Appendix to Lecture Notes on *Analysis I: Calculus of One Real Variable*

Peter Philip

January 24, 2016

Contents

A Axiomatic Set Theory **3**

A.1 Motivation, Russell's Antinomy 3

A.2 Set-Theoretic Formulas 3

A.3 The Axioms of Zermelo-Fraenkel Set Theory 5

A.3.1 Existence, Extensionality, Comprehension 6

A.3.2 Classes 9

A.3.3 Pairing, Union, Replacement 9

A.3.4 Infinity, Ordinals, Natural Numbers 13

A.3.5 Power Set 20

A.3.6 Foundation 21

A.4 The Axiom of Choice 22

A.5 Cardinality 27

A.5.1 Relations to Injective, Surjective, and Bijective Maps; Schröder-Bernstein Theorem 27

A.5.2 Finite Sets 34

A.5.3 Power Sets 38

B Commutativity and Associativity **38**

B.1 Commutativity 38

B.2 Associativity 41

C Algebraic Structures **44**

C.1 Groups 44

C.2 Rings 47

C.3 Fields 48

D Construction of the Real Numbers **50**

D.1 Natural Numbers 50

D.2 Interlude: Orders on Groups 54

D.3 Integers 55

D.4 Rational Numbers 58

D.5 Real Numbers 61

D.6 Uniqueness 67

E Series: Additional Material **72**

E.1 Riemann Rearrangement Theorem 72

E.2 b-Adic Representations of Real Numbers 74

F Cardinality of $\mathbb{R}$ and Some Related Sets **79**

G Partial Fraction Decomposition **83**

H Irrationality of e and π **88**

H.1 Irrationality of e 88

H.2 Irrationality of π 89

I Trigonometric Functions **91**

I.1 Additional Trigonometric Formulas 91

J Differential Calculus **91**

J.1 Continuous, But Nowhere Differentiable Functions 91

References **94**

A Axiomatic Set Theory

A.1 Motivation, Russell's Antinomy

As it turns out, *naive set theory*, founded on the definition of a set according to Cantor (as stated at the beginning of Sec. 1.3) is not suitable to be used in the foundation of mathematics. The problem lies in the possibility of obtaining contradictions such as *Russell's antinomy*, after Bertrand Russell, who described it in 1901.

Russell's antinomy is obtained when considering the set X of all sets that do not contain themselves as an element: When asking the question if $X \in X$, one obtains the contradiction that $X \in X \Leftrightarrow X \notin X$:

Suppose $X \in X$. Then X is a set that contains itself. But X was defined to contain only sets that do not contain themselves, i.e. $X \notin X$.

So suppose $X \notin X$. Then X is a set that does not contain itself. Thus, by the definition of X, $X \in X$.

Perhaps you think Russell's construction is rather academic, but it is easily translated into a practical situation. Consider a library. The catalog C of the library should contain all the library's books. Since the catalog itself is a book of the library, it should occur as an entry in the catalog. So there can be catalogs such as C that have themselves as an entry and there can be other catalogs that do not have themselves as an entry. Now one might want to have a catalog X of all catalogs that do not have themselves as an entry. As in Russell's antinomy, one is led to the contradiction that the catalog X must have itself as an entry if, and only if, it does not have itself as an entry.

One can construct arbitrarily many versions, which we will not do. Just one more: Consider a small town with a barber, who, each day, shaves all inhabitants, who do not shave themselves. The poor barber now faces a terrible dilemma: He will have to shave himself if, and only if, he does not shave himself.

To avoid contradictions such as Russell's antinomy, *axiomatic set theory* restricts the construction of sets via so-called axioms, as we will see below.

A.2 Set-Theoretic Formulas

The contradiction of Russell's antinomy is related to Cantor's sets not being hierarchical. Another source of contradictions in naive set theory is the imprecise nature of informal languages such as English. In (1.7), we said that

$$A := \{x \in B : P(x)\}$$

defines a subset of B if $P(x)$ is a statement about an element x of B. Now take $B := \mathbb{N} := \{1, 2, \dots\}$ to be the set of the natural numbers and let

$$P(x) := \text{“The number } x \text{ can be defined by fifty English words or less”}. \tag{A.1}$$

Then A is a finite subset of $\mathbb{N}$, since there are only finitely many English words (if you think there might be infinitely many English words, just restrict yourself to the words contained in some concrete dictionary). Then there is a smallest natural number n that is not in A. But then n is the smallest natural number that can not be defined by fifty English words or less, which, actually, defines n by less than fifty English words, in contradiction to $n \notin A$.

To avoid contradictions of this type, we require $P(x)$ to be a so-called *set-theoretic formula*.

Definition A.1. (a) The *language* of set theory consists precisely of the following symbols: $\wedge, \neg, \exists, (,), \in, =, v_j$, where $j = 1, 2, \ldots$.

(b) A *set-theoretic formula* is a finite string of symbols from the above language of set theory that can be built using the following recursive rules:

(i) $v_i \in v_j$ is a set-theoretic formula for $i, j = 1, 2, \ldots$.

(ii) $v_i = v_j$ is a set-theoretic formula for $i, j = 1, 2, \ldots$.

(iii) If ϕ and ψ are set-theoretic formulas, then $(\phi) \wedge (\psi)$ is a set-theoretic formula.

(iv) If ϕ is a set-theoretic formulas, then $\neg(\phi)$ is a set-theoretic formula.

(v) If ϕ is a set-theoretic formulas, then $\exists v_j(\phi)$ is a set-theoretic formula for $j = 1, 2, \ldots$.

Example A.2. Examples of set-theoretic formulas are $(v_3 \in v_5) \wedge (\neg(v_2 = v_3))$, $\exists v_1(\neg(v_1 = v_1))$; examples of symbol strings that are not set-theoretic formulas are $v_1 \in v_2 \in v_3$, $\exists\exists\neg$, and $\in v_3\exists$.

Remark A.3. It is noted that, for a given finite string of symbols, a computer can, in principle, check in finitely many steps, if the string constitutes a set-theoretic formula or not. The symbols that can occur in a set-theoretic formula are to be interpreted as follows: The variables $v_1, v_2, \ldots$ are variables for sets. The symbols $\wedge$ and $\neg$ are to be interpreted as the logical operators of conjunction and negation as described in Sec. 1.2.2. Similarly, $\exists$ stands for an existential quantifier as in Sec. 1.4: The statement $\exists v_j(\phi)$ means "there exists a set v_j that has the property ϕ". Parentheses (and) are used to make clear the scope of the logical symbols $\exists, \wedge, \neg$. Where the symbol $\in$ occurs, it is interpreted to mean that the set to the left of $\in$ is contained as an element in the set to the right of $\in$. Similarly, $=$ is interpreted to mean that the sets occurring to the left and to the right of $=$ are equal.

Remark A.4. A disadvantage of set-theoretic formulas as defined in Def. A.1 is that they quickly become lengthy and unreadable (at least to the human eye). To make formulas more readable and concise, one introduces additional symbols and notation. Formally, additional symbols and notation are always to be interpreted as abbreviations or transcriptions of actual set-theoretic formulas. For example, we use the rules of Th.

1.11 to *define* the additional logical symbols $\vee$, $\Rightarrow$, $\Leftrightarrow$ as abbreviations:

$$(\phi) \vee (\psi) \quad \text{is short for} \quad \neg((\neg(\phi)) \wedge (\neg(\psi))) \quad \text{(cf. Th. 1.11(j))}, \tag{A.2a}$$
$$(\phi) \Rightarrow (\psi) \quad \text{is short for} \quad (\neg(\phi)) \vee (\psi) \quad \text{(cf. Th. 1.11(a))}, \tag{A.2b}$$
$$(\phi) \Leftrightarrow (\psi) \quad \text{is short for} \quad ((\phi) \Rightarrow (\psi)) \wedge ((\psi) \Rightarrow (\phi)) \quad \text{(cf. Th. 1.11(b))}. \tag{A.2c}$$

Similarly, we use (1.18a) to define the universal quantifier:

$$\forall v_j(\phi) \quad \text{is short for} \quad \neg(\exists v_j(\neg(\phi))). \tag{A.2d}$$

Further abbreviations and transcriptions are obtained from omitting parentheses if it is clear from the context and/or from Convention 1.10 where to put them in, by writing variables bound by quantifiers under the respective quantifiers (as in Sec. 1.4), and by using other symbols than v_j for set variables. For example,

$$\underset{x}{\forall} \quad (\phi \Rightarrow \psi) \quad \text{transcribes} \quad \neg(\exists v_1(\neg((\neg(\phi)) \vee (\psi)))).$$

Moreover,

$$v_i \neq v_j \quad \text{is short for} \quad \neg(v_i = v_j); \qquad v_i \notin v_j \quad \text{is short for} \quad \neg(v_i \in v_j). \tag{A.2e}$$

Remark A.5. Even though axiomatic set theory requires the use of set-theoretic formulas as described above, the systematic study of formal symbolic languages is the subject of the field of *mathematical logic* and is beyond the scope of this class (see, e.g., [EFT07]). In Def. and Rem. 1.15, we defined a proof of statement B from statement A_1 as a finite sequence of statements $A_1, A_2, \ldots, A_n$ such that, for $1 \leq i < n$, A_i implies A_{i+1}, and A_n implies B. In the field of *proof theory*, also beyond the scope of this class, such proofs are formalized via a finite set of rules that can be applied to (set-theoretic) formulas (see, e.g., [EFT07, Sec. IV], [Kun12, Sec. II]). Once proofs have been formalized in this way, one can, in principle, *mechanically* check if a given sequence of symbols does, indeed, constitute a valid proof (without even having to understand the actual *meaning* of the statements). Indeed, several different computer programs have been devised that can be used for automatic proof checking, for example *Coq* [Wik15a], *HOL Light* [Wik15b], and *Isabelle* [Wik15c] to name just a few.

A.3 The Axioms of Zermelo-Fraenkel Set Theory

Axiomatic set theory seems to provide a solid and consistent foundation for conducting mathematics, and most mathematicians have accepted it as the basis of their everyday work. However, there do remain some deep, difficult, and subtle *philosophical issues* regarding the foundation of logic and mathematics (see, e.g., [Kun12, Sec. 0, Sec. III]).

Definition and Remark A.6. An *axiom* is a statement that is assumed to be true without any formal logical justification. The most basic axioms (for example, the standard axioms of set theory) are taken to be justified by common sense or some underlying philosophy. However, on a less fundamental (and less philosophical) level,

it is a common mathematical strategy to state a number of axioms (for example, the axioms defining the mathematical structure called a *group*), and then to study the logical consequences of these axioms (for example, *group theory* studies the statements that are true for all groups as a consequence of the group axioms). For a given system of axioms, the question if there exists an object satisfying all the axioms in the system (i.e. if the system of axioms is *consistent*, i.e. free of contradictions) can be extremely difficult to answer.

—

We are now in a position to formulate and discuss the axioms of axiomatic set theory. More precisely, we will present the axioms of *Zermelo-Fraenkel* set theory, usually abbreviated as ZF, which are Axiom 0 – Axiom 8 below. While there exist various set theories in the literature, each set theory defined by some collection of axioms, the axioms of ZFC, consisting of the axioms of ZF plus the axiom of choice (Axiom 9, see Sec. A.4 below), are used as the foundation of mathematics currently accepted by most mathematicians.

A.3.1 Existence, Extensionality, Comprehension

Axiom 0 *Existence:*

$$\underset{X}{\exists} \quad (X = X).$$

Recall that this is just meant to be a more readable transcription of the set-theoretic formula $\exists v_1(v_1 = v_1)$. The axiom of existence states that there exists (at least one) set X.

In Def. 1.18 two sets are defined to be equal if, and only if, they contain precisely the same elements. In axiomatic set theory, this is guaranteed by the axiom of extensionality:

Axiom 1 *Extensionality:*

$$\underset{X}{\forall} \quad \underset{Y}{\forall} \quad \left(\underset{z}{\forall} \quad (z \in X \;\Leftrightarrow\; z \in Y) \;\Rightarrow\; X = Y \right).$$

Following [Kun12], we assume than the substitution property of equality is part of the underlying logic, i.e. if $X = Y$, then X can be substituted for Y and vice versa without changing the truth value of a (set-theoretic) formula. In particular, this yields the converse to extensionality:

$$\underset{X}{\forall} \quad \underset{Y}{\forall} \quad \left(X = Y \;\Rightarrow\; \underset{z}{\forall} \quad (z \in X \;\Leftrightarrow\; z \in Y) \right).$$

Before we discuss further consequences of extensionality, we would like to have the existence of the empty set. However, Axioms 0 and 1 do not suffice to prove the existence of an empty set (see [Kun12, I.6.3]). This, rather, needs the additional axiom

of comprehension. More precisely, in the case of comprehension, we do not have a single axiom, but a scheme of infinitely many axioms, one for each set-theoretic formula. Its formulation makes use of the following definition:

Definition A.7. One obtains the *universal closure* of a set-theoretic formula ϕ, by writing $\underset{v_j}{\forall}$ in front of ϕ for each variable v_j that occurs as a free variable in ϕ (recall from Def. 1.31 that v_j is free in ϕ if, and only if, it is not bound by a quantifier in ϕ).

Axiom 2 *Comprehension Scheme:* For each set-theoretic formula ϕ, not containing Y as a free variable, the universal closure of

$$\underset{Y}{\exists} \ \underset{x}{\forall} \ \Big(x \in Y \ \Leftrightarrow \ (x \in X \wedge \ \phi) \Big)$$

is an axiom. Thus, the comprehension scheme states that, given the set X, there exists (at least one) set Y, containing precisely the elements of X that have the property ϕ.

Remark A.8. Comprehension does not provide uniqueness. However, if both Y and Y' are sets containing precisely the elements of X that have the property ϕ, then

$$\underset{x}{\forall} \ \Big(x \in Y \ \Leftrightarrow \ (x \in X \wedge \ \phi) \Leftrightarrow \ x \in Y' \Big),$$

and, then, extensionality implies $Y = Y'$. Thus, due to extensionality, the set Y given by comprehension is unique, justifying the notation

$$\{x : \ x \in X \wedge \ \phi\} := \{x \in X : \ \phi\} := Y \tag{A.3}$$

(this is the axiomatic justification for (1.7)).

Theorem A.9. *There exists a unique empty set (which we denote by $\emptyset$ or by 0 – it is common to identify the empty set with the number zero in axiomatic set theory).*

Proof. Axiom 0 provides the existence of a set X. Then comprehension allows us to define the empty set by

$$0 := \emptyset := \{x \in X : \ x \neq x\},$$

where, as explained in Rem. A.8, extensionality guarantees uniqueness. ■

Remark A.10. In Rem. A.4 we said that every formula with additional symbols and notation is to be regarded as an abbreviation or transcription of a set-theoretic formula as defined in Def. A.1(b). Thus, formulas containing symbols for defined sets (e.g. 0 or $\emptyset$ for the empty set) are to be regarded as abbreviations for formulas without such symbols. Some logical subtleties arise from the fact that there is some ambiguity in the way such abbreviations can be resolved: For example, $0 \in X$ can abbreviate either

$$\psi : \ \underset{y}{\exists} \Big(\phi(y) \wedge y \in X \Big) \quad \text{or} \quad \chi : \ \underset{y}{\forall} \Big(\phi(y) \Rightarrow y \in X \Big), \quad \text{where } \phi(y) \text{ stands for } \underset{v}{\forall} \, (v \notin y).$$

Then ψ and χ are equivalent if $\underset{y}{\exists!}\ \phi(y)$ is true (e.g., if Axioms 0 – 2 hold), but they can be nonequivalent, otherwise (see discussion between Lem. 2.9 and Lem. 2.10 in [Kun80]).

—

At first glance, the role played by the free variables in ϕ, which are allowed to occur in Axiom 2, might seem a bit obscure. So let us consider examples to illustrate that allowing free variables (i.e. set parameters) in comprehension is quite natural:

Example A.11. (a) Suppose ϕ in comprehension is the formula $x \in Z$ (having Z as a free variable), then the set given by the resulting axiom is merely the intersection of X and Z:

$$X \cap Z := \{x \in X : \phi\} = \{x \in X : x \in Z\}.$$

(b) Note that it is even allowed for ϕ in comprehension to have X as a free variable, so one can let ϕ be the formula $\underset{u}{\exists}\ \ (x \in u \wedge u \in X)$ to define the set

$$X^* := \left\{x \in X : \underset{u}{\exists}\ \ (x \in u \wedge u \in X)\right\}.$$

Then, if $0 := \emptyset$, $1 := \{0\}$, $2 := \{0, 1\}$, we obtain

$$2^* = \{0\} = 1.$$

—

It is a consequence of extensionality that the mathematical universe consists of sets and only of sets: Suppose there were other objects in the mathematical universe, for example a cow C and a monkey M (or any other object without elements, other than the empty set) – this would be equivalent to allowing a cow or a monkey (or any other object without elements, other than the empty set) to be considered a set, which would mean that our set-theoretic variables v_j were allowed to be a cow or a monkey as well. However, extensionality then implies the false statement $C = M = \emptyset$, thereby excluding cows and monkeys from the mathematical universe.

Similarly, $\{C\}$ and $\{M\}$ (or any other object that contains a non-set), can not be inside the mathematical universe. Indeed, otherwise we had

$$\underset{x}{\forall}\ \ \Big(x \in \{C\} \ \Leftrightarrow\ x \in \{M\}\Big)$$

(as C and M are non-sets) and, by extensionality, $\{C\} = \{M\}$ were true, in contradiction to a set with a cow inside not being the same as a set with a monkey inside. Thus, we see that all objects of the mathematical universe must be so-called *hereditary sets*, i.e. sets all of whose elements (thinking of the elements as being the children of the sets) are also sets.

A.3.2 Classes

As we need to avoid contradictions such as Russell's antinomy, we must not require the existence of a set $\{x : \phi\}$ for each set-theoretic formula ϕ. However, it can still be useful to think of a "collection" of all sets having the property ϕ. Such collections are commonly called *classes*:

Definition A.12. (a) If ϕ is a set-theoretic formula, then we call $\{x : \phi\}$ a *class*, namely the class of all sets that have the property ϕ (typically, ϕ will have x as a free variable). Even though one can identify the class $\{x : \phi\}$ with the formula ϕ, it is often useful to think of the class as being a collection of sets (reiterating the point already made above).

(b) If ϕ is a set-theoretic formula, then we say the class $\{x : \phi\}$ *exists* (as a set) if, and only if

$$\exists_X \quad \Big(\forall_x \quad \Big(x \in X \;\Leftrightarrow\; \phi\Big)\Big) \tag{A.4}$$

is true. Then X is actually unique by extensionality and we identify X with the class $\{x : \phi\}$. If (A.4) is false, then $\{x : \phi\}$ is called a *proper class* (and the usual interpretation is that the class is in some sense "too large" to be a set).

Example A.13. (a) Due to Russell's antinomy of Sec. A.1, we know that $\mathbf{R} := \{x : x \notin x\}$ forms a proper class.

(b) The *universal class* of all sets, $\mathbf{V} := \{x : x = x\}$, is a proper class. Once again, this is related to Russell's antinomy: If $\mathbf{V}$ were a set, then

$$\mathbf{R} = \{x : x \notin x\} = \{x : x = x \wedge x \notin x\} = \{x : x \in \mathbf{V} \wedge x \notin x\}$$

would also be a set by comprehension. However, this is in contradiction to $\mathbf{R}$ being a proper class by (a).

Remark A.14. From the perspective of formal logic, statements involving proper classes are to be regarded as abbreviations for statements without proper classes. For example, it turns out that the class $\mathbf{G}$ of all sets forming a group is a proper class. But we might write $G \in \mathbf{G}$ as an abbreviation for the statement "The set G is a group."

A.3.3 Pairing, Union, Replacement

Axioms 0 – 2 are still consistent with the empty set being the only set in existence (see [Kun12, I.6.13]). The next axiom provides the existence of nonempy sets:

Axiom 3 *Pairing:*

$$\forall_x \forall_y \exists_Z (x \in Z \;\wedge\; y \in Z).$$

Thus, the pairing axiom states that, for all sets x and y, there exists a set Z that contains x and y as elements.

In consequence of the pairing axiom, the sets

$$0 := \emptyset, \tag{A.5a}$$
$$1 := \{0\}, \tag{A.5b}$$
$$2 := \{0, 1\} \tag{A.5c}$$

all exist. More generally, we may define:

Definition A.15. If x, y are sets and Z is given by the pairing axiom, then we call

(a) $\{x, y\} := \{u \in Z : u = x \vee u = y\}$ the *unordered pair* given by x and y,

(b) $\{x\} := \{x, x\}$ the *singleton set* given by x,

(c) $(x, y) := \{\{x\}, \{x, y\}\}$ the *ordered pair* given by x and y (cf. Def. 2.1).

—

We can now show that ordered pairs behave as expected:

Lemma A.16. *The following holds true:*

$$\underset{x,y,x',y'}{\forall} \quad \Big((x, y) = (x', y') \;\Leftrightarrow\; (x = x') \wedge (y = y')\Big).$$

Proof. "$\Leftarrow$" is merely

$$(x, y) = \{\{x\}, \{x, y\}\} \overset{x=x',\, y=y'}{=} \{\{x'\}, \{x', y'\}\} = (x', y').$$

"$\Rightarrow$" is done by distinguishing two cases: If $x = y$, then

$$\{\{x\}\} = (x, y) = (x', y') = \{\{x'\}, \{x', y'\}\}.$$

Next, by extensionality, we first get $\{x\} = \{x'\} = \{x', y'\}$, followed by $x = x' = y'$, establishing the case. If $x \neq y$, then

$$\{\{x\}, \{x, y\}\} = (x, y) = (x', y') = \{\{x'\}, \{x', y'\}\},$$

where, by extensionality $\{x\} \neq \{x, y\} \neq \{x'\}$. Thus, using extensionality again, $\{x\} = \{x'\}$ and $x = x'$. Next, we conclude

$$\{x, y\} = \{x', y'\} = \{x, y'\}$$

and a last application of extensionality yields $y = y'$. ∎

While we now have the existence of the infinitely many different sets $0, \{0\}, \{\{0\}\}, \ldots$, we are not, yet, able to form sets containing more than two elements. This is remedied by the following axiom:

Axiom 4 *Union:*

$$\underset{\mathcal{M}}{\forall}\,\underset{Y}{\exists}\,\underset{x}{\forall}\,\underset{X}{\forall}\Big((x \in X \wedge X \in \mathcal{M}) \;\Rightarrow\; x \in Y\Big).$$

Thus, the union axiom states that, for each set of sets $\mathcal{M}$, there exists a set Y containing all elements of elements of $\mathcal{M}$.

Definition A.17. (a) If $\mathcal{M}$ is a set and Y is given by the union axiom, then define

$$\bigcup \mathcal{M} := \bigcup_{X\in\mathcal{M}} X := \left\{x \in Y : \underset{X\in\mathcal{M}}{\exists}\quad x \in X\right\}.$$

(b) If X and Y are sets, then define

$$X \cup Y := \bigcup\{X, Y\}.$$

(c) If x, y, z are sets, then define

$$\{x, y, z\} := \{x, y\} \cup \{z\}.$$

Remark A.18. (a) The definition of set-theoretic unions as

$$\bigcup_{i\in I} A_i := \left\{x : \underset{i\in I}{\exists}\; x \in A_i\right\}$$

in (1.26b) will be equivalent to the definition in Def. A.17(a) if we are allowed to form the set

$$\mathcal{M} := \{A_i : i \in I\}.$$

If I is a set and A_i is a set for each $i \in I$, then $\mathcal{M}$ as above will be a set by Axiom 5 below (the axiom of replacement).

(b) In contrast to unions, intersections can be obtained directly from comprehension without the introduction of an additional axiom: For example

$$X \cap Y := \{x \in X : x \in Y\},$$
$$\bigcap_{i\in I} A_i := \left\{x \in A_{i_0} : \underset{i\in I}{\forall}\, x \in A_i\right\},$$

where $i_0 \in I \neq \emptyset$ is an arbitrary fixed element of I.

(c) The union

$$\bigcup\emptyset = \bigcup_{X\in\emptyset} X = \bigcup_{i\in\emptyset} A_i = \emptyset$$

is the empty set – in particular, a set. However,

$$\bigcap\emptyset = \left\{x : \underset{X\in\emptyset}{\forall}\, x \in X\right\} = \mathbf{V} = \left\{x : \underset{i\in\emptyset}{\forall}\, x \in A_i\right\} = \bigcap_{i\in\emptyset} A_i,$$

i.e. the intersection over the empty set is the class of all sets – in particular, a proper class and not a set.

Definition A.19. We define the *successor function*

$$x \mapsto S(x) := x \cup \{x\} \quad \text{(for each set } x\text{)}.$$

Thus, recalling (A.5), we have $1 = S(0)$, $2 = S(1)$; and we can define $3 := S(2)$, ... In general, we call the set $S(x)$ the *successor* of the set x.

—

In Def. 2.3 and Def. 2.19, respectively, we define functions and relations in the usual manner, making use of the Cartesian product $A \times B$ of two sets A and B, which, according to (2.2) consists of all ordered pairs (x, y), where $x \in A$ and $y \in B$. However, Axioms 0 – 4 are not sufficient to justify the existence of Cartesian products. To obtain Cartesian products, we emply the axiom of replacement. Analogous to the axiom of comprehension, the following axiom of replacement actually consists of a scheme of infinitely many axioms, one for each set-theoretic formula:

Axiom 5 *Replacement Scheme:* For each set-theoretic formula, not containing Y as a free variable, the universal closure of

$$\left(\underset{x \in X}{\forall} \; \underset{y}{\exists!} \; \phi \right) \quad \Rightarrow \quad \left(\underset{Y}{\exists} \quad \underset{x \in X}{\forall} \quad \underset{y \in Y}{\exists} \quad \phi \right)$$

is an axiom. Thus, the replacement scheme states that if, for each $x \in X$, there exists a unique y having the property ϕ (where, in general, ϕ will depend on x), then there exists a set Y that, for each $x \in X$, contains this y with property ϕ. One can view this as obtaining Y by *replacing* each $x \in X$ by the corresponding $y = y(x)$.

Theorem A.20. *If A and B are sets, then the* Cartesian product *of A and B, i.e. the class*

$$A \times B := \left\{ x : \underset{a \in A}{\exists} \quad \underset{b \in B}{\exists} \quad x = (a, b) \right\}$$

exists as a set.

Proof. For each $a \in A$, we can use replacement with $X := B$ and $\phi := \phi_a$ being the formula $y = (a, x)$ to obtain the existence of the set

$$\{a\} \times B := \{(a, x) : x \in B\} \tag{A.6a}$$

(in the usual way, comprehension and extensionality were used as well). Analogously, using replacement again with $X := A$ and ϕ being the formula $y = \{x\} \times B$, we obtain the existence of the set

$$\mathcal{M} := \{\{x\} \times B : x \in A\}. \tag{A.6b}$$

In a final step, the union axiom now shows

$$\bigcup \mathcal{M} = \bigcup_{a \in A} \{a\} \times B = A \times B \tag{A.6c}$$

to be a set as well. ■

A.3.4 Infinity, Ordinals, Natural Numbers

The following axiom of infinity guarantees the existence of infinite sets (e.g., it will allow us to define the set of natural numbers $\mathbb{N}$, which is infinite by Th. A.46 below).

Axiom 6 *Infinity:*

$$\exists_X \quad \Big(0 \in X \ \wedge \ \forall_{x \in X} \quad (x \cup \{x\} \in X)\Big).$$

Thus, the infinity axiom states the existence of a set X containing $\emptyset$ (identified with the number 0), and, for each of its elements x, its successor $S(x) = x \cup \{x\}$.

In preparation for our official definition of $\mathbb{N}$ in Def. A.41 below, we will study so-called ordinals, which are special sets also of further interest to the field of set theory (the natural numbers will turn out to be precisely the finite ordinals). We also need some notions from the theory of relations, in particular, order relations (cf. Def. 2.19 and Def. 2.25).

Definition A.21. Let R be a relation on a set X.

(a) R is called *asymmetric* if, and only if,

$$\forall_{x,y \in X} \quad \Big(xRy \ \Rightarrow \ \neg(yRx)\Big), \tag{A.7}$$

i.e. if x is related to y only if y is not related to x.

(b) R is called a *strict partial order* if, and only if, R is asymmetric and transitive. It is noted that this is consistent with Not. 2.26, since, recalling the notation $\Delta(X) := \{(x, x) : x \in X\}$, R is a partial order on X if, and only if, $R \setminus \Delta(X)$ is a strict partial order on X. We extend the notions lower/upper bound, min, max, inf, sup of Def. 2.27 to strict partial orders R by applying them to $R \cup \Delta(X)$: We call $x \in X$ a lower bound of $Y \subseteq X$ with respect to R if, and only if, x is a lower bound of Y with respect to $R \cup \Delta(X)$, and analogous for the other notions.

(c) A strict partial order R is called a *strict total order* or a *strict linear order* if, and only if, for each $x, y \in X$, one has $x = y$ or xRy or yRx.

(d) R is called a *(strict) well-order* if, and only if, R is a (strict) total order and every nonempty subset of X has a min with respect to R (for example, the usual $\leq$ constitutes a well-order on $\mathbb{N}$ (see Th. D.5 below), but not on $\mathbb{R}$ (e.g., $\mathbb{R}^+$ does not have a min)).

(e) If $Y \subseteq X$, then the relation on Y defined by

$$xSy \ :\Leftrightarrow \ xRy$$

is called the *restriction* of R to Y, denoted $S = R|_Y$ (usually, one still writes R for the restriction).

Lemma A.22. *Let R be a relation on a set X and $Y \subseteq X$.*

(a) *If R is transitive, then $R{\restriction_Y}$ is transitive.*

(b) *If R is reflexive, then $R{\restriction_Y}$ is reflexive.*

(c) *If R is antisymmetric, then $R{\restriction_Y}$ is antisymmetric.*

(d) *If R is asymmetric, then $R{\restriction_Y}$ is asymmetric.*

(e) *If R is a (strict) partial order, then $R{\restriction_Y}$ is a (strict) partial order.*

(f) *If R is a (strict) total order, then $R{\restriction_Y}$ is a (strict) total order.*

(g) *If R is a (strict) well-order, then $R{\restriction_Y}$ is a (strict) well-order.*

Proof. (a): If $a, b, c \in Y$ with aRb and bRc, then aRc, since $a, b, c \in X$ and R is transitive on X.

(b): If $a \in Y$, then $a \in X$ and aRa, since R is reflexive on X.

(c): If $a, b \in Y$ with aRb and bRa, then $a = b$, since $a, b \in X$ and R is antisymmetric on X.

(d): If $a, b \in Y$ with aRb, then $\neg bRa$, since $a, b \in X$ and R is asymmetric on X.

(e) follows by combining (a) – (d).

(f): If $a, b \in Y$ with $a = b$ and $\neg aRb$, then bRa, since $a, b \in X$ and R is total on X. Combining this with (e) yields (f).

(g): Due to (f), it merely remains to show that every nonempty subset $Z \subseteq Y$ has a min. However, since $Z \subseteq X$ and R is a well-order on X, there is $m \in Z$ such that m is a min for R on X, implying m to be a min for R on Y as well. ■

Remark A.23. Since the universal class $\mathbf{V}$ is not a set, $\in$ is not a relation in the sense of Def. 2.19. It can be considered as a "class relation", i.e. a subclass of $\mathbf{V} \times \mathbf{V}$, but it is a proper class. However, $\in$ does constitute a relation in the sense of Def. 2.19 *on each set X* (recalling that each element of X must be a set as well). More precisely, if X is a set, then so is

$$R_\in := \{(x, y) \in X \times X : x \in y\}. \tag{A.8a}$$

Then

$$\underset{x,y \in X}{\forall} \quad (x, y) \in R_\in \quad \Leftrightarrow \quad x \in y. \tag{A.8b}$$

Definition A.24. A set X is called *transitive* if, and only if, every element of X is also a subset of X:

$$\underset{x \in X}{\forall} \quad x \subseteq X. \tag{A.9a}$$

Clearly, (A.9a) is equivalent to

$$\underset{x,y}{\forall} \quad \Big(x \in y \ \wedge \ y \in X \ \Rightarrow \ x \in X\Big). \tag{A.9b}$$

Lemma A.25. *If X, Y are transitive sets, then $X \cap Y$ is a transitive set.*

Proof. If $x \in X \cap Y$ and $y \in x$, then $y \in X$ (since X is transitive) and $y \in Y$ (since Y is transitive). Thus $y \in X \cap Y$, showing $X \cap Y$ is transitive. ■

Definition A.26. (a) A set α is called an *ordinal number* or just an *ordinal* if, and only if, α is transitive and $\in$ constitutes a strict well-order on α. An ordinal α is called a *successor ordinal* if, and only if, there exists an ordinal β such that $\alpha = S(\beta)$, where S is the successor function of Def. A.19. An ordinal $\alpha \neq 0$ is called a *limit ordinal* if, and only if, it is not a successor ordinal. We denote the class of all ordinals by **ON** (it is a proper class by Cor. A.33 below).

(b) We define

$$\underset{\alpha,\beta\in\mathbf{ON}}{\forall} \quad (\alpha < \beta \ :\Leftrightarrow \ \alpha \in \beta), \tag{A.10a}$$

$$\underset{\alpha,\beta\in\mathbf{ON}}{\forall} \quad (\alpha \leq \beta \ :\Leftrightarrow \ \alpha < \beta \vee \alpha = \beta). \tag{A.10b}$$

Example A.27. Using (A.5), $0 = \emptyset$ is an ordinal, and $1 = S(0)$, $2 = S(1)$ are both successor ordinals (in Prop. A.43, we will identify $\mathbb{N}_0$ as the smallest limit ordinal). Even though $X := \{1\}$ and $Y := \{0, 2\}$ are well-ordered by $\in$, they are not ordinals, since they are not transitive sets: $1 \in X$, but $1 \not\subseteq X$ (since $0 \in 1$, but $0 \notin X$); similarly, $1 \in 2 \in Y$, but $1 \notin Y$.

Lemma A.28. *No ordinal contains itself, i.e.*

$$\underset{\alpha\in\mathbf{ON}}{\forall} \quad \alpha \notin \alpha.$$

Proof. If α is an ordinal, then $\in$ is a strict order on α. Due to asymmetry of strict orders, $x \in x$ can not be true for any element of α, implying that $\alpha \in \alpha$ can not be true. ■

Proposition A.29. *Every element of an ordinal is an ordinal, i.e.*

$$\underset{\alpha\in\mathbf{ON}}{\forall} \quad \Big(X \in \alpha \ \Rightarrow \ X \in \mathbf{ON}\Big)$$

(in other words, **ON** *is a transitive class).*

Proof. Let $\alpha \in \mathbf{ON}$ and $X \in \alpha$. Since α is transitive, we have $X \subseteq \alpha$. As $\in$ is a strict well-order on α, it must also be a strict well-order on X by Lem. A.22(g). In consequence, it only remains to prove that X is transitive as well. To this end, let $x \in X$. Then $x \in \alpha$, as α is transitive. If $y \in x$, then, using transitivity of α again, $y \in \alpha$. Now $y \in X$, as $\in$ is transitive on α, proving $x \subseteq X$, i.e. X is transitive. ■

Proposition A.30. *If $\alpha, \beta \in \mathbf{ON}$, then $X := \alpha \cap \beta \in \mathbf{ON}$ (we will see in Th.* A.35(a) *below that, actually, $\alpha \cap \beta = \min\{\alpha, \beta\}$).*

Proof. X is transitive by Lem. A.25, and, since $X \subseteq \alpha$, $\in$ is a strict well-order on X by Lem. A.22(g). ■

Proposition A.31. *On the class* **ON***, the relation* $\leq$ *(as defined in* (A.10)*) is the same as the relation* $\subseteq$*, i.e.*

$$\bigvee_{\alpha,\beta\in\mathbf{ON}} \Big(\alpha \leq \beta \;\Leftrightarrow\; \alpha \subseteq \beta \;\Leftrightarrow\; (\alpha \in \beta \vee \alpha = \beta)\Big). \tag{A.11}$$

Proof. Let $\alpha, \beta \in \mathbf{ON}$.

Assume $\alpha \leq \beta$. If $\alpha = \beta$, then $\alpha \subseteq \beta$. If $\alpha \in \beta$, then $\alpha \subseteq \beta$, since β is transitive.

Conversely, assume $\alpha \subseteq \beta$ and $\alpha \neq \beta$. We have to show $\alpha \in \beta$. To this end, we set $X := \beta \setminus \alpha$. Then $X \neq \emptyset$ and, as $\in$ well-orders β, we can let $m := \min X$. We will show $m = \alpha$ (note that this will complete the proof, due to $\alpha = m \in X \subseteq \beta$). If $\mu \in m$, then $\mu \in \beta$ (since $m \in \beta$ and β is transitive) and $\mu \notin X$ (since $m = \min X$), implying $\mu \in \alpha$ (since $X = \beta \setminus \alpha$) and, thus, $m \subseteq \alpha$. Seeking a contradiction, assume $m \neq \alpha$. Then there must be some $\gamma \in \alpha \setminus m \subseteq \alpha \subseteq \beta$. In consequence $\gamma, m \in \beta$. As $\gamma \notin m$ and $\in$ is a total order on β, we must have either $m = \gamma$ or $m \in \gamma$. However, $m \neq \gamma$, since $\gamma \in \alpha$ and $m \notin \alpha$ (as $m \in X$). So it must be $m \in \gamma \in \alpha$, implying $m \in \alpha$, as β is transitive. This contradiction proves $m = \alpha$ and establishes the proposition. ■

Theorem A.32. *The class* **ON** *is well-odered by* $\in$*, i.e.*

(i) $\in$ *is transitive on* **ON***:*

$$\bigvee_{\alpha,\beta,\gamma\in\mathbf{ON}} \Big(\alpha < \beta \wedge \beta < \gamma \;\Rightarrow\; \alpha < \gamma\Big).$$

(ii) $\in$ *is asymmetric on* **ON***:*

$$\bigvee_{\alpha,\beta\in\mathbf{ON}} \Big(\alpha < \beta \;\Rightarrow\; \neg(\beta < \alpha)\Big).$$

(iii) *Ordinals are always comparable:*

$$\bigvee_{\alpha,\beta\in\mathbf{ON}} \Big(\alpha < \beta \vee \beta < \alpha \vee \alpha = \beta\Big).$$

(iv) *Every nonempty set of ordinals has a min.*

Proof. (i) is clear, as γ is a transitive set.

(ii): If $\alpha, \beta \in \mathbf{ON}$, then $\alpha \in \beta \in \alpha$ implies $\alpha \in \alpha$ by (i), which is a contradiction to Lem. A.28.

(iii): Let $\gamma := \alpha \cap \beta$. Then $\gamma \in \mathbf{ON}$ by Prop. A.30. Thus

$$\gamma \subseteq \alpha \wedge \gamma \subseteq \beta \quad \overset{\text{Lem. A.31}}{\Rightarrow} \quad (\gamma \in \alpha \vee \gamma = \alpha) \wedge (\gamma \in \beta \vee \gamma = \beta). \tag{A.12}$$

If $\gamma \in \alpha$ and $\gamma \in \beta$, then $\gamma \in \alpha \cap \beta = \gamma$, in contradiction to Lem. A.28. Thus, by (A.12), $\gamma = \alpha$ or $\gamma = \beta$. If $\gamma = \alpha$, then $\alpha \subseteq \beta$. If $\gamma = \beta$, then $\beta \subseteq \alpha$, completing the proof of (iii).

