Relationship between Divisibility and Modulo)**.** *The following are equivalent:*

- $m|n$
- $\equiv_m \supseteq \equiv_n$ *(i.e., for all x, y we have that $x \equiv_n y$ implies $x \equiv_m y$)*
- $n \equiv_m 0$

Remark A.22 (Modulo 0 and 1). In particular, the cases $m = 0$ and $m = 1$ are trivial again:

- $x \equiv_0 y$ iff $x = y$,
- $x \equiv_1 y$ always

Thus, just like 0 and 1 are greatest and least element for $|$, we have that $\equiv_0$ and $\equiv_1$ are the smallest and the largest equivalence relation on $\mathbb{Z}$.

Theorem A.23 (Modulo)**.** *The relation $\equiv_m$ has the following properties*

- *reflexive: $x \equiv_m x$*
- *transitive: if $x \equiv_m y$ and $y \equiv_m z$ then $x \equiv_m z$*
- *symmetric: if $x \equiv_m y$ then $y \equiv_m x$*

Thus, it is an equivalence relation.

It is also preserved by arithmetic operations: If $x \equiv_m x'$ and $y \equiv_m y'$, then

- *preserved by addition: $x + y \equiv_m x' + y'$*
- *preserved by subtraction: $x - y \equiv_m x' - y'$*
- *preserved by multiplication: $x \cdot y \equiv_m x' \cdot y'$*
- *preserved by division if $x/y \in \mathbb{Z}$ and $x'/y' \in \mathbb{Z}$: $x/y \equiv_m x'/y'$*
- *preserved by negation of both arguments: $-x \equiv_m -x'$*

A.3.3 Arithmetic Modulo

Definition A.24 (Modulus)**.** We write $x \bmod m$ for the smallest $y \in \mathbb{N}$ such that $x \equiv_m y$.
We also write $modulus_m$ for the function $x \mapsto x \bmod m$. We write $\mathbb{Z}_m$ for the image of $modulus_m$.

Remark A.25 (Modulo 0 and 1). The cases $m = 0$ and $m = 1$ are trivial again:
- $x \bmod 0 = x$ and $\mathbb{Z}_0 = \mathbb{Z}$
- $x \bmod 1 = 0$ and $\mathbb{Z}_1 = \{0\}$

Remark A.26 (Possible Values). For $m \neq 0$, we have $x \bmod m \in \{0, \dots, m-1\}$. In particular, there are m possible values for $x \bmod m$.
For example, we have $x \bmod 1 \in \{0\}$. And we have $x \bmod 2 = 0$ if x is even and $x \bmod 2 = 1$ if x is odd.

Definition A.27 (Arithmetic Modulo m)**.** For $x, y \in \mathbb{Z}$, we define arithmetic operations modulo m by

$$x \circ_m y = (x \circ y) \bmod m \qquad \text{for} \qquad \circ \in \{+, -, \cdot\}$$

Moreover, if there is a unique $q \in \mathbb{Z}_m$ such that $q \cdot x \equiv_m y$, we define $x/_m y = q$.

Note that the condition $y|x$ is neither necessary nor sufficient for $x/_m y$ to de defined. For example, $2/_4 2$ is undefined because $1 \cdot 2 \equiv_4 3 \cdot 2 \equiv_4 2$. Conversely, $2/_4 3$ is defined, namely 2.

Theorem A.28 (Arithmetic Modulo m)**.** *For $x, y \in \mathbb{Z}$,* mod *commutes with arithmetic operations in the sense that*

$$(x \circ y) \bmod m = (x \bmod m) \circ_m (y \bmod m) \qquad \text{for} \qquad \circ \in \{+, -, \cdot\}$$

Moreover, $x/_m y$ is defined iff $\gcd(y, m) = 1$ *and*

$$(x/y) \bmod m = (x \bmod m)/_m (y \bmod m) \qquad \text{if} \qquad y|x$$

$$x/_m y = x \cdot_m a \qquad \text{if } ay + bm = 1 \text{ as in Thm. A.17}$$

Theorem A.29 (Fermat's Little Theorem)**.** *For all prime numbers p and $x \in \mathbb{Z}$, we have that $x^p \equiv_p x$.*
If x and p are coprime, that is equivalent to $x^{p-1} \equiv_p 1$.

A.3.4 Digit-Base Representations

Fix $m \in \mathbb{N} \setminus \{0\}$, which we call the base.

Theorem A.30 (Div-Mod Representation)**.** *Every $x \in \mathbb{Z}$ can be uniquely represented as $a \cdot m + b$ for $a \in \mathbb{Z}$ and $b \in \mathbb{Z}_m$.*
Moreover, $b = x \bmod m$. We write $b \operatorname{div} m$ for a.