(iv): Let X be a nonempty set of ordinals and consider $\alpha \in X$. If $\alpha = \min X$, then we are already done. Otherwise, $Y := \alpha \cap X = \{\beta \in X : \beta \in \alpha\} \neq \emptyset$. Since α is well-ordered by $\in$, there is $m := \min Y$. If $\beta \in X$, then either $\beta < \alpha$ or $\alpha \leq \beta$ by (iii). If $\beta < \alpha$, then $\beta \in Y$ and $m \leq \beta$. If $\alpha \leq \beta$, then $m < \alpha \leq \beta$. Thus, $m = \min X$, proving (iv). ■

Corollary A.33. ON *is a proper class (i.e. there is no set containing all the ordinals).*

Proof. If there is a set X containing all ordinals, then, by comprehension, $\beta := \mathbf{ON} = \{\alpha \in X : \alpha \text{ is an ordinal}\}$ must be a set as well. But then Prop. A.29 says that the *set* β is transitive and Th. A.32 yields that the *set* β is well-ordered by $\in$, implying β to be an ordinal, i.e. $\beta \in \beta$ in contradiction to Lem. A.28. ■

Corollary A.34. *For each set X of ordinals, we have:*

(a) *X is well-ordered by $\in$.*

(b) *X is an ordinal if, and only if, X is transitive. Note: A transitive set of ordinals X is sometimes called an* initial segment *of* **ON***, since, here, transitivity can be restated in the form*

$$\underset{\alpha\in\mathbf{ON}}{\forall} \quad \underset{\beta\in X}{\forall} \quad \Big(\alpha < \beta \;\Rightarrow\; \alpha \in X\Big). \tag{A.13}$$

Proof. (a) is a simple consequence of Th. A.32(i)-(iv).

(b) is immediate from (a). ■

Theorem A.35. *Let X be a nonempty set of ordinals.*

(a) *Then $\gamma := \bigcap X$ is an ordinal, namely $\gamma = \min X$. In particular, if $\alpha, \beta \in \mathbf{ON}$, then $\min\{\alpha, \beta\} = \alpha \cap \beta$.*

(b) *Then $\delta := \bigcup X$ is an ordinal, namely $\delta = \sup X$. In particular, if $\alpha, \beta \in \mathbf{ON}$, then $\max\{\alpha, \beta\} = \alpha \cup \beta$.*

Proof. (a): Let $m := \min X$. Then $\gamma \subseteq m$, since $m \in X$. Conversely, if $\alpha \in X$, then $m \leq \alpha$ implies $m \subseteq \alpha$ by Prop. A.31, i.e. $m \subseteq \gamma$. Thus, $m = \gamma$, proving (a).

(b): To show $\delta \in \mathbf{ON}$, we need to show δ is transitive (then δ is an ordinal by Cor. A.34(b)). If $\alpha \in \delta$, then there is $\beta \in X$ such that $\alpha \in \beta$. Thus, if $\gamma \in \alpha$, then $\gamma \in \beta$, since β is transitive. As $\gamma \in \beta$ implies $\gamma \in \delta$, we see that δ is transitive, as needed. It remains to show $\delta = \sup X$. If $\alpha \in X$, then $\alpha \subseteq \delta$, i.e. $\alpha \leq \delta$, showing δ to be an upper bound for X. Now let $u \in \mathbf{ON}$ be an arbitrary upper bound for X, i.e.

$$\underset{\alpha\in X}{\forall} \quad \alpha \subseteq u.$$

Thus, $\delta \subseteq u$, i.e. $\delta \leq u$, proving $\delta = \sup X$. ■

Next, we obtain some results regarding the successor function of Def. A.19 in the context of ordinals.

Lemma A.36. *We have*

$$\underset{\alpha\in\mathbf{ON}}{\forall}\ \Big(x, y \in S(\alpha) \wedge x \in y \quad \Rightarrow \quad x \neq \alpha\Big).$$

Proof. Seeking a contradiction, we reason as follows:

$$x = \alpha \overset{\alpha\notin\alpha}{\Rightarrow} y \neq \alpha \overset{y\in S(\alpha)}{\Rightarrow} y \in \alpha \overset{\alpha \text{ transitive}}{\Rightarrow} y \subseteq \alpha \overset{x\in y}{\Rightarrow} \alpha \in \alpha.$$

This contradiction to $\alpha \notin \alpha$ yields $x \neq \alpha$, concluding the proof. ■

Proposition A.37. *For each $\alpha \in \mathbf{ON}$, the following holds:*

(a) $S(\alpha) \in \mathbf{ON}$.

(b) $\alpha < S(\alpha)$.

(c) *For each ordinal β, $\beta < S(\alpha)$ holds if, and only if, $\beta \leq \alpha$.*

(d) *For each ordinal β, if $\beta < \alpha$, then $S(\beta) < S(\alpha)$.*

(e) *For each ordinal β, if $S(\beta) < S(\alpha)$, then $\beta < \alpha$.*

Proof. (a): Due to Prop. A.29, $S(\alpha)$ is a set of ordinals. Thus, by Cor. A.34(b), it merely remains to prove that $S(\alpha)$ is transitive. Let $x \in S(\alpha)$. If $x = \alpha$, then $x = \alpha \subseteq \alpha \cup \{\alpha\} = S(\alpha)$. If $x \neq \alpha$, then $x \in \alpha$ and, since α is transitive, this implies $x \subseteq \alpha \subseteq S(\alpha)$, showing $S(\alpha)$ to be transitive, thereby completing the proof of (a).

(b) holds, as $\alpha \in S(\alpha)$ holds by the definition of $S(\alpha)$.

(c) is clear, since, for each ordinal β,

$$\beta < S(\alpha) \ \Leftrightarrow\ \beta \in S(\alpha) \ \Leftrightarrow\ \beta \in \alpha \vee \beta = \alpha \ \Leftrightarrow\ \beta \leq \alpha.$$

(d): If $\beta < \alpha$, then $S(\beta) = \beta \cup \{\beta\} \subseteq \alpha$, i.e. $S(\beta) \leq \alpha < S(\alpha)$.

(e) follows from (d) using contraposition: If $\neg(\beta < \alpha)$, then $\beta = \alpha$ or $\alpha < \beta$, implying $S(\beta) = S(\alpha)$ or $S(\alpha) < S(\beta)$, i.e. $\neg(S(\beta) < S(\alpha))$. ■

We now proceed to define the natural numbers:

Definition A.38. An ordinal n is called a *natural number* if, and only if,

$$n \neq 0 \ \wedge \underset{m\in\mathbf{ON}}{\forall}\ \Big(m \leq n \ \Rightarrow\ m = 0 \vee m \text{ is successor ordinal}\Big).$$

Proposition A.39. *If $n = 0$ or n is a natural number, then $S(n)$ is a natural number and every element of n is a natural number or 0.*

Proof. Suppose n is 0 or a natural number. If $m \in n$, then m is an ordinal by Prop. A.29. Suppose $m \neq 0$ and $k \in m$. Then $k \in n$, since n is transitive. Since n is a natural number, $k = 0$ or k is a successor ordinal. Thus, m is a natural number. It remains to show that $S(n)$ is a natural number. By definition, $S(n) = n \cup \{n\} \neq 0$. Moreover, $S(n) \in \mathbf{ON}$ by Prop. A.37(a), and, thus, $S(n)$ is a successor ordinal. If $m \in S(n)$, then $m \leq n$, implying $m = 0$ or m is a successor ordinal, completing the proof that $S(n)$ is a natural number. ■

Theorem A.40 (Principle of Induction)**.** *If X is a set satisfying*

$$0 \in X \ \wedge \ \underset{x \in X}{\forall} \quad S(x) \in X, \tag{A.14}$$

then X contains 0 and all natural numbers.

Proof. Let X be a set satisfying (A.14). Then $0 \in X$ is immediate. Let n be a natural number and, seeking a contradiction, assume $n \notin X$. Consider $N := S(n) \setminus X$. According to Prop. A.39, $S(n)$ is a natural number and all nonzero elements of $S(n)$ are natural numbers. Since $N \subseteq S(n)$ and $0 \in X$, $0 \notin N$ and all elements of N must be natural numbers. As $n \in N$, $N \neq 0$. Since $S(n)$ is well-ordered by $\in$ and $0 \neq N \subseteq S(n)$, N must have a min $m \in N$, $0 \neq m \leq n$. Since m is a natural number, there must be k such that $m = S(k)$. Then $k < m$, implying $k \notin N$. On the other hand

$$k < m \wedge m \leq n \ \Rightarrow \ k \leq n \ \Rightarrow \ k \in S(n).$$

Thus, $k \in X$, implying $m = S(k) \in X$, in contradiction to $m \in N$. This contradiction proves $n \in X$, thereby establishing the case. ■

Definition A.41. If the set X is given by the axiom of infinity, then we use comprehension to define the set

$$\mathbb{N}_0 := \{n \in X : \ n = 0 \ \vee \ n \text{ is a natural number}\}$$

and note $\mathbb{N}_0$ to be unique by extensionality. We also denote $\mathbb{N} := \mathbb{N}_0 \setminus \{0\}$. In set theory, it is also very common to use the symbol ω for the set $\mathbb{N}_0$.

Corollary A.42. $\mathbb{N}_0$ *is the set of all natural numbers and* 0*, i.e.*

$$\underset{n}{\forall} \quad \Big(n \in \mathbb{N}_0 \ \Leftrightarrow \ n = 0 \ \vee \ n \textit{ is a natural number}\Big).$$

Proof. "⇒" is clear from Def. A.41 and "⇐" is due to Th. A.40. ■

Proposition A.43. $\omega = \mathbb{N}_0$ *is the smallest limit ordinal.*

Proof. Since ω is a set of ordinals and ω is transitive by Prop. A.39, ω is an ordinal by Cor. A.34(b). Moreover $\omega \neq 0$, since $0 \in \omega$; and ω is not a successor ordinal (if $\omega = S(n) = n \cup \{n\}$, then $n \in \omega$ and $S(n) \in \omega$ by Prop. A.39, in contradiction to $\omega = S(n)$), implying it is a limit ordinal. To see that ω is the smallest limit ordinal, let $\alpha \in \mathbf{ON}$, $\alpha < \omega$. Then $\alpha \in \omega$, that means $\alpha = 0$ or α is a natural number (in particular, a successor ordinal). ■

In the following Th. A.44, we will prove that $\mathbb{N}$ satisfies the Peano axioms P1 – P3 of Sec. 3.1 (if one prefers, one can show the same for $\mathbb{N}_0$, where 0 takes over the role of 1).

Theorem A.44. *The set of natural numbers* $\mathbb{N}$ *satisfies the Peano axioms* P1 – P3 *of Sec.* 3.1*.*

Proof. For P1 and P2, we have to show that, for each $n \in \mathbb{N}$, one has $S(n) \in \mathbb{N} \setminus \{1\}$ and that $S(m) \neq S(n)$ for each $m, n \in \mathbb{N}$, $m \neq n$. Let $n \in \mathbb{N}$. Then $S(n) \in \mathbb{N}$ by Prop. A.39. If $S(n) = 1$, then $n < S(n) = 1$ by Prop. A.37(b), i.e. $n = 0$, in contradiction to $n \in \mathbb{N}$. If $m, n \in \mathbb{N}$ with $m \neq n$, then $S(m) \neq S(n)$ is due to Prop. A.37(d). To prove P3, suppose $A \subseteq \mathbb{N}$ has the property that $1 \in A$ and $S(n) \in A$ for each $n \in A$. We need to show $A = \mathbb{N}$ (i.e. $\mathbb{N} \subseteq A$, as $A \subseteq \mathbb{N}$ is assumed). Let $X := A \cup \{0\}$. Then X satisfies (A.14) and Th. A.40 yields $\mathbb{N}_0 \subseteq X$. Thus, if $n \in \mathbb{N}$, then $n \in X \setminus \{0\} = A$, showing $\mathbb{N} \subseteq A$. ■

Notation A.45. For each $n \in \mathbb{N}_0$, we introduce the notation $n + 1 := S(n)$ (more generally, one also defines $\alpha + 1 := S(\alpha)$ for each ordinal α).

Theorem A.46. *Let* $n \in \mathbb{N}_0$*. Then* $A := \mathbb{N}_0 \setminus n$ *is infinite (see Def.* 3.12(b)*). In particular,* $\mathbb{N}_0$ *and* $\mathbb{N} = \mathbb{N}_0 \setminus \{0\} = \mathbb{N}_0 \setminus 1$ *are infinite.*

Proof. Since $n \notin n$, we have $n \in A \neq \emptyset$. Thus, if A were finite, then there were a bijection $f : A \longrightarrow A_m := \{1, \ldots, m\} = \{k \in \mathbb{N} : k \leq m\}$ for some $m \in \mathbb{N}$. However, we will show by induction on $m \in \mathbb{N}$ that there is no injective map $f : A \longrightarrow A_m$. Since $S(n) \notin n$, we have $S(n) \in A$. Thus, if $f : A \longrightarrow A_1 = \{1\}$, then $f(n) = f(S(n))$, showing that f is not injective and proving the cases $m = 1$. For the induction step, we proceed by contraposition and show that the existence of an injective map $f : A \longrightarrow A_{m+1}$, $m \in \mathbb{N}$, (cf. Not. A.45) implies the existence of an injective map $g : A \longrightarrow A_m$. To this end, let $m \in \mathbb{N}$ and $f : A \longrightarrow A_{m+1}$ be injective. If $m + 1 \notin f(A)$, then f itself is an injective map into A_m. If $m + 1 \in f(A)$, then there is a unique $a \in A$ such that $f(a) = m + 1$. Define

$$g : A \longrightarrow A_m, \quad g(k) := \begin{cases} f(k) & \text{for } k < a, \\ f(k+1) & \text{for } a \leq k. \end{cases} \tag{A.15}$$

Then g is well-defined: If $k \in A$ and $a \leq k$, then $k + 1 \in A \setminus \{a\}$, and, since f is injective, g does, indeed, map into A_m. We verify g to be injective: If $k, l \in A$, $k < l$, then also $k < l + 1$ and $k + 1 \neq l + 1$ (by Peano axiom P2 – $k + 1 < l + 1$ then also follows, but we do not make use of that here). In each case, $g(k) \neq g(l)$, proving g to be injective. ■

For more basic information regarding ordinals see, e.g., [Kun12, Sec. I.8].

A.3.5 Power Set

There is one more basic construction principle for sets that is not covered by Axioms 0 – 6, namely the formation of power sets. This needs another axiom:

Axiom 7 *Power Set*:

$$\underset{X}{\forall}\ \underset{\mathcal{M}}{\exists}\ \underset{Y}{\forall}\quad \Big(Y \subseteq X \ \Rightarrow\ Y \in \mathcal{M}\Big).$$

Thus, the power set axiom states that, for each set X, there exists a set $\mathcal{M}$ that contains all subsets Y of X as elements.

Definition A.47. If X is a set and $\mathcal{M}$ is given by the power set axiom, then we call

$$\mathcal{P}(X) := \{Y \in \mathcal{M} : Y \subseteq X\}$$

the *power set* of X. Another common notation for $\mathcal{P}(X)$ is 2^X (cf. Prop. 2.18).

A.3.6 Foundation

Foundation is, perhaps, the least important of the axioms in ZF. It basically cleanses the mathematical universe of unnecessary "clutter", i.e. of certain pathological sets that are of no importance to standard mathematics anyway.

Axiom 8 *Foundation:*

$$\underset{X}{\forall}\left(\underset{x}{\exists}\,(x \in X)\ \Rightarrow\ \underset{x\in X}{\exists}\ \neg\underset{z}{\exists}\left(z \in x\ \wedge\ z \in X\right)\right).$$

Thus, the foundation axiom states that every nonempty set X contains an element x that is disjoint to X.

Theorem A.48. *Due to the foundation axiom, the $\in$ relation can have no cycles, i.e. there do not exist sets $x_1, x_2, \ldots, x_n$, $n \in \mathbb{N}$, such that*

$$x_1 \in x_2 \in \cdots \in x_n \in x_1. \tag{A.16a}$$

In particular, sets can not be members of themselves:

$$\neg\underset{x}{\exists}\, x \in x. \tag{A.16b}$$

Proof. If there were sets $x_1, x_2, \ldots, x_n$, $n \in \mathbb{N}$, such that (A.16a) were true, then, by using the pairing axiom and the union axiom, we could form the set

$$X := \{x_1, \ldots, x_n\}.$$

Then, in contradiction to the foundation axiom, $X \cap x_i \neq \emptyset$, for each $i = 1, \ldots, n$: Indeed, $x_n \in X \cap x_1$, and $x_{i-1} \in X \cap x_i$ for each $i = 2, \ldots, n$. ∎

For a detailed explanation, why "sets" forbidden by foundation do not occur in standard mathematics, anyway, see, e.g., [Kun12, Sec. I.14].

A.4 The Axiom of Choice

In addition to the axioms of ZF discussed in the previous section, there is one more axiom, namely the axiom of choice (AC) that, together with ZF, makes up ZFC, the axiom system at the basis of current standard mathematics. Even though AC is used and accepted by most mathematicians, it does have the reputation of being somewhat less "natural". Thus, many mathematicians try to avoid the use of AC, where possible, and it is often pointed out explicitly, if a result depends on the use of AC (but this practise is by no means consistent, neither in the literature nor in this class, and one might sometimes be surprised, which seemingly harmless result does actually depend on AC in some subtle nonobvious way). We will now state the axiom:

Axiom 9 *Axiom of Choice (AC):*

$$\forall_{\mathcal{M}} \left(\emptyset \notin \mathcal{M} \Rightarrow \exists_{f:\mathcal{M}\longrightarrow \bigcup_{N\in\mathcal{M}} N} \left(\forall_{M\in\mathcal{M}} f(M) \in M \right) \right).$$

Thus, the axiom of choice postulates, for each nonempty set $\mathcal{M}$, whose elements are all nonempty sets, the existence of a *choice function*, that means a function that assigns, to each $M \in \mathcal{M}$, an element $m \in M$.

Example A.49. For example, the axiom of choice postulates, for each nonempty set A, the existence of a choice function on $\mathcal{P}(A) \setminus \{\emptyset\}$ that assigns each subset of A one of its elements.

—

The axiom of choice is remarkable since, at first glance, it seems so natural that one can hardly believe it is not provable from the axioms in ZF. However, one can actually show that it is neither provable nor disprovable from ZF (see, e.g., [Jec73, Th. 3.5, Th. 5.16] – such a result is called an *independence proof*, see [Kun80] for further material). If you want to convince yourself that the existence of choice functions is, indeed, a tricky matter, try to define a choice function on $\mathcal{P}(\mathbb{R}) \setminus \{\emptyset\}$ without AC (but do not spend too much time on it – one can show this is actually impossible to accomplish).

Theorem A.52 below provides several important equivalences of AC. Its statement and proof needs some preparation. We start by introducing some more relevant notions from the theory of partial orders:

Definition A.50. Let X be a set and let $\leq$ be a partial order on X.

(a) An element $m \in X$ is called *maximal* (with respect to $\leq$) if, and only if, there exists no $x \in X$ such that $m < x$ (note that a maximal element does not have to be a max and that a maximal element is not necessarily unique).

(b) A nonempty subset C of X is called a *chain* if, and only if, C is *totally* ordered by $\leq$. Moreover, a chain C is called *maximal* if, and only if, no strict superset Y of C (i.e. no $Y \subseteq X$ such that $C \subsetneq Y$) is a chain.

—

The following lemma is a bit technical and will be used to prove the implication AC $\Rightarrow$ (ii) in Th. A.52 (other proofs in the literature often make use of so-called *transfinite recursion*, but that would mean further developing the theory of ordinals, and we will not pursue this route in this class).

Lemma A.51. *Let X be a set and let $\emptyset \neq \mathcal{M} \subseteq \mathcal{P}(X)$ be a nonempty set of subsets of X. We let $\mathcal{M}$ be partially ordered by inclusion, i.e. setting $A \leq B :\Leftrightarrow A \subseteq B$ for each $A, B \in \mathcal{M}$. Moreover, define*

$$\underset{\mathcal{S}\subseteq\mathcal{M}}{\forall} \quad \bigcup \mathcal{S} := \bigcup_{S\in\mathcal{S}} S \tag{A.17}$$

and assume

$$\underset{\mathcal{C}\subseteq\mathcal{M}}{\forall} \quad \Big(\mathcal{C} \text{ is a chain } \Rightarrow \bigcup \mathcal{C} \in \mathcal{M}\Big). \tag{A.18}$$

If the function $g : \mathcal{M} \longrightarrow \mathcal{M}$ has the property that

$$\underset{M\in\mathcal{M}}{\forall} \quad \Big(M \subseteq g(M) \ \wedge \ \#(g(M) \setminus M) \leq 1\Big), \tag{A.19}$$

then g has a fixed point, i.e.

$$\underset{M\in\mathcal{M}}{\exists} \quad g(M) = M. \tag{A.20}$$

Proof. Fix some arbitrary $M_0 \in \mathcal{M}$. We call $\mathcal{T} \subseteq \mathcal{M}$ an M_0-tower if, and only if, $\mathcal{T}$ satisfies the following three properties

(i) $M_0 \in \mathcal{T}$.

(ii) If $\mathcal{C} \subseteq \mathcal{T}$ is a chain, then $\bigcup \mathcal{C} \in \mathcal{T}$.

(iii) If $M \in \mathcal{T}$, then $g(M) \in \mathcal{T}$.

Let $\mathbb{T} := \{\mathcal{T} \subseteq \mathcal{M} : \mathcal{T} \text{ is an } M_0\text{-tower}\}$. If $\mathcal{T}_1 := \{M \in \mathcal{M} : M_0 \subseteq M\}$, then, clearly, $\mathcal{T}_1$ is an M_0-tower and, in particular, $\mathbb{T} \neq \emptyset$. Next, we note that the intersection of all M_0-towers, i.e. $\mathcal{T}_0 := \bigcap_{\mathcal{T}\in\mathbb{T}} \mathcal{T}$, is also an M_0-tower. Clearly, no strict subset of $\mathcal{T}_0$ can be an M_0-tower and

$$M \in \mathcal{T}_0 \ \Rightarrow \ M \in \mathcal{T}_1 \ \Rightarrow \ M_0 \subseteq M. \tag{A.21}$$

The main work of the rest of the proof consists of showing that $\mathcal{T}_0$ is a chain. To show $\mathcal{T}_0$ to be a chain, define

$$\Gamma := \left\{ M \in \mathcal{T}_0 : \underset{N\in\mathcal{T}_0}{\forall} \quad (M \subseteq N \ \vee \ N \subseteq M) \right\}. \tag{A.22}$$

We intend to show that $\Gamma = \mathcal{T}_0$ by verifying that Γ is an M_0-tower. As an intermediate step, we define

$$\underset{M\in\Gamma}{\forall} \quad \Phi(M) := \{N \in \mathcal{T}_0 : N \subseteq M \ \vee \ g(M) \subseteq N\}$$

and also show each $\Phi(M)$ to be an M_0-tower. Actually, Γ and each $\Phi(M)$ satisfy (i) due to (A.21). To verify Γ satisfies (ii), let $\mathcal{C} \subseteq \Gamma$ be a chain and $U := \bigcup \mathcal{C}$. Then $U \in \mathcal{T}_0$, since $\mathcal{T}_0$ satisfies (ii). If $N \in \mathcal{T}_0$, and $C \subseteq N$ for each $C \in \mathcal{C}$, then $U \subseteq N$. If $N \in \mathcal{T}_0$, and there is $C \in \mathcal{C}$ such that $C \not\subseteq N$, then $N \subseteq C$ (since $C \in \Gamma$), i.e. $N \subseteq U$, showing $U \in \Gamma$ and Γ satisfying (ii). Now, let $M \in \Gamma$. To verify $\Phi(M)$ satisfies (ii), let $\mathcal{C} \subseteq \Phi(M)$ be a chain and $U := \bigcup \mathcal{C}$. Then $U \in \mathcal{T}_0$, since $\mathcal{T}_0$ satisfies (ii). If $U \subseteq M$, then $U \in \Phi(M)$ as desired. If $U \not\subseteq M$, then there is $x \in U$ such that $x \notin M$. Thus, there is $C \in \mathcal{C}$ such that $x \in C$ and $g(M) \subseteq C$ (since $C \in \Phi(M)$), i.e. $g(M) \subseteq U$, showing $U \in \Phi(M)$ also in this case, and $\Phi(M)$ satisfies (ii). We will verify that $\Phi(M)$ satisfies (iii) next. For this purpose, fix $N \in \Phi(M)$. We need to show $g(N) \in \Phi(M)$. We already know $g(N) \in \mathcal{T}_0$, as $\mathcal{T}_0$ satisfies (iii). As $N \in \Phi(M)$, we can now distinguish three cases. Case 1: $N \subsetneq M$. In this case, we cannot have $M \subsetneq g(N)$ (otherwise, $\#(g(N) \setminus N) \geq 2$ in contradiction to (A.19)). Thus, $g(N) \subseteq M$ (since $M \in \Gamma$), showing $g(N) \in \Phi(M)$. Case 2: $N = M$. Then $g(N) = g(M) \in \Phi(M)$ (since $g(M) \in \mathcal{T}_0$ and $g(M) \subseteq g(M)$). Case 3: $g(M) \subseteq N$. Then $g(M) \subseteq g(N)$ by (A.19), again showing $g(N) \in \Phi(M)$. Thus, we have verified that $\Phi(M)$ satisfies (iii) and, therefore, is an M_0-tower. Then, by the definition of $\mathcal{T}_0$, we have $\mathcal{T}_0 \subseteq \Phi(M)$. As we also have $\Phi(M) \subseteq \mathcal{T}_0$ (from the definition of $\Phi(M)$), we have shown

$$\underset{M \in \Gamma}{\forall} \quad \Phi(M) = \mathcal{T}_0.$$

As a consequence, if $N \in \mathcal{T}_0$ and $M \in \Gamma$, then $N \in \Phi(M)$ and this means $N \subseteq M \subseteq g(M)$ or $g(M) \subseteq N$, i.e. each $N \in \mathcal{T}_0$ is comparable to $g(M)$, showing $g(M) \in \Gamma$ and Γ satisfying (iii), completing the proof that Γ is an M_0-tower. As with the $\Phi(M)$ above, we conclude $\Gamma = \mathcal{T}_0$, as desired. To conclude the proof of the lemma, we note $\Gamma = \mathcal{T}_0$ implies $\mathcal{T}_0$ is a chain. We claim that

$$M := \bigcup \mathcal{T}_0$$

satisfies (A.20): Indeed, $M \in \mathcal{T}_0$, since $\mathcal{T}_0$ satisfies (ii). Then $g(M) \in \mathcal{T}_0$, since $\mathcal{T}_0$ satisfies (iii). We then conclude $g(M) \subseteq M$ from the definition of M. As we always have $M \subseteq g(M)$ by (A.19), we have established $g(M) = M$ and proved the lemma. ■

Theorem A.52 (Equivalences to the Axiom of Choice)**.** *The following statements* (i) – (v) *are equivalent to the axiom of choice (as stated as Axiom* 9 *above).*

- **(i)** *Every Cartesian product* $\prod_{i \in I} A_i$ *of nonempty sets* A_i*, where* I *is a nonempty index set, is nonempty (cf. Def.* 2.15(c)*).*
- **(ii)** Hausdorff's Maximality Principle: *Every nonempty partially ordered set* X *contains a maximal chain (i.e. a maximal totally ordered subset).*
- **(iii)** Zorn's Lemma: *Let* X *be a nonempty partially ordered set. If every chain* $C \subseteq X$ *(i.e. every nonempty totally ordered subset of* X*) has an upper bound in* X *(such chains with upper bounds are sometimes called* inductive*), then* X *contains a maximal element (cf. Def.* A.50(a)*).*

(iv) Zermelo's Well-Ordering Theorem: *Every set can be well-ordered (recall the definition of a well-order from Def.* A.21(d)*).*

(v) *Every vector space V over a field F has a basis $B \subseteq V$.*

Proof. "(i) $\Leftrightarrow$ AC": Assume (i). Given a nonempty set of nonempty sets $\mathcal{M}$, let $I := \mathcal{M}$ and, for each $M \in \mathcal{M}$, let $A_M := M$. If $f \in \prod_{M\in I} A_M$, then, according to Def. 2.15(c), for each $M \in I = \mathcal{M}$, one has $f(M) \in A_M = M$, proving AC holds. Conversely, assume AC. Consider a family $(A_i)_{i\in I}$ such that $I \neq \emptyset$ and each $A_i \neq \emptyset$. Let $\mathcal{M} := \{A_i : i \in I\}$. Then, by AC, there is a map $g : \mathcal{M} \longrightarrow \bigcup_{N\in\mathcal{M}} N = \bigcup_{j\in I} A_j$ such that $g(M) \in M$ for each $M \in \mathcal{M}$. Then we can define

$$f : I \longrightarrow \bigcup_{j\in I} A_j, \quad f(i) := g(A_i) \in A_i,$$

to prove (i).

Next, we will show AC $\Rightarrow$ (ii) $\Rightarrow$ (iii) $\Rightarrow$ (iv) $\Rightarrow$ AC.

"AC $\Rightarrow$ (ii)": Assume AC and let X be a nonempty partially ordered set. Let $\mathcal{M}$ be the set of all chains in X (i.e. the set of all nonempty totally ordered subsets of X). Then $\emptyset \notin \mathcal{M}$ and $\mathcal{M} \neq \emptyset$ (since $X \neq \emptyset$ and $\{x\} \in \mathcal{M}$ for each $x \in X$). Moreover, $\mathcal{M}$ satisfies the hypothesis of Lem. A.51, since, if $\mathcal{C} \subseteq \mathcal{M}$ is a chain of totally ordered subsets of X, then $\bigcup\mathcal{C}$ is a totally ordered subset of X, i.e. in $\mathcal{M}$ (here we have used the notation of (A.17); also note that we are dealing with two different types of chains here, namely those with respect to the order on X and those with respect to the order given by $\subseteq$ on $\mathcal{M}$). Let $f : \mathcal{P}(X) \setminus \{\emptyset\} \longrightarrow X$ be a choice function given by AC, i.e. such that

$$\forall_{Y\in\mathcal{P}(X)\setminus\{\emptyset\}} \quad f(Y) \in Y.$$

As an auxiliary notation, we set

$$\forall_{M\in\mathcal{M}} \quad M^* := \big\{x \in X \setminus M : M \cup \{x\} \in \mathcal{M}\big\}.$$

With the intention of applying Lem. A.51, we define

$$g : \mathcal{M} \longrightarrow \mathcal{M}, \quad g(M) := \begin{cases} M \cup \{f(M^*)\} & \text{if } M^* \neq \emptyset, \\ M & \text{if } M^* = \emptyset. \end{cases}$$

Since g clearly satisfies (A.19), Lem. A.51 applies, providing an $M \in \mathcal{M}$ such that $g(M) = M$. Thus, $M^* = \emptyset$, i.e. M is a maximal chain, proving (ii).

"(ii) $\Rightarrow$ (iii)": Assume (ii). To prove Zorn's lemma, let X be a nonempty set, partially ordered by $\leq$, such that every chain $C \subseteq X$ has an upper bound. Due to Hausdorff's maximality principle, we can assume $C \subseteq X$ to be a *maximal* chain. Let $m \in X$ be an upper bound for the maximal chain C. We claim that m is a maximal element: Indeed, if there were $x \in X$ such that $m < x$, then $x \notin C$ (since m is upper bound for C) and $C \cup \{x\}$ would constitute a strict superset of C that is also a chain, contradicting the maximality of C.

"(iii) $\Rightarrow$ (iv)": Assume (iii) and let X be a nonempty set. We need to construct a well-order on X. Let $\mathcal{W}$ be the set of all well-orders on subsets of X, i.e.

$$\mathcal{W} := \big\{(Y,W) : Y \subseteq X \wedge W \subseteq Y \times Y \subseteq X \times X \text{ is a well-order on } Y\big\}.$$

We define a partial order $\leq$ on $\mathcal{W}$ by setting

$$\underset{(Y,W),(Y',W')\in\mathcal{W}}{\forall} \Big((Y,W) \leq (Y',W') \;:\Leftrightarrow Y \subseteq Y' \wedge W = W'\restriction_Y \\ \wedge\; (y \in Y,\, y' \in Y',\, y'W'y \;\Rightarrow\; y' \in Y)\Big)$$

(recall the definition of the restriction of a relation from Def. A.21(e)). To apply Zorn's lemma to $(\mathcal{W},\leq)$, we need to check that every chain $\mathcal{C} \subseteq \mathcal{W}$ has an upper bound. To this end, if $\mathcal{C} \subseteq \mathcal{W}$ is a chain, let

$$U_{\mathcal{C}} := (Y_{\mathcal{C}}, W_{\mathcal{C}}), \quad \text{where} \quad Y_{\mathcal{C}} := \bigcup_{(Y,W)\in\mathcal{C}} Y, \quad W_{\mathcal{C}} := \bigcup_{(Y,W)\in\mathcal{C}} W.$$

We need to verify $U_{\mathcal{C}} \in \mathcal{W}$: If $aW_{\mathcal{C}}b$, then there is $(Y,C) \in \mathcal{C}$ such that aWb. In particular, $(a,b) \in Y \times Y \subseteq Y_{\mathcal{C}} \times Y_{\mathcal{C}}$, showing $W_{\mathcal{C}}$ to be a relation on $Y_{\mathcal{C}}$. Clearly, $W_{\mathcal{C}}$ is a total order on $Y_{\mathcal{C}}$ (one just uses that, if $a,b \in Y_{\mathcal{C}}$, then, as $\mathcal{C}$ is a chain, there is $(Y,W) \in \mathcal{C}$ such that $a,b \in Y$ and $W = W_{\mathcal{C}}\restriction_Y$ is a total order on Y). To see that $W_{\mathcal{C}}$ is a well-order on $Y_{\mathcal{C}}$, let $\emptyset \neq A \subseteq Y_{\mathcal{C}}$. If $a \in A$, then there is $(Y,W) \in \mathcal{C}$ such that $a \in Y$. Since $W = W_{\mathcal{C}}\restriction_Y$ is a well-order on Y, we can let $m := \min Y \cap A$. We claim that $m = \min A$ as well: Let $b \in A$. Then there is $(B,U) \in \mathcal{C}$ such that $b \in B$. If $B \subseteq Y$, then $b \in Y \cap A$ and mWb. If $Y \subseteq B$, then $m,b \in B$. If mUb, then we are done. If bUm, then $b \in Y$ (since $(Y,W) \leq (B,U)$), i.e., again, $b \in Y \cap A$ and mWb (actually $m = b$ in this case), proving $m = \min A$. This completes the proof that $W_{\mathcal{C}}$ is a well-order on $Y_{\mathcal{C}}$ and, thus, shows $U_{\mathcal{C}} \in \mathcal{W}$. Next, we check $U_{\mathcal{C}}$ to be an upper bound for $\mathcal{C}$: If $(Y,W) \in \mathcal{C}$, then $Y \subseteq Y_{\mathcal{C}}$ and $W = W_{\mathcal{C}}\restriction_Y$ are immediate. If $y \in Y$, $y' \in Y_{\mathcal{C}}$, and $y'W_{\mathcal{C}}y$, then $y' \in Y$ (otherwise, $y' \in A$ with $(A,U) \in \mathcal{C}$, $(Y,W) \leq (A,U)$, $y'Uy$, in contradiction to $y' \notin Y$). Thus, $(Y,W) \leq U_{\mathcal{C}}$, showing $U_{\mathcal{C}}$ to be an upper bound for $\mathcal{C}$. By Zorn's lemma, we conclude that $\mathcal{W}$ contains a maximal element (M,W_M). But then $M = X$ and W_M is the desired well-order on X: Indeed, if there is $x \in X \setminus M$, then we can let $Y := M \cup \{x\}$ and,

$$\underset{a,b\in Y}{\forall} \Big(aWb \;:\Leftrightarrow\; (a,b \in M \wedge aW_Mb) \;\vee\; b = x\Big).$$

Then $(Y,W) \in \mathcal{W}$ with $(M,W_M) < (Y,W)$ in contradiction to the maximality of (M,W_M).

"(iv) $\Rightarrow$ AC": Assume (iv). Given a nonempty set of nonempty sets $\mathcal{M}$, let $X := \bigcup_{M\in\mathcal{M}} M$. By (iv), there exists a well-order R on X. Then every nonempty $Y \subseteq X$ has a unique min. As every $M \in \mathcal{M}$ is a nonempty subset of X, we can define a choice function

$$f : \mathcal{M} \longrightarrow X, \quad f(M) := \min M \in M,$$

proving AC.

"(v) $\Leftrightarrow$ AC": That every vector space has a basis is usually proved in textbooks on Linear Algebra by the use of Zorn's lemma (see, e.g., [Str08, Lem. 11.3]). That, conversely, (v) implies AC was first shown in [Bla84], but the proof needs more algebraic tools than we have available in this class. ■

A.5 Cardinality

A.5.1 Relations to Injective, Surjective, and Bijective Maps; Schröder-Bernstein Theorem

Theorem A.53. *Let $\mathcal{M}$ be a set of sets. Then the relation $\sim$ on $\mathcal{M}$, defined by*

$$A \sim B \ :\Leftrightarrow \ A \text{ and } B \text{ have the same cardinality,} \tag{A.23}$$

constitutes an equivalence relation on $\mathcal{M}$.

Proof. According to Def. 2.23, we have to prove that $\sim$ is reflexive, symmetric, and transitive. According to Def. 3.12(a), $A \sim B$ holds for $A, B \in \mathcal{M}$ if, and only if, there exists a bijective map $f : A \longrightarrow B$. Thus, since the identity $\mathrm{Id} : A \longrightarrow A$ is bijective, $A \sim A$, showing $\sim$ is reflexive. If $A \sim B$, then there exists a bijective map $f : A \longrightarrow B$, and f^{-1} is a bijective map $f^{-1} : B \longrightarrow A$, showing $B \sim A$ and that $\sim$ is symmetric. If $A \sim B$ and $B \sim C$, then there are bijective maps $f : A \longrightarrow B$ and $g : B \longrightarrow C$. Then, according to Th. 2.14, the composition $(g \circ f) : A \longrightarrow C$ is also bijective, proving $A \sim C$ and that $\sim$ is transitive. ■

The next theorem provides two interesting, and sometimes useful, characterizations of infinite sets:

Theorem A.54. *Let A be a set. Using the axiom of choice (AC) of Sec.* A.4*, the following statements* (i) – (iii) *are equivalent. More precisely,* (ii) *and* (iii) *are equivalent even without AC (a set A is sometimes called* Dedekind-infinite *if, and only if, it satisfies* (iii)*),* (iii) *implies* (i) *without AC, but AC is needed to show* (i) *implies* (ii), (iii).

(i) *A is infinite.*

(ii) *There exists $M \subseteq A$ and a bijective map $f : M \longrightarrow \mathbb{N}$.*

(iii) *There exists a strict subset $B \subsetneq A$ and a bijective map $g : A \longrightarrow B$.*

One sometimes expresses the equivalence between (i) *and* (ii) *by saying that a set is infinite if, and only if, it contains a copy of the natural numbers. The property stated in* (iii) *might seem strange at first, but infinite sets are, indeed, precisely those identical in size to some of their strict subsets (as an example think of the natural bijection $n \mapsto 2n$ between all natural numbers and the even numbers).*

Proof. We first prove, without AC, the equivalence between (ii) and (iii).