Definition A.31 (Base-m-Notation)**.** For $d_i \in \mathbb{Z}_m$, we define $(d_k \dots d_0)_m = d_k \cdot m^k + \dots + d_1 \cdot k + d_0$.
The d_i are called **digits**.

Theorem A.32 (Base-m Representation)**.** *Every $x \in \mathbb{N}$ can be uniquely represented as $(0)_m$ or $(d_k \dots d_0)_m$ such that $d_k \neq 0$.*
Moreover, we have $k = \lfloor \log_m x \rfloor$ and $d_0 = x \bmod m$, $d_1 = (x \operatorname{div} m) \bmod m$, $d_2 = ((x \operatorname{div} m) \operatorname{div} m) \bmod m$ and so on.

Example A.33 (Important Bases). We call $(d_k \dots d_0)_m$ the binary/octal/decimal/hexadecimal representation if $m = 2, 8, 10, 16$, respectively.
In case $m = 16$, we write the elements of $\mathbb{Z}_m$ as $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, a, b, c, d, e, f\}$

A.3.5 Finite Fields

In this section, let $m = p$ be prime.

Construction Because p is prime, $x/_p y$ is defined for all $x, y \in \mathbb{Z}_p$ with $y \neq 0$. Moreover, $\mathbb{Z}_p$ is a field.
Up to isomorphism, all finite fields are obtained as n-dimensional vector spaces $\mathbb{Z}_p^n$ for some prime p and $n \geq 1$. This field is usually called F_{p^n} because it has p^n elements. From now on, let $q = p^n$.
The elements of F_q are vectors $(a_0, \dots, a_{n-1})$ for $a_i \in \mathbb{Z}_p$. Addition and subtraction are component-wise, the 0-element is $(0, \dots, 0)$, the 1-elements is $(1, 0, \dots, 0)$.
However, multiplication in F_q is tricky if $n > 1$. To multiply two elements, we think of the vectors $(a_0, \dots, a_{n-1})$ as polynomials $a_{n-1}X^{n-1} + \dots + a_1X + a_0$ and multiply the polynomials. This can introduce powers X^n and higher, which we eliminate using $X^n = k_{n-1}X^{n-1} + \dots + k_1X + k_0$ for certain k_i. The resulting polynomial has degree at most $n - 1$, and its coefficients (modulo p) yield the result.
The values k_i always exists but are non-trivial to find. They must be such that the polynomial $X^n - k_{n-1}X^{n-1} - \dots - k_1X - k_0$ has no roots in $\mathbb{Z}_p$. There may be multiple such polynomials, which may lead to different multiplication operations. However, all of them yield isomorphic fields.

Binary Fields The operations become particularly easy if $p = 2$. The elements of F_{2^n} are just the bit vectors of length n. Addition and subtraction are the same operation and can be computed by component-wise XOR. Multiplication is a bit more complex but can be obtained as a sequence of bit-shifts and XORs.

Exponentiation and Logarithm Because F_q has multiplication, we can define natural powers in the usual way:

Definition A.34. For $x \in F^q$ and $l \in \mathbb{N}$, we define $x^l \in F_q$ by $x^0 = 1$ and $x^{l+1} = x \cdot x^l$.
If l is the smallest number such that $x^l = y$, we write $l = \log_x y$ and call n the **discrete q-logarithm** of y with base x.

The powers $1, x, x^2, \dots \in F_q$ of x can take only $q - 1$ different values because F_q has only q elements and x^l can never be 0 (unless $x = 0$). Therefore, they must be periodic:

Theorem A.35. *For every $x \in F_q$, we have $x^q = x$. If $x \neq 0$, that is equivalently to $x^{q-1} = 1$.*

For some x, the period is indeed $q - 1$, i.e., we have $\{1, x, x^2, \dots, x^{q-1}\} = F_q \setminus \{0\}$. Such an x is called a **primitive element** of F_q. But the period may be smaller. For example, the powers of 1 are $1, \dots, 1$, i.e., 1 has period 1. For a non-trivial example consider $p = 5$, $n = 1$, (i.e., $q = 5$): The powers of 4 are $4^0 = 1$, $4^1 = 4$, $4^2 = 16 \bmod 5 = 1$, and $4^3 = 4$.
If the period is smaller than $q - 1$, x^l does not take all possible values in F_q. In that case $\log_x y$ is not defined for all $y \in F_q$.

Computing x^l is straightforward and can be done efficiently. (If $n > 1$, we first have to find the values k_i needed to do the multiplication, but we can precompute them once and for all.)
Determining whether $\log_x y$ is defined and computing its value is also straightforward: We can enumerate all powers $x, x^2, \dots$ until $x^l = 1$ (in which case the logarithm is undefined) or $x^l = y$ (in which case the logarithm is l). However, no efficient algorithm is known.

A.3.6 Infinity

Occasionally, it is useful to compute also with infinity ∞ or $-\infty$. When adding infinity, some but not all arithmetic operations still behave nicely.

Positive Infinity We write $\mathbb{N}^\infty = \mathbb{N} \cup \{\infty\}$.
The order $\leq$ works as usual. ∞ is the greatest element.
Addition works as usual. ∞ is an attractive element.
Subtraction is introduced as usual, i.e., $a - b = x$ whenever x is the unique value such that $a = x + b$. Thus, $\infty - n = \infty$ for $n \in N$. $x - \infty$ is undefined. The law $x - x = 0$ does not hold anymore.
Multiplication becomes partial because $\infty \cdot 0$ is undefined. For $x \neq 0$, we put $\infty \cdot x = \infty$.
Divisibility $|$ is defined as usual. Thus, we have $x|\infty$ for all $x \neq 0$, and $\infty|x$ iff $x = \infty$. There is no greatest element anymore because: 0 and ∞ are both greater than every other element except for each other.

Negative Infinity We write $\mathbb{Z}^\infty = \mathbb{Z} \cup \{\infty, -\infty\}$.
The order $\leq$ works as usual. $-\infty$ is the least and ∞ the greatest element.
Addition becomes partial because $-\infty + \infty$ is undefined. We put $-\infty + z = -\infty$ for $z \neq \infty$.
Subtraction is introduced as usual. Thus, $z - \infty = -\infty - z = -\infty$ for $z \in \mathbb{Z}$. $\infty - \infty$ is undefined.
Multiplication works similarly to $\mathbb{N}^\infty$. $-\infty \cdot 0$ is undefined. And for $x \neq 0$, we define $\infty \cdot x$ and $-\infty \cdot x$ as ∞ or $-\infty$ depending on the signs.

A.4 Size of Sets

The size $|S|$ of a set S is a very complex topic of mathematics because there are different degrees of infinity. Specifically, we have that $|\mathcal{P}(S)| > |S|$, i.e., we have infinitely many degrees of infinity.
In computer science, we are only interested in countable sets. We use a very simple definition that writes C for countable and merges all greater sizes into uncountable sets, whose size we write as U.

Definition A.36 (Size of sets)**.** The size $|S| \in \mathbb{N} \cup \{C, U\}$ of a set S is defined by:

- if S is finite: $|S|$ is the number of elements of S
- if S is infinite and bijective to $\mathbb{N}$: $|S| = C$, and we say that S is countable
- if S is infinite and not bijective to $\mathbb{N}$: $|S| = U$, and we say that S is uncountable

We can compute with set sizes as follows:

Definition A.37 (Computing with Sizes)**.** For two sizes $s, t \in \mathbb{N} \cup \{C, U\}$, we define addition, multiplication, and exponentiation by the following tables:

s \ t	$n \in \mathbb{N}$	C	U
$m \in \mathbb{N}$	$m + n$	C	U
C	C	C	U
U	U	U	U

s \ t	$n \in \mathbb{N}$	C	U
$m \in \mathbb{N}$	$m * n$	C	U
C	C	C	U
U	U	U	U

s \ t	0	1	$n \in \mathbb{N} \setminus \{0\}$	C	U
0	1	0	0	0	0
1	1	1	1	1	1
$m \in \mathbb{N} \setminus \{0\}$	1	m	m^n	U	U
C	1	C	C	U	U
U	1	U	U	U	U

Because exponentiation s^t is not commutative, the order matters: s is given by the row and t by the column.

The intuition behind these rules is given by the following:

Theorem A.38. *For all sets S, T, we have for the size of the*

- *disjoint union:*
$$|S \uplus T| = |S| + |T|$$
- *Cartesian product:*
$$|S \times T| = |S| * |T|$$
- *set of functions from T to S:*
$$|S^T| = |S|^{|T|}$$

Thus, we can understand the rules for exponentiation as follows. Let us first consider the 4 cases where one of the arguments has size 0 or 1: For every set A

1. there is exactly one function from the empty set (namely the empty function): $|A^{\varnothing}| = 1$,
2. there are as many functions from a singleton set as there are elements of A: $|A^{\{x\}}| = |A|$,
3. there are no functions to the empty set (unless A is empty): $|\varnothing^A| = 0$ if $A \neq \varnothing$,
4. there is exactly one function into a singleton set (namely the constant function): $|\{x\}^A| = 1$,

Now we need only one more rule: The set of functions from a non-empty finite set to a finite/countable/uncountable set is again finite/countable/uncountable. In all other cases, the set of functions is uncountable.