"(ii) $\Rightarrow$ (iii)": Let E denote the even numbers. Then $E \subsetneq \mathbb{N}$ and $h : \mathbb{N} \longrightarrow E$, $h(n) := 2n$, is a bijection, showing that (iii) holds for the natural numbers. According to (ii), there exists $M \subseteq A$ and a bijective map $f : M \longrightarrow \mathbb{N}$. Define $B := (A \setminus M) \dot\cup f^{-1}(E)$ and

$$h : A \longrightarrow B, \quad h(x) := \begin{cases} x & \text{for } x \in A \setminus M, \\ f^{-1} \circ h \circ f(x) & \text{for } x \in M. \end{cases} \tag{A.24}$$

Then $B \subsetneq A$ since B does not contain the elements of M that are mapped to odd numbers under f. Still, h is bijective, since $h\restriction_{A\setminus M} = \mathrm{Id}_{A\setminus M}$ and $h\restriction_M = f^{-1} \circ h \circ f$ is the composition of the bijective maps f, h, and $f^{-1}\restriction_E : E \longrightarrow f^{-1}(E)$.

"(iii) $\Rightarrow$ (ii)": As (iii) is assumed, there exist $B \subseteq A$, $a \in A \setminus B$, and a bijective map $g : A \longrightarrow B$. Set

$$M := \{a_n := g^n(a) : n \in \mathbb{N}\}.$$

We show that $a_n \neq a_m$ for each $m, n \in \mathbb{N}$ with $m \neq n$: Indeed, suppose $m, n \in \mathbb{N}$ with $n > m$ and $a_n = a_m$. Then, since g is bijective, we can apply g^{-1} m times to $a_n = a_m$ to obtain

$$a = (g^{-1})^m(a_m) = (g^{-1})^m(a_n) = g^{n-m}(a).$$

Since $l := n - m \geq 1$, we have $a = g(g^{l-1}(a))$, in contradiction to $a \in A \setminus B$. Thus, all the $a_n \in M$ are distinct and we can define $f : M \longrightarrow \mathbb{N}$, $f(a_n) := n$, which is clearly bijective, proving (ii).

"(iii) $\Rightarrow$ (i)": The proof is conducted by contraposition, i.e. we assume A to be finite and proof that (iii) does not hold. If $A = \emptyset$, then there is nothing to prove. If $\emptyset \neq A$ is finite, then, by Def. 3.12(b), there exists $n \in \mathbb{N}$ and a bijective map $f : A \longrightarrow \{1, \ldots, n\}$. If $B \subsetneq A$, then, according to Th. A.63(a), there exists $m \in \mathbb{N}_0$, $m < n$, and a bijective map $h : B \longrightarrow \{1, \ldots, m\}$. If there were a bijective map $g : A \longrightarrow B$, then $h \circ g \circ f^{-1}$ were a bijective map from $\{1, \ldots, n\}$ onto $\{1, \ldots, m\}$ with $m < n$ in contradiction to Th. A.61.

"(i) $\Rightarrow$ (ii)": Inductively, we construct a strictly increasing sequence $M_1 \subseteq M_2 \subseteq \ldots$ of subsets M_n of A $n \in \mathbb{N}$, and a sequence of functions $f_n : M_n \longrightarrow \{1, \ldots, n\}$ satisfying

$$\forall_{n\in\mathbb{N}} \quad f_n \text{ is bijective,} \tag{A.25a}$$

$$\forall_{m,n\in\mathbb{N}} \quad \Big(m \leq n \quad \Rightarrow \quad f_n\restriction_{M_m} = f_m\Big) : \tag{A.25b}$$

Since $A \neq \emptyset$, there exists $m_1 \in A$. Set $M_1 := \{m_1\}$ and $f_1 : M_1 \longrightarrow \{1\}$, $f_1(m_1) := 1$. Then $M_1 \subseteq A$ and f_1 bijective are trivially clear. Now let $n \in \mathbb{N}$ and suppose $M_1, \ldots, M_n$ and $f_1, \ldots, f_n$ satisfying (A.25) have already been constructed. Since A is infinite, there must be $m_{n+1} \in A \setminus M_n$ (otherwise $M_n = A$ and the bijectivity of $f_n : M_n \longrightarrow \{1, \ldots, n\}$ shows A is finite with $\#A = n$; AC is used to select the $m_{n+1} \in A \setminus M_n$). Set $M_{n+1} := M_n \cup \{m_{n+1}\}$ and

$$f_{n+1} : M_{n+1} \longrightarrow \{1, \ldots, n+1\}, \quad f_{n+1}(x) := \begin{cases} f_n(x) & \text{for } x \in M_n, \\ n+1 & \text{for } x = m_{n+1}. \end{cases} \tag{A.26}$$

Then the bijectivity of f_n implies the bijectivity of f_{n+1}, and, since $f_{n+1}{\restriction}_{M_n} = f_n$ holds by definition of f_{n+1}, the implication

$$m \leq n+1 \quad \Rightarrow \quad f_{n+1}{\restriction}_{M_m} = f_m$$

holds true as well. An induction also shows $M_n = \{m_1, \ldots, m_n\}$ and $f_n(m_n) = n$ for each $n \in \mathbb{N}$. We now define

$$M := \bigcup_{n\in\mathbb{N}} M_n = \{m_n : n \in \mathbb{N}\}, \quad f : M \longrightarrow \mathbb{N}, \quad f(m_n) := f_n(m_n) = n. \tag{A.27}$$

Clearly, $M \subseteq A$, and f is bijective with $f^{-1} : \mathbb{N} \longrightarrow M$, $f^{-1}(n) = m_n$. ■

Theorem A.55 (Schröder-Bernstein)**.** *Let A, B be sets. The following statements are equivalent (even without assuming the axiom of choice):*

(i) *The sets A and B have the same cardinality (i.e. there exists a bijective map $\phi : A \longrightarrow B$).*

(ii) *There exist an injective map $f : A \longrightarrow B$ and an injective map $g : B \longrightarrow A$.*

We will give two proofs of the Schröder-Bernstein theorem. The first proof is rather elegant, but also quite abstract. The second proof is longer, but less abstract. Even though it is still nonconstructive in the general situation, in many concrete cases, it does provide a method for actually constructing a bijective map from two injective maps. The first proof is based on the following lemma:

Lemma A.56. *Let A be a set. Consider $\mathcal{P}(A)$ to be endowed with the partial order given by set inclusion, i.e., for each $X, Y \in \mathcal{P}(A)$, $X \leq Y$ if, and only if, $X \subseteq Y$. If $F : \mathcal{P}(A) \longrightarrow \mathcal{P}(A)$ is isotone with respect to that order, then F has a fixed point, i.e. $F(X_0) = X_0$ for some $X_0 \in \mathcal{P}(A)$.*

Proof. Define

$$\mathcal{A} := \{X \in \mathcal{P}(A) : F(X) \subseteq X\}, \quad X_0 := \bigcap_{X\in\mathcal{A}} X \tag{A.28}$$

(X_0 is well-defined, since $F(A) \subseteq A$). Suppose $X \in \mathcal{A}$, i.e. $F(X) \subseteq X$ and $X_0 \subseteq X$. Then $F(X_0) \subseteq F(X) \subseteq X$ due to the isotonicity of F. But, then, $F(X_0) \subseteq X_0$, since $X \in \mathcal{A}$. Using the isotonicity of F again shows $F(F(X_0)) \subseteq F(X_0)$, implying $F(X_0) \in \mathcal{A}$ and $X_0 \subseteq F(X_0)$, i.e. $F(X_0) = X_0$ as desired. ■

First Proof of Th. A.55. (i) trivially implies (ii), as one can simply set $f := \phi$ and $g := \phi^{-1}$. It remains to show (ii) implies (i). Thus, let $f : A \longrightarrow B$ and $g : B \longrightarrow A$ be injective. To apply Lem. A.56, define

$$F : \mathcal{P}(A) \longrightarrow \mathcal{P}(A), \quad F(X) := A \setminus g\big(B \setminus f(X)\big),$$

and note

$$\begin{aligned} X \subseteq Y \subseteq A \;&\Rightarrow\; f(X) \subseteq f(Y) \;\Rightarrow\; B \setminus f(Y) \subseteq B \setminus f(X) \\ &\Rightarrow\; g\big(B \setminus f(Y)\big) \subseteq g\big(B \setminus f(X)\big) \;\Rightarrow\; F(X) \subseteq F(Y). \end{aligned}$$

Thus, by Lem. A.56, F has a fixed point X_0. We claim that a bijection is obtained via setting

$$\phi : A \longrightarrow B, \quad \phi(x) := \begin{cases} f(x) & \text{for } x \in X_0, \\ g^{-1}(x) & \text{for } x \notin X_0. \end{cases}$$

First, ϕ is well-defined, since $x \notin X_0 = F(X_0)$ implies $x \in g\big(B \setminus f(X_0)\big)$. To verify that ϕ is injective, let $x, y \in A$, $x \neq y$. If $x, y \in X_0$, then $\phi(x) \neq \phi(y)$, as f is injective. If $x, y \in A \setminus X_0$, then $\phi(x) \neq \phi(y)$, as g^{-1} is well-defined. If $x \in X_0$ and $y \notin X_0$, then $\phi(x) \in f(X_0)$ and $\phi(y) \in B \setminus f(X_0)$, once again, implying $\phi(x) \neq \phi(y)$. If remains to prove surjectivity. If $b \in f(X_0)$, then $\phi(f^{-1}(b)) = b$. If $b \in B \setminus f(X_0)$, then $g(b) \notin X_0 = F(X_0)$, i.e. $\phi(g(b)) = b$, showing ϕ to be surjective. ■

Second Proof of Th. A.55. As in the first proof, we only need to show (ii) implies (i). We first assume that A and B are disjoint. To define ϕ, we first construct a suitable partition of $A \dot\cup B$, where the subsets of the partition are given via sequences defined by using f and g. The idea is to assign a unique sequence $\sigma(a)$ to each $a \in A$ and a unique sequence $\sigma(b)$ to each $b \in B$ by alternately applying f and g to advance the sequence to the right and by alternately applying f^{-1} and g^{-1} to advance the sequence to the left, if possible (for a given $a \in A$, $g^{-1}(a)$ might not be defined and, for a given $b \in B$, $f^{-1}(a)$ might not be defined). Thus, for $a \in A$, $\sigma(a)$ has the form

$$\ldots,\; f^{-1}\big(g^{-1}(a)\big),\; g^{-1}(a),\; a,\; f(a),\; g\big(f(a)\big),\; \ldots \tag{A.29}$$

More precisely, for each $a \in A$, we define $\sigma(a) = (\sigma_i(a))_{i \in I_a}$ recursively by

$$\begin{aligned}
&\sigma_i(a) := a && \text{for } i = 0, && \text{(A.30a)}\\
&\sigma_i(a) := f\big(\sigma_{i-1}(a)\big) && \text{for } i > 0 \text{ odd}, && \text{(A.30b)}\\
&\sigma_i(a) := g\big(\sigma_{i-1}(a)\big) && \text{for } i > 0 \text{ even}, && \text{(A.30c)}\\
&\sigma_i(a) := g^{-1}\big(\sigma_{i+1}(a)\big) && \text{for } i < 0 \text{ odd and } \sigma_{i+1}(a) \in g(B), && \text{(A.30d)}\\
&\quad m_a := i+1,\; I_a := \{k \in \mathbb{Z} : m_a \leq k\} && \text{for } i < 0 \text{ odd and } \sigma_{i+1}(a) \notin g(B), && \text{(A.30e)}\\
&\sigma_i(a) := f^{-1}\big(\sigma_{i+1}(a)\big) && \text{for } i < 0 \text{ even and } \sigma_{i+1}(a) \in f(A), && \text{(A.30f)}\\
&\quad m_a := i+1,\; I_a := \{k \in \mathbb{Z} : m_a \leq k\} && \text{for } i < 0 \text{ even and } \sigma_{i+1}(a) \notin f(A), && \text{(A.30g)}
\end{aligned}$$

where the conditions in (A.30e) and (A.30g) are meant to implicitly require $\sigma_{i+1}(a)$ to be defined for $i+1$. By induction, one shows $\sigma_{i-1}(a) \in A$ for each $i > 0$ odd, $\sigma_{i-1}(a) \in B$ for each $i > 0$ even, $\sigma_{i+1}(a) \in A$ for each $m_a \leq i < 0$ odd, and $\sigma_{i+1}(a) \in B$ for each $m_a \leq i < 0$ even, such that $\sigma_i(a)$ is well-defined by (A.30) for each $i \in I_a$ (with $I_a = \mathbb{Z}$ if (A.30e) and (A.30g) are never satisfied). Analogously, for each $b \in B$, we

define $\sigma(b) = (\sigma_i(b))_{i\in I_b}$ recursively by

$$\begin{aligned}
\sigma_i(b) &:= b && \text{for } i = 0, && \text{(A.31a)}\\
\sigma_i(b) &:= g\big(\sigma_{i-1}(b)\big) && \text{for } i > 0 \text{ odd}, && \text{(A.31b)}\\
\sigma_i(b) &:= f\big(\sigma_{i-1}(b)\big) && \text{for } i > 0 \text{ even}, && \text{(A.31c)}\\
\sigma_i(b) &:= f^{-1}\big(\sigma_{i+1}(b)\big) && \text{for } i < 0 \text{ odd and } \sigma_{i+1}(b) \in f(A), && \text{(A.31d)}\\
m_b &:= i+1,\ I_b := \{k \in \mathbb{Z} :\ m_b \le k\} && \text{for } i < 0 \text{ odd and } \sigma_{i+1}(b) \notin f(A), && \text{(A.31e)}\\
\sigma_i(b) &:= g^{-1}\big(\sigma_{i+1}(b)\big) && \text{for } i < 0 \text{ even and } \sigma_{i+1}(b) \in g(B), && \text{(A.31f)}\\
m_b &:= i+1,\ I_b := \{k \in \mathbb{Z} :\ m_b \le k\} && \text{for } i < 0 \text{ even and } \sigma_{i+1}(b) \notin g(B), && \text{(A.31g)}
\end{aligned}$$

where the conditions in (A.31e) and (A.31g) are meant to implicitly require $\sigma_{i+1}(b)$ to be defined for $i+1$. By induction, one shows $\sigma_{i-1}(b) \in B$ for each $i > 0$ odd, $\sigma_{i-1}(b) \in A$ for each $i > 0$ even, $\sigma_{i+1}(b) \in B$ for each $m_b \le i < 0$ odd, and $\sigma_{i+1}(b) \in A$ for each $m_b \le i < 0$ even, such that $\sigma_i(b)$ is well-defined by (A.31) for each $i \in I_b$ (with $I_b = \mathbb{Z}$ if (A.31e) and (A.31g) are never satisfied). The $\sigma(a)$ and $\sigma(b)$ now allow us to define the sets

$$\underset{x\in A\,\dot\cup\, B}{\forall} \quad S_x := \{\sigma_i(x) :\ i \in I_x\} \subseteq A\,\dot\cup\, B. \tag{A.32}$$

Moreover, we call $x \in A\,\dot\cup\, B$ an A-stopper if, and only if, $\sigma(x)$ terminates to the left with some element in A; a B-stopper, if, and only if, $\sigma(x)$ terminates to the left with some element in B; and a non-stopper, if $\sigma(x)$ does never terminate to the left – thus,

$$\begin{aligned}
x\ A\text{-stopper} \quad &\Leftrightarrow \quad \Big(I_x \neq \mathbb{Z} \ \wedge\ \big((x \in A \wedge m_x \text{ even}) \ \vee\ (x \in B \wedge m_x \text{ odd})\big)\Big),\\
x\ B\text{-stopper} \quad &\Leftrightarrow \quad \Big(I_x \neq \mathbb{Z} \ \wedge\ \big((x \in A \wedge m_x \text{ odd}) \ \vee\ (x \in B \wedge m_x \text{ even})\big)\Big),\\
x \text{ non-stopper} \quad &\Leftrightarrow \quad I_x = \mathbb{Z}.
\end{aligned} \tag{A.33}$$

Next, we prove that the S_x form a partition of $A\,\dot\cup\, B$. Since, for each $x \in A\,\dot\cup\, B$, $x = \sigma_0(x) \in S_x$, it only remains to show

$$\underset{x,y\in A\,\dot\cup\, B}{\forall} \quad \Big(S_x = S_y \quad \vee \quad S_x \cap S_y = \emptyset\Big). \tag{A.34}$$

To prove (A.34), it clearly suffices to show

$$\underset{x,z\in A\,\dot\cup\, B}{\forall} \quad \Big(z \in S_x \quad \Rightarrow \quad S_x = S_z\Big). \tag{A.35}$$

To verify (A.35), let $z \in S_x$. Then there exists $i \in I_x$ such that $z = \sigma_0(z) = \sigma_i(x)$ and a simple inductions show $\sigma_k(z) = \sigma_{k+i}(x)$ for each $k \in I_z$ and $\sigma_{k-i}(z) = \sigma_k(x)$ for each $k \in I_x$ (in particular, $i + I_z = I_x$), proving $S_x = S_z$.

We are now in a position to define the desired bijection $\phi :\ A \longrightarrow B$:

$$\phi :\ A \longrightarrow B, \quad \phi(a) := \begin{cases} f(a) & \text{if } a \text{ is an } A\text{-stopper or a non-stopper,}\\ g^{-1}(a) & \text{if } a \text{ is a } B\text{-stopper.}\end{cases} \tag{A.36}$$

Indeed, ϕ is injective: If $a_1, a_2 \in \{a \in A : a \text{ } A\text{-stopper or non-stopper}\}$ with $a_1 \neq a_2$, then $\phi(a_1) \neq \phi(a_2)$ due to f being injective; if $a_1, a_2 \in \{a \in A : a \text{ } B\text{-stopper}\}$ with $a_1 \neq a_2$, then $\phi(a_1) \neq \phi(a_2)$ due to g^{-1} being injective; and $a_1, a_2 \in A$ with a_2 a B-stopper and a_1 not a B-stopper, $S_{a_1} = S_{f(a_1)}$ and $S_{a_2} = S_{g^{-1}(a_2)}$, i.e. $\phi(a_2)$ is also a B-stopper, whereas $\phi(a_1)$ is not a B-stopper, in particular, $\phi(a_1) \neq \phi(a_2)$. Moreover, ϕ is also surjective: If $b \in B$ is a B-stopper, then, due to $S_b = S_{g(b)}$, so is $g(b)$, and $b = g^{-1}(g(b)) = \phi(g(b))$; if $b \in B$ is not a B-stopper, then $f^{-1}(b)$ is defined and in S_b, i.e. $f^{-1}(b)$ is not a B-stopper, either, and $b = f(f^{-1}(b)) = \phi(f^{-1}(b))$.

To conclude, the proof, we consider the case that A and B are not necessarily disjoint. Since $A \times \{0\}$ and $B \times \{1\}$ are always disjoint with

$$\tilde{f} : A \times \{0\} \longrightarrow B \times \{1\}, \quad \tilde{f}(a, 0) := (f(a), 1), \tag{A.37a}$$
$$\tilde{g} : B \times \{1\} \longrightarrow A \times \{0\}, \quad \tilde{g}(b, 0) := (g(b), 0), \tag{A.37b}$$

still being injective if f, g are, the first part of the proof yields a bijective function $\tilde{\phi} : A \times \{0\} \longrightarrow B \times \{1\}$. Then, using the clearly bijective functions

$$\alpha : A \longrightarrow A \times \{0\}, \quad \alpha(a) := (a, 0), \tag{A.38a}$$
$$\beta : B \longrightarrow B \times \{1\}, \quad \beta(b) := (b, 1), \tag{A.38b}$$

$\phi := \beta^{-1} \circ \tilde{\phi} \circ \alpha : A \longrightarrow B$ is also bijective. ■

Remark A.57. In general, the second proof of the Schröder-Bernstein Th. A.55 is still nonconstructive, since one has, in general, no algorithm to determine if a given element is an A-stopper, a B-stopper, or a non-stopper. However, as the following Ex. A.58 shows, in particular situations, determining A-stoppers, B-stoppers, and non-stoppers does not have to be difficult.

Example A.58. Let $A := \mathbb{N}_0$, $B := \{n \in \mathbb{N}_0 : n \text{ even}\}$. We consider A and B as being made disjoint (for example, by using the trick employed in the last part of the proof of Th. A.55 above), but, for the sake of readability, we will not reflect this in the used notation. Define the maps

$$f : A \longrightarrow B, \quad f(n) := 4n, \tag{A.39a}$$
$$g : B \longrightarrow A, \quad g(n) := n, \tag{A.39b}$$

both being clearly injective, but not surjective. The goal is to, explicitly, find the bijective map $\phi : A \longrightarrow B$, given by (A.36). As an intermediate step, we determine which elements of A are non-stoppers, A-stoppers, and B-stoppers, and likewise for the elements of B. Clearly $0 \in A$ and $0 \in B$ are non-stoppers. We will see that all other elements are either A-stoppers or B-stoppers. The precise claim is

$$A_1 := \{a \in A : a \text{ is } A\text{-stopper}\} = \{a \in A : a = n\,4^k,\ n \text{ odd},\ k \in \mathbb{N}_0\}, \tag{A.40a}$$
$$A_2 := \{a \in A : a \text{ is } B\text{-stopper}\} = A \setminus (A_1 \cup \{0\}), \tag{A.40b}$$
$$B_1 := \{b \in B : b \text{ is } A\text{-stopper}\} = B \setminus (B_2 \cup \{0\}), \tag{A.40c}$$
$$B_2 := \{b \in B : b \text{ is } B\text{-stopper}\} = \{b \in B : b = n\,2^k;\ n, k \text{ odd};\ n, k \geq 1\}. \tag{A.40d}$$

To prove (A.40), denote the sets on the right-hand side of (A.40) by C_1, C_2, D_1, D_2, respectively. If $c = n\,4^k \in C_1$, then $(f^{-1} \circ g^{-1})^k(c) = n$ is odd, i.e. $n \notin g(B)$, showing c is an A-stopper, proving $C_1 \subseteq A_1$. If $d = n\,2^k \in D_2$, then $k - 1 = 2m$ with $m \in \mathbb{N}_0$, i.e. $d = n\,2 \cdot 4^m$ and $(g^{-1} \circ f^{-1})^m(d) = 2n$ is not divisable by 4, i.e. $2n \notin f(A)$, showing d is a B-stopper, proving $D_2 \subseteq B_2$. Clearly, each $a \in \mathbb{N}$ either has the form $a = n\,4^k$ with n odd and $k \in \mathbb{N}_0$ (i.e. $a \in C_1$) or $a = 2 \cdot n4^k$ with n odd and $k \in \mathbb{N}_0$, i.e.

$$\begin{aligned} C_2 &= \{a \in A :\ a = 2 \cdot n(2 \cdot 2)^k;\ n \text{ odd};\ k \in \mathbb{N}_0\} \\ &= \{a \in A :\ a = n\,2^k;\ n, k \text{ odd};\ n, k \geq 1\} = g(D_2). \end{aligned} \tag{A.41}$$

Since $D_2 \subseteq B_2$, all elements of D_2 are B-stoppers, and, thus, so are all elements of C_2, proving $C_2 \subseteq A_2$. Since $A = C_1 \,\dot\cup\, C_2 \,\dot\cup \{0\}$, we then also obtain $A_1 = C_1$ and $A_2 = C_2$. Clearly, each even $b \in \mathbb{N}$ either has the form $b = n\,2^k$ with odd $n, k \geq 1$ (i.e. $b \in D_2$) or $b = n4^k$ with n odd and $k \in \mathbb{N}$, i.e.

$$D_1 = \{b \in B :\ b = n\,4^k,\ n \text{ odd},\ k \in \mathbb{N}\} = f(C_1). \tag{A.42}$$

Since $C_1 = A_1$, all elements of C_1 are A-stoppers, and, thus, so are all elements of D_1. Since $B = D_1 \,\dot\cup\, D_2 \,\dot\cup \{0\}$, we then also obtain $B_1 = D_1$ and $B_2 = D_2$.

Now that we have identified explicit formulas for A_1 and A_2, we can write the assignment rule for the bijective $\phi :\ A \longrightarrow B$, given by (A.36), in the explicit form

$$\phi(a) := \begin{cases} 0 & \text{if } a = 0, \\ 4a & \text{if } a = n\,4^k \text{ with } n \text{ odd and } k \in \mathbb{N}_0, \\ a & \text{if } a = 2 \cdot n4^k \text{ with } n \text{ odd and } k \in \mathbb{N}_0. \end{cases} \tag{A.43}$$

Thus, ϕ starts out with the assignments

$$\begin{array}{rccccccccccc} & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & \\ \phi : & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \ldots \\ & 0 & 4 & 2 & 12 & 16 & 20 & 6 & 28 & 8 & \end{array} \tag{A.44}$$

Theorem A.59. *Let A, B be nonempty sets. Then the following statements are equivalent (where the implication "(ii) $\Rightarrow$ (i)" makes use of the axiom of choice (AC) of Sec.* A.4*).*

(i) *There exists an injective map $f :\ A \longrightarrow B$.*

(ii) *There exists a surjective map $g :\ B \longrightarrow A$.*

Proof. According to Th. 2.13(b), (i) is equivalent to f having a left inverse $g :\ B \longrightarrow A$ (i.e. $g \circ f = \mathrm{Id}_A$), which is equivalent to g having a right inverse, which, according to Th. 2.13(a), is equivalent to (ii) (AC is used in the proof of Th. 2.13(a) to show each surjective map has a right inverse). ■

Corollary A.60. *Let A, B be nonempty sets. Using AC, we can expand the two equivalent statements of Th.* A.55 *to the following list of equivalent statements:*

(i) *The sets A and B have the same cardinality (i.e. there exists a bijective map $\phi : A \longrightarrow B$).*

(ii) *There exist an injective map $f : A \longrightarrow B$ and an injective map $g : B \longrightarrow A$.*

(iii) *There exist a surjective map $f : A \longrightarrow B$ and a surjective map $g : B \longrightarrow A$.*

(iv) *There exist an injective map $f_1 : A \longrightarrow B$ and a surjective map $f_2 : A \longrightarrow B$.*

(v) *There exist an injective map $g_1 : B \longrightarrow A$ and a surjective map $g_2 : B \longrightarrow A$.*

Proof. The equivalences are an immediate consequence of combining Th. A.55 with Th. A.59. ■

A.5.2 Finite Sets

It is intuitively clear that finite cardinalities are uniquely determined. Still one has to provide a rigorous proof. The key is the following theorem:

Theorem A.61. *If $m, n \in \mathbb{N}$ and the map $f : \{1, \ldots, m\} \longrightarrow \{1, \ldots, n\}$ is bijective, then $m = n$.*

Proof. We conduct the proof via induction on m. If $m = 1$, then the surjectivity of f implies $n = 1$. For the induction step, we now consider $m > 1$. From the bijective map f, we define the map

$$g : \{1, \ldots, m\} \longrightarrow \{1, \ldots, n\}, \quad g(x) := \begin{cases} n & \text{for } x = m, \\ f(m) & \text{for } x = f^{-1}(n), \\ f(x) & \text{otherwise.} \end{cases} \tag{A.45}$$

Then g is bijective, since it is the composition $g = h \circ f$ of the bijective map f with the bijective map

$$h : \{f(m), n\} \longrightarrow \{f(m), n\}, \quad h(f(m)) := n, \quad h(n) := f(m). \tag{A.46}$$

Thus, the restriction $g \restriction_{\{1,\ldots,m-1\}} : \{1, \ldots, m-1\} \longrightarrow \{1, \ldots, n-1\}$ must also be bijective, such that the induction hypothesis yields $m - 1 = n - 1$, which, in turn, implies $m = n$ as desired. ■

Corollary A.62. *Let $m, n \in \mathbb{N}$ and let A be a set. If $\#A = m$ and $\#A = n$, then $m = n$.*

Proof. If $\#A = m$, then, according to Def. 3.12(b), there exists a bijective map $f : A \longrightarrow \{1, \ldots, m\}$. Analogously, if $\#A = n$, then there exists a bijective map $g : A \longrightarrow \{1, \ldots, n\}$. In consequence, we have the bijective map $(g \circ f^{-1}) : \{1, \ldots, m\} \longrightarrow \{1, \ldots, n\}$, such that Th. A.61 yields $m = n$. ■

Theorem A.63. *Let $A \neq \emptyset$ be a finite set.*

(a) *If $B \subseteq A$ with $A \neq B$, then B is finite with $\#B < \#A$.*

(b) *If $a \in A$, then $\#\big(A \setminus \{a\}\big) = \#A - 1$.*

Proof. For $\#A = 0$, i.e. $A = \emptyset$, (a) and (b) are trivially true, since A has neither strict subsets nor elements. For $\#A = n \in \mathbb{N}$, we use induction to prove (a) and (b) simultaneously, i.e. we show

$$\underset{n\in\mathbb{N}}{\forall} \underbrace{\left(\#A = n \;\Rightarrow\; \underset{B\in\mathcal{P}(A)\setminus\{A\}}{\forall}\; \underset{a\in A}{\forall} \#B \in \{0,\ldots,n-1\} \wedge \#\big(A\setminus\{a\}\big) = n-1\right)}_{\phi(n)}.$$

Base Case ($n = 1$): In this case, A has precisely one element, i.e. $B = A \setminus \{a\} = \emptyset$, and $\#\emptyset = 0 = n - 1$ proves $\phi(1)$.

Induction Step: For the induction hypothesis, we assume $\phi(n)$ to be true, i.e. we assume (a) and (b) hold for each A with $\#A = n$. We have to prove $\phi(n+1)$, i.e., we consider A with $\#A = n+1$. From $\#A = n+1$, we conclude the existence of a bijective map $\varphi : A \longrightarrow \{1,\ldots,n+1\}$. We have to construct a bijective map $\psi : A \setminus \{a\} \longrightarrow \{1,\ldots,n\}$. To this end, set $k := \varphi(a)$ and define the auxiliary function

$$f : \{1,\ldots,n+1\} \longrightarrow \{1,\ldots,n+1\}, \quad f(x) := \begin{cases} n+1 & \text{for } x = k, \\ k & \text{for } x = n+1, \\ x & \text{for } x \notin \{k, n+1\}. \end{cases}$$

Then $f \circ \varphi : A \longrightarrow \{1,\ldots,n+1\}$ is bijective by Th. 2.14, and

$$(f \circ \varphi)(a) = f(\varphi(a)) = f(k) = n+1.$$

Thus, the restriction $\psi := f\restriction_{A\setminus\{a\}}$ is the desired bijective map $\psi : A\setminus\{a\} \longrightarrow \{1,\ldots,n\}$, proving $\#\big(A \setminus \{a\}\big) = n$. It remains to consider the strict subset B of A. Since B is a strict subset of A, there exists $a \in A \setminus B$. Thus, $B \subseteq A \setminus \{a\}$ and, as we have already shown $\#\big(A \setminus \{a\}\big) = n$, the induction hypothesis applies and yields B is finite with $\#B \leq \#\big(A \setminus \{a\}\big) = n$, i.e. $\#B \in \{0,\ldots,n\}$, proving $\phi(n+1)$, thereby completing the induction. ∎

Theorem A.64. *For $\#A = \#B = n \in \mathbb{N}$ and $f : A \longrightarrow B$, the following statements are equivalent:*

(i) *f is injective.*

(ii) *f is surjective.*

(iii) *f is bijective.*

Proof. It suffices to prove the equivalence of (i) and (ii).

If f is injective, then $f : A \longrightarrow f(A)$ is bijective. Since $\#A = n$, there exists a bijective map $\varphi : A \longrightarrow \{1, \ldots, n\}$. Then $(\varphi \circ f^{-1}) : f(A) \longrightarrow \{1, \ldots, n\}$ is also bijective, showing $\#f(A) = n$, i.e., according to Th. A.63(a), $f(A)$ can not be a strict subset of B, i.e. $f(A) = B$, proving f is surjective.

If f is surjective, then f has a right inverse $g : B \longrightarrow A$: One can obtain this from Th. 2.13(a), but, here, we can actually construct g without the axiom of choice: We let $\varphi : A \longrightarrow \{1, \ldots, n\}$ be the bijective map from above and, for $b \in B$, we let $g(b)$ be the unique $a \in C := f^{-1}(\{b\})$ such that $\varphi(a) = \min \varphi(C)$. Then, clearly, $f \circ g = \mathrm{Id}_B$. But this also means f is a left inverse for g, such that g must be injective by Th. 2.13(b). According to what we have already proved above, g injective implies g surjective, i.e. g must be bijective. From Th. 2.13(c), we then know the left inverse of g is unique, implying $f = g^{-1}$. In particular, f is injective. ∎

Lemma A.65. *For each finite set A (i.e. $\#A = n \in \mathbb{N}_0$) and each $B \subseteq A$, one has $\#(A \setminus B) = \#A - \#B$.*

Proof. For $B = \emptyset$, the assertion is true since $\#(A \setminus B) = \#A = \#A - 0 = \#A - \#B$.

For $B \neq \emptyset$, the proof is conducted over the size of B, i.e. as a finite induction (cf. Cor. 3.6) over the set $\{1, \ldots, n\}$, showing

$$\forall_{m \in \{1,\ldots,n\}} \underbrace{\big(\#B = m \;\Rightarrow\; \#(A \setminus B) = \#A - \#B\big)}_{\phi(m)}.$$

Base Case ($m = 1$): $\phi(1)$ is precisely the statement provided by Th. A.63(b).

Induction Step: For the induction hypothesis, we assume $\phi(m)$ with $1 \leq m < n$. To prove $\phi(m+1)$, consider $B \subseteq A$ with $\#B = m + 1$. Fix an element $b \in B$ and set $B_1 := B \setminus \{b\}$. Then $\#B_1 = m$ by Th. A.63(b), $A \setminus B = (A \setminus B_1) \setminus \{b\}$, and we compute

$$\begin{aligned} \#(A \setminus B) &= \#\big((A \setminus B_1) \setminus \{b\}\big) \overset{\text{Th. A.63(b)}}{=} \#(A \setminus B_1) - 1 \overset{(\phi(m))}{=} \#A - \#B_1 - 1 \\ &= \#A - \#B, \end{aligned}$$

proving $\phi(m+1)$ and completing the induction. ∎

Theorem A.66. *If A, B are finite sets, then $\#(A \cup B) = \#A + \#B - \#(A \cap B)$.*

Proof. The assertion is clearly true if A or B is empty. If A and B are nonempty, then there exist $m, n \in \mathbb{N}$ such that $\#A = m$ and $\#B = n$, i.e. there are bijective maps $f : A \longrightarrow \{1, \ldots, m\}$ and $g : B \longrightarrow \{1, \ldots, n\}$.

We first consider the case $A \cap B = \emptyset$. We need to construct a bijective map $h : A \cup B \longrightarrow \{1, \ldots, m+n\}$. To this end, we define

$$h : A \cup B \longrightarrow \{1, \ldots, m+n\}, \quad h(x) := \begin{cases} f(x) & \text{for } x \in A, \\ g(x) + m & \text{for } x \in B. \end{cases}$$

The bijectivity of f and g clearly implies the bijectivity of h, proving $\#(A \cup B) = m + n = \#A + \#B$.

Finally, we consider the case of arbitrary A, B. Since $A \cup B = A \dot\cup (B \setminus A)$ and $B \setminus A = B \setminus (A \cap B)$, we can compute

$$\begin{aligned}\#(A \cup B) &= \#\big(A \dot\cup (B \setminus A)\big) = \#A + \#(B \setminus A) \\ &= \#A + \#\big(B \setminus (A \cap B)\big) \overset{\text{Lem. A.65}}{=} \#A + \#B - \#(A \cap B),\end{aligned}$$

thereby establishing the case. ∎

Theorem A.67. *If $(A_1, \ldots, A_n)$, $n \in \mathbb{N}$, is a finite sequence of finite sets, then*

$$\#\prod_{i=1}^{n} A_i = \#\big(A_1 \times \cdots \times A_n\big) = \prod_{i=1}^{n} \#A_i. \tag{A.47}$$

Proof. If at least one A_i is empty, then (A.47) is true, since both sides are 0.

The case where all A_i are nonempty is proved by induction over n, i.e. we know $k_i := \#A_i \in \mathbb{N}$ for each $i \in \{1, \ldots, n\}$ and show by induction

$$\underset{n\in\mathbb{N}}{\forall} \underbrace{\#\prod_{i=1}^{n} A_i = \prod_{i=1}^{n} k_i}_{\phi(n)}.$$

Base Case ($n = 1$): $\prod_{i=1}^{1} A_i = \#A_1 = k_1 = \prod_{i=1}^{1} k_i$, i.e. $\phi(1)$ holds.

Induction Step: From the induction hypothesis $\phi(n)$, we obtain a bijective map $\varphi : A \longrightarrow \{1, \ldots, N\}$, where $A := \prod_{i=1}^{n} A_i$ and $N := \prod_{i=1}^{n} k_i$. To prove $\phi(n+1)$, we need to construct a bijective map $h : A \times A_{n+1} \longrightarrow \{1, \ldots, N \cdot k_{n+1}\}$. Since $\#A_{n+1} = k_{n+1}$, there exists a bijective map $f : A_{n+1} \longrightarrow \{1, \ldots, k_{n+1}\}$. We define

$$\begin{aligned} h : A \times A_{n+1} &\longrightarrow \{1, \ldots, N \cdot k_{n+1}\}, \\ h(a_1, \ldots, a_n, a_{n+1}) &:= \big(f(a_{n+1}) - 1\big) \cdot N + \varphi(a_1, \ldots, a_n). \end{aligned}$$

Since φ and f are bijective, and since every $m \in \{1, \ldots, N \cdot k_{n+1}\}$ has a unique representation in the form $m = a \cdot N + r$ with $a \in \{0, \ldots, k_{n+1} - 1\}$ and $r \in \{1, \ldots, N\}$ (exercise), h is also bijective. This proves $\phi(n+1)$ and completes the induction. ∎

Theorem A.68. *For each finite set A (i.e. $\#A = n \in \mathbb{N}_0$), one has $\#\mathcal{P}(A) = 2^n$.*

Proof. The proof is conducted by induction by showing

$$\underset{n\in\mathbb{N}_0}{\forall} \underbrace{\big(\#A = n \;\Rightarrow\; \#\mathcal{P}(A) = 2^n\big)}_{\phi(n)}.$$

Base Case ($n = 0$): For $n = 0$, we have $A = \emptyset$, i.e. $\mathcal{P}(A) = \{\emptyset\}$. Thus, $\#\mathcal{P}(A) = 1 = 2^0$, proving $\phi(0)$.

Induction Step: Assume $\phi(n)$ and consider A with $\#A = n+1$. Then A contains at least one element a. For $B := A \setminus \{a\}$, we then know $\#B = n$ from Th. A.63(b). Moreover, setting $\mathcal{M} := \big\{C \cup \{a\} : C \in \mathcal{P}(B)\big\}$, we have the disjoint decomposition $\mathcal{P}(A) = \mathcal{P}(B) \,\dot\cup\, \mathcal{M}$. As the map $\varphi : \mathcal{P}(B) \longrightarrow \mathcal{M}$, $\varphi(C) := C \cup \{a\}$, is clearly bijective, $\mathcal{P}(B)$ and $\mathcal{M}$ have the same cardinality. Thus,

$$\#\mathcal{P}(A) \overset{\text{Th. A.66}}{=} \#\mathcal{P}(B) + \#\mathcal{M} = \#\mathcal{P}(B) + \#\mathcal{P}(B) \overset{(\phi(n))}{=} 2 \cdot 2^n = 2^{n+1},$$

thereby proving $\phi(n+1)$ and completing the induction. ■

A.5.3 Power Sets

Theorem A.69. *Let A be a set. There can never exist a surjective map from A onto $\mathcal{P}(A)$ (in this sense, the size of $\mathcal{P}(A)$ is always strictly bigger than the size of A; in particular, A and $\mathcal{P}(A)$ can never have the same size).*

Proof. If $A = \emptyset$, then there is nothing to prove. For nonempty A, the idea is to conduct a proof by contradiction. To this end, assume there does exist a surjective map $f : A \longrightarrow \mathcal{P}(A)$ and define

$$B := \{x \in A : x \notin f(x)\}. \tag{A.48}$$

Now B is a subset of A, i.e. $B \in \mathcal{P}(A)$ and the assumption that f is surjective implies the existence of $a \in A$ such that $f(a) = B$. If $a \in B$, then $a \notin f(a) = B$, i.e. $a \in B$ implies $a \in B \wedge \neg(a \in B)$, so that the principle of contradiction tells us $a \notin B$ must be true. However, $a \notin B$ implies $a \in f(a) = B$, i.e., this time, the principle of contradiction tells us $a \in B$ must be true. In conclusion, we have shown our original assumption that there exists a surjective map $f : A \longrightarrow \mathcal{P}(A)$ implies $a \in B \wedge \neg(a \in B)$, i.e., according to the principle of contradiction, no surjective map from A into $\mathcal{P}(A)$ can exist. ■

B General Forms of the Laws of Commutativity and Associativity

B.1 Commutativity

In the present section, we will generalize the law of commutativity $ab = ba$ to a finite number of factors. For this purpose, we introduce the notion of *permutation*, also useful in many other mathematical contexts.

Definition and Remark B.1. Let $n \in \mathbb{N}$. Each bijective map $\pi : \{1, \ldots, n\} \longrightarrow \{1, \ldots, n\}$ is called a *permutation* of $\{1, \ldots, n\}$. The set of permutations of $\{1, \ldots, n\}$ forms a group with respect to the composition of maps, the so-called *symmetric group* S_n: Indeed, the composition of maps is associative by Prop. 2.10(a); the neutral element

is the identity map $e : \{1, \dots, n\} \longrightarrow \{1, \dots, n\}$, $e(i) = i$; and, for each $\pi \in S_n$, its inverse map π^{-1} is also its inverse element in the group S_n. Caveat: Simple examples show that S_n is *not* commutative.

Theorem B.2 (General Law of Commutativity)**.** *Let A be a set with a composition $\cdot : A \times A \longrightarrow A$ (which we write as a multiplication, but, clearly, this is not essential, and we could also write it as an addition or with some other symbol). If the composition is commutative, i.e. if*

$$\forall_{a,b \in A} \quad ab = ba, \tag{B.1}$$

then

$$\forall_{n \in \mathbb{N}} \quad \forall_{\pi \in S_n} \quad \forall_{a_1, \dots, a_n \in A} \quad \prod_{i=1}^{n} a_i = \prod_{i=1}^{n} a_{\pi(i)}. \tag{B.2}$$

Before we can carry out the proof, we need to learn a bit more about permutations.

Definition B.3. Let $k, n \in \mathbb{N}$, $k \leq n$. A permutation $\pi \in S_n$ is called a k-cycle if, and only if, there exist k distinct numbers $i_1, \dots, i_k \in \{1, \dots, n\}$ such that

$$\pi(i) = \begin{cases} i_{j+1} & \text{if } i = i_j,\ j \in \{1, \dots, k-1\}, \\ i_1 & \text{if } i = i_k, \\ i & \text{if } i \notin \{i_1, \dots, i_k\}. \end{cases} \tag{B.3}$$

If π is a cycle as in (B.3), then one writes

$$\pi = (i_1\ i_2\ \dots\ i_k). \tag{B.4}$$

Each 2-cycle is also known as a *transposition.*

Theorem B.4. *Let $n \in \mathbb{N}$.*

(a) *Each permutation can be decomposed into finitely many disjoint cycles: For each $\pi \in S_n$, there exists a decomposition of $\{1, \dots, n\}$ into disjoint sets $A_1, \dots, A_N$, $N \in \mathbb{N}$, i.e.*

$$\{1, \dots, n\} = \bigcup_{i=1}^{N} A_i \quad \text{and} \quad A_i \cap A_j = \emptyset \quad \text{for } i \neq j, \tag{B.5}$$

such that A_i consists of the distinct elements $a_{i1}, \dots, a_{i,N_i}$ and

$$\pi = (a_{N1}\ \dots\ a_{N,N_N}) \cdots (a_{11}\ \dots\ a_{1,N_1}). \tag{B.6}$$

The decomposition (B.6) *is unique up to the order of the cycles.*

(b) *If $n \geq 2$, then every permutation $\pi \in S_n$ is the composition of finitely many transpositions, where each transposition permutes two juxtaposed elements, i.e.*

$$\forall_{\pi \in S_n} \quad \exists_{N \in \mathbb{N}} \quad \exists_{\tau_1, \dots, \tau_N \in T} \quad \pi = \tau_N \circ \cdots \circ \tau_1, \tag{B.7}$$

where $T := \big\{(i\ \ i+1) : i \in \{1, \dots, n-1\}\big\}$.

Proof. (a): We prove the statement by induction on n. For $n = 1$, there is nothing to prove. Let $n > 1$ and choose $i \in \{1, \dots, n\}$. We claim that

$$\underset{k\in\mathbb{N}}{\exists} \quad \left(\pi^k(i) = i \ \wedge \ \underset{l\in\{1,\dots,k-1\}}{\forall} \quad \pi^l(i) \neq i\right). \tag{B.8}$$

Indeed, since $\{1, \dots, n\}$ is finite, there must be a smallest $k \in \mathbb{N}$ such that $\pi^k(i) \in A_1 := \{i, \pi(i), \dots, \pi^{k-1}(i)\}$. Since π is bijective, it must be $\pi^k(i) = i$ and $(i\ \pi(i)\ \dots, \pi^{k-1}(i))$ is a k-cycle. We are already done in case $k = n$. If $k < n$, then consider $B := \{1, \dots, n\} \setminus A_1$. Then, again using the bijectivity of π, $\pi\restriction_B$ is a permutation on B with $1 \leq \#B < n$. By induction, there are disjoint sets $A_2, \dots, A_N$ such that $B = \bigcup_{j=2}^N A_j$, A_j consists of the distinct elements $a_{j1}, \dots, a_{j,N_j}$ and

$$\pi\restriction_B = (a_{N1}\ \dots\ a_{N,N_N}) \cdots (a_{21}\ \dots\ a_{2,N_2}).$$

Since $\pi = (i\ \pi(i)\ \dots, \pi^{k-1}(i)) \circ \pi\restriction_B$, this finishes the proof of (B.6). If there were another, different, decomposition of π into cycles, say, given by disjoint sets $B_1, \dots, B_M$, $\{1, \dots, n\} = \bigcup_{i=1}^M B_i$, $M \in \mathbb{N}$, then there were $A_i \neq B_j$ and $k \in A_i \cap B_j$. But then k were in the cycle given by A_i and in the cycle given by B_j, implying $A_i = \{\pi^l(k) : l \in \mathbb{N}\} = B_j$, in contradiction to $A_i \neq B_j$.

(b): We first show that every $\pi \in S_n$ is a composition of finitely many transpositions (not necessarily transpositions from the set T): According to (a), it suffices to show that every cycle is a composition of finitely many transpositions. Since each 1-cycle is the identity, it is $(i) = \mathrm{Id} = (1\ 2)\,(1\ 2)$ for each $i \in \{1, \dots, n\}$. If $(i_1\ \dots\ i_k)$ is a k-cycle, $k \in \{2, \dots, n\}$, then

$$(i_1\ \dots\ i_k) = (i_1\ i_2)\,(i_2\ i_3) \cdots (i_{k-1}\ i_k) : \tag{B.9}$$

Indeed,

$$\underset{i\in\{1,\dots,n\}}{\forall} \quad (i_1\ i_2)\,(i_2\ i_3) \cdots (i_{k-1}\ i_k)(i) = \begin{cases} i_1 & \text{for } i = i_k, \\ i_{l+1} & \text{for } i = i_l,\ l \in \{1, \dots, k-1\}, \\ i & \text{for } i \notin \{i_1, \dots, i_k\}, \end{cases} \tag{B.10}$$

proving (B.9). To finish the proof of (b), we observe that every transposition is a composition of finitely many elements of T: If $i, j \in \{1, \dots, n\}$, $i < j$, then

$$(i\ j) = (i\ \ i+1) \cdots (j-2\ j-1)(j-1\ \ j) \cdots (i+1\ \ i+2)(i\ \ i+1) : \tag{B.11}$$

Indeed,

$$\begin{aligned} &\underset{k\in\{1,\dots,n\}}{\forall} \quad (i\ \ i+1) \cdots (j-2\ j-1)(j-1\ \ j) \cdots (i+1\ \ i+2)(i\ \ i+1)(k) \\ &\qquad = \begin{cases} j & \text{for } k = i, \\ i & \text{for } k = j, \\ k & \text{for } i < k < j, \\ k & \text{for } k \notin \{i, i+1, \dots, j\}, \end{cases} \end{aligned} \tag{B.12}$$

proving (B.11). ∎

Proof of Th. B.2. For $n = 1$, there is nothing to prove. So let $n > 1$. For $l \in 1, \ldots, n-1$, let $\tau_l : \{1, \ldots, n\} \longrightarrow \{1, \ldots, n\}$ be the transposition that interchanges l and $l+1$ and leaves all other elements fixed (i.e. $\tau_l(l) = l+1$, $\tau_l(l+1) = l$, $\tau(\alpha) = \alpha$ for each $\alpha \in \{1, \ldots, n\} \setminus \{l, l+1\}$) and let $T := \{\tau_1, \ldots, \tau_{n-1}\}$. Then (B.1) directly implies that the theorem holds for $\pi = \tau$ for each $\tau \in T$. For a general permutation $\pi \in S_n$, Th. B.4(b) provides a finite sequence $(\tau^1, \ldots, \tau^N)$, $N \in \mathbb{N}$, of elements of T such that $\pi = \tau^N \circ \cdots \circ \tau^1$. Thus, as we already know that the theorem holds for $N = 1$, the case $N > 1$ follows by induction. ■

B.2 Associativity

In the literature, the general law of associativity is often stated in the form that $a_1 a_2 \cdots a_n$ gives the same result "for every admissible way of inserting parentheses into $a_1 a_2 \cdots a_n$", but a completely precise formulation of what that actually means seems to be rare. As a warm-up, we first prove a special case of the general law:

Proposition B.5. *Let A be a set with a composition $\cdot : A \times A \longrightarrow A$ (which we write as a multiplication, but, clearly, this is not essential, and we could also write it as an addition or with some other symbol). If the composition is associative, i.e. if*

$$\underset{a,b,c \in A}{\forall} \quad (ab)c = a(bc), \tag{B.13}$$

then

$$\underset{\substack{n \in \mathbb{N}, \\ n \geq 2}}{\forall} \quad \underset{a_1, \ldots, a_n \in A}{\forall} \quad \underset{k \in \{2, \ldots, n\}}{\forall} \quad \left(\prod_{i=k}^{n} a_i \right) \left(\prod_{i=1}^{k-1} a_i \right) = \prod_{i=1}^{n} a_i, \tag{B.14}$$

where the product symbol is defined according to (3.20a).

Proof. If $k = n$, then (B.14) is immediate from (3.20a). For $2 \leq k < n$, we prove (B.14) by induction on n: For the base case, $n = 2$, there is nothing to prove. For $n > 2$, one computes

$$\begin{aligned} \left(\prod_{i=k}^{n} a_i \right) \left(\prod_{i=1}^{k-1} a_i \right) &\overset{(3.20a)}{=} \left(a_n \cdot \prod_{i=k}^{n-1} a_i \right) \left(\prod_{i=1}^{k-1} a_i \right) \overset{(B.13)}{=} a_n \cdot \left(\left(\prod_{i=k}^{n-1} a_i \right) \left(\prod_{i=1}^{k-1} a_i \right) \right) \\ &\overset{\text{ind.hyp.}}{=} a_n \cdot \prod_{i=1}^{n-1} a_i \overset{(3.20a)}{=} \prod_{i=1}^{n} a_i, \end{aligned} \tag{B.15}$$

completing the induction and the proof of the proposition. ■

The difficulty in stating the general form of the law of associativity lies in giving a precise definition of what one means by "an admissible way of inserting parentheses into $a_1 a_2 \cdots a_n$". So how does one actually proceed to calculate the value of $a_1 a_2 \cdots a_n$, given that parentheses have been inserted in an admissible way? The answer is that one does it in $n-1$ steps, where, in each step, one combines two juxtaposed elements,

consistent with the inserted parentheses. There can still be some ambiguity: For example, for $(a_1a_2)(a_3(a_4a_5))$, one has the freedom of first combining a_1, a_2, or of first combining a_4, a_5. In consequence, our general law of associativity will show that, for each admissible sequence of $n-1$ directives for combining two juxtaposed elements, the final result is the same (under the hypothesis that (B.13) holds). This still needs some preparatory work.

In the following, one might see it as a slight notational inconvenience that we have defined $\prod_{i=1}^{n} a_i$ as $a_n \cdots a_1$ rather than $a_1 \cdots a_n$. For this reason, we will enumerate the elements to be combined by composition from right to left rather then from left to right.

Definition and Remark B.6. Let A be a (nonempty) set with a composition $\cdot : A \times A \longrightarrow A$, let $n \in \mathbb{N}$, $n \geq 2$, and let I be a totally ordered index set, $\#I = n$, $I = \{i_1, \ldots, i_n\}$ with $i_1 < \cdots < i_n$. Moreover, let $F := (a_{i_n}, \ldots, a_{i_1})$ be a family of n elements of A.

(a) An *admissible composition directive* (for combining two juxtaposed elements of the family) is an index $i_k \in I$ with $1 \leq k \leq n-1$. It transforms the family F into the family $G := (a_{i_n}, \ldots, a_{i_{k+1}}a_{i_k}, \ldots, a_{i_1})$. In other words, $G = (b_j)_{j\in J}$, where $J := I \setminus \{i_{k+1}\}$, $b_j = a_j$ for each $j \in J \setminus \{i_k\}$, and $b_{i_k} = a_{i_{k+1}}a_{i_k}$. We can write this transformation as two maps

$$F \mapsto \delta^{(1)}_{i_k}(F) := G = (a_{i_n}, \ldots, a_{i_{k+1}}a_{i_k}, \ldots, a_{i_1}) = (b_j)_{j\in J}, \tag{B.16a}$$

$$I \mapsto \delta^{(2)}_{i_k}(I) := J = I \setminus \{i_{k+1}\}. \tag{B.16b}$$

Thus, an application of an admissible composition directive reduces the length of the family and the number indices by one.

(b) Recursively, we define (finite) sequences of families, index sets, and indices as follows:

$$F_n := F, \quad I_n := I, \tag{B.17a}$$

$$\underset{\alpha\in\{2,\ldots,n\}}{\forall} \quad F_{\alpha-1} := \delta^{(1)}_{j_\alpha}(F_\alpha), \quad I_{\alpha-1} := \delta^{(2)}_{j_\alpha}(I_\alpha), \quad \text{where } j_\alpha \in I_\alpha \setminus \{\max I_\alpha\}. \tag{B.17b}$$

The corresponding sequence of indices $\mathcal{D} := (j_n, \ldots, j_2)$ in I is called an *admissible evaluation directive*. Clearly,

$$\underset{\alpha\in\{1,\ldots,n\}}{\forall} \quad \#I_\alpha = \alpha, \quad \text{i.e. } F_\alpha \text{ has length } \alpha. \tag{B.18}$$

In particular, $I_1 = \{j_2\} = \{i_1\}$ (where the second equality follows from (B.16b)), $F_1 = (a)$, and we call

$$\mathcal{D}(F) := a \tag{B.19}$$

the *result* of the admissible evaluation directive $\mathcal{D}$ applied to F.

Theorem B.7 (General Law of Associativity)**.** *Let A be a (nonempty) set with a composition $\cdot : A \times A \longrightarrow A$, let $n \in \mathbb{N}$, $n \geq 2$, and let I be a totally ordered index set, $\#I = n$, $I = \{i_1, \ldots, i_n\}$ with $i_1 < \cdots < i_n$. Moreover, let $F := (a_{i_n}, \ldots, a_{i_1})$ be a family of n elements of A. If the composition is associative, i.e. if* (B.13) *holds, then, for each admissible evaluation directive as defined in Def. and Rem.* B.6(b)*, the result is the same, namely*

$$\mathcal{D}(F) = \prod_{k=1}^{n} a_{i_k}. \tag{B.20}$$

Proof. We conduct the proof via induction on n. For $n = 3$, there are only two possible directives and (B.13) guarantees that they yield the same result. For the induction step, let $n > 3$. As in Def. and Rem. B.6(b), we write $\mathcal{D} = (j_n, \ldots, j_2)$ and obtain some $I_2 = \{i_1, i_m\}$, $1 < m \leq n$, as the corresponding penultimate index set. Depending on i_m, we partition $\mathcal{D}$ as follows: Set

$$J_1 := \big\{k \in \{3, \ldots, n\} : j_k < i_m\big\}, \quad J_2 := \big\{k \in \{3, \ldots, n\} : j_k \geq i_m\big\}. \tag{B.21}$$

Then J_1 and J_2 might or might not be the empty set; however they cannot both be the empty set. If $J_1 \neq \emptyset$, then $\mathcal{D}_1 := (j_k)_{k \in J_1}$ is an admissible evaluation directive for $(a_{i_1}, \ldots, a_{i_{m-1}})$ – this follows from

$$K \subseteq \{i_1, \ldots, i_{m-1}\} \wedge j_k \in J_1 \cap K \;\Rightarrow\; \delta^{(2)}_{j_k}(K) \subseteq \{i_1, \ldots, i_{m-1}\}. \tag{B.22}$$

Since $m - 1 < n$, the induction hypothesis applies and yields

$$\mathcal{D}_1(a_{i_1}, \ldots, a_{i_{m-1}}) = \prod_{k=1}^{m-1} a_{i_k}. \tag{B.23}$$

Analogously, if $J_2 \neq \emptyset$, then $\mathcal{D}_2 := (j_k)_{k \in J_2}$ is an admissible evaluation directive for $(a_{i_m}, \ldots, a_{i_n})$ – this follows from

$$K \subseteq \{i_m, \ldots, i_n\} \wedge j_k \in J_2 \cap K \;\Rightarrow\; \delta^{(2)}_{j_k}(K) \subseteq \{i_m, \ldots, i_n\}. \tag{B.24}$$

Since $m > 1$, the induction hypothesis applies and yields

$$\mathcal{D}_2(a_{i_m}, \ldots, a_{i_n}) = \prod_{k=m}^{n} a_{i_k}. \tag{B.25}$$

Thus, if $J_1 \neq \emptyset$ and $J_2 \neq \emptyset$, then we obtain

$$\mathcal{D}(F) \overset{j_2 = i_1}{=} \mathcal{D}_2(a_{i_m}, \ldots, a_{i_n}) \cdot \mathcal{D}_1(a_{i_1}, \ldots, a_{i_{m-1}}) = \left(\prod_{k=m}^{n} a_{i_k}\right)\left(\prod_{k=1}^{m-1} a_{i_k}\right) \overset{\text{Prop. B.5}}{=} \prod_{k=1}^{n} a_{i_k} \tag{B.26}$$

as desired. If $J_1 = \emptyset$, then $j_k \neq i_1$ for each $k \in \{3, \ldots, n\}$, implying $m = 2$. Thus, in this case,

$$\mathcal{D}(F) \overset{j_2 = i_1}{=} \mathcal{D}_2(a_{i_2}, \ldots, a_{i_n}) \cdot a_{i_1} = \left(\prod_{k=2}^{n} a_{i_k}\right) \cdot a_{i_1} \overset{\text{Prop. B.5}}{=} \prod_{k=1}^{n} a_{i_k} \tag{B.27}$$

as needed. Finally, if $J_2 = \emptyset$, then, in each of the $n-2$ steps to obtain I_2, an i_k with $k < m$ was removed from I, implying $m = n$. Thus, in this case,

$$\mathcal{D}(F) \overset{j_2=i_1}{=} a_{i_n} \cdot \mathcal{D}_1(a_{i_1}, \ldots, a_{i_{n-1}}) = \prod_{k=1}^{n} a_{i_k}, \tag{B.28}$$

again, as desired, and completing the induction. ■

C Algebraic Structures

C.1 Groups

Definition C.1. Let G be a nonempty set with a map

$$\circ : G \times G \longrightarrow G, \quad (x, y) \mapsto x \circ y \tag{C.1}$$

(called a *composition* on G, the examples we have in mind are addition and multiplication on $\mathbb{R}$). Then $(G, \circ)$ (or just G, if the composition $\circ$ is understood) is called a *group* if, and only if, the following three conditions are satisfied:

(i) Associativity: $x \circ (y \circ z) = (x \circ y) \circ z$ holds for all $x, y, z \in G$.

(ii) There exists a *neutral element* $e \in G$, i.e. an element $e \in G$ such that

$$\underset{x \in G}{\forall} \; x \circ e = x. \tag{C.2}$$

(iii) For each $x \in G$, there exists an *inverse element* $\overline{x} \in G$, i.e. an element $\overline{x} \in G$ such that

$$x \circ \overline{x} = e.$$

G is called a *commutative* or *abelian* group if, and only if, it is a group and satisfies the additional condition:

(iv) Commutativity: $x \circ y = y \circ x$ holds for all $x, y \in G$.

Theorem C.2. *The following statements and rules are valid in every group $(G, \circ)$:*

(a) *If* (C.2) *holds for $e \in G$, then*

$$\underset{x \in G}{\forall} \; e \circ x = x. \tag{C.3}$$

also holds.

(b) *The neutral element is unique: If $e, f \in G$, then*

$$\left(\left(\underset{x \in G}{\forall} \; x \circ e = x \right) \wedge \left(\underset{x \in G}{\forall} \; x \circ f = x \right) \right) \quad \Rightarrow \quad e = f. \tag{C.4}$$

(c) *If $x, a \in G$ and $x \circ a = e$ (where $e \in G$ is the neutral element), then $a \circ x = e$ as well. Moreover, inverse elements are unique (for each $x \in G$, the unique inverse is then denoted by x^{-1}).*

(d) *$(x^{-1})^{-1} = x$ holds for each $x \in G$.*

(e) *$y^{-1} \circ x^{-1} = (x \circ y)^{-1}$ holds for each $x, y \in G$.*

(f) *$x \circ a = y \circ a \;\Rightarrow\; x = y$ holds for each $x, y, a \in G$.*

Proof. (a): Let $x \in G$. By Def. C.1(iii), there exists $y \in G$ such that $x \circ y = e$ and, in turn, $z \in G$ such that $y \circ z = e$. Thus,

$$e \circ z = (x \circ y) \circ z = x \circ (y \circ z) = x \circ e = x, \tag{C.5}$$

implying

$$x = e \circ z = (e \circ e) \circ z = e \circ (e \circ z) = e \circ x \tag{C.6}$$

as desired.

(b): If e, f are both neutral elements, then, using (a), $f = e \circ f = e$.

(c): Assume $x \circ a = e$. Then there is b such that $a \circ b = e$. One computes

$$\begin{aligned} e = a \circ b &= (a \circ e) \circ b = (a \circ (x \circ a)) \circ b = a \circ ((x \circ a) \circ b) = a \circ (x \circ (a \circ b)) \\ &= a \circ (x \circ e) = a \circ x, \end{aligned} \tag{C.7}$$

establishing the case. Now let a, b be inverses to x. Then $a = a \circ e = a \circ x \circ b = e \circ b = b$.

(d): $x^{-1} \circ x = e$ holds according to (c) and shows that x is the inverse to x^{-1}. Thus, $(x^{-1})^{-1} = x$ as claimed.

(e) is due to $y^{-1} \circ x^{-1} \circ x \circ y = y^{-1} \circ e \circ y = e$.

(f): If $x \circ a = y \circ a$, then $x = x \circ a \circ a^{-1} = y \circ a \circ a^{-1} = y$ as claimed. ■

Definition C.3. Let $(G, \circ)$ and $(H, \circ)$ be groups. A map $\phi : G \longrightarrow H$ is called a (group) *homomorphism* if, and only if,

$$\underset{a,b \in G}{\forall} \quad \phi(a \circ b) = \phi(a) \circ \phi(b). \tag{C.8}$$

Proposition C.4. *Let $(G, \circ)$ and $(H, \circ)$ be groups and let $\phi : G \longrightarrow H$ be a group homomorphism. Let e, e' denote the neutral elements of G and H, respectively. Then the following holds:*

(a) *$\phi(e) = e'$.*

(b) *$\phi(a^{-1}) = (\phi(a))^{-1}$ for each $a \in G$.*

(c) *If ϕ is bijective, then ϕ^{-1} is also a group homomorphism.*

Proof. (a): We compute

$$\phi(e) \circ e' = \phi(e) = \phi(e \circ e) = \phi(e) \circ \phi(e).$$

Applying $(\phi(e))^{-1}$ to both sides of the above equality then proves $\phi(e) = e'$.

(b): We compute

$$\phi(a^{-1}) \circ \phi(a) = \phi(a^{-1} \circ a) = \phi(e) = e',$$

proving (b).

(c): Applying ϕ^{-1} to (C.8) yields

$$\underset{a,b \in G}{\forall} \quad a \circ b = \phi^{-1}(\phi(a) \circ \phi(b)). \tag{C.9}$$

Thus, for each $x, y \in H$, we obtain

$$\phi^{-1}(x \circ y) = \phi^{-1}\big(\phi(\phi^{-1}(x)) \circ \phi(\phi^{-1}(y))\big) \overset{(C.9)}{=} \phi^{-1}(x) \circ \phi^{-1}(y),$$

establishing the case and completing the proof of the proposition. ■

Notation C.5. Exponentiation with Integer Exponents: Let G be a nonempty set with a composition $\cdot : G \times G \longrightarrow G$. Assume there exists a (unique) neutral element $1 \in G$ (satisfying $x \cdot 1 = x$ for each $x \in G$). Define recursively for each $x \in G$ and each $n \in \mathbb{N}_0$:

$$x^0 := 1, \quad \underset{n \in \mathbb{N}_0}{\forall} \quad x^{n+1} := x \cdot x^n. \tag{C.10a}$$

Moreover, if $(G, \cdot)$ constitutes a group, then also define for each $x \in G$ and each $n \in \mathbb{N}$:

$$x^{-n} := (x^{-1})^n. \tag{C.10b}$$

Theorem C.6. Exponentiation Rules: *Let G be a nonempty set with a composition $\cdot : G \times G \longrightarrow G$. Assume that the composition satisfies the law of associativity and that there exists a (unique) neutral element $1 \in G$ (satisfying $x \cdot 1 = x$ for each $x \in G$). Let $x, y \in G$. Then the following rules hold for each $m, n \in \mathbb{N}_0$. If $(G, \cdot)$ is a group, then the rules even hold for every $m, n \in \mathbb{Z}$.*

(a) $x^{m+n} = x^m \cdot x^n$.

(b) $(x^m)^n = x^{m\,n}$.

(c) *If the composition is commutative (i.e. $xy = yx$ for each $x, y \in G$), then it holds that $x^n y^n = (xy)^n$.*

Proof. (a): First, we fix $n \in \mathbb{N}_0$ and prove the statement for each $m \in \mathbb{N}_0$ by induction: The base case ($m = 0$) is $x^n = x^n$, which is true. For the induction step, we compute

$$x^{m+1+n} \overset{(C.10a)}{=} x \cdot x^{m+n} \overset{\text{ind. hyp.}}{=} x \cdot x^m \cdot x^n \overset{(C.10a)}{=} x^{m+1} x^n,$$

completing the induction step. Now assume G to be a group. Consider $m \geq 0$ and $n < 0$. If $m + n \geq 0$, then, using what we have already shown,

$$x^m x^n \overset{\text{(C.10b)}}{=} x^m (x^{-1})^{-n} = x^{m+n} x^{-n} (x^{-1})^{-n} = x^{m+n}.$$

Similarly, if $m + n < 0$, then

$$x^m x^n \overset{\text{(C.10b)}}{=} x^m (x^{-1})^{-n} = x^m (x^{-1})^m (x^{-1})^{-n-m} \overset{\text{(C.10b)}}{=} x^{m+n}.$$

The case $m < 0$, $n \geq 0$ is treated completely analogously. It just remains to consider $m < 0$ and $n < 0$. In this case,

$$x^{m+n} = x^{-(-m-n)} \overset{\text{(C.10b)}}{=} (x^{-1})^{-m-n} = (x^{-1})^{-m} \cdot (x^{-1})^{-n} \overset{\text{(C.10b)}}{=} x^m \cdot x^n.$$

(b): First, we prove the statement for each $n \in \mathbb{N}_0$ by induction (for $m < 0$, we assmue G to be a group): The base case ($n = 0$) is $(x^m)^0 = 1 = x^0$, which is true. For the induction step, we compute

$$(x^m)^{n+1} \overset{\text{(C.10a)}}{=} x^m \cdot (x^m)^n \overset{\text{ind. hyp.}}{=} x^m \cdot x^{m\,n} \overset{\text{(a)}}{=} x^{m\,n+m} = x^{m\,(n+1)},$$

completing the induction step. Now, let G be a group and $n < 0$. We already know $(x^m)^{-1} = x^{-m}$. Thus, using what we have already shown,

$$(x^m)^n \overset{\text{(C.10b)}}{=} \big((x^m)^{-1}\big)^{-n} = (x^{-m})^{-n} = x^{(-m)(-n)} = x^{m\,n}.$$

(c): For $n \in \mathbb{N}_0$, the statement is proved by induction: The base case ($n = 0$) is $x^0 y^0 = 1 = (xy)^0$, which is true. For the induction step, we compute

$$x^{n+1} y^{n+1} \overset{\text{(C.10a)}}{=} x \cdot x^n \cdot y \cdot y^n \overset{\text{ind. hyp.}}{=} xy \cdot (xy)^n \overset{\text{(C.10a)}}{=} (xy)^{n+1},$$

completing the induction step. If G is a group and $n < 0$, then, using what we have already shown,

$$x^n y^n \overset{\text{(C.10b)}}{=} (x^{-1})^{-n} (y^{-1})^{-n} = (x^{-1} y^{-1})^{-n} \overset{\text{Th. C.2(e)}}{=} \big((xy)^{-1}\big)^{-n} \overset{\text{(C.10b)}}{=} (xy)^n,$$

which completes the proof. ■

C.2 Rings

Definition C.7. Let R be a nonempty set with two maps

$$\begin{aligned} + &: R \times R \longrightarrow R, \quad (x, y) \mapsto x + y, \\ \cdot &: R \times R \longrightarrow R, \quad (x, y) \mapsto x \cdot y \end{aligned} \tag{C.11}$$

($+$ is called *addition* and $\cdot$ is called *multiplication*; often one writes xy instead of $x \cdot y$). Then $(R, +, \cdot)$ (or just R, if $+$ and $\cdot$ are understood) is called a *ring* if, and only if, the following two conditions are satisfied:

(i) R is a commutative group with respect to $+$.

(ii) Multiplication is associative.

(iii) Distributivity:

$$\underset{x,y,z\in R}{\forall} x\cdot(y+z)=x\cdot y+x\cdot z, \tag{C.12a}$$

$$\underset{x,y,z\in R}{\forall} (y+z)\cdot x=y\cdot x+z\cdot x. \tag{C.12b}$$

A ring R is called *commutative* if, and only if, its multiplication is commutative. Moreover, a ring called a *ring with unity* if, and only if, R contains a neutral element with respect to multiplication (i.e. there is $1 \in R$ such that $1 \cdot x = x \cdot 1 = x$ for each $x \in R$) – some authors always require a ring to have a neutral element with respect to multiplication.

Theorem C.8. *The following statements and rules are valid in every ring with unity* $(R,+,\cdot)$ *(let* $x,y,z\in R$*):*

(a) $x\cdot 0=0=0\cdot x$.

(b) $x(-y)=-(xy)=(-x)y$.

(c) $(-x)(-y)=xy$.

(d) $x(y-z)=xy-xz$.

Proof. (a): One computes

$$x\cdot 0+x\cdot 1 \overset{\text{(C.12a)}}{=} x\cdot(0+1)=x\cdot 1=0+x\cdot 1,$$

i.e. $x \cdot 0 = 0$ follows since $(R,+)$ is a group. The second equality follows analogously using (C.12b).

(b): $xy+x(-y)=x(y-y)=x\cdot 0=0$, where we used (C.12a) and (a). This shows $x(-y)$ is the additive inverse to xy. The second equality follows analogously using (C.12b).

(c): $xy=-(-(xy))=-(x(-y))=(-x)(-y)$, where (b) was used twice.

(d): $x(y-z)=x(y+(-z))=xy+x(-z)=xy-xz$. ■

C.3 Fields

Definition C.9. Let $(F,+,\cdot)$ be a ring with unity. Then $(F,+,\cdot)$ (or just F, if $+$ and $\cdot$ are understood) is called a *field* if, and only if, $F\setminus\{0\}$ is a commutative group with respect to $\cdot$.

Theorem C.10. *The following statements and rules are valid in every field* $(F,+,\cdot)$*:*

(a) *Inverse elements are unique. For each $x \in F$, the unique inverse with respect to addition is denoted by $-x$. Also define $y-x := y+(-x)$. For each $x \in F \setminus \{0\}$, the unique inverse with respect to multiplication is denoted by x^{-1}. For $x \neq 0$, define the* fractions $\frac{y}{x} := y/x := yx^{-1}$ *with* numerator y *and* denominator x.

(b) $-(-x) = x$ *and* $(x^{-1})^{-1} = x$ *for* $x \neq 0$.

(c) $(-x) + (-y) = -(x + y)$ *and* $x^{-1}y^{-1} = (xy)^{-1}$ *for* $x, y \neq 0$.

(d) $x + a = y + a \;\Rightarrow\; x = y$ *and, for* $a \neq 0$, $xa = ya \;\Rightarrow\; x = y$.

(e) $x \cdot 0 = 0$.

(f) $x(-y) = -(xy)$.

(g) $(-x)(-y) = xy$.

(h) $x(y - z) = xy - xz$.

(i) $xy = 0 \;\Rightarrow\; x = 0 \vee y = 0$.

(j) Rules for Fractions:

$$\frac{a}{c} + \frac{b}{d} = \frac{ad + bc}{cd}, \quad \frac{a}{c} \cdot \frac{b}{d} = \frac{ab}{cd}, \quad \frac{a/c}{b/d} = \frac{ad}{bc},$$

where all denominators are assumed $\neq 0$.

Proof. (a) follows by applying Th. C.2(c) to the groups $(F, +)$ and $\big(F \setminus \{0\}, \cdot\big)$.

(b) follows by applying Th. C.2(d) to the groups $(F, +)$ and $\big(F \setminus \{0\}, \cdot\big)$.

(c) follows by applying Th. C.2(e) to the groups $(F, +)$ and $\big(F \setminus \{0\}, \cdot\big)$, plus then using commutativity of the groups.

(d) follows by applying Th. C.2(f) to the groups $(F, +)$ and $\big(F \setminus \{0\}, \cdot\big)$ (in the latter situation, the case $x = y = 0$ is also clear).

(e) follows by applying Th. C.8(a) to the ring with unity $(F, +, \cdot)$.

(f) follows by applying Th. C.8(b) to the ring with unity $(F, +, \cdot)$.

(g) follows by applying Th. C.8(c) to the ring with unity $(F, +, \cdot)$.

(h) follows by applying Th. C.8(d) to the ring with unity $(F, +, \cdot)$.

(i): If $xy = 0$ and $x \neq 0$, then $y = 1 \cdot y = x^{-1}xy = x^{-1} \cdot 0 = 0$.

(j): One computes

$$\frac{a}{c} + \frac{b}{d} = ac^{-1} + bd^{-1} = add^{-1}c^{-1} + bcc^{-1}d^{-1} = (ad + bc)(cd)^{-1} = \frac{ad + bc}{cd}$$

and

$$\frac{a}{c} \cdot \frac{b}{d} = ac^{-1}bd^{-1} = ab(cd)^{-1} = \frac{ab}{cd}$$

and

$$\frac{a/c}{b/d} = ac^{-1}(bd^{-1})^{-1} = ac^{-1}b^{-1}d = ad(bc)^{-1} = \frac{ad}{bc},$$

completing the proof. ■

D Construction of the Real Numbers

In Th. 4.4, we have defined the set of real numbers $\mathbb{R}$ as a complete totally ordered field and we claimed that such a complete totally ordered field does actually exist. In the following, we will describe how $\mathbb{R}$ can be constructed. We will follow [EHH$^+$95, Chs. 1,2], which contains several different approaches for the construction of $\mathbb{R}$.

D.1 Natural Numbers

In the first step, one starts with the natural numbers $\mathbb{N}$. The set of natural numbers $\mathbb{N}$ was defined in Def. A.41 and it was shown in Th. A.44 that $\mathbb{N}$ satisfies the Peano axioms P1 – P3 of Sec. 3.1. We denote natural numbers using the usual symbols $0 := \emptyset$, $1 := S(0) = \{0\}$, $2 := S(1) = \{0, 1\}$, $3 := S(2) = \{0, 1, 2\}$, $\ldots$, $n + 1 := S(n) = n \cup \{n\} = \{0, 1, \ldots, n\}$ (which is consistent with previous definitions in (A.5) and Not. A.45).

Theorem 3.7 allows to define *addition* and *multiplication* on $\mathbb{N}_0$ via recursion:

Definition D.1. (a) For each $m, n \in \mathbb{N}_0$, $m + n$ is defined recursively by

$$m + 0 := m, \quad m + 1 := S(m), \quad \underset{n\in\mathbb{N}}{\forall} m + S(n) := S(m + n). \tag{D.1}$$

This fits into the framework of Th. 3.7, using $A := \mathbb{N}_0$, $x_1 := S(m)$, and, for each $n \in \mathbb{N}$, $f_n : A^n \longrightarrow A$, $f_n(x_1, \ldots, x_n) := S(x_n)$ (due to the different initializations, one obtains a different recursion for each $m \in \mathbb{N}_0$).

(b) For each $m, n \in \mathbb{N}_0$, $mn := m \cdot n$ is defined recursively by

$$m \cdot 0 := 0, \quad m \cdot 1 := m, \quad \underset{n\in\mathbb{N}}{\forall} m \cdot (n + 1) := m \cdot n + m. \tag{D.2}$$

This fits into the framework of Th. 3.7, using $A := \mathbb{N}_0$, $x_1 := m$, and, for each $m, n \in \mathbb{N}$, $f_{m,n} : A^n \longrightarrow A$, $f_{m,n}(x_1, \ldots, x_n) := x_n + m$.