A.5 Important Sets and Functions

The meaning and purpose of a data structure is to describe a set in the sense of mathematics. Similarly, the meaning and purpose of an algorithm is to describe a function between two sets.

Thus, it is helpful to collect some sets and functions as examples. These are typically among the first data structures and algorithms implemented in any programming language and they serve as test cases for evaluating our languages.

A.5.1 Base Sets

When building sets, we have to start somewhere with some sets that are assumed to exist. These are called the *bases sets* or the *primitive sets*.

The following table gives an overview, where we also list the size of each set according to Def. A.36:

set	description/definition	size
typical base sets of mathematics[3]		
$\varnothing$	empty set	0
$\mathbb{N}$	natural numbers	C
$\mathbb{Z}$	integers	C
$\mathbb{Z}_m$ for $m > 0$	integers modulo m, $\{0, \ldots, m-1\}$ [4]	m
$\mathbb{Q}$	rational numbers	C
$\mathbb{R}$	real numbers	U
additional or alternative base sets used in computer science		
void	alternative name for $\varnothing$	0
unit	unit type, $\{()\}$, equivalent to $\mathbb{Z}_1$	1
$\mathbb{B}$	booleans, $\{false, true\}$, equivalent to $\mathbb{Z}_2$	2
int	primitive integers, $-2^{n-1}, \ldots, 2^{n-1}-1$ for machine-dependent n, equivalent to $\mathbb{Z}_{2^n}$ [5]	2^n
float	IEEE floating point approximations of real numbers	C
char	characters	finite[6]
string	lists of characters	C

A.5.2 Functions on the Base Sets

For every base set, we can define some basic operations. These are usually built-in features of programming languages whenever the respective base set is built-in.
We only list a few examples here.

Numbers

For all number sets, we can define addition, subtraction, multiplication, and division in the usual way.
Some care must be taken when subtracting or dividing because the result may be in a different set. For example, the difference of two natural numbers is not in general a natural number but only an integer (e.g., $3 - 5 \notin \mathbb{N}$). Moreover, division by 0 is always forbidden.

Quotients of the Integers

The function $modulus_m$ (see Sect. A.3.3) for $m \in \mathbb{N}$ maps $x \in \mathbb{Z}$ to $x \bmod m \in \mathbb{Z}_m$.
In programming languages, the set $\mathbb{Z}_m$ is usually not provided. Instead, $x \bmod y$ is built-in as a function on int. [7]

Booleans

On booleans, we can define the usual boolean operations conjunction (usually written & or &&), disjunction (usually written | or ||), and negation (usually written !).
Moreover, we have the equality and inequality functions, which take two objects x, y and return a boolean. These are usually written $x == y$ and $x! = y$ in text files and $x = y$ and $x \neq y$ on paper.

A.5.3 Set Constructors

From the base sets, we build all other sets by applying set constructors. Those are operations that take sets and return new sets.
The following table gives an overview, where we also list the size of each set according to Def. A.37:

[3] All of mathematics can be built by using $\varnothing$ as the only base set because the others are definable. But it is common to assume at least the number sets as primitives.
[4] $\mathbb{Z}_0$ also exists but is trivial: $\mathbb{Z}_0 = \mathbb{Z}$.
[5] Primitive integers are the 2^n possible values for a sequence of n bits. Old machines used $n = 8$ (and the integers were called "bytes"), later machines used $n = 16$ (called "words"). Modern machines typically use 32-bit or 64-bit integers. Modern programmers usually—but dangerously—assume that 2^n is much bigger than any number that comes up in practice so that essentially $int = \mathbb{Z}$. Some programming languages (e.g., Python) correctly implement $int = \mathbb{Z}$.
[6] The ASCII standard defined 2^7 or 2^8 characters. Nowadays, we use Unicode characters, which is a constantly growing set containing the characters of virtually any writing system, many scientific symbols, emojis, etc. Many programming languages assume that there is one character for every primitive integers, e.g., typically 2^{32} characters.
[7] Some care must be taken if x is negative because not all programming languages agree.