Theorem D.2. *The set $\mathbb{N}_0$ of the natural numbers (including 0) with the maps of addition and multiplication*

$$\begin{aligned} + &: \mathbb{N}_0 \times \mathbb{N}_0 \longrightarrow \mathbb{N}_0, & (x, y) &\mapsto x + y, \\ \cdot &: \mathbb{N}_0 \times \mathbb{N}_0 \longrightarrow \mathbb{N}_0, & (x, y) &\mapsto x \cdot y, \end{aligned}$$

as defined in Def. D.1(a) *and Def.* D.1(b)*, respectively, satisfies Def.* C.1(i),(ii),(iv) *for both addition and multiplication, i.e. associativity, commutativity, and the existence of a*

neutral element. This can be summarized as the statement that $\mathbb{N}_0$ *forms a commutative* semigroup *with respect to both addition and multiplication (however, no group, as the existence of inverse elements is lacking). Moreover, distributivity, i.e. Def.* C.7(iii) *is also satisfied.*

Proof. Associativity of Addition: We have to show

$$\underset{k,m,n\in\mathbb{N}_0}{\forall} \quad (k+m)+n = k+(m+n). \tag{D.3a}$$

The proof of (D.3a) is carried out by induction on n. The base case ($n = 0$) follows from the first definition in (D.1): $(k+m)+0 = k+m = k+(m+0)$ for every $k, m \in \mathbb{N}_0$. For the induction step, one computes, for every $k, m, n \in \mathbb{N}_0$,

$$\begin{aligned}(k+m)+(n+1) &\overset{\text{(D.1)}}{=} (k+m)+S(n) \overset{\text{(D.1)}}{=} S((k+m)+n) \overset{\text{ind. hyp.}}{=} S(k+(m+n))\\ &\overset{\text{(D.1)}}{=} k+S(m+n) \overset{\text{(D.1)}}{=} k+(m+S(n))\\ &\overset{\text{(D.1)}}{=} k+(m+(n+1)),\end{aligned} \tag{D.3b}$$

completing the induction.

Neutral Element of Addition: That 0 is the neutral element of addition is immediate from (D.1).

Commutativity of Addition: We have to show

$$\underset{m,n\in\mathbb{N}_0}{\forall} \quad m+n = n+m. \tag{D.4a}$$

The proof of (D.4a) is also carried out by induction on n. More precisely, we prove $n = 0$ separately, and then carry out the induction for $n \in \mathbb{N}$. The case $n = 0$ is proved by induction on m: The base case ($m = 0$) is the true statement $0+0 = 0 = 0+0$. For the induction step, one computes $(m+1)+0 = m+1 = S(m) = S(m+0) = S(0+m) = 0+S(m) = 0+(m+1)$. The base case for the induction on n, i.e. $n = 1$ is also proved by induction on m: The base case ($m = 0$) is the true statement $0+1 = S(0) = 1 = 1+0$. For the induction step, one computes, for every $m \in \mathbb{N}_0$,

$$\begin{aligned}(m+1)+1 &\overset{\text{(D.1)}}{=} S(m+1) \overset{\text{ind. hyp.}}{=} S(1+m) \overset{\text{(D.1)}}{=} (1+m)+1\\ &\overset{\text{(D.3a)}}{=} 1+(m+1).\end{aligned} \tag{D.4b}$$

Now, for the induction step of the induction on n, one computes, for every $(m, n) \in \mathbb{N}_0 \times \mathbb{N}$,

$$\begin{aligned}m+(n+1) &\overset{\text{(D.1)}}{=} m+S(n) \overset{\text{(D.1)}}{=} S(m+n) \overset{\text{ind. hyp.}}{=} S(n+m) \overset{\text{(D.1)}}{=} n+S(m)\\ &\overset{\text{(D.1)}}{=} n+(m+1) \overset{\text{base case}}{=} n+(1+m) \overset{\text{(D.3a)}}{=} (n+1)+m,\end{aligned} \tag{D.4c}$$

completing the induction.

Neutral Element of Multiplication: That 1 is the neutral element of addition is immediate from (D.2).

Commutativity of Multiplication: We have to show

$$\forall_{m,n\in\mathbb{N}_0} \quad m \cdot n = n \cdot m. \tag{D.5a}$$

We start with some preparatory steps: We first show

$$\forall_{m\in\mathbb{N}_0} \quad m = m \cdot 1 = 1 \cdot m. \tag{D.5b}$$

We have $m \cdot 1 = m$ for each $m \in \mathbb{N}_0$ directly from (D.2). We prove $1 \cdot m = m$ for each $m \in \mathbb{N}_0$ via induction on m: $1 \cdot 0 = 0$ and $1 \cdot 1 = 1$ are immediate from (D.2). For the induction step, one computes, for every $m \in \mathbb{N}$,

$$1 \cdot (m+1) \overset{\text{(D.2)}}{=} 1 \cdot m + 1 \overset{\text{ind. hyp.}}{=} m + 1.$$

Next, we show

$$\forall_{m,n\in\mathbb{N}_0} \quad n \cdot m + m = (n+1) \cdot m \tag{D.5c}$$

via induction on m. For the base case ($m = 0$), we note $(n+1) \cdot 0 = 0$ by (D.2), and $n \cdot 0 + 0 = 0$ by (D.1) and (D.2). For the induction step, we compute

$$\begin{aligned} n \cdot (m+1) + m + 1 &\overset{\text{(D.2)}}{=} n \cdot m + n + m + 1 \overset{\text{(D.4a)}}{=} n \cdot m + m + n + 1 \\ &\overset{\text{ind. hyp.}}{=} (n+1) \cdot m + n + 1 \overset{\text{(D.2)}}{=} (n+1) \cdot (m+1). \end{aligned}$$

We are now in a position to carry out the proof of (D.5a) by induction on n. More precisely, we prove $n = 0$ separately, and then carry out the induction for $n \in \mathbb{N}$. Let $n = 0$. We have $m \cdot 0 = 0$ for each $m \in \mathbb{N}_0$ directly from (D.2). We prove $0 \cdot m = 0$ for each $m \in \mathbb{N}_0$ via induction on m: $0 \cdot 0 = 0$ and $0 \cdot 1 = 0$ are immediate from (D.2). For the induction step, one computes, for every $m \in \mathbb{N}$,

$$0 \cdot (m+1) \overset{\text{(D.2)}}{=} 0 \cdot m + 0 \overset{\text{ind. hyp., (D.1)}}{=} 0.$$

The base case for the induction on $n \in \mathbb{N}$ is provided by (D.5b). For the induction step, one computes, for every $(m, n) \in \mathbb{N}_0 \times \mathbb{N}$,

$$m \cdot (n+1) \overset{\text{(D.2)}}{=} m \cdot n + m \overset{\text{ind. hyp.}}{=} n \cdot m + m \overset{\text{(D.5c)}}{=} (n+1) \cdot m,$$

completing the proof of (D.5a).

Distributivity: As we have commutativity of multiplication, we only need to show

$$\forall_{k,m,n\in\mathbb{N}_0} \quad (k+m) \cdot n = k \cdot n + m \cdot n. \tag{D.6a}$$

The proof of (D.6a) is carried out by induction on n. The base case ($n = 0$) follows from (D.2): $(k + m) \cdot 0 = 0 = k \cdot 0 + m \cdot 0$. For the induction step, one computes, for every $k, m, n \in \mathbb{N}_0$,

$$\begin{aligned}(k+m)\cdot(n+1) &\overset{\text{(D.2)}}{=} (k+m)\cdot n + k + m \\ &\overset{\text{ind. hyp.}}{=} k\cdot n + m\cdot n + k + m \\ &\overset{\text{(D.4a)}}{=} k\cdot n + k + m\cdot n + m \\ &\overset{\text{(D.2)}}{=} k\cdot(n+1) + m\cdot(n+1),\end{aligned} \tag{D.6b}$$

completing the induction.

Associativity of Multiplication: We have to show

$$\forall_{k,m,n\in\mathbb{N}_0} \quad (k\cdot m)\cdot n = k\cdot(m\cdot n). \tag{D.7a}$$

The proof of (D.7a) is carried out by induction on n. The base case ($n = 0$) follows from (D.2): $(k \cdot m) \cdot 0 = 0 = k \cdot (m \cdot 0)$ for every $k, m \in \mathbb{N}_0$. For the induction step, one computes, for every $k, m, n \in \mathbb{N}_0$,

$$\begin{aligned}(k\cdot m)\cdot(n+1) &\overset{\text{(D.2)}}{=} (k\cdot m)\cdot n + k\cdot m \overset{\text{ind. hyp.}}{=} k\cdot(m\cdot n) + k\cdot m \\ &\overset{\text{(D.6a)}}{=} k\cdot(m\cdot n + m) \overset{\text{(D.2)}}{=} k\cdot(m\cdot(n+1)),\end{aligned} \tag{D.7b}$$

completing the induction. ■

Lemma D.3. *We have*

$$\forall_{m\in\mathbb{N}_0} \quad \forall_{n\in\mathbb{N}} \quad m + n \neq m. \tag{D.8}$$

Proof. We fix $n \in \mathbb{N}$ and conduct an induction over $m \in \mathbb{N}_0$. The base case ($m = 0$) is clear, since $0 + n = n \neq 0$. The induction step is also clear, since $m + n \neq m$ implies $m + 1 + n = m + n + 1 = S(m + n) \neq S(m) = m + 1$, where we used that S is injective by Peano axiom P2. ■

Next, one defines an order $\leq$ on $\mathbb{N}_0$:

Definition D.4. For each $n, m \in \mathbb{N}_0$, let

$$n \leq m \quad :\Leftrightarrow \quad \exists_{k\in\mathbb{N}_0} \; n + k = m. \tag{D.9}$$

Theorem D.5. *The relation defined in* (D.9) *constitutes a well-order on* $\mathbb{N}_0$ *(in particular, a total order) that is compatible with addition and multiplication, i.e. it satisfies* (4.3).

Proof. $\leq$ is Reflexive: For each $n \in \mathbb{N}_0$, we have $n + 0 = n$, showing $n \leq n$.

$\leq$ is Antisymmetric: If $m \leq n$ and $n \leq m$, then $m + k = n$ and $n + l = m$. Thus, $n = m + k = n + l + k$. Thus, $l + k = 0$ by Lem. D.3, implying $l = k = 0$ (since $0 \notin S(\mathbb{N}_0)$) and $m = n$.

$\leq$ is Transitive: If $n \leq m$ and $m \leq l$, then there are $k_n, k_m \in \mathbb{N}_0$ such that $n + k_n = m$ and $m + k_m = l$. Then $n + k_n + k_m = m + k_m = l$, showing $m \leq l$.

$\leq$ is Total Order: We have to show that, if $m, n \in \mathbb{N}_0$, then $m \leq n$ or $n \leq m$. To this end, fix $n \in \mathbb{N}_0$ and conduct an induction over m. If $m = 0$, then $m + n = n$, showing $m \leq n$. For the induction step, let $m \in \mathbb{N}_0$. If $m \leq n$, then $m + k = n$. If $k = 0$, then $n \leq n + 1 = m + 1$. If $k \neq 0$, then $k = S(l)$, i.e. $n = m + l + 1 = m + 1 + l$, showing $m + 1 \leq n$. If $n \leq m$, then $n \leq m + 1$ is immediate, completing the induction.

$\leq$ is Well-Order: We first show that, for each $n \in \mathbb{N}_0$, the set $A_n := \{m \in \mathbb{N}_0 : m \leq n\}$ is finite: Indeed, this follows by an induction over n if we can show $A_{n+1} = A_n \cup \{n+1\}$. Indeed, if $m \leq n$, then $m \leq n \leq n + 1$, i.e. $m \leq n + 1$ by transitivity, showing $A_n \cup \{n+1\} \leq A_{n+1}$. If $m \leq n+1$ and $m \not\leq n$, then $n < m$, i.e. $n+k = m$ with $k \neq 0$, i.e. $n+1+l = m$, showing $n+1 \leq m$, i.e. $m = n+1$ and $A_{n+1} \subseteq A_n \cup \{n+1\}$. One now finishes the proof that $\leq$ is a well-order as in the proof of Th. 3.13(b): Let $\emptyset \neq A \subseteq \mathbb{N}_0$. We have to show A has a min. If A is finite, then A has a min by Th. 3.13(a). If A is infinite, let n be an element from A. Then the finite set $B := \{k \in A : k \leq n\} = A_n \cap B$ must have a min m by Th. 3.13(a). Since $m \leq x$ for each $x \in B$ and $m \leq n < x$ for each $x \in A \setminus B$, we have $m = \min A$.

Compatibility with Addition: We have to show

$$\underset{k,m,n \in \mathbb{N}_0}{\forall} \big(k \leq m \;\Rightarrow\; k + n \leq m + n\big). \tag{D.10}$$

To this end, note that $k \leq m$ means that there is $l \in \mathbb{N}_0$ such that $k + l = m$. But then $k + n + l = k + l + n = m + n$, showing $k + n \leq m + n$.

Compatibility with Multiplication: As $0 \leq n$ holds for every $n \in \mathbb{N}_0$, there is nothing to prove. ■

D.2 Interlude: Orders on Groups

In the succeeding sections, we will construct the set of integers $\mathbb{Z}$, the set of rational numbers $\mathbb{Q}$, and the set of real numbers $\mathbb{R}$. In each case, we will use the same method to define a total order on the constructed set, making use of the algebraic structure of its additive group. It is therefore economical as well as mathematically interesting, to study this construction once in its abstract form, which is the purpose of the present section.

Recall the definition of a group from Def. C.1.

Theorem D.6. *Assume $(G, +)$ to be a group. Moreover, assume we have a disjoint decomposition*

$$G = P \dot\cup \{0\} \dot\cup (-P), \quad -P := \{x \in G : -x \in P\}, \tag{D.11}$$

where $-x$ denotes the inverse of x with respect to $+$. If P is closed under $+$ (i.e. $x, y \in P$ implies $x + y \in P$), then

$$y \leq x \quad :\Leftrightarrow \quad x - y \in P \cup \{0\} \tag{D.12}$$

defines a total order on G that is compatible with addition, i.e. it satisfies (4.3a). *Moreover, if a multiplication is also defined on G and $P \cup \{0\}$ is closed under this multiplication, then $\leq$ is also compatible with multiplication, i.e. it satisfies* (4.3b). *Of course, one refers to the elements of P as* positive *and to the elements of $-P$ as* negative.

Proof. For each $x \in G$, one has $x - x = 0 \in P \cup \{0\}$, i.e. $x \leq x$ and the relation is reflexive. If $x, y \in G$, $x \leq y$ and $y \leq x$, then $x - y \in P \cup \{0\}$ and $-(x - y) = y - x \in P \cup \{0\}$, and the disjointness of the union in (D.11) implies $x - y = 0$, i.e. $x = y$, showing the relation is antisymmetric. If $x, y, z \in G$ with $x \leq y$ and $y \leq z$, then $y - x \in P \cup \{0\}$, $z - y \in P \cup \{0\}$, and $z - x = z - y + y - x \in P \cup \{0\}$ since P is closed under $+$, showing the relation is transitive. So we have shown $\leq$ constitutes a partial order on G. It remains to show the order is total. However, given the decomposition in (D.11), for each $x, y \in G$, precisely one of the statements $x - y \in P$ (i.e. $y < x$), $x - y = 0$ (i.e. $x = y$), $x - y \in -P$ (i.e. $x < y$) must be true, proving that the order is total. To see $\leq$ satisfies (4.3a), let $x, y, z \in G$. If $x \leq y$, then $y - x \in P \cup \{0\}$, i.e. $y + z - (x + z) = y + z - z - x \in P \cup \{0\}$, showing $x + z \leq y + z$. The proof is completed by noting (4.3b) is precisely the statement that $P \cup \{0\}$ is closed under multiplication. ■

D.3 Integers

As compared to our goal, the set of real numbers $\mathbb{R}$, the set $\mathbb{N}_0$ still has three deficiencies, namely the lack of inverse elements for addition, the lack of inverse elements for multiplication, and that the order $\leq$ lacks completeness. The construction of the integers will remedy (only) the first of the three deficiencies by providing the inverse elements of addition.

Definition and Remark D.7. The relation $\sim$ on $\mathbb{N}_0 \times \mathbb{N}_0$ defined by

$$(a, b) \sim (c, d) \quad :\Leftrightarrow \quad a + d = b + c, \tag{D.13}$$

constitutes an equivalence relation on $\mathbb{N}_0 \times \mathbb{N}_0$ (cf. Def. 2.23).

Definition D.8. (a) Define the set of *integers* $\mathbb{Z}$ as the set of equivalence classes of the equivalence relation $\sim$ defined in (D.13), i.e.

$$\mathbb{Z} := (\mathbb{N}_0 \times \mathbb{N}_0)/\sim \; = \; \big\{[(a, b)] : (a, b) \in \mathbb{N}_0 \times \mathbb{N}_0\big\} \tag{D.14}$$

is the quotient set of $\mathbb{N}_0 \times \mathbb{N}_0$ with respect to $\sim$ (cf. Ex. 2.24(c)). To simplify notation, in the following, we will write

$$[a, b] := [(a, b)] \tag{D.15}$$

for the equivalence class of (a, b) with respect to $\sim$.

(b) *Addition* on $\mathbb{Z}$ is defined by

$$+ : \mathbb{Z} \times \mathbb{Z} \longrightarrow \mathbb{Z}, \quad \big([a,b],[c,d]\big) \mapsto [a,b] + [c,d] := [a+c, b+d]. \tag{D.16}$$

Subtraction on $\mathbb{Z}$ is defined by

$$- : \mathbb{Z} \times \mathbb{Z} \longrightarrow \mathbb{Z}, \quad \big([a,b],[c,d]\big) \mapsto [a,b] - [c,d] := [a,b] + [d,c]. \tag{D.17}$$

—

For the definitions in Def. D.8(b) to make sense, one needs to check that they do not depend on the chosen representatives of the equivalence classes. Moreover, one needs to convince oneself that these definitions yield the desired familiar operations of addition and subtraction. Let us start by verifying the independence of the representatives is the following Lem. D.9.

Lemma D.9. *The definitions in Def.* D.8(b) *do not depend on the chosen representatives, i.e.*

$$\forall_{a,b,c,d,\tilde{a},\tilde{b},\tilde{c},\tilde{d}\in\mathbb{N}_0} \quad \Big([a,b] = [\tilde{a},\tilde{b}] \wedge [c,d] = [\tilde{c},\tilde{d}] \quad \Rightarrow \quad [a+c, b+d] = [\tilde{a}+\tilde{c}, \tilde{b}+\tilde{d}]\Big) \tag{D.18}$$

and

$$\forall_{a,b,c,d,\tilde{a},\tilde{b},\tilde{c},\tilde{d}\in\mathbb{N}_0} \quad \Big([a,b] = [\tilde{a},\tilde{b}] \wedge [c,d] = [\tilde{c},\tilde{d}] \quad \Rightarrow \quad [a,b] - [c,d] = [\tilde{a},\tilde{b}] - [\tilde{c},\tilde{d}]\Big). \tag{D.19}$$

Proof. (D.18): $[a,b] = [\tilde{a},\tilde{b}]$ means $a + \tilde{b} = b + \tilde{a}$, $[c,d] = [\tilde{c},\tilde{d}]$ means $c + \tilde{d} = d + \tilde{c}$, implying $a + c + \tilde{b} + \tilde{d} = b + \tilde{a} + d + \tilde{c}$, i.e. $[a+c, b+d] = [\tilde{a}+\tilde{c}, \tilde{b}+\tilde{d}]$.

(D.19) is just (D.17) combined with (D.18). ■

Theorem D.10. *The set of integers $\mathbb{Z}$ forms a commutative group with respect to addition as defined in Def.* D.8(b)*, where $[0,0]$ is the neutral element, $[b,a]$ is the inverse element of $[a,b]$ for each $a, b \in \mathbb{N}_0$, and, denoting the inverse element of $[a,b]$ by $-[a,b]$ in the usual way, $[a,b] - [c,d] = [a,b] + (-[c,d])$ for each $a, b, c, d \in \mathbb{N}_0$.*

Proof. One easily verifies that associativity and commutativity of the addition on $\mathbb{N}_0$ imply the respective laws on $\mathbb{Z}$. For every $a, b \in \mathbb{N}_0$, one obtains $[a,b] + [0,0] = [a+0, b+0] = [a,b]$, proving neutrality of $[0,0]$, whereas $[a,b] + [b,a] = [a+b, b+a] = [a+b, a+b] = [0,0]$ (since $(a+b, a+b) \sim (0,0)$) shows $[b,a] = -[a,b]$. Now $[a,b] - [c,d] = [a,b] + (-[c,d])$ is immediate from (D.17). ■

Remark D.11. The map

$$\iota : \mathbb{N}_0 \longrightarrow \mathbb{Z}, \quad \iota(n) := [n,0], \tag{D.20}$$

is a monomorphism, i.e. it is injective (since $\iota(m) = [m,0] = \iota(n) = [n,0]$ implies $m + 0 = 0 + n$, i.e. $m = n$) and satisfies

$$\forall_{m,n\in\mathbb{N}_0} \quad \iota(m+n) = [m+n, 0] = [m,0] + [n,0] = \iota(m) + \iota(n). \tag{D.21}$$

It is customary to identify $\mathbb{N}_0$ with $\iota(\mathbb{N}_0)$, as it usually does not cause any confusion. One then just writes n instead of $[n,0]$ and $-n$ instead of $[0,n] = -[n,0]$.

Lemma D.12. *We have the disjoint decomposition*

$$\mathbb{Z} = \mathbb{N} \dot\cup \{0\} \dot\cup \mathbb{Z}^-, \quad \mathbb{Z}^- := -\mathbb{N} = \{n \in \mathbb{Z} : \, -n \in \mathbb{N}\}. \tag{D.22}$$

Proof. Note that, due to (D.13), an equivalence class remains the same if a natural number is added or subtracted in both components: $[a, b] = [a + m, b + m]$. Thus, for each $x = [a, b] \in \mathbb{Z}$, if $a > b$, then $x = [a - b, 0] \in \mathbb{N}$; if $a = b$, then $x = [0, 0] = 0$; if $a < b$, then $x = [0, b - a] = -[b - a, 0] \in \mathbb{Z}^-$. It just remains to verify that the union in (D.22) is disjoint. However, if $[n, 0] = [0, m]$ with $m, n \in \mathbb{N}_0$, then $n + m = 0$, proving $n = m = 0$, completing the proof. ∎

Remark D.13. In the above construction, we obtained the commutative group $(\mathbb{Z}, +)$ from the commutative semigroup $(\mathbb{N}_0, +)$. It is worth pointing out that the same construction always works when, instead of with $\mathbb{N}_0$, one starts with any commutative semigroup $(H, +)$ that satisfies the *cancellation law* $a + c = b + c \;\Rightarrow\; a = b$, to obtain a commutative group $(G, +)$ and a monomorphism $\iota : H \longrightarrow G$.

—

To obtain the expected laws of arithmetic, multiplication on $\mathbb{Z}$ needs to be defined such that $(a - b) \cdot (c - d) = (ac + bd) - (ad + bc)$, which leads to the following definition.

Definition D.14. *Multiplication* on $\mathbb{Z}$ is defined by

$$\cdot : \mathbb{Z} \times \mathbb{Z} \longrightarrow \mathbb{Z}, \quad \big([a, b], [c, d]\big) \mapsto [a, b] \cdot [c, d] := [ac + bd, ad + bc]. \tag{D.23}$$

Lemma D.15. *The definition in Def.* D.14 *does not depend on the chosen representatives, i.e.*

$$\underset{a,b,c,d,\tilde a,\tilde b,\tilde c,\tilde d\in\mathbb{N}_0}{\forall} \quad \Big([a, b] = [\tilde a, \tilde b] \wedge [c, d] = [\tilde c, \tilde d] \quad \Rightarrow \quad [ac+bd, ad+bc] = [\tilde a\tilde c+\tilde b\tilde d, \tilde a\tilde d+\tilde b\tilde c]\Big). \tag{D.24}$$

Proof. As mentioned before, due to (D.13), an equivalence class remains the same if a natural number is added or subtracted in both components. Thus, one computes

$$\begin{aligned}
[ac + bd, ad + bc] &\overset{(D.13)}{=} [ac + bd + \tilde b c, ad + bc + \tilde b c] = [(a + \tilde b)c + bd, ad + bc + \tilde b c] \\
&= [(\tilde a + b)c + bd, ad + bc + \tilde b c] \overset{(D.13)}{=} [\tilde a\tilde d + \tilde a c + bd, \tilde a\tilde d + ad + \tilde b c] \\
&= [\tilde a(\tilde d + c) + bd, \tilde a\tilde d + ad + \tilde b c] = [\tilde a(d + \tilde c) + bd, \tilde a\tilde d + ad + \tilde b c] \\
&= [\tilde a\tilde c + (\tilde a + b)d, \tilde a\tilde d + ad + \tilde b c] = [\tilde a\tilde c + (a + \tilde b)d, \tilde a\tilde d + ad + \tilde b c] \\
&\overset{(D.13)}{=} [\tilde a\tilde c + \tilde b d + \tilde b\tilde c, \tilde a\tilde d + \tilde b c + \tilde b\tilde c] = [\tilde a\tilde c + \tilde b(d + \tilde c), \tilde a\tilde d + \tilde b c + \tilde b\tilde c] \\
&= [\tilde a\tilde c + \tilde b(\tilde d + c), \tilde a\tilde d + \tilde b c + \tilde b\tilde c] = [\tilde a\tilde c + \tilde b\tilde d, \tilde a\tilde d + \tilde b\tilde c],
\end{aligned} \tag{D.25}$$

completing the proof. ∎

Theorem D.16. *The set of integers $\mathbb{Z}$ is associative and commutative with respect to the multiplication defined in Def.* D.14. *Moreover, distributivity, i.e. Def.* C.7(iii) *is satisfied,* $[1,0]$ *is the neutral element of multiplication, and there are no* zero divisors, *i.e.*

$$\forall_{a,b,c,d\in\mathbb{N}_0} \quad \Big([a,b]\cdot[c,d]=[ac+bd,ad+bc]=[0,0] \quad \Rightarrow \quad [a,b]=[0,0]\vee[c,d]=[0,0]\Big). \tag{D.26}$$

Algebraically, the theorem can be summarized by saying that $(\mathbb{Z},+,\cdot)$ *constitutes a* principal ideal domain.

Proof. Associativity and commutativity of multiplication as well as distributivity are easily verified, while $[a,b]\cdot[1,0]=[a\cdot 1+b\cdot 0, a\cdot 0+b\cdot 1]=[a,b]$ proves neutrality of $[1,0]$. It remains to prove (D.26). Note that, due to (D.13), the conclusion is equivalent to $a=b$ or $c=d$. We assume $0\le a<b$ and have to prove $c=d$. According to Def. D.5, $a<b$ means $b=a+k$ for some $k\in\mathbb{N}$. Thus, $[ac+bd,ad+bc]=[0,0]$ implies

$$ac+(a+k)d=ac+bd=ad+bc=ad+(a+k)c \quad \Rightarrow \quad kd=kc \quad \overset{k\ge 0}{\Rightarrow} \quad c=d, \tag{D.27}$$

establishing the case. ■

Definition D.17. For each $k,l\in\mathbb{Z}$, let

$$l\le k \quad :\Leftrightarrow \quad k-l\in\mathbb{N}_0. \tag{D.28}$$

Theorem D.18. (a) *The relation defined in* (D.28) *constitutes a total order on* $\mathbb{Z}$ *that is compatible with addition and multiplication, i.e. it satisfies* (4.3).

(b) *The map* ι *from* (D.20) *is strictly increasing.*

Proof. (a) follows from (D.28), (D.22), and Th. D.6 since $\mathbb{N}_0$ is closed under addition and multiplication.

(b): According to Def. D.5, if $m,n\in\mathbb{N}$ with $n<m$, then $m=n+k$ for some $k\in\mathbb{N}$. In consequence $\iota(m)=\iota(n)+\iota(k)$ by (D.21), i.e. $\iota(m)-\iota(n)=\iota(k)\in\mathbb{N}$, proving $\iota(n)<\iota(m)$. ■

D.4 Rational Numbers

The remaining two deficiencies of the set of integers $\mathbb{Z}$ (as compared with $\mathbb{R}$) are the lack of inverse elements for multiplication and that the order $\le$ lacks completeness. We proceed to the construction of the rational numbers, which will provide the inverse elements for multiplication. The completion of the order will then be achieved in the last step in the next section.

Definition and Remark D.19. The relation $\sim$ on $\mathbb{Z}\times(\mathbb{Z}\setminus\{0\})$ defined by

$$(a,b)\sim(c,d) \quad :\Leftrightarrow \quad ad=bc, \tag{D.29}$$

constitutes an equivalence relation on $\mathbb{Z}\times(\mathbb{Z}\setminus\{0\})$ (cf. Def. 2.23).

Definition D.20. (a) Define the set of *rational numbers* $\mathbb{Q}$ as the set of equivalence classes of the equivalence relation $\sim$ defined in (D.29), i.e.

$$\mathbb{Q} := \big(\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})\big)/\sim \;\; = \big\{[(a,b)] : (a,b) \in \mathbb{Z} \times (\mathbb{Z} \setminus \{0\})\big\} \tag{D.30}$$

is the quotient set of $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ with respect to $\sim$ (cf. Ex. 2.24(c)). As is common, we will write

$$\frac{a}{b} := a/b := [(a,b)] \tag{D.31}$$

for the equivalence class of (a,b) with respect to $\sim$.

(b) *Addition* on $\mathbb{Q}$ is defined by

$$+ : \mathbb{Q} \times \mathbb{Q} \longrightarrow \mathbb{Q}, \quad \left(\frac{a}{b}, \frac{c}{d}\right) \mapsto \frac{a}{b} + \frac{c}{d} := \frac{ad+bc}{bd}. \tag{D.32}$$

Multiplication on $\mathbb{Q}$ is defined by

$$\cdot : \mathbb{Q} \times \mathbb{Q} \longrightarrow \mathbb{Q}, \quad \left(\frac{a}{b}, \frac{c}{d}\right) \mapsto \frac{a}{b} \cdot \frac{c}{d} := \frac{ac}{bd}. \tag{D.33}$$

—

For the definitions in Def. D.20(b) to make sense, one needs to check that they do not depend on the chosen representatives of the equivalence classes, and that the results of both addition and multiplication are always elements of $\mathbb{Q}$. All this is provided by the following lemma.

Lemma D.21. *The definitions in Def.* D.20(b) *do not depend on the chosen representatives, i.e.*

$$\forall_{a,c,\tilde{a},\tilde{c}\in\mathbb{Z}} \quad \forall_{b,d,\tilde{b},\tilde{d}\in\mathbb{Z}\setminus\{0\}} \quad \left(\frac{a}{b} = \frac{\tilde{a}}{\tilde{b}} \wedge \frac{c}{d} = \frac{\tilde{c}}{\tilde{d}} \quad \Rightarrow \quad \frac{ad+bc}{bd} = \frac{\tilde{a}\tilde{d}+\tilde{b}\tilde{c}}{\tilde{b}\tilde{d}}\right) \tag{D.34}$$

and

$$\forall_{a,c,\tilde{a},\tilde{c}\in\mathbb{Z}} \quad \forall_{b,d,\tilde{b},\tilde{d}\in\mathbb{Z}\setminus\{0\}} \quad \left(\frac{a}{b} = \frac{\tilde{a}}{\tilde{b}} \wedge \frac{c}{d} = \frac{\tilde{c}}{\tilde{d}} \quad \Rightarrow \quad \frac{ac}{bd} = \frac{\tilde{a}\tilde{c}}{\tilde{b}\tilde{d}}\right). \tag{D.35}$$

Furthermore, the results of both addition and multiplication are always elements of $\mathbb{Q}$.

Proof. (D.34): $a/b = \tilde{a}/\tilde{b}$ means $a\tilde{b} = \tilde{a}b$, $c/d = \tilde{c}/\tilde{d}$ means $c\tilde{d} = \tilde{c}d$, implying

$$(ad+bc)\tilde{b}\tilde{d} = bd(\tilde{a}\tilde{d}+\tilde{b}\tilde{c}), \quad \text{i.e.} \quad \frac{ad+bc}{bd} = \frac{\tilde{a}\tilde{d}+\tilde{b}\tilde{c}}{\tilde{b}\tilde{d}} \tag{D.36}$$

and

$$ac\tilde{b}\tilde{d} = bd\tilde{a}\tilde{c}, \quad \text{i.e.} \quad \frac{ac}{bd} = \frac{\tilde{a}\tilde{c}}{\tilde{b}\tilde{d}}. \tag{D.37}$$

That the results of both addition and multiplication are always elements of $\mathbb{Q}$ follows from (D.26), i.e. from the fact that $\mathbb{Z}$ has no zero divisors. In particular, if $b, d \neq 0$, then $bd \neq 0$, showing $(ad+bc)/(bd) \in \mathbb{Q}$ and $(ac)/(bd) \in \mathbb{Q}$. ■

Theorem D.22. (a) *The set of rational numbers $\mathbb{Q}$ with addition and multiplication as defined in Def.* D.20 *forms a field, where $0/1$ and $1/1$ are the neutral elements with respect to addition and multiplication, respectively, $(-a/b)$ is the additive inverse to a/b, whereas b/a is the multiplicative inverse to a/b with $a \neq 0$.*

(b) *Defining* subtraction *and* division *in the usual way, for each $r, s \in \mathbb{Q}$, by $s - r := s + (-r)$ and $s/r := sr^{-1}$, respectively, with $-r$ denoting the additive inverse of r and r^{-1} denoting the multiplicative inverse of $r \neq 0$, all the rules stated in Th.* C.10 *are valid in $\mathbb{Q}$.*

(c) *The map*

$$\iota : \mathbb{Z} \longrightarrow \mathbb{Q}, \quad \iota(k) := \frac{k}{1}, \tag{D.38}$$

is a monomorphism, i.e. it is injective and satisfies

$$\underset{k,l\in\mathbb{Z}}{\forall} \quad \iota(k+l) = \iota(k) + \iota(l), \tag{D.39a}$$

$$\underset{k,l\in\mathbb{Z}}{\forall} \quad \iota(kl) = \iota(k) \cdot \iota(l). \tag{D.39b}$$

It is customary to identify $\mathbb{Z}$ with $\iota(\mathbb{Z})$, as it usually does not cause any confusion. One then just writes k instead of $\frac{k}{1}$.

Proof. A detailed proof of (a) is provided in [Lan65, Ch. 2.3,2.4]. Let us check the claims regarding neutral and inverse elements:

$$\frac{a}{b} + \frac{0}{1} = \frac{a \cdot 1 + b \cdot 0}{b \cdot 1} = \frac{a}{b}, \tag{D.40a}$$

$$\frac{a}{b} + \frac{-a}{b} = \frac{ab + b(-a)}{b^2} \overset{\text{Def. C.7(iii) for } \mathbb{Z}}{=} \frac{(a-a)b}{b^2} = \frac{0}{b^2} \overset{\text{(D.29)}}{=} \frac{0}{1}, \tag{D.40b}$$

$$\frac{a}{b} \cdot \frac{1}{1} = \frac{a \cdot 1}{b \cdot 1} = \frac{a}{b}, \tag{D.40c}$$

$$\frac{a}{b} \cdot \frac{b}{a} = \frac{ab}{ba} \overset{\text{(D.29)}}{=} \frac{1}{1}. \tag{D.40d}$$

(b) is a consequence of (a), since Th. C.10 and its proof are valid in every field.

(c): The map ι is injective, as $\iota(k) = k/1 = \iota(l) = l/1$ implies $k \cdot 1 = l \cdot 1$, i.e. $k = l$. Moreover,

$$\iota(k) + \iota(l) = \frac{k}{1} + \frac{l}{1} = \frac{k \cdot 1 + 1 \cdot l}{1} = \frac{k+l}{1} = \iota(k+l), \tag{D.41a}$$

$$\iota(k) \cdot \iota(l) = \frac{k}{1} \cdot \frac{l}{1} = \frac{kl}{1} = \iota(kl), \tag{D.41b}$$

completing the proof. ■

Definition and Remark D.23. Define

$$\mathbb{Q}^+ := \left\{ r \in \mathbb{Q} : \exists_{a,b\in\mathbb{N}} \, r = \frac{a}{b} \right\}. \tag{D.42}$$

We then have the decomposition

$$\mathbb{Q} = \mathbb{Q}^+ \dot\cup \{0\} \dot\cup \mathbb{Q}^-, \quad \mathbb{Q}^- := -\mathbb{Q}^+ = \{r \in \mathbb{Q} : \, -r \in \mathbb{Q}^+\}, \tag{D.43}$$

since

$$a/b \in \mathbb{Q}^+ \quad \Leftrightarrow \quad \big((a > 0 \wedge b > 0) \vee (a < 0 \wedge b < 0)\big), \tag{D.44a}$$
$$a/b = 0 \quad \Leftrightarrow \quad a = 0, \tag{D.44b}$$
$$a/b \in \mathbb{Q}^- \quad \Leftrightarrow \quad \big((a > 0 \wedge b < 0) \vee (a < 0 \wedge b > 0)\big). \tag{D.44c}$$

Definition D.24. For each $r, s \in \mathbb{Q}$, let

$$s \le r \quad :\Leftrightarrow \quad r - s \in \mathbb{Q}_0^+ := \mathbb{Q}^+ \cup \{0\}. \tag{D.45}$$

Theorem D.25. (a) *The relation defined in* (D.45) *constitutes a total order on* $\mathbb{Q}$ *that is compatible with addition and multiplication, i.e. it satisfies* (4.3)*; in other words* $(\mathbb{Q}, +, \cdot, \le)$ *constitutes a totally ordered field.*

(b) *All the rules stated in Th.* 4.5 *are valid in* $\mathbb{Q}$.

(c) *The map* ι *from* (D.38) *is strictly increasing.*

Proof. (a) follows from (D.45), (D.43), and Th. D.6, since it is immediate from (D.32) and (D.33) that $\mathbb{Q}^+$ is closed under addition and multiplication.

(b) is a consequence of (a), since Th. 4.5 and its proof are valid in every totally ordered field.

(c): According to Def. D.25, if $k, l \in \mathbb{Z}$ with $l < k$, then $n := k - l \in \mathbb{N}$. In consequence $\iota(k) = \iota(l) + \iota(n)$ by (D.39a), i.e. $\iota(k) - \iota(l) = \iota(n) = n/1 \in \mathbb{Q}^+$, proving $\iota(l) < \iota(k)$. ■

D.5 Real Numbers

In the previous section, the construction of the rational numbers $\mathbb{Q}$ yielded a totally ordered field. However, the order on $\mathbb{Q}$ is not complete – for example, Rem. and Def. 7.62 shows that the set $M := \{r \in \mathbb{Q} : \, r^2 < 2\}$, which is bounded from above (for example by 2), has no supremum in $\mathbb{Q}$ (otherwise, we had a rational number $q = \sup M$ with $q^2 = 2$). Finally, in the present section, we will start out from $\mathbb{Q}$ to construct the set of real numbers $\mathbb{R}$ such that it becomes a complete totally ordered field. There are several different important constructions to obtain $\mathbb{R}$ from $\mathbb{Q}$. We will describe the construction that defines real numbers as equivalence classes of rational Cauchy sequences following [EHH$^+$95, Ch. 2.3]. The construction using so-called Dedekind cuts can be found in [EHH$^+$95, Ch. 2.2], the construction via nested intervals in [EHH$^+$95, Ch. 2.4].

Definition D.26. (a) Let $\mathcal{S}$ denote the set of all Cauchy sequences in $\mathbb{Q}$, where we call a sequence $(r_n)_{n\in\mathbb{N}}$ in $\mathbb{Q}$ a Cauchy sequence if, and only if,

$$\forall_{\epsilon\in\mathbb{Q}^+}\ \exists_{N\in\mathbb{N}}\ \forall_{n,m>N}\ |r_n - r_m| < \epsilon, \tag{D.46}$$

which defers from (7.25) in that ϵ has to be from $\mathbb{Q}^+$ rather than from $\mathbb{R}^+$.

(b) *Addition* on $\mathcal{S}$ is defined by

$$+ : \mathcal{S}\times\mathcal{S} \longrightarrow \mathcal{S}, \quad ((r_n)_{n\in\mathbb{N}}, (s_n)_{n\in\mathbb{N}}) \mapsto (r_n)_{n\in\mathbb{N}} + (s_n)_{n\in\mathbb{N}} := (r_n + s_n)_{n\in\mathbb{N}}. \tag{D.47}$$

Multiplication on $\mathcal{S}$ is defined by

$$\cdot : \mathcal{S}\times\mathcal{S} \longrightarrow \mathcal{S}, \quad ((r_n)_{n\in\mathbb{N}}, (s_n)_{n\in\mathbb{N}}) \mapsto (r_n)_{n\in\mathbb{N}} \cdot (s_n)_{n\in\mathbb{N}} := (r_n s_n)_{n\in\mathbb{N}}. \tag{D.48}$$

As a consequence of the following Lem. D.27, addition and multiplication are well-defined on $\mathcal{S}$.

Lemma D.27. *If $(r_n)_{n\in\mathbb{N}}$ and $(s_n)_{n\in\mathbb{N}}$ are Cauchy sequences in $\mathbb{Q}$, so are $(r_n + s_n)_{n\in\mathbb{N}}$ and $(r_n s_n)_{n\in\mathbb{N}}$.*

Proof. The proofs are analogous to the proofs of Th. 7.13(7.11b),(7.11c):

Given $\epsilon \in \mathbb{Q}^+$, there exists $N \in \mathbb{N}$ such that, for each $n, m > N$, $|r_n - r_m| < \epsilon/2$ and $|s_n - s_m| < \epsilon/2$, implying

$$\forall_{n,m>N} |r_n + s_n - (r_m + s_m)| \le |r_n - r_m| + |s_n - s_m| < \epsilon/2 + \epsilon/2 = \epsilon, \tag{D.49}$$

proving $(r_n + s_n)_{n\in\mathbb{N}}$ is Cauchy.

The proof of Th. 7.29 shows both $(r_n)_{n\in\mathbb{N}}$ and $(s_n)_{n\in\mathbb{N}}$ are bounded, i.e. there exists $M \in \mathbb{Q}^+$ that is an upper bound for the sets $\{|r_n| : n \in \mathbb{N}\}$ and $\{|s_n| : n \in \mathbb{N}\}$. Moreover, given $\epsilon \in \mathbb{Q}^+$, there exists $N \in \mathbb{N}$ such that, for each $n, m > N$, $|r_n - r_m| < \epsilon/(2M)$ and $|s_n - s_m| < \epsilon/(2M)$, implying

$$\forall_{n,m>N} \left(\begin{aligned} |r_n s_n - r_m s_m| &= \big|(r_n - r_m)s_n + r_m(s_n - s_m)\big| \\ &\le |s_n| \cdot |r_n - r_m| + |r_m| \cdot |s_n - s_m| < \frac{M\,\epsilon}{2M} + \frac{M\,\epsilon}{2M} = \epsilon \end{aligned} \right), \tag{D.50}$$

completing the proof of the lemma. ■

Theorem D.28. $(\mathcal{S}, +)$ *is a group and, in addition, $\mathcal{S}$ is associative and commutative with respect to multiplication. Moreover, distributivity also holds in $\mathcal{S}$. In algebraic terms, this can be summarized as the statement that $(\mathcal{S}, +, \cdot)$ constitutes a commutative ring.*

Proof. Note that, since the rational sequence $(r_n)_{n\in\mathbb{N}}$ is nothing but the function $f : \mathbb{N} \longrightarrow \mathbb{Q}$, $f(n) = r_n$, addition and multiplication as defined in Def. D.26(b) is analogous to the definition of addition and multiplication of real-valued functions in (6.1a), (6.1c), respectively. It is an easy exercise to verify that these function operations always inherit associativity, commutativity, and distributivity if these rules hold for the operations defined on the function range (i.e. for $+$ and $\cdot$ on $\mathbb{Q}$ in our present situation of rational sequences). The constant sequence $(0, 0, \dots)$ is the neutral element of addition on $\mathcal{S}$ and $-(r_n)_{n\in\mathbb{N}} = (-r_n)_{n\in\mathbb{N}}$ is the additive inverse of $(r_n)_{n\in\mathbb{N}}$. ■

The reason that we need another step in our construction of $\mathbb{R}$ is the fact that $\mathcal{S}$ is not a field: As soon as 0 occurs, even just once, in the sequence $(r_n)_{n\in\mathbb{N}} \in \mathcal{S}$, the sequence does not have a multiplicative inverse (where the neutral element of multiplication is obviously the constant sequence $(1, 1, \dots)$). The solution to this problem consists of *factoring out* all sequences converging to 0.

Definition and Remark D.29. Let

$$\mathcal{N} := \Big\{(r_n)_{n\in\mathbb{N}} \in \mathcal{S} : \lim_{n\to\infty} r_n = 0\Big\}. \tag{D.51}$$

be the set of rational sequences converging to zero. The relation $\sim$ on $\mathcal{S}$ defined by

$$(r_n)_{n\in\mathbb{N}} \sim (s_n)_{n\in\mathbb{N}} \quad \Leftrightarrow \quad (r_n)_{n\in\mathbb{N}} - (s_n)_{n\in\mathbb{N}} \in \mathcal{N}, \tag{D.52}$$

constitutes an equivalence relation on $\mathcal{S}$ (cf. Def. 2.23): Indeed, $\sim$ is reflexive, as $f \in \mathcal{S}$ implies $f - f = 0 \in \mathcal{N}$; $\sim$ is symmetric, since $f, g \in \mathcal{S}$ with $f - g \in \mathcal{N}$ implies $g - f \in \mathcal{N}$; $\sim$ is transitive, as $f, g, h \in \mathcal{S}$ with $f - g \in \mathcal{N}$ and $g - h \in \mathcal{N}$ implies $f - h = f - g + g - h \in \mathcal{N}$.

Definition D.30. (a) Define the set of *real numbers* $\mathbb{R}$ as the set of equivalence classes of the equivalence relation $\sim$ defined in (D.52), i.e.

$$\mathbb{R} := \mathcal{S}/\sim \;\; = \; \{[(r_n)_{n\in\mathbb{N}}] : (r_n)_{n\in\mathbb{N}} \in \mathcal{S}\} \tag{D.53}$$

is the quotient set of $\mathcal{S}$ with respect to $\sim$ (cf. Ex. 2.24(c)).

(b) *Addition* on $\mathbb{R}$ is defined by

$$+ : \mathbb{R}\times\mathbb{R} \longrightarrow \mathbb{R}, \quad \big([f], [g]\big) \mapsto [f] + [g] := [f + g]. \tag{D.54}$$

Multiplication on $\mathbb{R}$ is defined by

$$\cdot : \mathbb{R}\times\mathbb{R} \longrightarrow \mathbb{R}, \quad \big([f], [g]\big) \mapsto [f] \cdot [g] := [fg]. \tag{D.55}$$

—

Once again, for the definitions in Def. D.30(b) to make sense, one needs to check that they do not depend on the chosen representatives of the equivalence classes, and once again, we provide a lemma providing this check:

Lemma D.31. *The definitions in Def.* D.30(b) *do not depend on the chosen representatives, i.e.*

$$\underset{f,g,\tilde{f},\tilde{g}\in\mathcal{S}}{\forall} \quad \big(f-\tilde{f}\in\mathcal{N}\wedge g-\tilde{g}\in\mathcal{N} \quad\Rightarrow\quad f+g-(\tilde{f}+\tilde{g})\in\mathcal{N}\big) \tag{D.56}$$

and

$$\underset{f,g,\tilde{f},\tilde{g}\in\mathcal{S}}{\forall} \quad \big(f-\tilde{f}\in\mathcal{N}\wedge g-\tilde{g}\in\mathcal{N} \quad\Rightarrow\quad fg-(\tilde{f}\tilde{g})\in\mathcal{N}\big). \tag{D.57}$$

Proof. Let $f=(r_n)_{n\in\mathbb{N}}$, $g=(s_n)_{n\in\mathbb{N}}$, $\tilde{f}=(\tilde{r}_n)_{n\in\mathbb{N}}$, $\tilde{g}=(\tilde{s}_n)_{n\in\mathbb{N}}$ be elements of $\mathcal{S}$ such that $f-\tilde{f}\in\mathcal{N}$ and $g-\tilde{g}\in\mathcal{N}$, i.e. $\lim_{n\to\infty}(r_n-\tilde{r}_n)=\lim_{n\to\infty}(s_n-\tilde{s}_n)=0$.

Then (7.11b) implies $0=\lim_{n\to\infty}\big(r_n+s_n-(\tilde{r}_n+\tilde{s}_n)\big)$, proving (D.56).

To prove (D.57), one computes

$$\lim_{n\to\infty}\big(r_ns_n-\tilde{r}_n\tilde{s}_n\big)=\lim_{n\to\infty}\big(r_n(s_n-\tilde{s}_n)-\tilde{s}_n(r_n-\tilde{r}_n)\big)=0, \tag{D.58}$$

where the last equality follows from the boundedness of $(r_n)_{n\in\mathbb{N}}$ and $(\tilde{s}_n)_{n\in\mathbb{N}}$ together with Prop. 7.11(b). ■

We will also use the following auxiliary result:

Proposition D.32. *If $(r_n)_{n\in\mathbb{N}}\in\mathcal{S}$, then precisely one of the following statements is correct:*

$$(r_n)_{n\in\mathbb{N}}\in\mathcal{N}, \tag{D.59a}$$

$$\underset{\epsilon\in\mathbb{Q}^+}{\exists}\ \#\{n\in\mathbb{N}:\ r_n\le\epsilon\}\in\mathbb{N}_0, \tag{D.59b}$$

$$\underset{\epsilon\in\mathbb{Q}^+}{\exists}\ \#\{n\in\mathbb{N}:\ r_n\ge-\epsilon\}\in\mathbb{N}_0. \tag{D.59c}$$

Proof. Let us first verify that the three statements in (D.59) are mutually exclusive. If (D.59a) holds, then, for every $\epsilon\in\mathbb{Q}^+$, $-\epsilon<r_n<\epsilon$ holds for almost all (in particular, for infinitely many) $n\in\mathbb{N}$, i.e. (D.59b) and (D.59c) are both false. If (D.59b) holds, then (D.59a) must be false as we have just seen. Moreover, if $r_n\le\epsilon$ holds for at most finitely many $n\in\mathbb{N}$, then $r_n>\epsilon>0$ must hold for infinitely many $n\in\mathbb{N}$, i.e. (D.59c) is false.

Now suppose (D.59a) and (D.59b) are false. We have to show that (D.59c) is true. Since (D.59a) is false, there exists $\delta>0$ and an increasing sequence of indices $(n_k)_{k\in\mathbb{N}}$ with $|r_{n_k}|>\delta$ for each $k\in\mathbb{N}$. Since (D.59b) is false, there is an increasing sequence of indices $(m_k)_{k\in\mathbb{N}}$ with $r_{m_k}<1/k$. Thus, since $(r_n)_{n\in\mathbb{N}}$ is a Cauchy sequence, only finitely many $r_{n_k}>\delta$ and infinitely many $r_{n_k}<-\delta$. Now, if $N\in\mathbb{N}$ is such that $|r_n-r_m|<\delta/2$ for all $n,m>N$ and $k_0\in\mathbb{N}$ such that $n_{k_0}>N$, then $r_n<-\delta/2$ for each $n>N$ (since $|r_n-r_{n_{k_0}}|<\delta/2$). Thus, (D.59c) holds with $\epsilon:=\delta/2$. ■

Theorem D.33. (a) *The set of real numbers $\mathbb{R}$ with addition and multiplication as defined in Def.* D.30 *forms a field, where $[(0,0,\dots)]$ and $[(1,1,\dots)]$ are the neutral elements with respect to addition and multiplication, respectively.*

(b) *The map*

$$\iota : \mathbb{Q} \longrightarrow \mathbb{R}, \quad \iota(r) := \big[(r, r, \dots)\big], \tag{D.60}$$

is a monomorphism, i.e. it is injective and satisfies

$$\underset{r,s\in\mathbb{Q}}{\forall} \quad \iota(r+s) = \iota(r) + \iota(s), \tag{D.61a}$$

$$\underset{r,s\in\mathbb{Q}}{\forall} \quad \iota(rs) = \iota(r) \cdot \iota(s). \tag{D.61b}$$

It is customary to identify $\mathbb{Q}$ with $\iota(\mathbb{Q})$, as it usually does not cause any confusion. One then just writes r instead of $\big[(r, r, \dots)\big]$.

Proof. (a): Clearly, Def. D.30(b) ensures the laws of associativity and commutativity of addition and multiplication valid in $\mathcal{S}$ are preserved in $\mathbb{R}$, and, likewise, the law of distributivity. It is also immediate from (D.54) and (D.55), respectively, that $[(0, 0, \dots)]$ and $[(1, 1, \dots)]$ are the respective neutral elements of addition and multiplication. Moreover, if $-f$ is the additive inverse of $f \in \mathcal{S}$, then $[-f]$ is the additive inverse of $[f] \in \mathbb{R}$. It remains to show that each $x = [(r_n)_{n\in\mathbb{N}}] \neq [(0, 0, \dots)]$ has a multiplicative inverse x^{-1} in $\mathbb{R}$. We claim $x^{-1} = [(s_n)_{n\in\mathbb{N}}]$, where

$$\underset{n\in\mathbb{N}}{\forall} \quad s_n := \begin{cases} r_n^{-1} & \text{for } r_n \neq 0, \\ 1 & \text{for } r_n = 0. \end{cases} \tag{D.62}$$

We need to verify $[(s_n)_{n\in\mathbb{N}}] \in \mathbb{R}$, i.e. $(s_n)_{n\in\mathbb{N}}$ is a Cauchy sequence. We know $(r_n)_{n\in\mathbb{N}}$ is a Cauchy sequence that does not converge to 0. Thus, according to Prop. D.32, there exists $\delta > 0$ and $M \in \mathbb{N}$ such that, for each $n > M$, we have $|r_n| > \delta$ (in particular, $r_n \neq 0$). Let $\epsilon > 0$. As $(r_n)_{n\in\mathbb{N}}$ is a Cauchy sequence, there exists $N \in \mathbb{N}$ such that $N \geq M$ and, for each $n, m > N$, $|r_n - r_m| < \epsilon\,\delta^2$. Thus,

$$\underset{n,m>N}{\forall} |s_n - s_m| = \left|\frac{1}{r_n} - \frac{1}{r_m}\right| = \left|\frac{r_n - r_m}{r_n\, r_m}\right| < \frac{\epsilon\,\delta^2}{\delta^2} = \epsilon, \tag{D.63}$$

proving $(s_n)_{n\in\mathbb{N}}$ is a Cauchy sequence. Moreover,

$$\big[(r_n)_{n\in\mathbb{N}}\big] \cdot \big[(s_n)_{n\in\mathbb{N}}\big] = \big[(r_n s_n)_{n\in\mathbb{N}}\big] = \big[(1, 1, \dots,)\big], \tag{D.64}$$

since $r_n s_n = 1$ for almost all $n \in \mathbb{N}$, and the proof of (a) is complete.

(b): The map ι is injective, since $\iota(r) = [(r, r, \dots)] = \iota(s) = [(s, s, \dots)]$ implies $\lim_{n\to\infty}(r - s) = 0$, i.e. $r = s$. Moreover,

$$\iota(r) + \iota(s) = \big[(r, r, \dots)\big] + \big[(s, s, \dots)\big] = \big[(r+s, r+s, \dots)\big] = \iota(r+s), \tag{D.65a}$$

$$\iota(r) \cdot \iota(s) = \big[(r, r, \dots)\big] \cdot \big[(s, s, \dots)\big] = \big[(rs, rs, \dots)\big] = \iota(rs), \tag{D.65b}$$

completing the proof. ■

Definition D.34. We define $\mathbb{R}^+$ to consist of all real numbers represented by sequences $(r_n)_{n\in\mathbb{N}}$ such that there exists $\epsilon \in \mathbb{Q}^+$ satisfying $r_n > \epsilon$ for almost all $n \in \mathbb{N}$, i.e.

$$\mathbb{R}^+ := \left\{ \big[(r_n)_{n\in\mathbb{N}}\big] \in \mathbb{R} : \underset{\epsilon\in\mathbb{Q}^+}{\exists} \#\{n \in \mathbb{N} :\, r_n \leq \epsilon\} \in \mathbb{N}_0 \right\}. \tag{D.66}$$

Proposition D.35. (a) *The definition in* (D.66) *does not depend on the chosen representatives* $(r_n)_{n\in\mathbb{N}}$.

(b) *We have the decomposition*

$$\mathbb{R} = \mathbb{R}^+ \dot\cup \{0\} \dot\cup \mathbb{R}^-, \quad \mathbb{R}^- := -\mathbb{R}^+ = \{x \in \mathbb{R} : -x \in \mathbb{R}^+\}. \tag{D.67}$$

Proof. (a): If $(s_n)_{n\in\mathbb{N}} \in \mathcal{S}$ with $\lim_{n\to\infty}(r_n - s_n) = 0$, then $|r_n - s_n| < \epsilon/2$ for almost all $n \in \mathbb{N}$. Thus, since $|s_n| \geq |r_n| - |r_n - s_n|$, we obtain $s_n > \epsilon/2$ for almost all $n \in \mathbb{N}$, i.e. $\#\{n \in \mathbb{N} : s_n \leq \frac{\epsilon}{2}\} \in \mathbb{N}_0$.

(b) is an immediate consequence of Prop. D.32. ■

Definition D.36. For each $x, y \in \mathbb{R}$, let

$$y \leq x \quad :\Leftrightarrow \quad x - y \in \mathbb{R}_0^+ := \mathbb{R}^+ \cup \{0\}. \tag{D.68}$$

Theorem D.37. (a) *The relation defined in* (D.68) *constitutes a total order on* $\mathbb{R}$ *that is compatible with addition and multiplication, i.e. it satisfies* (4.3)*; in other words* $(\mathbb{R}, +, \cdot, \leq)$ *constitutes a totally ordered field.*

(b) *The map* ι *from* (D.60) *is strictly increasing.*

Proof. (a) follows from (D.68), (D.67), and Th. D.6, once we have shown that $\mathbb{R}^+$ is closed under addition and multiplication. Let $(r_n)_{n\in\mathbb{N}} \in \mathcal{S}$, $(s_n)_{n\in\mathbb{N}} \in \mathcal{S}$. If $r_n > \epsilon_1 \in \mathbb{Q}^+$ for almost all $n \in \mathbb{N}$ and $s_n > \epsilon_2 \in \mathbb{Q}^+$ for almost all $n \in \mathbb{N}$, then $r_n + s_n > \epsilon_1 + \epsilon_2$, showing $\mathbb{R}^+$ is closed under addition. Moreover, $r_n s_n > \epsilon_1\epsilon_2$, showing $\mathbb{R}^+$ is closed under multiplication.

(b): According to Def. D.37, if $r, s \in \mathbb{Q}$ with $s < r$, then $q := r - s \in \mathbb{Q}^+$. In consequence $\iota(r) = \iota(s) + \iota(q)$ by (D.61a), i.e. $\iota(r) - \iota(s) = \iota(q) = [(q, q, \dots)] \in \mathbb{R}^+$, proving $\iota(s) < \iota(r)$. ■

Finally, we will show in Th. D.39 below that the order $\leq$ on $\mathbb{R}$ is complete. However, we first need some additional auxiliary results.

Proposition D.38. (a) *For each* $x \in \mathbb{R}$*, there is* $(r_n)_{n\in\mathbb{N}} \in \mathcal{S}$ *satisfying* $\lim_{n\to\infty} r_n = x$.

(b) *Every* $(r_n)_{n\in\mathbb{N}} \in \mathcal{S}$ *converges in* $\mathbb{R}$ *– more precisely,* $\lim_{n\to\infty} r_n = [(r_n)_{n\in\mathbb{N}}]$.

(c) *Every Cauchy sequence in* $\mathbb{R}$ *converges in* $\mathbb{R}$.

Proof. (a) and (b): If $x = [(r_n)_{n\in\mathbb{N}}]$ with $(r_n)_{n\in\mathbb{N}} \in \mathcal{S}$, then, given $\epsilon > 0$, choose $N \in \mathbb{N}$ such that, for each $m, n > N$, one has $|r_n - r_m| < \epsilon/2$. Then, for each $k > N$, one has $|x - r_k| = |[(r_n - r_k)_{n\in\mathbb{N}}]| < \epsilon$, since $|r_n - r_k| < \epsilon/2$ for all $n \geq k$, showing $\lim_{n\to\infty} r_n = x$.

(c): Let $(x_n)_{n\in\mathbb{N}}$ be a Cauchy sequence in $\mathbb{R}$. According to (a), for each $n \in \mathbb{N}$, there exists $r_n \in \mathbb{Q}$ such that $|x_n - r_n| < \frac{1}{n}$. Then $(r_n)_{n\in\mathbb{N}}$ is a Cauchy sequence: Given $\epsilon > 0$, choose $k \in \mathbb{N}$ such that $\frac{1}{k} < \frac{\epsilon}{3}$ and $|x_n - x_m| < \frac{\epsilon}{3}$ for each $n, m > k$. Then

$$\forall_{n,m>k} |r_n - r_m| \leq |r_n - x_n| + |x_n - x_m| + |x_m - r_m| < \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon, \tag{D.69}$$

showing $(r_n)_{n\in\mathbb{N}}$ is Cauchy. Thus, from (b), we obtain $x \in \mathbb{R}$ with $\lim_{n\to\infty} r_n = x$. We can now show, $\lim_{n\to\infty} x_n = x$ as well: Given $\epsilon > 0$, choose $N \in \mathbb{N}$ such that $\frac{1}{N} < \frac{\epsilon}{2}$ and $|x - r_n| < \frac{\epsilon}{2}$ for each $n > N$. Then

$$\forall_{n>N} |x - x_n| \leq |x - r_n| + |r_n - x_n| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon, \tag{D.70}$$

showing $\lim_{n\to\infty} x_n = x$ and completing the proof. ■

Theorem D.39. *The order $\leq$ on $\mathbb{R}$ is complete, i.e. $(\mathbb{R}, +, \cdot, \leq)$ constitutes a complete totally ordered field.*

Proof. Let $\emptyset \neq A \subseteq \mathbb{R}$ and let $M \in \mathbb{R}$ be an upper bound for A. We have to show that A has a supremum in $\mathbb{R}$. To this end, we recursively construct two Cauchy sequences $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ in $\mathbb{R}$ such that $(x_n)_{n\in\mathbb{N}}$ is increasing, $(y_n)_{n\in\mathbb{N}}$ is decreasing, $x_n < y_n$, and $\lim_{n\to\infty}(y_n - x_n) = 0$. Let $x_1 \in A$ be arbitrary and $y_1 := M$. Define

$$\forall_{n\in\mathbb{N}} \left(\begin{array}{l} x_{n+1} := \begin{cases} (x_n + y_n)/2 & \text{if } (x_n + y_n)/2 \text{ is not an upper bound for } A, \\ x_n & \text{otherwise,} \end{cases} \\ \\ y_{n+1} := \begin{cases} (x_n + y_n)/2 & \text{if } (x_n + y_n)/2 \text{ is an upper bound for } A, \\ y_n & \text{otherwise.} \end{cases} \end{array} \right) \tag{D.71}$$

Then, clearly, the x_n are increasing, the y_n are decreasing, and $x_n \leq y_n$ holds for each $n \in \mathbb{N}$. Moreover, letting $d := M - x_1 \geq 0$, a simple induction shows $y_n - x_n = d/2^{n-1}$ and $\lim_{n\to\infty}(y_n - x_n) = 0$. Also, for $m > n$,

$$x_m - x_n = \sum_{i=n}^{m-1}(x_{i+1} - x_i) \leq d\sum_{i=n}^{m-1} 2^{-i} = \frac{d}{2^n}\sum_{i=n}^{m-1} 2^{-i+n} = \frac{d}{2^n}\sum_{i=0}^{m-1-n} 2^{-i} \leq \frac{2d}{2^n}, \tag{D.72}$$

showing $(x_n)_{n\in\mathbb{N}}$ is a Cauchy sequence. Analogous, one sees that $(y_n)_{n\in\mathbb{N}}$ is a Cauchy sequence. By Prop. D.38(c), we obtain $s \in \mathbb{R}$ such that $s = \lim_{n\to\infty} x_n = \lim_{n\to\infty}(y_n - x_n + x_n) = \lim_{n\to\infty} y_n$. We claim $s = \sup A$. If $s < y$, then there is $n \in \mathbb{N}$ with $s \leq y_n < y$, showing $y \notin A$, i.e. s is an upper bound for A. If $y < s$, then there is $n \in \mathbb{N}$ with $y < x_n \leq s$, showing y is not an upper bound for A. Thus, s is the smallest upper bound for A, i.e. $s = \sup A$. ■

D.6 Uniqueness

We will show in Th. D.43 below that, up to a unique isomorphism, $\mathbb{R}$ is the only complete totally ordered field.

Notation D.40. Let $(A, +, \cdot, \leq)$ be a complete totally ordered field. The neutral elements with respect to $+$ and $\cdot$, we denote with 0_A and 1_A, respectively. We recursively define $(n+1)_A := n_A + 1_A$ for each $n \in \mathbb{N}$. Then $\mathbb{N}_A := \{n_A : n \in \mathbb{N}\}$, $\mathbb{Z}_A := \mathbb{N}_A \cup \{0_A\} \cup \{-n_A : n \in \mathbb{N}\}$, $\mathbb{Q}_A := \{0_A\} \cup \{\frac{k}{l} : k, l \in \mathbb{Z}_A \setminus \{0_A\}\}$.

Proposition D.41. *Let $(A, +, \cdot, \leq)$ and $(B, +, \cdot, \leq)$ be complete totally ordered fields. Moreover, let $\phi : A \longrightarrow B$ be a field isomorphism, i.e. a bijective map, satisfying*

$$\bigvee_{x,y \in A} \phi(x+y) = \phi(x) + \phi(y), \tag{D.73a}$$

$$\bigvee_{x,y \in A} \phi(xy) = \phi(x)\phi(y). \tag{D.73b}$$

(a) *$\phi(n_A) = n_B$ holds for each $n \in \mathbb{N}$.*

(b) *ϕ is strictly isotone, i.e.*

$$\bigvee_{x,y \in A} \big(x < y \;\Rightarrow\; \phi(x) < \phi(y)\big). \tag{D.74}$$

Proof. (a): As (D.73a) and (D.73b) state ϕ to be a group homomorphism with respect to addition and multiplication, respectively, Prop. C.4(a) yields $\phi(0_A) = 0_B$ and $\phi(1_A) = 1_B$. If $n \in \mathbb{N}$, then (D.73a) implies $\phi(n_A + 1_A) = \phi(n_A) + 1_B$ and, thus, an induction shows $\phi(n_A) = n_B$ for each $n \in \mathbb{N}$.

(b): If $x, y \in A$ with $x < y$, then, by Rem. and Def. 7.61, there exists a unique $z \in A$ such that $z^2 = y - x$. Thus, $(\phi(z))^2 = \phi(y) - \phi(x)$. By Th. 4.5(c), we have $(\phi(z))^2 > 0$ and, thus, $\phi(x) < \phi(y)$, proving the strict isotonicity of ϕ. ■

Proposition D.42. *Let $(A, +, \cdot, \leq)$ be a complete totally ordered field. If $\phi : A \longrightarrow A$ is (field) automorphism, i.e. a bijective map, satisfying* (D.73)*, then ϕ is the identity on A.*

Proof. From Prop. D.41(a), we already know $\phi(n) = n$ for each $n \in \mathbb{N}_A$. Next, if $n \in \mathbb{N}_A$, then, using Prop. C.4(b), we obtain $\phi(-n) = -\phi(n) = -n$, showing $\phi(k) = k$ for each $k \in \mathbb{Z}_A$. If $k, l \in \mathbb{Z}_A \setminus \{0_A\}$, then $\phi(k/l) = \phi(k \cdot l^{-1}) = \phi(k) \cdot (\phi(l))^{-1} = kl^{-1}$, where Prop. C.4(b) was used again. Thus, we already have $\phi(q) = q$ for each $q \in \mathbb{Q}_A$. From Prop. D.41(b), we know ϕ to be strictly isotone. Finally, if $x \in A$, then, by Th. 7.68(c), there exist a sequences $(r_n)_{n\in\mathbb{N}}$ in $\mathbb{Q}_A$ and $(s_n)_{n\in\mathbb{N}}$ in $\mathbb{Q}_A$ such that $(r_n)_{n\in\mathbb{N}}$ is strictly increasing, $(s_n)_{n\in\mathbb{N}}$ is strictly decreasing, and

$$\lim_{n\to\infty} r_n = \lim_{n\to\infty} s_n = x.$$

As ϕ is isotone, we obtain

$$\bigvee_{n\in\mathbb{N}} \quad \phi(r_n) = r_n \leq f(x) \leq s_n = \phi(s_n).$$

But then $f(x) = x$ follows from Th. 7.16, proving $\phi = \mathrm{Id}_A$. ■

Theorem D.43. *Let $(A,+,\cdot,\leq)$ and $(B,+,\cdot,\leq)$ be complete totally ordered fields. Then there exists a unique isomorphism $\phi : A \longrightarrow B$, i.e. a unique bijective $\phi : A \longrightarrow B$, satisfying* (D.73a), (D.73b), *and* (D.74).

Proof. Uniqueness: Suppose, $\phi : A \longrightarrow B$ and $\psi : A \longrightarrow B$ are both isomorphisms. Then, according to Prop. D.42, $\phi^{-1} \circ \psi = \mathrm{Id}_A$, where Prop. C.4(c) was used as well. However, this already shows $\psi = \phi$.

Existence: Due to Prop. D.41(b), it suffices to show there exists a bijective $\phi : A \longrightarrow B$, satisfying (D.73). We define ϕ and verify (D.73) in several steps. In the first step, set

$$\underset{n\in\mathbb{N}_0}{\forall} \quad \phi(n_A) := n_B. \tag{D.75}$$

Then $\phi : \mathbb{N}_A \cup \{0_A\} \longrightarrow \mathbb{N}_B \cup \{0_B\}$ is bijective with $\phi^{-1} : \mathbb{N}_B \cup \{0_B\} \longrightarrow \mathbb{N}_A \cup \{0_A\}$, $\phi^{-1}(n_B) = n_A$. We first verify

$$\underset{n\in\mathbb{N}_0}{\forall} \quad \phi(1_A + n_A) = \phi(1_A) + \phi(n_A) : \tag{D.76}$$

Indeed,

$$\phi(1_A + n_A) = \phi((n+1)_A) = (n+1)_B = 1_B + n_B = \phi(1_A) + \phi(n_A).$$

Next, we verify

$$\underset{m,n\in\mathbb{N}_0}{\forall} \quad \phi(m_A + n_A) = \phi(m_A) + \phi(n_A) \tag{D.77}$$

via induction on m: The case $m = 0$ holds, due to $\phi(0_A + n_A) = \phi(n_A) = n_B = 0_B + n_B = \phi(0_A) + \phi(n_A)$. The case $m = 1$ holds, due to (D.76). For the induction step, we compute

$$\begin{aligned} \phi((m+1)_A + n_A) \quad &= \quad \phi(m_A + 1_A + n_A) \overset{\text{ind. hyp.}}{=} \phi(m_A) + \phi(1_A + n_A) \\ &\overset{\text{(D.76)}}{=} \quad \phi(m_A) + \phi(1_A) + \phi(n_A) \overset{\text{(D.76)}}{=} \phi((m+1)_A) + \phi(n_A), \end{aligned}$$

proving (D.77). We now prove

$$\underset{m,n\in\mathbb{N}_0}{\forall} \quad \phi(m_A n_A) = \phi(m_A)\phi(n_A) \tag{D.78}$$

via induction on m: The case $m = 0$ holds, due to

$$\phi(0_A n_A) = \phi(0_A) = 0_B = \phi(0_A)\phi(n_A).$$

For the induction step, we compute

$$\begin{aligned} \phi((m+1)_A\, n_A) \quad &= \quad \phi((m_A + 1_A)\, n_A) = \phi(m_A n_A + n_A) \overset{\text{(D.77)}}{=} \phi(m_A n_A) + \phi(n_A) \\ &\overset{\text{ind. hyp.}}{=} \quad \phi(m_A)\phi(n_A) + \phi(n_A) = \big(\phi(m_A) + 1_B\big)\, \phi(n_A) \\ &\overset{\text{(D.77)}}{=} \quad \phi(m_A + 1_A)\, \phi(n_A) = \phi((m+1)_A)\, \phi(n_A). \end{aligned}$$

In particular, according to (D.77) and (D.78), (D.73) holds for each $x, y \in \mathbb{N}_A$.

In the second step, set

$$\underset{n\in\mathbb{N}}{\forall} \quad \phi(-n_A) := -n_B. \tag{D.79}$$

Then $\phi : \mathbb{Z}_A \longrightarrow \mathbb{Z}_B$ is still bijective, where, for each $m, n \in \mathbb{N}$, we have $\phi^{-1}(-n_B) = -n_A$.

Let $m, n \in \mathbb{N}_0$. If $m \leq n$, then

$$\phi(-m_A + n_A) = \phi((n-m)_A) = (n-m)_B = -m_B + n_B = \phi(-m_A) + \phi(n_A).$$

If $m > n$, then

$$\phi(-m_A + n_A) = \phi(-(m-n)_A) = -(m-n)_B = -m_B + n_B = \phi(-m_A) + \phi(n_A).$$

Now, for arbitrary $m, n \in \mathbb{N}_0$, $\phi(m_A + (-n_A)) = \phi(-n_A + m_A) = \phi(-n_A) + \phi(m_A) = \phi(m_A) + \phi(-n_A)$ and $\phi(-m_A + (-n_A)) = \phi(-(m_A + n_A) = -\phi(m_A + n_A) = -(\phi(m_A) + \phi(n_A)) = -\phi(m_A) - \phi(n_A) = \phi(-m_A) + \phi(-n_A)$. We now consider multiplication, still for $m, n \in \mathbb{N}_0$:

$$\phi((-m_A)\, n_A) = -\phi(m_A n_A) \overset{(D.78)}{=} -\phi(m_A)\phi(n_A) = \phi(-m_A)\phi(n_A).$$

Then $\phi(m_A\,(-n_A)) = \phi(m_A)\phi(-n_A)$ also follows and

$$\phi((-m_A)(-n_A) = \phi(m_A n_A) \overset{(D.78)}{=} \phi(m_A)\phi(n_A) = \phi(-m_A)\phi(-n_A).$$

Thus, we have verified (D.73) for each $x, y \in \mathbb{Z}_A$.

In the third step, set

$$\underset{k,l\in\mathbb{Z}_A\setminus\{0_A\}}{\forall} \quad \phi(k/l) := \phi(k)/\phi(l). \tag{D.80}$$

We verify that (D.80) well-defines ϕ for each $q \in \mathbb{Q}_A$: If $m, n, k, l \in \mathbb{Z}_A$ with $n, l \neq 0_A$, then $m/n = k/l$ implies $ml = kn$ and $\phi(m)\phi(l) = \phi(ml) = \phi(kn) = \phi(k)\phi(n)$. Thus,

$$\phi(m/n) = \phi(m)/\phi(n) = \phi(k)/\phi(l) = \phi(k/l).$$

We show $\phi : \mathbb{Q}_A \longrightarrow \mathbb{Q}_B$ to be bijective by providing the inverse map: Define $\psi : \mathbb{Q}_B \longrightarrow \mathbb{Q}_A$ by setting

$$\underset{k,l\in\mathbb{Z}_B\setminus\{0_B\}}{\forall} \quad \psi(k/l) := \phi^{-1}(k)/\phi^{-1}(l).$$

We claim that $\psi = \phi^{-1}$ on $\mathbb{Q}_B$: Indeed, for each $k, l \in \mathbb{Z}_A \setminus \{0_A\}$ and for each $m, n \in \mathbb{Z}_B \setminus \{0_B\}$

$$\begin{aligned}
\psi(\phi(k/l)) &= \psi\big(\phi(k)/\phi(l)\big) = \phi^{-1}(\phi(k))/\phi^{-1}(\phi(l)) = k/l, \\
\phi(\psi(m/n)) &= \phi\big(\phi^{-1}(m)/\phi^{-1}(n)\big) = \phi(\phi^{-1}(m))/\phi(\phi^{-1}(n)) = m/n.
\end{aligned}$$

Moreover, we have

$$\phi\left(\frac{m}{n}+\frac{k}{l}\right)=\phi\left(\frac{ml+kn}{nl}\right)=\frac{\phi(ml+kn)}{\phi(nl)}\overset{\text{(D.73) for } \mathbb{Z}_A}{=}\frac{\phi(m)\phi(l)+\phi(k)\phi(n)}{\phi(n)\phi(l)}$$
$$=\frac{\phi(m)}{\phi(n)}+\frac{\phi(k)}{\phi(l)}=\phi\left(\frac{m}{n}\right)+\phi\left(\frac{k}{l}\right)$$

and

$$\phi\left(\frac{m}{n}\cdot\frac{k}{l}\right)=\phi\left(\frac{mk}{nl}\right)=\frac{\phi(mk)}{\phi(nl)}\overset{\text{(D.73) for } \mathbb{Z}_A}{=}\frac{\phi(m)\phi(k)}{\phi(n)\phi(l)}$$
$$=\frac{\phi(m)}{\phi(n)}\cdot\frac{\phi(k)}{\phi(l)}=\phi\left(\frac{m}{n}\right)\cdot\phi\left(\frac{k}{l}\right).$$

Thus, we have verified (D.73) for each $x, y \in \mathbb{Q}_A$.

We now show $\phi : \mathbb{Q}_A \longrightarrow \mathbb{Q}_B$ to be strictly isotone, i.e.

$$\underset{r,s\in\mathbb{Q}_A}{\forall}\quad \Big(r<s\ \Rightarrow\ \phi(r)<\phi(s)\Big): \tag{D.81}$$

Let $r, s \in \mathbb{Q}_A$ such that $r < s$. Then $d := s - r > 0_A$, i.e. there are $m, n \in \mathbb{N}_A$ satisfying $d = \frac{m}{n}$. Then $\phi(s) - \phi(r) = \phi(d) = \frac{\phi(m)}{\phi(n)} > 0_B$, proving $\phi(r) < \phi(s)$.