set	description/definition	size
typical constructors in mathematics		
$A \uplus B$	disjoint union	$\|A\| + \|B\|$
$A \times B$	(Cartesian) product	$\|A\| * \|B\|$
A^n for $n \in \mathbb{N}$	n-dimensional vectors over A	$\|A\|^n$
B^A or $A \to B$	functions from A to B	$\|B\|^{\|A\|}$
$\mathcal{P}(A)$	power set, equivalent to $\mathbb{B}^A$	$2^{\|A\|} = \begin{cases} 2^n & \text{if } \|A\| = n \\ U & \text{otherwise} \end{cases}$
$\{x \in A \mid P(x)\}$	subset of A given by property P	$\leq \|A\|$
$\{f(x) : x \in A\}$	image of operation f when applied to elements of A	$\leq \|A\|$
A/r	quotient set for an equivalence relation r on A	$\leq \|A\|$
selected additional constructors often used in computer science		
A^*	lists over A	$\begin{cases} 1 & \text{if } \|A\| = 0 \\ U & \text{if } \|A\| = U \\ C & \text{otherwise} \end{cases}$
$A^?$	optional element[8] of A	$1 + \|A\|$
for new names $l_1, \ldots, l_n$		
$enum\{l_1, \ldots, l_n\}$	enumeration: like $\mathbb{Z}_n$ but also introduces named elements l_i of the enumeration	n
$l_1(A_1) \mid \ldots \mid l_n(A_n)$	labeled union: like $A_1 \uplus \ldots \uplus A_n$ but also introduces named injections l_i from A_i into the union	$\|A_1\| + \ldots + \|A_n\|$
$\{l_1 : A_1, \ldots, l_n : A_n\}$	record: like $A_1 \times \ldots \times A_n$ but also introduces named projections l_i from the record into A_i	$\|A_1\| * \ldots * \|A_n\|$
inductive data types[9]		C
classes[10]		U

A.5.4 Characteristic Functions of the Set Constructors

Every set constructor comes systematically with characteristic functions into and out of the constructed sets C. These functions allow building elements of C or using elements of C for other computations.

For some sets, these functions do not have standard notations in mathematics. In those cases, different programming languages may use slightly different notations.

The following table gives an overview:

set C	build an element of C	use an element x of C
$A_1 \uplus A_2$	$inj_1(a_1)$ or $inj_2(a_2)$ for $a_i \in A_i$	pattern-matching
$A_1 \times A_2$	(a_1, a_2) for $a_i \in A_i$	$x.i \in A_i$ for $i = 1, 2$
A^n	$(a_1, \ldots, a_n)$ for $a_i \in A$	$x.i \in A$ for $i = 1, \ldots, n$
B^A	$(a \in A) \mapsto b(a)$	$x(a)$ for $a \in A$
A^*	$[a_0, \ldots, a_{l-1}]$[11] for $a_i \in A$	pattern-matching
$A^?$	$None$ or $Some(a)$ for $a \in A$	pattern-matching
$enum\{l_1, \ldots, l_n\}$	l_1 or $\ldots$ or l_n	switch statement or pattern-matching
$l_1(A_1) \mid \ldots \mid l_n(A_n)$	$l_1(a_1)$ or $\ldots$ or $l_n(a_n)$ for $a_i \in A_i$	pattern-matching
$\{l_1 : A_1, \ldots, l_n : A_n\}$	$\{l_1 = a_1, \ldots, l_n = a_n\}$ for $a_i \in A_i$	$x.l_i \in A_i$
inductive data type A	$l(u_1, \ldots, u_n)$ for a constructor l of A	pattern-matching
class A	**new** A	$x.l(u_1, \ldots, u_n)$ for a field l of A

[8] An optional element of A is either absent or an element of A.

[9] These are too complex to define at this point. They are a key feature of functional programming languages like SML.

[10] These are too complex to define at this point. They are a key feature of object-oriented programming languages like Java.

[11] Mathematicians start counting at 1 and would usually write a list of length n as $[a_1, \ldots, a_n]$. However, computer scientists always start counting at 0 and therefore write it as $[a_0, \ldots, a_{n-1}]$. We use the computer science numbering here.

Bibliography

[CLR10] T. Cormen, C. Leiserson, and R. Rivest. *Introduction to Algorithms*. MIT Press, 2010.

[EucBC] Euclid. *Elements*. around 300 BC. English translation by T. Heath (1956) available online.

[Hil00] D. Hilbert. Mathematische Probleme. *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen*, pages 253–297, 1900.

[Hil26] D. Hilbert. Über das Unendliche. *Mathematische Annalen*, 95:161–90, 1926.

[Knu73] D. Knuth. *The Art of Computer Programming*. Addison-Wesley, 1973.