In the fourth (and last) step, for each $x \in A$, we choose a sequence $(r_n)_{n\in\mathbb{N}}$ in $\mathbb{Q}_A$ such that $x = \lim_{n\to\infty} r_n$ and set

$$\phi(x) := \lim_{n\to\infty} \phi(r_n). \tag{D.82}$$

To show that ϕ is well-defined by (D.82), we have to verify that $(\phi(r_n))_{n\in\mathbb{N}}$ does, indeed, converge in B, and that $\phi(x)$ does not depend on the chosen sequence $(r_n)_{n\in\mathbb{N}}$. As $(r_n)_{n\in\mathbb{N}}$ converges to x, it has to be a Cauchy sequence by Th. 7.29. We show that $(\phi(r_n))_{n\in\mathbb{N}}$ must be a Cauchy sequence as well: Let $\epsilon \in B$, $\epsilon > 0_B$ and choose $\tilde{\epsilon} \in \mathbb{Q}_B$ such that $0_B < \tilde{\epsilon} < \epsilon$. Then

$$\underset{N\in\mathbb{N}}{\exists}\quad \underset{n,m>N}{\forall}\quad |r_n - r_m| < \phi^{-1}(\tilde{\epsilon}).$$

As ϕ is strictly isotone, we obtain

$$\underset{n,m>N}{\forall}\quad |\phi(r_n) - \phi(r_m)| < \tilde{\epsilon} < \epsilon,$$

proving $(\phi(r_n))_{n\in\mathbb{N}}$ to be a Cauchy sequence. Now Th. 7.29 implies the convergence of $(\phi(r_n))_{n\in\mathbb{N}}$. Next, we show $\lim_{n\to\infty} r_n = 0_A$ implies $\lim_{n\to\infty} \phi(r_n) = 0_B$ for each sequence in $\mathbb{Q}_A$: Indeed, as above, let $\epsilon > 0_B$ and choose $\tilde{\epsilon} \in \mathbb{Q}_B$ such that $0_B < \tilde{\epsilon} < \epsilon$. Then

$$\underset{N\in\mathbb{N}}{\exists}\quad \underset{n>N}{\forall}\quad |r_n| < \phi^{-1}(\tilde{\epsilon}).$$

As ϕ is strictly isotone, we obtain

$$\underset{n>N}{\forall}\quad |\phi(r_n)| < \tilde{\epsilon} < \epsilon,$$

proving $\lim_{n\to\infty}\phi(r_n) = 0_B$. Thus, if $(r_n)_{n\in\mathbb{N}}$ and $(s_n)_{n\in\mathbb{N}}$ are sequences in $\mathbb{Q}_A$ such that $\lim_{n\to\infty} r_n = x = \lim_{n\to\infty} s_n$, then

$$\lim_{n\to\infty}\phi(r_n) = \lim_{n\to\infty}\phi(r_n - s_n + s_n) = 0_B + \lim_{n\to\infty}\phi(s_n) = \lim_{n\to\infty}\phi(s_n),$$

showing ϕ to be well-defined by (D.82). To see that ϕ is injective, let $x, y \in A$ with $x < y$ and choose $r, s \in \mathbb{Q}_A$ such that $x < r < s < y$. If $(r_n)_{n\in\mathbb{N}}$ and $(s_n)_{n\in\mathbb{N}}$ are sequences in $\mathbb{Q}_A$ such that $x = \lim_{n\to\infty} r_n$ and $y = \lim_{n\to\infty} s_n$, then

$$\underset{N\in\mathbb{N}}{\exists}\quad \underset{n>N}{\forall}\quad \Big(r_n < r < s < s_n\Big).$$

As ϕ is strictly isotone, we obtain

$$\underset{n>N}{\forall}\quad \Big(\phi(r_n) < \phi(r) < \phi(s) < \phi(s_n)\Big),$$

showing $\phi(x) \neq \phi(y)$ and the injectivity of ϕ. To see that ϕ is surjective, let $b \in B$ and let $(r_n)_{n\in\mathbb{N}}$ be an increasing sequence in $\mathbb{Q}_B$ such that $b = \lim_{n\to\infty} r_n$. Then $(\phi^{-1}(r_n))_{n\in\mathbb{N}}$ is an increasing sequence in $\mathbb{Q}_A$ that is bounded, i.e. it must converge to some $a \in A$. Then

$$\phi(a) = \lim_{n\to\infty}\phi(\phi^{-1}(r_n)) = \lim_{n\to\infty} r_n = b,$$

showing ϕ to be surjective. Finally, if $x, y \in A$, then let $(r_n)_{n\in\mathbb{N}}$ and $(s_n)_{n\in\mathbb{N}}$ be sequences in $\mathbb{Q}_A$ such that $x = \lim_{n\to\infty} r_n$ and $y = \lim_{n\to\infty} s_n$. Then

$$\phi(x+y) = \lim_{n\to\infty}\phi(r_n + s_n) = \lim_{n\to\infty}\phi(r_n) + \lim_{n\to\infty}\phi(s_n) = \phi(x) + \phi(y)$$

and

$$\phi(xy) = \lim_{n\to\infty}\phi(r_n s_n) = \lim_{n\to\infty}\phi(r_n)\ \lim_{n\to\infty}\phi(s_n) = \phi(x)\phi(y).$$

Thus, we have verified (D.73) for each $x, y \in A$, and, thereby completed the proof. ■

E Series: Additional Material

E.1 Riemann Rearrangement Theorem

Here, we provide the details for the proof of the Riemann rearrangement Th. 7.93, that was merely sketched in Sec. 7.3.3.

Proof of Th. 7.93. As already stated in the sketch, we define

$$\underset{k\in\mathbb{N}}{\forall}\quad x_k := \begin{cases} -k & \text{for } x = -\infty, \\ x & \text{for } x \in \mathbb{R}, \\ k & \text{for } x = \infty, \end{cases} \qquad y_k := \begin{cases} -k & \text{for } y = -\infty, \\ y & \text{for } y \in \mathbb{R}, \\ k & \text{for } y = \infty, \end{cases} \tag{E.1}$$

noting $x_k \leq y_k$ for almost all $k \in \mathbb{N}$. Next, we observe

$$\mathbb{N} = I^+ \dot\cup I^-, \quad \text{where} \tag{E.2a}$$
$$I^+ := \{j \in \mathbb{N} : a_j \geq 0\}, \tag{E.2b}$$
$$I^- := \{j \in \mathbb{N} : a_j < 0\}. \tag{E.2c}$$

We have to define a suitable bijective map $\phi : \mathbb{N} \longrightarrow \mathbb{N}$ such that

$$\underset{j\in\mathbb{N}}{\forall} \quad b_j := a_{\phi(j)}, \tag{E.3a}$$
$$\underset{n\in\mathbb{N}}{\forall} \quad t_n := \sum_{j=1}^{n} b_j. \tag{E.3b}$$

The definition of ϕ will be recursive, and we will also need to recursively define an auxiliary sequence $(\sigma_j)_{j\in\mathbb{N}}$ taking values in $\{-1, 1\}$, serving as an accounting tool to keep track if we are in the process of moving right (i.e. adding a_j^+) or moving left (i.e. subtracting a_j^-). Moreover, we need a recursively defined auxiliary function $\kappa : \mathbb{N} \longrightarrow \mathbb{N}$ to update the left and right boundaries x_k and y_k, respectively, to handle the first and third case of (E.1) if need be. The recursion is initialized by

$$\phi(1) := 1, \tag{E.4a}$$
$$\sigma_1 := \begin{cases} 1 & \text{if } t_1 \leq y_1, \\ -1 & \text{if } t_1 > y_1, \end{cases} \tag{E.4b}$$
$$\kappa(1) := \begin{cases} 1 & \text{if } t_1 \leq y_1, \\ 2 & \text{if } t_1 > y_1, \end{cases} \tag{E.4c}$$

and completed by

$$\underset{j>1}{\forall} \quad \phi(j) := \begin{cases} \min\left(I^+ \setminus \phi\{1, \ldots, j-1\}\right) & \text{if } \sigma_{j-1} = 1, \\ \min\left(I^- \setminus \phi\{1, \ldots, j-1\}\right) & \text{if } \sigma_{j-1} = -1, \end{cases} \tag{E.5a}$$
$$\underset{j>1}{\forall} \quad \sigma_j := \begin{cases} 1 & \text{if } \sigma_{j-1} = 1 \text{ and } t_j \leq y_{\kappa(j-1)}, \\ -1 & \text{if } \sigma_{j-1} = 1 \text{ and } t_j > y_{\kappa(j-1)}, \\ -1 & \text{if } \sigma_{j-1} = -1 \text{ and } t_j \geq x_{\kappa(j-1)}, \\ 1 & \text{if } \sigma_{j-1} = -1 \text{ and } t_j < x_{\kappa(j-1)}, \end{cases} \tag{E.5b}$$
$$\underset{j>1}{\forall} \quad \kappa(j) := \begin{cases} \kappa(j-1) & \text{if } \sigma_{j-1} = 1 \text{ and } t_j \leq y_{\kappa(j-1)}, \\ 1 + \kappa(j-1) & \text{if } \sigma_{j-1} = 1 \text{ and } t_j > y_{\kappa(j-1)}, \\ \kappa(j-1) & \text{if } \sigma_{j-1} = -1 \text{ and } t_j \geq x_{\kappa(j-1)}, \\ 1 + \kappa(j-1) & \text{if } \sigma_{j-1} = -1 \text{ and } t_j < x_{\kappa(j-1)}. \end{cases} \tag{E.5c}$$

We note that ϕ is well-defined, since, according to (7.86), both I^+ and I^- must have infinitely many elements. Moreover, ϕ is injective, since, for $j_1 < j_2$, $\phi(j_2) \neq \phi(j_1)$ is immediate from (E.5a). Finally, ϕ is also surjective: Otherwise, there is a smallest

$n \in \mathbb{N} \setminus \{1\}$ such that $n \notin \phi(\mathbb{N})$. Suppose $n \in I^+$. Then, according to (E.5a), there must be $j_0 \in \mathbb{N}$ such that $\sigma_j = -1$ for every $j > j_0$, i.e., according to (E.5b) and (E.5c), $t_j \geq x_{\kappa(j_0)} \in \mathbb{R}$ for each $j > j_0$, which is in contradiction to the $\sum_{j=1}^{\infty} a_j^- = \infty$ part of (7.86). Analogously, $n \in I^-$ leads to a contradiction to the $\sum_{j=1}^{\infty} a_j^+ = \infty$ part of (7.86), completing the proof of surjectivity of ϕ. So we have shown that $\sum_{j=1}^{\infty} b_j$ is a rearrangement of $\sum_{j=1}^{\infty} a_j$ as desired. We still need to verify that $\sum_{j=1}^{\infty} b_j$ (i.e. $(t_n)_{n\in\mathbb{N}}$) has precisely all elements of $[x, y]$ as cluster points. To this end, first note that, due to (7.86) and (E.1), $\lim_{j\to\infty} x_{\kappa(j)} = -\infty$ holds if, and only if, $x = -\infty$; and $\lim_{j\to\infty} x_{\kappa(j)} = \infty$ holds if, and only if, $x = \infty$; and likewise for the $y_{\kappa(j)}$ and y. If $x = -\infty$, then $\lim_{j\to\infty} x_{\kappa(j)} = -\infty$ and the bijectivity of ϕ together with (E.5b) and (E.5c) implies

$$\underset{N\in\mathbb{N}}{\forall} \quad \underset{j\in\mathbb{N}}{\exists} \quad t_j < x_{\kappa(j-1)} \leq -N, \tag{E.6}$$

showing $-\infty$ is a cluster point of $(t_n)_{n\in\mathbb{N}}$. Analogously, if $y = \infty$, then $\lim_{j\to\infty} y_{\kappa(j)} = \infty$ and the bijectivity of ϕ together with (E.5b) and (E.5c) implies

$$\underset{N\in\mathbb{N}}{\forall} \quad \underset{j\in\mathbb{N}}{\exists} \quad t_j > y_{\kappa(j-1)} \geq N, \tag{E.7}$$

showing ∞ is a cluster point of $(t_n)_{n\in\mathbb{N}}$. Now let $\xi \in [x, y] \cap \mathbb{R}$ and $\epsilon > 0$. Due to $\lim_{j\to\infty} a_j^+ = \lim_{j\to\infty} a_j^- = 0$, we have

$$\underset{N\in\mathbb{N}}{\exists} \quad \underset{j>N}{\forall} \quad t_j - t_{j-1} < \epsilon. \tag{E.8}$$

Due to the bijectivity of ϕ together with (E.5b) and (E.5c), for each $j_0 \in \mathbb{N}$, there exists $j > \max\{j_0, N\}$ such that $t_{j-1} \leq \xi \leq t_j$, showing ξ is a cluster point of $(t_n)_{n\in\mathbb{N}}$. On the other hand, if $\xi \in]-\infty, x[$, then $x \neq -\infty$. If $x = \infty$, then $\lim_{j\to\infty} t_j = \infty$ and ξ is not a cluster point of $(t_n)_{n\in\mathbb{N}}$. If $\xi < x < \infty$, then let $\epsilon := (x - \xi)/2$ and choose N as in (E.8). Then, by (E.5b) and (E.5c), for each $j > N$, $t_j > x - \epsilon = \xi + \epsilon$, showing ξ is not a cluster point of $(t_n)_{n\in\mathbb{N}}$. Analogously, one sees that $\xi \in]y, \infty[$ can not be a cluster point of $(t_n)_{n\in\mathbb{N}}$. ■

E.2 b-Adic Representations of Real Numbers

The main goal of this section is to provide a proof of Th. 7.99. We begin with some preparatory lemmas.

Lemma E.1. *Given a natural number $b \geq 2$, consider the b-adic series given by* (7.96)*. Then*

$$\sum_{\nu=0}^{\infty} d_{N-\nu}\, b^{N-\nu} \leq b^{N+1}, \tag{E.9}$$

and, in particular, the b-adic series converges to some $x \in \mathbb{R}_0^+$. Moreover, equality in (E.9) *holds if, and only if, $d_n = b - 1$ for every $n \in \{N, N-1, N-2, \dots\}$.*

Proof. One estimates, using the formula for the value of a geometric series:

$$\sum_{\nu=0}^{\infty} d_{N-\nu}\, b^{N-\nu} \leq \sum_{\nu=0}^{\infty} (b-1)\, b^{N-\nu} = (b-1) b^N \sum_{\nu=0}^{\infty} b^{-\nu} = (b-1) b^N \, \frac{1}{1-\frac{1}{b}} = b^{N+1}. \tag{E.10}$$

Note that (E.10) also shows that equality is achieved if all d_n are equal to $b-1$. Conversely, if there is $n \in \{N, N-1, N-2, \dots\}$ such that $d_n < b-1$, then there is $\tilde{n} \in \mathbb{N}$ such that $d_{N-\tilde{n}} < b-1$ and one estimates

$$\sum_{\nu=0}^{\infty} d_{N-\nu}\, b^{N-\nu} < \sum_{\nu=0}^{\tilde{n}-1} d_{N-\nu}\, b^{N-\nu} + (b-1) b^{N-\tilde{n}} + \sum_{\nu=\tilde{n}+1}^{\infty} d_{N-\nu}\, b^{N-\nu} \leq b^{N+1}, \tag{E.11}$$

showing that the inequality in (E.9) is strict. ∎

Lemma E.2. *Given a natural number $b \geq 2$, consider two b-adic series*

$$x := \sum_{\nu=0}^{\infty} d_{N-\nu}\, b^{N-\nu} = \sum_{\nu=0}^{\infty} e_{N-\nu}\, b^{N-\nu}, \tag{E.12}$$

$N \in \mathbb{Z}$ and $d_n, e_n \in \{0, \dots, b-1\}$ for each $n \in \{N, N-1, N-2, \dots\}$. If $d_N < e_N$, then $e_N = d_N + 1$, $d_n = b-1$ for each $n < N$ and $e_n = 0$ for each $n < N$.

Proof. By subtracting $d_N b^N$ from both series, one can assume $d_N = 0$ without loss of generality. From Lem. E.1, we know

$$x = \sum_{\nu=0}^{\infty} d_{N-\nu}\, b^{N-\nu} = \sum_{\nu=0}^{\infty} d_{N-1-\nu}\, b^{N-1-\nu} \leq b^N. \tag{E.13a}$$

On the other hand:

$$x = \sum_{\nu=0}^{\infty} e_{N-\nu}\, b^{N-\nu} \geq b^N. \tag{E.13b}$$

Combining (E.13a) and (E.13b) yields $x = b^N$. Once again employing Lem. E.1, (E.13a) also shows that $d_n = b-1$ for each $n \leq N-1$ as claimed. Since $e_N > 0$ and $e_n \geq 0$ for each n, equality in (E.13b) can only occur for $e_N = 1$ and $e_n = 0$ for each $n < N$, thereby completing the proof of the lemma. ∎

Notation E.3. For each $x \in \mathbb{R}$, we let

$$\lfloor x \rfloor := \max\{k \in \mathbb{Z} : \, k \leq x\} \tag{E.14}$$

denote the *integral part* of x (also called *floor* of x or x *rounded down*).

Proof of Th. 7.99. We start by constructing numbers N and d_n, $n \in \{N, N-1, N-2, \dots\}$, such that (7.97) holds. For $x = 0$, one chooses an arbitrary $N \in \mathbb{Z}$ and $d_n = 0$ for each $n \in \{N, N-1, N-2, \dots\}$. Thus, for the remainder of the proof, fix $x > 0$. Let

$$N := \max\{n \in \mathbb{Z} : \, b^n \leq x\}. \tag{E.15}$$

The numbers $d_{N-n} \in \{0, \dots, b-1\}$ and $x_n \in \mathbb{R}^+$, $n \in \mathbb{N}_0$, are defined inductively by letting

$$d_N := \left\lfloor \frac{x}{b^N} \right\rfloor, \qquad x_0 := d_N b^N, \tag{E.16a}$$

$$d_{N-n} := \left\lfloor \frac{x - x_{n-1}}{b^{N-n}} \right\rfloor, \qquad x_n := x_{n-1} + d_{N-n}\, b^{N-n} \quad \text{for } n \geq 1. \tag{E.16b}$$

Claim 1. One can verify by induction on n that the numbers d_{N-n} and x_n enjoy the following properties for each $n \in \mathbb{N}_0$:

$$d_{N-n} \in \{0, \dots, b-1\}, \tag{E.17a}$$

$$0 < x_n = \sum_{\nu=0}^{n} d_{N-\nu}\, b^{N-\nu} \leq x, \tag{E.17b}$$

$$x - x_n < b^{N-n}. \tag{E.17c}$$

Proof. The induction is carried out for all three statements of (E.17) simultaneously. From (E.15), we know $b^N \leq x < b^{N+1}$, i.e. $1 \leq \frac{x}{b^N} < b$. Using (E.16a), this yields $d_N \in \{1, \dots, b-1\}$ and $0 < x_0 = d_N b^N = b^N d_N \leq b^N \frac{x}{b^N} = x$ as well as $x - x_0 = x - d_N b^N = b^N(\frac{x}{b^N} - d_N) < b^N$. For $n \geq 1$, by induction, one obtains $0 \leq x - x_{n-1} < b^{1+N-n}$, i.e. $0 \leq \frac{x - x_{n-1}}{b^{N-n}} < b$. Using (E.16b), this yields $d_{N-n} \in \{0, \dots, b-1\}$ and $x_n = x_{n-1} + d_{N-n} b^{N-n} \leq x_{n-1} + b^{N-n} \frac{x - x_{n-1}}{b^{N-n}} = x$. Moreover, by induction, $0 < x_{n-1} = \sum_{\nu=0}^{n-1} d_{N-\nu}\, b^{N-\nu}$, such that (E.16b) implies $x_n = x_{n-1} + d_{N-n}\, b^{N-n} \geq x_{n-1} > 0$ and $x_n = x_{n-1} + d_{N-n}\, b^{N-n} = d_{N-n}\, b^{N-n} + \sum_{\nu=0}^{n-1} d_{N-\nu}\, b^{N-\nu} = \sum_{\nu=0}^{n} d_{N-\nu}\, b^{N-\nu}$. Finally, $x - x_n = x - x_{n-1} - d_{N-n}\, b^{N-n} = b^{N-n}(\frac{x - x_{n-1}}{b^{N-n}} - d_{N-n}) \leq b^{N-n}$, completing the proof of the claim. ▲

Since, for each $n \in \mathbb{N}_0$,

$$0 \overset{\text{(E.17b)}}{\leq} x - x_n \overset{\text{(E.17c)}}{<} b^{N-n}, \tag{E.18}$$

and $\lim_{n\to\infty} b^{N-n} = 0$, we have $\lim_{n\to\infty} x_n = x$, thereby establishing (7.97).

It remains to verify the equivalence of (i) – (iv).

(ii) ⇒ (i) is trivial.

"(iii) ⇒ (i)": Assume (iii) holds. Without loss of generality, we can assume that n_0 is the largest index such that $d_n = 0$ for each $n \leq n_0$. We distinguish two cases. If $n_0 < N - 1$ or $d_N \neq 1$, then

$$\sum_{\nu=0}^{N-n_0-2} d_{N-\nu}\, b^{N-\nu} + (d_{n_0+1} - 1) b^{n_0+1} + \sum_{\nu=N-n_0}^{\infty} (b-1)\, b^{N-\nu}$$

is a different b-adic representation of x and its first coefficient is nonzero. If $n_0 = N - 1$ and $d_N = 1$, then

$$\sum_{\nu=1}^{\infty} (b-1)\, b^{N-\nu} = \sum_{\nu=0}^{\infty} (b-1)\, b^{N-1-\nu}$$

is a different b-adic representation of x and its first coefficient is nonzero.

"(iv) $\Rightarrow$ (i)": Assume (iv) holds. Without loss of generality, we can assume that n_0 is the largest index such that $d_n = b-1$ for each $n \le n_0$. Then

$$\sum_{\nu=0}^{N-n_0-2} d_{N-\nu}\, b^{N-\nu} + (d_{n_0+1}+1)b^{n_0+1} + \sum_{\nu=N-n_0}^{\infty} 0\, b^{N-\nu}$$

is a different b-adic representation of x and its first coefficient is nonzero.

We will now show that, conversely, (i) implies (ii), (iii), and (iv). To that end, let $x > 0$ and suppose that x has two different b-adic representations

$$x = \sum_{\nu=0}^{\infty} d_{N_1-\nu}\, b^{N_1-\nu} = \sum_{\nu=0}^{\infty} e_{N_2-\nu}\, b^{N_2-\nu} \tag{E.19}$$

with $N_1, N_2 \in \mathbb{Z}$; $d_n, e_n \in \{0, \dots, b-1\}$; and $d_{N_1}, e_{N_2} > 0$. This implies

$$x \ge b^{N_1}, \quad x \ge b^{N_2}. \tag{E.20a}$$

Moreover, Lem. E.1 yields

$$x \le b^{N_1+1}, \quad x \le b^{N_2+1}. \tag{E.20b}$$

If $N_2 > N_1$, then (E.20) imply $N_2 = N_1 + 1$ and $b^{N_2} \le x \le b^{N_1+1} = b^{N_2}$, i.e. $x = b^{N_2} = b^{N_1+1}$. Since $e_{N_2} > 0$, one must have $e_{N_2} = 1$, and, in turn, $e_n = 0$ for each $n < N_2$. Moreover, $x = b^{N_1+1}$ and Lem. E.1 imply that $d_n = b-1$ for each $n \in \{N_1, N_1-1, \dots\}$. Thus, for $N_2 > N_1$, the value of N_1 is determined by N_2 and the values of all d_n and e_n are also completely determined, showing that there are precisely two b-adic representations of x. Moreover, the d_n have the property required in (iv) and the e_n have the property required in (iii). The argument also shows that, for $N_1 > N_2$, one must have $N_1 = N_2 + 1$ with the e_n taking the values of the d_n and vice versa. Once again, there are precisely two b-adic representations of x; now the d_n have the property required in (iii) and the e_n have the property required in (iv).

It remains to consider the case $N := N_1 = N_2$. Since, by hypothesis, the two b-adic representations of x in (E.19) are not identical, there must be a largest index $n \le N$ such that $d_n \ne e_n$. Thus, (E.19) implies

$$y := \sum_{\nu=0}^{\infty} d_{n-\nu}\, b^{n-\nu} = \sum_{\nu=0}^{\infty} e_{n-\nu}\, b^{n-\nu}. \tag{E.21}$$

Now Lem. E.2 shows that there are precisely two b-adic representations of x, one having the property required in (iii) and the other having property required in (iv).

Thus, in each case ($N_2 > N_1$, $N_1 > N_2$, and $N_1 = N_2$), we find that (i) implies (ii), (iii), and (iv), thereby concluding the proof of the theorem. ■

In most cases, it is understood that we work only with decimal representations such that there is no confusion about the meaning of symbol strings like 101.01. However,

in general, 101.01 could also be meant with respect to any other base, and, the number represented by the same string of symbols does obviously depend on the base used. Thus, when working with different representations, one needs some notation to keep track of the base.

Notation E.4. Given a natural number $b \geq 2$ and finite sequences

$$(d_{N_1}, d_{N_1-1}, \ldots, d_0) \in \{0, \ldots, b-1\}^{N_1+1}, \tag{E.22a}$$

$$(e_1, e_2, \ldots, e_{N_2}) \in \{0, \ldots, b-1\}^{N_2}, \tag{E.22b}$$

$$(p_1, p_2, \ldots, p_{N_3}) \in \{0, \ldots, b-1\}^{N_3}, \tag{E.22c}$$

$N_1, N_2, N_3 \in \mathbb{N}_0$ (where $N_2 = 0$ or $N_3 = 0$ is supposed to mean that the corresponding sequence is empty), the respective string

$$\begin{array}{ll} (d_{N_1} d_{N_1-1} \ldots d_0)_b & \text{for } N_2 = N_3 = 0, \\ (d_{N_1} d_{N_1-1} \ldots d_0 \,.\, e_1 \ldots e_{N_2} \overline{p_1 \ldots p_{N_3}})_b & \text{for } N_2 + N_3 > 0 \end{array} \tag{E.23}$$

represents the number

$$\sum_{\nu=0}^{N_1} d_\nu \, b^\nu + \sum_{\nu=1}^{N_2} e_\nu \, b^{-\nu} + \sum_{\alpha=0}^{\infty} \sum_{\nu=1}^{N_3} p_\nu \, b^{-N_2-\alpha N_3-\nu}. \tag{E.24}$$

Example E.5. For the number from (7.95), we get

$$x = (131.\overline{6})_{10} = (10000011.\overline{10})_2 = (83.\overline{\mathrm{A}})_{16} \tag{E.25}$$

(for the hexadecimal system, it is customary to use the symbols 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F).

—

One frequently needs to convert representations with respect to one base into representations with respect to another base. When working with digital computers, conversions between bases 10 and 2 and vice versa are the most obvious ones that come up. Converting representations is related to the following elementary remainder theorem and the well-known long division algorithm.

Theorem E.6. *For each pair of numbers $(a, b) \in \mathbb{N}^2$, there exists a unique pair of numbers $(q, r) \in \mathbb{N}_0^2$ satisfying the two conditions $a = qb + r$ and $0 \leq r < b$.*

Proof. Existence: Define

$$q := \max\{n \in \mathbb{N}_0 : \, nb \leq a\}, \tag{E.26a}$$

$$r := a - qb. \tag{E.26b}$$

Then $q \in \mathbb{N}_0$ by definition and (E.26b) immediately yields $a = qb + r$ as well as $r \in \mathbb{Z}$. Moreover, from (E.26a), $qb \leq a = qb + r$, i.e. $0 \leq r$, in particular, $r \in \mathbb{N}_0$. Since (E.26a) also implies $(q+1)b > a = qb + r$, we also have $b > r$ as required.

Uniqueness: Suppose $(q_1, r_1) \in \mathbb{N}_0$, satisfying the two conditions $a = q_1 b + r_1$ and $0 \leq r_1 < b$. Then $q_1 b = a - r_1 \leq a$ and $(q_1 + 1)b = a - r_1 + b > a$, showing $q_1 = \max\{n \in \mathbb{N}_0 : \, nb \leq a\} = q$. This, in turn, implies $r_1 = a - q_1 b = a - qb = r$, thereby establishing the case. ∎

F Cardinality of $\mathbb{R}$ and Some Related Sets

Theorem F.1. (a) *The set of natural numbers $\mathbb{N}$ is countable.*

(b) *The set of integers $\mathbb{Z}$ is countable: $\#\mathbb{Z} = \#\mathbb{N}$.*

(c) *The set of rational numbers $\mathbb{Q}$ is countable: $\#\mathbb{Q} = \#\mathbb{N}$.*

Proof. (a): The identity $\mathrm{Id} : \mathbb{N} \longrightarrow \mathbb{N}$ shows $\mathbb{N}$ is countable.

(b): Using (D.22), the map

$$\phi : \mathbb{N} \longrightarrow \mathbb{Z}, \quad \phi(n) := \begin{cases} n/2 & \text{if } n \text{ is even,} \\ 0 & \text{if } n = 1, \\ -(n-1)/2 & \text{if } n \text{ is odd,} \end{cases} \tag{F.1}$$

is clearly bijective, proving $\#\mathbb{Z} = \#\mathbb{N}$.

(c): According to (b), $\mathbb{Z}$ and $\mathbb{Z} \setminus \{0\}$ are countable. Then Th. 3.16 implies that $A := \mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ is countable and there is a bijective map $f : \mathbb{N} \longrightarrow A$. It is then immediate from Def. D.20(a) that the map

$$\phi : \mathbb{N} \longrightarrow \mathbb{Q}, \quad \phi(n) := \big[f(n)\big], \tag{F.2}$$

where $[f(n)]$ denotes the equivalence class of $f(n)$ with respect to $\sim$ from (D.29), is surjective. Thus, $\mathbb{Q}$ is countable by Prop. 3.15. ■

In the following theorem and its two corollaries, we will see that the set $\mathbb{R}$ of real numbers is not countable, but has the same cardinality as the power set of $\mathbb{N}$. Moreover, the same is true for every nontrivial interval of real numbers.

Theorem F.2. *Let $a, b \in \mathbb{R}$ with $a < b$. Recalling the notations $\mathcal{F}\big(\mathbb{N}, \{0,1\}\big) = \{0,1\}^{\mathbb{N}}$ for the set of sequences in $\{0,1\}$, we obtain the following equalities of cardinalities:*

$$\#\mathbb{R} = \#]a,b[= \#\{0,1\}^{\mathbb{N}} = \#\mathcal{P}(\mathbb{N}). \tag{F.3}$$

Proof. We devide the proof into the following steps:

(i) $\#\{0,1\}^{\mathbb{N}} = \#\mathcal{P}(\mathbb{N})$.

(ii) $\#]0,1[= \#\{0,1\}^{\mathbb{N}}$.

(iii) $\#]-1,1[= \#\mathbb{R}$.

(iv) $\#]a,b[= \#]0,1[$.

(i): To prove $\#\{0,1\}^{\mathbb{N}} = \#\mathcal{P}(\mathbb{N})$, we have to show the existence of a bijective map $f : \{0,1\}^{\mathbb{N}} \longrightarrow \mathcal{P}(\mathbb{N})$. Given $\sigma \in \{0,1\}^{\mathbb{N}}$, i.e. σ is a function $\sigma : \mathbb{N} \longrightarrow \{0,1\}$, define

$$f(\sigma) := \sigma^{-1}\{1\} = \{n \in \mathbb{N} : \sigma(n) = 1\}. \tag{F.4}$$

Then, indeed, $f : \{0,1\}^{\mathbb{N}} \longrightarrow \mathcal{P}(\mathbb{N})$. It remains to show f is bijective. To verify f is injective, consider $\sigma, \tau \in \{0,1\}^{\mathbb{N}}$. If $\sigma \neq \tau$, then there exists $n \in \mathbb{N}$ with $\sigma(n) \neq \tau(n)$. If $\sigma(n) = 1$, then $\tau(n) = 0$, i.e. $n \in f(\sigma)$, but $n \notin f(\tau)$, showing $f(\sigma) \neq f(\tau)$. Analogously, if $\sigma(n) = 0$, then $\tau(n) = 1$, i.e. $n \in f(\tau)$, but $n \notin f(\sigma)$, again showing $f(\sigma) \neq f(\tau)$, concluding the proof that f is injective. To verify f is surjective, for each $A \in \mathcal{P}(\mathbb{N})$, define

$$\sigma_A : \mathbb{N} \longrightarrow \{0,1\}, \quad \sigma_A(n) := \begin{cases} 1 & \text{if } n \in A, \\ 0 & \text{if } n \notin A. \end{cases} \tag{F.5}$$

Then $\sigma_A \in \{0,1\}^{\mathbb{N}}$ and $f(\sigma_A) = \sigma_A^{-1}\{1\} = A$, proving f is surjective.

(ii): To prove $\#\{0,1\}^{\mathbb{N}} = \#]0,1[$, we have to show the existence of a bijective map $f : \{0,1\}^{\mathbb{N}} \longrightarrow]0,1[$. The map

$$g : \{0,1\}^{\mathbb{N}} \longrightarrow [0,1], \quad g\big((x_i)_{i\in\mathbb{N}}\big) := \sum_{i=1}^{\infty} x_i \, 2^{-i}, \tag{F.6}$$

is well-defined by Lem. E.1 (i.e. $0 \leq g \leq 1$). Moreover, according to Th. 7.99, g is surjective, but not injective, as there are numbers $x \in]0,1[$, that have two different dual (i.e. 2-adic) representations. However, as there are only countably many such numbers, we can use a modification to obtain our desired f. In preparation, we define, for each $n \in \mathbb{N}$, the sequences $e_n := (e_{ni})_{i\in\mathbb{N}}$ and $f_n := (f_{ni})_{i\in\mathbb{N}}$, where

$$\underset{n,i\in\mathbb{N}}{\forall} \quad e_{ni} := \begin{cases} 1 & \text{for } i = n, \\ 0 & \text{for } i \neq n, \end{cases} \tag{F.7a}$$

$$\underset{n,i\in\mathbb{N}}{\forall} \quad f_{ni} := \begin{cases} 1 & \text{for } i > n, \\ 0 & \text{for } i \leq n, \end{cases} \tag{F.7b}$$

and we note

$$g\big((0,0,\dots)\big) = 0, \tag{F.8a}$$

$$g\big((1,1,\dots)\big) = 1, \tag{F.8b}$$

$$g(e_n) = g(f_n) = 2^{-n} \quad \text{for each } n \in \mathbb{N}. \tag{F.8c}$$

We are now in a position to define

$$f : \{0,1\}^{\mathbb{N}} \longrightarrow]0,1[, \quad f\big((x_i)_{i\in\mathbb{N}}\big) := \begin{cases} 2^{-1} & \text{if } (x_i)_{i\in\mathbb{N}} = (0,0,\dots), \\ 2^{-2} & \text{if } (x_i)_{i\in\mathbb{N}} = (1,1,\dots), \\ 2^{-(2n+1)} & \text{if } x_i = e_{ni} \text{ for each } i \in \mathbb{N}, \\ 2^{-(2n+2)} & \text{if } x_i = f_{ni} \text{ for each } i \in \mathbb{N}, \\ \sum_{i=1}^{\infty} x_i \, 2^{-i} & \text{otherwise.} \end{cases} \tag{F.9}$$

Introducing the auxiliary sets

$$A := \{(0,0,\dots),(1,1,\dots)\} \cup \{e_n : n \in \mathbb{N}\} \cup \{f_n : n \in \mathbb{N}\}, \tag{F.10a}$$

$$B := \{2^{-n} : n \in \mathbb{N}\}, \tag{F.10b}$$

it follows from Th. 7.99 that (the following restrictions of f which, to simplify notation, we also denote by f)

$$f : \{0,1\}^{\mathbb{N}} \setminus A \longrightarrow]0,1[\setminus B, \tag{F.11a}$$

and

$$f : A \longrightarrow B \tag{F.11b}$$

are bijective, i.e. the full f of (F.9) is itself bijective, completing the proof of (ii).

(iii): To prove $\#]-1,1[= \#\mathbb{R}$, we have to show the existence of a bijective map $f : \mathbb{R} \longrightarrow]-1,1[$. Since we know from Def. and Rem. 8.27 that $\arctan : \mathbb{R} \longrightarrow]-\pi/2, \pi/2[$ is bijective, we can define

$$f : \mathbb{R} \longrightarrow]0,1[, \quad f(x) := \frac{2 \arctan x}{\pi}. \tag{F.12}$$

However, even though this provides a valid proof, arctan is a somewhat complicated function (as it is defined via sin and cos, which are defined via power series). Thus, it might be desirable to see an alternative proof, using a more elementary f. We claim that

$$f : \mathbb{R} \longrightarrow]-1,1[, \quad f(x) := \frac{x}{|x|+1}, \tag{F.13}$$

is also bijective. Since f is clearly continuous, according to the intermediate value Th. 7.57, it suffices to show

$$\forall_{\epsilon \in]0,1[} \quad \exists_{x_1, x_2 \in \mathbb{R}} \; f(x_1) < -1+\epsilon < 1-\epsilon < f(x_2). \tag{F.14}$$

However, for each $\epsilon \in]0,1[$,

$$x_1 < \frac{-1+\epsilon}{\epsilon} = -\epsilon^{-1} + 1 \;\Rightarrow\; x_1 < x_1 - 1 - \epsilon\, x_1 + \epsilon \;\Rightarrow\; f(x_1) = \frac{x_1}{-x_1+1} < -1+\epsilon,$$

$$x_2 > \frac{1-\epsilon}{\epsilon} = \epsilon^{-1} - 1 \;\Rightarrow\; x_2 > 1 + x_2 - \epsilon - \epsilon\, x_2 \;\Rightarrow\; f(x_2) = \frac{x}{x+1} > 1-\epsilon,$$

proving (F.14) and the surjectivity of f. To verify f is injective, it suffices to show that f is strictly increasing. Since

$$\begin{aligned}
x_1 \le 0 \le x_2 \;\wedge\; x_1 < x_2 \quad &\Rightarrow \quad f(x_1) = \frac{x_1}{-x_1+1} \le 0 \le \frac{x_2}{x_2+1} = f(x_2) \\
&\qquad \wedge\; f(x_1) < f(x_2), \\
x_1 < x_2 \le 0 \quad &\Rightarrow \quad -x_1\, x_2 + x_1 < -x_1\, x_2 + x_2 \\
&\Rightarrow \quad f(x_1) = \frac{x_1}{-x_1+1} < \frac{x_2}{-x_2+1} = f(x_2), \\
0 \le x_1 < x_2 \quad &\Rightarrow \quad x_1\, x_2 + x_1 < x_1\, x_2 + x_2 \\
&\Rightarrow \quad f(x_1) = \frac{x_1}{x_1+1} < \frac{x_2}{x_2+1} = f(x_2),
\end{aligned}$$

showing f is strictly increasing and, hence, injective.

(iv): To prove $\#]a,b[= \#]0,1[$, we have to show the existence of a bijective map $f :]a,b[\longrightarrow]0,1[$. Such a bijective map is given by the (restriction of an) affine map

$$f :]a,b[\longrightarrow]0,1[, \quad f(x) := \frac{x-a}{b-a}. \tag{F.15}$$

The proof that f is bijective can be conducted analogous to (but much simpler than) the proof in (iii), or one can use (for example, from Linear Algebra) that every nonconstant affine map from $\mathbb{R}$ into $\mathbb{R}$ is bijective. ■

Corollary F.3. $\#\mathbb{R} = \#\mathcal{P}(\mathbb{N})$ – *in particular,* $\mathbb{R}$ *is not countable.*

Proof. $\#\mathbb{R} = \#\mathcal{P}(\mathbb{N})$ was proved in Th. F.2 and $\mathcal{P}(\mathbb{N})$ is uncountable by Th. A.69. ■

Corollary F.4. *If* $a, b \in \mathbb{R}$ *with* $a < b$, *then* $\#(\mathbb{Q} \cap]a,b[) = \#\mathbb{N}$ *and* $\#(]a,b[\setminus \mathbb{Q}) = \#\mathbb{R}$, *i.e.* $]a,b[$ *contains countably many rational and uncountably many irrational numbers.*

Proof. Since $\mathbb{Q} \cap]a,b[\subseteq \mathbb{Q}$, the claim $\#(\mathbb{Q} \cap]a,b[) = \#\mathbb{N}$ follows from Th. F.1(c), Prop. 3.14, and Th. 7.68(a).

To prove $\#(]a,b[\setminus \mathbb{Q}) = \#\mathbb{R}$, a bijection between $]a,b[\setminus \mathbb{Q}$ and $\mathbb{R}$ can be constructed analogous to the construction of f in step (ii) of the proof of Th. F.2, making use of the fact that $\#]a,b[= \#\mathbb{R}$ and $\#\mathbb{Q} = \#\mathbb{N}$. ■

Theorem F.5. *The set of complex numbers* $\mathbb{C} = \mathbb{R} \times \mathbb{R}$ *has the same cardinality as* $\mathbb{R}$: $\#(\mathbb{R} \times \mathbb{R}) = \#\mathbb{R} = \#\mathcal{P}(\mathbb{N})$.

Proof. Let

$$A := \{0,1\}^{\mathbb{N}}. \tag{F.16}$$

By an application of Th. F.2, it suffices to prove $\#A = \#(A \times A)$, which is accomplished by showing the existence of a bijective map $f : A \longrightarrow A \times A$. We define

$$f : A \longrightarrow A \times A, \quad f\big((x_j)_{j\in\mathbb{N}}\big) := \big((y_j)_{j\in\mathbb{N}}, (z_j)_{j\in\mathbb{N}}\big), \tag{F.17a}$$

where

$$\bigvee_{j\in\mathbb{N}} \quad y_j := x_{2j-1}, \tag{F.17b}$$

$$\bigvee_{j\in\mathbb{N}} \quad z_j := x_{2j}, \tag{F.17c}$$

and

$$g : A \times A \longrightarrow A, \quad g\big((y_j)_{j\in\mathbb{N}}, (z_j)_{j\in\mathbb{N}}\big) := (x_j)_{j\in\mathbb{N}}, \tag{F.18a}$$

where

$$\bigvee_{j\in\mathbb{N}} \quad x_j := \begin{cases} y_{(j+1)/2} & \text{for } j \text{ odd}, \\ z_{j/2} & \text{for } j \text{ even}. \end{cases} \tag{F.18b}$$

Clearly, $g = f^{-1}$, proving that f is bijective as desired. ■

G Partial Fraction Decomposition

We consider $\mathbb{C}$-valued rational functions of the form

$$z \mapsto R(z) := \frac{P(z)}{Q(z)}, \tag{G.1}$$

where $P, Q : \mathbb{C} \longrightarrow \mathbb{C}$ are polynomials such that $\deg(P) < \deg(Q) =: n$. Using Cor. 8.33 as well as Rem. 6.7, we write Q in the form

$$Q(z) = c \prod_{j=1}^{k} (z - \lambda_j)^{m_j}, \tag{G.2}$$

where $c \in \mathbb{C}$, and $\lambda_1, \ldots, \lambda_k \in \mathbb{C}$, $k \in \{1, \ldots, n\}$, are the *distinct* zeros of Q, $m_j \in \mathbb{N}$ with $\sum_{j=1}^{k} m_j = n$ being their respective multiplicities.

In can be useful to write R as a linear combination of the so-called *partial fractions*

$$\frac{1}{z - \lambda_j}, \ \frac{1}{(z - \lambda_j)^2}, \ \ldots, \ \frac{1}{(z - \lambda_j)^{m_j}} \quad (j = 1, \ldots, k) \tag{G.3}$$

(for example, for the computation of the antiderivative of R, cf. Ex. 10.22(b)). The following Th. G.1 guarantees this is always possible:

Theorem G.1. *Let $P, Q : \mathbb{C} \longrightarrow \mathbb{C}$ be polynomials such that $\deg(P) < \deg(Q) =: n$. Moreover, let $\mathcal{N}(Q)$ denote the set of zeros of Q, and assume Q to have the form of (G.2). Then there exists a unique family of coefficients*

$$a_{jl} \in \mathbb{C} \quad (j = 1, \ldots, k, \ l = 1, \ldots, m_j), \tag{G.4}$$

such that

$$\underset{z \in \mathbb{C} \setminus \mathcal{N}(Q)}{\forall} \quad R(z) = \frac{P(z)}{Q(z)} = \sum_{j=1}^{k} \sum_{l=1}^{m_j} \frac{a_{jl}}{(z - \lambda_j)^l}. \tag{G.5}$$

Proof. We first prove the existence of the decomposition (G.5) via induction on $n = \deg(Q)$. If $n = 1$, then P must be constant and there is nothing to prove. For the induction step, consider $n \geq 2$. Let ζ be a zero of Q with multiplicity $m \in \{1, \ldots, n\}$. Then, according to Rem. 6.7, there exists a polynomial $S : \mathbb{C} \longrightarrow \mathbb{C}$ such that $Q(z) = (z - \zeta)^m S(z)$ and $S(\zeta) \neq 0$. Noting

$$\tilde{R}(z) := \frac{P(z)}{S(z)} - \frac{P(\zeta)}{S(\zeta)} = \frac{P(z)S(\zeta) - S(z)P(\zeta)}{S(z)S(\zeta)} \tag{G.6}$$

and that $P(z)S(\zeta) - S(z)P(\zeta)$ vanishes for $z = \zeta$, there exists a polynomial $T : \mathbb{C} \longrightarrow \mathbb{C}$, $\deg T \leq n - 2$, such that

$$\tilde{R}(z) = \frac{(z - \zeta)T(z)}{S(z)}. \tag{G.7}$$

Thus, for each $z \in \mathbb{C} \setminus \mathcal{N}(Q)$, we have

$$R(z) - \frac{P(\zeta)}{(z-\zeta)^m\, S(\zeta)} = \frac{\tilde{R}(z)}{(z-\zeta)^m} \overset{(G.7)}{=} \frac{T(z)}{(z-\zeta)^{m-1} S(z)}. \tag{G.8}$$

We will now apply (G.8) with $\zeta = \lambda_k$ and $m = m_k$. Since $\deg(T) < n-1 = \deg\big((z-\zeta)^{m-1}S(z)\big) < \deg(Q)$, the induction hypothesis applies to the function in (G.8), yielding coefficients $a_{jl} \in \mathbb{C}$, $j = 1, \ldots, k$, $l = 1, \ldots, m_j$ for $j < k$, $l = 1, \ldots, m_j - 1$ for $j = k$, satisfying

$$R(z) - \frac{P(\lambda_k)}{(z-\lambda_k)^{m_k}\, S(\lambda_k)} = \sum_{j=1}^{k-1} \sum_{l=1}^{m_j} \frac{a_{jl}}{(z-\lambda_j)^l} + \sum_{l=1}^{m_k-1} \frac{a_{kl}}{(z-\lambda_k)^l}, \tag{G.9}$$

thereby completing the induction for the existence proof.

It remains to prove the uniqueness of the coefficients a_{jl} in (G.5). Thus, suppose one has $b_{jl} \in \mathbb{C}$, $j = 1, \ldots, k$, $l = 1, \ldots, m_j$, such that

$$\sum_{j=1}^{k} \sum_{l=1}^{m_j} \frac{a_{jl}}{(z-\lambda_j)^l} = \sum_{j=1}^{k} \sum_{l=1}^{m_j} \frac{b_{jl}}{(z-\lambda_j)^l}. \tag{G.10}$$

We fix j_0 and prove $a_{j_0 l} = b_{j_0 l}$ via induction on $l = 1, \ldots, m_{j_0}$: Let $l \in \{1, \ldots, m_{j_0}\}$ and assume $a_{j_0 \alpha} = b_{j_0 \alpha}$ has already been shown for each $\alpha > l$ (the induction does, indeed, start at $l = m_{j_0}$, working itself down to $l = 1$). Then (G.10) implies

$$\sum_{\beta=1}^{l} \frac{a_{j_0\beta}}{(z-\lambda_{j_0})^\beta} + \sum_{\substack{j=1,\\ j\neq j_0}}^{k} \sum_{\beta=1}^{m_j} \frac{a_{j\beta}}{(z-\lambda_j)^\beta} = \sum_{\beta=1}^{l} \frac{b_{j_0\beta}}{(z-\lambda_{j_0})^\beta} + \sum_{\substack{j=1,\\ j\neq j_0}}^{k} \sum_{\beta=1}^{m_j} \frac{b_{j\beta}}{(z-\lambda_j)^\beta}. \tag{G.11}$$

One now multiplies (G.11) by $(z-\lambda_{j_0})^l$. Then taking the limit for $z \to \lambda_{j_0}$ on both sides yields $a_{j_0 l} = b_{j_0 l}$ as desired. ■

If, in (G.1), $P, Q : \mathbb{R} \longrightarrow \mathbb{R}$, then the partial fraction decomposition (G.5) of Th. G.1 is not quite satisfactory, since, even though P and Q are both real, the a_{jl} will typically be nonreal elements of $\mathbb{C}$. As the following Th. G.2 shows, if P, Q are real, then it is always possible to obain a partial fraction decomposition with only real coefficients, however its form is somewhat more complicated.

We start by using the real factorization of $Q : \mathbb{R} \longrightarrow \mathbb{R}$, $\deg(Q) = n \in \mathbb{N}$, according to (8.58), where, as in (G.2), we combine identical factors, obtaining

$$Q(x) = c \prod_{j=1}^{k_1} (x-\lambda_j)^{m_j} \prod_{j=1}^{k_2} (x^2 + \alpha_j x + \beta_j)^{n_j}, \tag{G.12}$$

where $c \in \mathbb{R}$; $\lambda_1, \ldots, \lambda_{k_1} \in \mathbb{R}$, $k_1 \in \{0, \ldots, n\}$, are the *distinct* real zeros of Q (if any), $m_j \in \mathbb{N}$ being their respective multiplicities; and $(\alpha_1, \beta_1), \ldots, (\alpha_{k_2}, \beta_{k_2}) \in \mathbb{R}^2$

$k_2 \in \{0, \dots, n\}$, are *distinct* pairs of real numbers, each pair arising from combining two conjugate nonreal zeros of Q according to (8.60), $n_j \in \mathbb{N}$ being their respective multiplicities;

$$\sum_{j=1}^{k_1} m_j + 2\sum_{j=1}^{k_2} n_j = n. \tag{G.13}$$

Theorem G.2. *Let $P, Q : \mathbb{R} \longrightarrow \mathbb{R}$ be polynomials such that $\deg(P) < \deg(Q) =: n$. Moreover, let $\mathcal{N}(Q)$ denote the set of zeros of Q, and assume Q to have the form of* (G.12)*. Then there exist families of coefficients*

$$a_{jl} \in \mathbb{R} \quad (j = 1, \dots, k_1,\ l = 1, \dots, m_j), \tag{G.14a}$$
$$b_{jl} \in \mathbb{R} \quad (j = 1, \dots, k_2,\ l = 1, \dots, n_j), \tag{G.14b}$$
$$c_{jl} \in \mathbb{R} \quad (j = 1, \dots, k_2,\ l = 1, \dots, n_j) \tag{G.14c}$$

such that

$$\underset{x \in \mathbb{R} \setminus \mathcal{N}(Q)}{\forall} \quad R(x) = \frac{P(x)}{Q(x)} = \sum_{j=1}^{k_1} \sum_{l=1}^{m_j} \frac{a_{jl}}{(x - \lambda_j)^l} + \sum_{j=1}^{k_2} \sum_{l=1}^{n_j} \frac{b_{jl}\, x + c_{jl}}{(x^2 + \alpha_j x + \beta_j)^l}. \tag{G.15}$$

Proof. We show that, if P, Q are real, where Q is as in (G.12), then (G.5) can be rewritten in the form (G.15): First, consider $\lambda_{j_0} \in \mathbb{R}$ to be a real zero of Q. Then all corresponding coefficients in (G.5) are real: We prove $a_{j_0 l} \in \mathbb{R}$ via induction on $l = 1, \dots, m_{j_0}$: Let $l \in \{1, \dots, m_{j_0}\}$ and assume $a_{j_0 \alpha} \in \mathbb{R}$ has already been shown for each $\alpha > l$. Then (G.10) (with z replaced by x) implies

$$\underset{x \in \mathbb{R} \setminus \mathcal{N}(Q)}{\forall} \quad S(x) := R(x) - \sum_{\beta=l+1}^{m_{j_0}} \frac{a_{j_0 \beta}}{(x - \lambda_{j_0})^\beta} = \sum_{\beta=1}^{l} \frac{a_{j_0 \beta}}{(x - \lambda_{j_0})^\beta} + \sum_{\substack{j=1,\\ j \neq j_0}}^{k} \sum_{\beta=1}^{m_j} \frac{a_{j\beta}}{(x - \lambda_j)^\beta} \in \mathbb{R}. \tag{G.16}$$

One now multiplies (G.16) by $(x - \lambda_{j_0})^l$. Then taking the limit for $x \to \lambda_{j_0}$ on both sides yields $a_{j_0 l} \in \mathbb{R}$ as desired (the limit on the right-hand side is clearly $a_{j_0 l}$ and all values and, thus, the limit on the left-hand side are clearly in $\mathbb{R}$).

Thus, the summands corresponding to real zeros of Q are identical in (G.5) and (G.15). It remains to show that terms in (G.5), corresponding to conjugate nonreal zeros of Q, can be combined to result in the summands involving the b_{jl} and c_{jl} in (G.5). To this end, consider $\lambda_{j_0}, \lambda_{j_1} \in \mathbb{C}$ to be conjugate nonreal zeros of Q, $\lambda_{j_1} = \overline{\lambda}_{j_0}$. Then all corresponding coefficients in (G.5) are conjugate: We prove $a_{j_0 l} = \overline{a}_{j_1 l}$ via induction on $l = 1, \dots, m_{j_0} = m_{j_1}$: Let $l \in \{1, \dots, m_{j_0}\}$ and assume $a_{j_0 \alpha} = \overline{a}_{j_1 \alpha}$ has already been shown for each $\alpha > l$. We once again have the formula (G.16) for $S(x)$ (even for each $x \in \mathbb{C} \setminus \mathcal{N}(Q)$, but we can no longer expect $S(x) \in \mathbb{R}$). As before, after multiplying (G.16) by $(x - \lambda_{j_0})^l$, we obtain

$$\lim_{x \to \lambda_{j_0}} \left(S(x)(x - \lambda_{j_0})^l \right) = a_{j_0 l}. \tag{G.17}$$

Analogously, we also have

$$\forall_{x\in\mathbb{C}\setminus\mathcal{N}(Q)} \quad R(x) - \sum_{\beta=l+1}^{m_{j_1}} \frac{a_{j_1\beta}}{(x-\lambda_{j_1})^\beta} = \sum_{\beta=1}^{l} \frac{a_{j_1\beta}}{(x-\lambda_{j_1})^\beta} + \sum_{\substack{j=1,\\ j\neq j_1}}^{k} \sum_{\beta=1}^{m_j} \frac{a_{j\beta}}{(x-\lambda_j)^\beta}. \tag{G.18}$$

Taking complex conjugates in (G.16) and using the induction hypothesis as well as $\overline{R(x)} = R(\overline{x})$ (since the coefficients of P, Q are real) yields

$$\forall_{x\in\mathbb{C}\setminus\mathcal{N}(Q)} \quad \overline{S(x)} = R(\overline{x}) - \sum_{\beta=l+1}^{m_{j_1}} \frac{a_{j_1\beta}}{(\overline{x}-\lambda_{j_1})^\beta} \overset{(G.18)}{=} \sum_{\beta=1}^{l} \frac{a_{j_1\beta}}{(\overline{x}-\lambda_{j_1})^\beta} + \sum_{\substack{j=1,\\ j\neq j_1}}^{k} \sum_{\beta=1}^{m_j} \frac{a_{j\beta}}{(\overline{x}-\lambda_j)^\beta}. \tag{G.19}$$

If we multiply (G.19) by $(\overline{x}-\lambda_{j_1})^l$, we obtain

$$\lim_{\overline{x}\to\lambda_{j_1}} \left(\overline{S(x)}(\overline{x}-\lambda_{j_1})^l\right) = a_{j_1 l}. \tag{G.20}$$

Thus,

$$\overline{a}_{j_0 l} \overset{(G.17)}{=} \overline{\lim_{x\to\lambda_{j_0}} \left(S(x)(x-\lambda_{j_0})^l\right)} = \lim_{x\to\lambda_{j_0}} \left(\overline{S(x)}(\overline{x}-\lambda_{j_1})^l\right) \overset{(G.20)}{=} a_{j_1 l}, \tag{G.21}$$

as needed.

We now combine two corresponding summands of (G.5) (for $x \in \mathbb{R}\setminus\mathcal{N}(Q)$):

$$\sigma_l := \frac{a_{j_0 l}}{(x-\lambda_{j_0})^l} + \frac{\overline{a}_{j_0 l}}{(x-\overline{\lambda}_{j_0})^l} = \frac{a_{j_0 l}(x-\overline{\lambda}_{j_0})^l + \overline{a}_{j_0 l}(x-\lambda_{j_0})^l}{(x^2 - 2x\operatorname{Re}\lambda_{j_0} + |\lambda_{j_0}|^2)^l} = \frac{a(x-\overline{\lambda})^l + \overline{a}(x-\lambda)^l}{(x^2+bx+c)^l}, \tag{G.22}$$

where we have set

$$a := a_{j_0 l}, \quad \lambda := \lambda_{j_0}, \quad b := -2\operatorname{Re}\lambda_{j_0}, \quad c := |\lambda_{j_0}|^2, \tag{G.23}$$

to simplify notation. To finish the proof of (G.15), it remains to show there are real coefficients $s_{1l}, \ldots, s_{ll}$ and $t_{1l}, \ldots, t_{ll}$ such that

$$\forall_{l\in\{1,\ldots,m_{j_0}\}} \quad \sigma_l = \sum_{\beta=1}^{l} \frac{s_{\beta l}x + t_{\beta l}}{(x^2+bx+c)^\beta}, \tag{G.24}$$

which we prove via induction on l: For $l = 1$, we have

$$\sigma_1 = \frac{a(x-\overline{\lambda}) + \overline{a}(x-\lambda)}{x^2+bx+c} = \frac{(a+\overline{a})x - (\lambda+\overline{\lambda})}{x^2+bx+c}, \tag{G.25}$$

which fits the requirements of (G.24). For the induction step, we consider, for $l = 1, \ldots, m_{j_0} - 1$,

$$\sigma_{l+1} = \frac{a(x-\overline{\lambda})^{l+1} + \overline{a}(x-\lambda)^{l+1}}{(x^2+bx+c)^{l+1}}. \tag{G.26}$$

The numerator can be rewritten as

$$\begin{aligned} a(x-\overline{\lambda})^{l+1} + \overline{a}(x-\lambda)^{l+1} = a(x-\overline{\lambda})^{l+1} + \overline{a}(x-\lambda)^{l}(x-\overline{\lambda}) \\ - \overline{a}(x-\lambda)^{l}(x-\overline{\lambda}) - a(x-\overline{\lambda})^{l}(x-\lambda) \\ + a(x-\overline{\lambda})^{l}(x-\lambda) + \overline{a}(x-\lambda)^{l+1}. \end{aligned} \tag{G.27}$$

Thus, $\sigma_{l+1} = S_1 + S_2 + S_3$, where

$$S_1 = \frac{\big(a(x-\overline{\lambda})^{l} + \overline{a}(x-\lambda)^{l}\big)(x-\overline{\lambda})}{(x^2+bx+c)^{l+1}} = \frac{\sigma_l(x-\overline{\lambda})}{x^2+bx+c}, \tag{G.28}$$

$$S_2 = -\frac{\big(\overline{a}(x-\lambda)^{l-1} + a(x-\overline{\lambda})^{l-1}\big)(x-\lambda)(x-\overline{\lambda})}{(x^2+bx+c)^{l+1}} = \frac{\sigma_{l-1}}{x^2+bx+c}, \tag{G.29}$$

$$S_3 = \frac{\big(a(x-\overline{\lambda})^{l} + \overline{a}(x-\lambda)^{l}\big)(x-\lambda)}{(x^2+bx+c)^{l+1}} = \frac{\sigma_l(x-\lambda)}{x^2+bx+c}, \tag{G.30}$$

where, for (G.29) to hold for $l = 1$, we set $\sigma_0 := a + \overline{a} \in \mathbb{R}$. Using the induction hypothesis, S_2 clearly has the form required by (G.24). Using the induction hypothesis together with the elementary equality

$$\frac{sx^2}{x^2+bx+c} = s - \frac{sbx+sc}{x^2+bx+c}, \tag{G.31}$$

we also obtain

$$\begin{aligned} S_1 + S_3 &= \frac{\sigma_l\,(2x+b)}{x^2+bx+c} = \frac{2x+b}{x^2+bx+c}\sum_{\beta=1}^{l}\frac{s_{\beta l}x + t_{\beta l}}{(x^2+bx+c)^{\beta}} \\ &= \sum_{\beta=1}^{l}\frac{2s_{\beta l}x^2 + (bs_{\beta l}+2t_{\beta l})x + bt_{\beta l}}{(x^2+bx+c)^{\beta}(x^2+bx+c)} \\ &\overset{(G.31)}{=} \sum_{\beta=1}^{l}\frac{1}{(x^2+bx+c)^{\beta}}\left(2s_{\beta l} - \frac{2s_{\beta l}(bx+c)}{x^2+bx+c}\right) \\ &\quad + \sum_{\beta=1}^{l}\frac{(bs_{\beta l}+2t_{\beta l})x + bt_{\beta l}}{(x^2+bx+c)^{\beta+1}} \end{aligned} \tag{G.32}$$

to have the form required by (G.24), thereby finishing the induction and the proof of the theorem. ■

Remark G.3. Given a rational function $R = P/Q$ as in Th. G.1 (or Th. G.2), there remains the question of how to actually compute the coefficients a_{jl} of the partial fraction decomposition (G.5) (or a_{jl}, b_{jl}, c_{jl} of the partial fraction decomposition (G.15) in the real case)? First, one always needs to obtain the zeros λ_j and their respective multiplicities m_j, which, for $\deg(Q)$ large, can be very difficult. Then there are basically three different possibilities to proceed, where, in practise, the most efficient way in a concrete situation might be to mix the three strategies:

(a) Linear System: To determine k unknown coefficients, one can plug k different values for z into (G.5) (or for x into (G.15)) to obtain a linear system for the unknown coefficients.

(b) One can multiply (G.5) (or (G.15)) by Q, obtaining a polynomial on both sides of the equation. As the polynomials need to be equal, the coefficients of equal powers need to be equal on both sides, yielding a system of equations for the unknown coefficients.

(c) Multiplying (G.5) (or (G.15)) by $(z-\lambda_j)^{m_j}$ and setting $z=\lambda_j$ yields the coefficient of $\frac{1}{(z-\lambda_j)^{m_j}}$, etc.

H Irrationality of e and π

H.1 Irrationality of e

The following Prop. H.1, which will then be used to prove the irrationality of e in Th. H.2, shows, in particular, that the series (8.26) can be used to efficiently compute accurate approximations of e.

Proposition H.1. *Defining*

$$\underset{n\in\mathbb{N}}{\forall}\quad\underset{z\in\mathbb{C}}{\forall}\quad R_n(z) := e^z - \sum_{j=0}^{n-1}\frac{z^j}{j!}, \tag{H.1}$$

we have

$$\underset{n\in\mathbb{N}}{\forall}\quad\left(|z|\le 1\quad\Rightarrow\quad\left|R_n(z)\right|\le\frac{2\,|z|^n}{n!}\right), \tag{H.2}$$

i.e. the error made when approximating e^z by the partial sum (for $|z|\le 1$) is at most as large as twice the modulus of the first missing summand.

Proof. One estimates, for each $n\in\mathbb{N}$ and each $z\in\mathbb{C}$ with $|z|\le 1$,

$$\begin{aligned}\left|R_n(z)\right| &\overset{(8.24),(7.81)}{\le} \sum_{j=n}^{\infty}\frac{|z|^j}{j!}\overset{(7.73)}{=}\frac{|z|^n}{n!}\left(1+\frac{|z|}{n+1}+\frac{|z|^2}{(n+1)(n+2)}+\dots\right)\\ &\overset{|z|\le 1}{\le} \frac{|z|^n}{n!}\left(1+\frac{1}{2}+\frac{1}{2^2}+\dots\right)\overset{(7.71)}{=}\frac{2\,|z|^n}{n!},\end{aligned} \tag{H.3}$$

which establishes the case. ∎

Theorem H.2. *Euler's number e is irrational.*

Proof. Seeking a contradiction, we assume e to be rational. Then there exist $m, n \in \mathbb{N}$ with $n \geq 2$ such that $e = \frac{m}{n}$. Then $n!e \in \mathbb{N}$ and, thus,

$$n!\, R_{n+1}(1) \overset{(\mathrm{H}.1)}{=} n!\, e - n! \sum_{j=0}^{n} \frac{1}{j!} \in \mathbb{Z}, \tag{H.4}$$

in contradiction to $0 < |R_{n+1}(1)| < \frac{2}{n+1} < 1$, which holds according to (H.2) (recalling $n \geq 2$). ■

H.2 Irrationality of π

Theorem H.3. *π^2 is irrational (then, in particular, π must be irrational as well).*

Proof. Seeking a contradiction, we assume π^2 to be rational. Then

$$\underset{a,b\in\mathbb{N}}{\exists} \quad \pi^2 = \frac{a}{b}. \tag{H.5}$$

We can then choose some even $n \in \mathbb{N}$ satisfying

$$0 < \frac{\pi\, a^n}{n!} < 1. \tag{H.6}$$

We now consider the function

$$f : \mathbb{R} \longrightarrow \mathbb{R}, \quad f(x) := \frac{x^n\,(1-x)^n}{n!} \overset{(*)}{=} \frac{1}{n!} \sum_{k=n}^{2n} (-1)^k \binom{n}{k-n} x^k, \tag{H.7}$$

where the equality at $(*)$ is proved by

$$\frac{x^n\,(1-x)^n}{n!} \overset{(5.22)}{=} \frac{x^n}{n!} \sum_{k=0}^{n} (-1)^k \binom{n}{k} x^k \overset{n \text{ even}}{=} \frac{1}{n!} \sum_{k=n}^{2n} (-1)^k \binom{n}{k-n} x^k. \tag{H.8}$$

Thus, for the polynomial f, we obtain the derivatives

$$f^{(j)}(0) = \begin{cases} 0 & \text{for } 0 \leq j < n, \\ \frac{j!}{n!}(-1)^j \binom{n}{j-n} & \text{for } n \leq j \leq 2n, \\ 0 & \text{for } 2n < j. \end{cases} \tag{H.9}$$

In consequence, since, for $n \leq j \leq 2n$, $\frac{j!}{n!} \in \mathbb{N}$ and $\binom{n}{j-n} \in \mathbb{N}$,

$$\underset{j\in\mathbb{N}_0}{\forall} \quad f^{(j)}(0) \in \mathbb{Z}. \tag{H.10}$$

Moreover, since $f(1-x) = f(x)$ for each $x \in \mathbb{R}$, and, thus, $f^{(j)}(1-x) = (-1)^j f^{(j)}(x)$ for each $x \in \mathbb{R}$, we also have

$$\underset{j\in\mathbb{N}_0}{\forall} \quad f^{(j)}(1) \in \mathbb{Z}. \tag{H.11}$$

Next, we consider another polynomial, namely

$$g : \mathbb{R} \longrightarrow \mathbb{R}, \quad g(x) := b^n \sum_{k=0}^{n} (-1)^k \pi^{2(n-k)} f^{(2k)}(x). \tag{H.12}$$

Due to (H.5), (H.10), (H.11), and (H.12), we have

$$\forall_{j \in \mathbb{N}_0} \quad \Big(g(0) \in \mathbb{Z} \ \wedge \ g(1) \in \mathbb{Z} \Big). \tag{H.13}$$

For each $x \in \mathbb{R}$, one calculates

$$\begin{aligned} g''(x) + \pi^2 \, g(x) &= b^n \sum_{k=0}^{n} (-1)^k \pi^{2(n-k)} f^{(2(k+1))}(x) + b^n \sum_{k=0}^{n} (-1)^k \pi^{2(n-(k-1))} f^{(2k)}(x) \\ &= b^n \sum_{k=1}^{n+1} (-1)^{k-1} \pi^{2(n-(k-1))} f^{(2k)}(x) + b^n \sum_{k=0}^{n} (-1)^k \pi^{2(n-(k-1))} f^{(2k)}(x) \\ &= b^n \, (-1)^n \, f^{(2n+2)}(x) + b^n \, \pi^{2n+2} \, f(x) = b^n \, \pi^{2n+2} \, f(x), \end{aligned} \tag{H.14}$$

and, thus, for

$$h : \mathbb{R} \longrightarrow \mathbb{R}, \quad h(x) := g'(x) \sin(\pi x) - \pi g(x) \cos(\pi x), \tag{H.15}$$

one obtains, for each $x \in \mathbb{R}$,

$$\begin{aligned} h'(x) &= g''(x) \sin(\pi x) + \pi g'(x) \cos(\pi x) - \pi g'(x) \cos(\pi x) + \pi^2 g(x) \sin(\pi x) \\ &= \big(g''(x) + \pi^2 g(x) \big) \, \sin(\pi x) \overset{(H.14)}{=} b^n \, \pi^{2n+2} \, f(x) \, \sin(\pi x) \\ &\overset{(H.5)}{=} \pi^2 \, a^n \, f(x) \, \sin(\pi x), \end{aligned} \tag{H.16}$$

implying the function h is the antiderivative of the function $x \mapsto \pi^2 \, a^n \, f(x) \, \sin(\pi x)$. This, together with the fundamental theorem of calculus in the form Th. 10.20(b) implies

$$I := \frac{\pi^2 \, a^n}{\pi} \int_0^1 f(x) \, \sin(\pi x) \, \mathrm{d}x \, = \frac{h(1) - h(0)}{\pi} = \frac{\pi g(1) + \pi g(0)}{\pi} = g(1) + g(0) \in \mathbb{Z}. \tag{H.17}$$

On the other hand, the definition of f in (H.7) yields

$$\forall_{0<x<1} \quad 0 < f(x) < \frac{1}{n!}, \tag{H.18}$$

and, thus, by (10.30) (i.e. by the monotonicity of the integral),

$$0 < I < \frac{\pi \, a^n}{n!} \overset{(H.6)}{<} 1. \tag{H.19}$$

The contradiction between (H.19) and (H.17) establishes the case. ■

I Trigonometric Functions

I.1 Additional Trigonometric Formulas

Proposition I.1. *We have the following identities:*

$$\forall_{z\in\mathbb{C}} \quad \sin(2z) = 2\sin z\cos z, \tag{I.1a}$$

$$\forall_{z\in\mathbb{C}} \quad \cos(2z) = (\cos z)^2 - (\sin z)^2, \tag{I.1b}$$

$$\forall_{z\in\mathbb{C}} \quad \frac{1-\cos z}{2} = \Big(\sin\frac{z}{2}\Big)^2, \tag{I.1c}$$

$$\forall_{z\in\mathbb{C}\setminus\{(2k+1)\pi:\, k\in\mathbb{Z}\}} \quad \tan\frac{z}{2} = \frac{\sin z}{\cos z+1}. \tag{I.1d}$$

$$\forall_{z\in\mathbb{C}\setminus\{(2k+1)\pi:\, k\in\mathbb{Z}\}} \quad \cos z = \frac{1-(\tan\frac{z}{2})^2}{1+(\tan\frac{z}{2})^2}. \tag{I.1e}$$

Proof. (I.1a) is immediate from (8.44c), (I.1b) is immediate from (8.44d).

(I.1c): For each $z\in\mathbb{C}$, one computes

$$\frac{1-\cos z}{2} \overset{\text{(I.1b)}}{=} \frac{1-(\cos\frac{z}{2})^2+(\sin\frac{z}{2})^2}{2} \overset{\text{(8.44e)}}{=} \frac{2\,(\sin\frac{z}{2})^2}{2} = \Big(\sin\frac{z}{2}\Big)^2, \tag{I.2}$$

thereby establishing the case.

(I.1d): Note that, according to (8.47d), it is

$$\cos\frac{z}{2} = 0 \quad\Leftrightarrow\quad \exists_{k\in\mathbb{Z}} \quad z = (2k+1)\,\pi. \tag{I.3}$$

Thus, for each $z\in\mathbb{C}\setminus\{(2k+1)\pi:\, k\in\mathbb{Z}\}$, one computes

$$\tan\frac{z}{2} = \frac{2\,\sin\frac{z}{2}\cos\frac{z}{2}}{2\,(\cos\frac{z}{2})^2} \overset{\text{(I.1a),(8.44e)}}{=} \frac{\sin z}{(\cos\frac{z}{2})^2-(\sin\frac{z}{2})^2+1} = \frac{\sin z}{\cos z+1}, \tag{I.4}$$

thereby establishing the case.

(I.1e): Once again, using (I.3), one computes for each $z\in\mathbb{C}\setminus\{(2k+1)\pi:\, k\in\mathbb{Z}\}$:

$$\cos z \overset{\text{(I.1b),(8.44e)}}{=} \frac{(\cos\frac{z}{2})^2-(\sin\frac{z}{2})^2}{(\cos\frac{z}{2})^2+(\sin\frac{z}{2})^2} = \frac{1-(\tan\frac{z}{2})^2}{1+(\tan\frac{z}{2})^2}, \tag{I.5}$$

as claimed. ∎

J Differential Calculus

J.1 Continuous, But Nowhere Differentiable Functions

The following Ex. J.1 provides functions from $f:\mathbb{R}\longrightarrow\mathbb{R}$ that are continuous, but nowhere differentiable.

Example J.1. We start by defining the *triangle wave* function

$$g : \mathbb{R} \longrightarrow \mathbb{R}, \quad g(x) := \begin{cases} x - k & \text{for } k \leq x \leq k + \frac{1}{2},\ k \in \mathbb{Z}, \\ -x + k + 1 & \text{for } k + \frac{1}{2} \leq x \leq k + 1,\ k \in \mathbb{Z}. \end{cases} \tag{J.1}$$

Then g is well-defined and continuous, since, clearly, g is piecewise affine, $k + \frac{1}{2} - k = \frac{1}{2} = -k - \frac{1}{2} + k + 1$, and $-(k+1) + k + 1 = 0 = k + 1 - (k+1)$. Moreover, for each $k \in \mathbb{Z}$, g is clearly strictly increasing on $[k, k+\frac{1}{2}]$ and clearly decreasing on $[k+\frac{1}{2}, k+1]$, implying

$$\forall_{x \in \mathbb{R}} \quad 0 \leq g(x) \leq \frac{1}{2}. \tag{J.2}$$

Clearly, (J.1) implies g to be periodic with period 1, i.e.

$$\forall_{x \in \mathbb{R}} \quad g(x+1) = g(x). \tag{J.3}$$

Now fix $q \in \mathbb{R}$, $a \in \mathbb{N}$ such that

$$0 < q < 1 \quad \wedge \quad a \geq 4 \quad \wedge \quad aq > 2 \tag{J.4}$$

(clearly, $q = \frac{1}{2}$ and $a = 5$ satisfy (J.4), and there are (uncountably) many other admissible choices for a and q). We now claim that

$$f : \mathbb{R} \longrightarrow \mathbb{R}, \quad f(x) := \sum_{n=0}^{\infty} q^n \, g(a^n x), \tag{J.5}$$

is continuous and nowhere differentiable. We first note that, as $\sum_{n=0}^{\infty} q^n$ converges and

$$\forall_{n \in \mathbb{N}} \quad \forall_{x \in \mathbb{R}} \quad 0 \overset{\text{(J.2)}}{\leq} q^n \, g(a^n x) \overset{\text{(J.2)}}{\leq} \frac{q^n}{2} < q^n, \tag{J.6}$$

Cor. 8.7(b) implies the series in (J.5) to converge uniformly. Then, since each function $f_n : \mathbb{R} \longrightarrow \mathbb{R}$, $f_n(x) := q^n g(a^n x)$, is continuous, f must be continuous by Cor. 8.7(c).

In preparation for showing f to be nowhere differentiable, we have to further investigate the properties of g. We proceed by showing g to be Lipschitz continuous with Lipschitz constant 1, i.e.

$$\forall_{x,y \in \mathbb{R}} \quad |g(x) - g(y)| \leq |x - y| : \tag{J.7}$$

If $|x - y| \geq \frac{1}{2}$, then, using (J.2),

$$|g(x) - g(y)| \leq \frac{1}{2} \leq |x - y|.$$

If $|x - y| < \frac{1}{2}$, then we distinguish four cases, where, without loss of generality, we let x denote the smaller of the two points and y the larger, i.e. $x \leq y$. Case (i): There is $k \in \mathbb{Z}$ such that $k \leq x, y \leq k + \frac{1}{2}$. Then

$$|g(x) - g(y)| = |x - k - y + k| = |x - y|.$$

Case (ii): There is $k \in \mathbb{Z}$ such that $k + \frac{1}{2} \leq x, y \leq k + 1$. Then

$$|g(x) - g(y)| = | - x + k + 1 + y - k - 1| = |x - y|.$$

Case (iii): There is $k \in \mathbb{Z}$ such that $k \leq x \leq k + \frac{1}{2}$ and $k + \frac{1}{2} \leq y \leq k + 1$. Then $x - k - \frac{1}{2} \leq 0$ and $y - k - \frac{1}{2} \geq 0$, implying

$$\begin{aligned} |g(x) - g(y)| = |x - k + y - k - 1| &= \left|x - k - \frac{1}{2} + \frac{1}{2} + y - k - 1\right| \\ &\leq \left|x - k - \frac{1}{2} - y + k + \frac{1}{2}\right| = |x - y|. \end{aligned}$$

Case (iv): There is $k \in \mathbb{Z}$ such that $k - \frac{1}{2} \leq x \leq k$ and $k \leq y \leq k + \frac{1}{2}$. Then $-x + k \geq 0$ and $-y + k \leq 0$, implying

$$|g(x) - g(y)| = | - x + k - y + k| \leq | - x + k + y - k| = |x - y|,$$

finishing the proof of (J.7).

For each $c \in \mathbb{R}$, we also consider the following modified versions of g:

$$g_c : \mathbb{R} \longrightarrow \mathbb{R}, \quad g_c(x) := g(cx) = \begin{cases} cx - k & \text{for } k \leq cx \leq k + \frac{1}{2},\ k \in \mathbb{Z}, \\ -cx + k + 1 & \text{for } k + \frac{1}{2} \leq cx \leq k + 1,\ k \in \mathbb{Z}. \end{cases} \tag{J.8}$$

Then, for $c \neq 0$, g_c is periodic with period c^{-1}:

$$\forall_{x \in \mathbb{R}} \quad g_c(x + c^{-1}) = g(cx + 1) \overset{(J.3)}{=} g(cx) = g_c(x). \tag{J.9}$$

Moreover, g_c is Lipschitz continuous with Lipschitz constant $|c|$:

$$\forall_{x,y \in \mathbb{R}} \quad |g_c(x) - g_c(y)| = |g(cx) - g(cy)| \overset{(J.7)}{\leq} |cx - cy| = |c|\,|x - y|. \tag{J.10}$$

To show that f is nowhere differentiable, we will now study suitable difference quotients. Let $(h_k)_{k \in \mathbb{N}}$ be a sequence such that

$$\forall_{k \in \mathbb{N}} \quad h_k = \pm \frac{1}{a^{k+1}}. \tag{J.11}$$

Then (J.4) implies $\lim_{k \to \infty} h_k = 0$. Let $x \in \mathbb{R}$ be arbitrary. Define

$$\forall_{k,n \in \mathbb{N}} \quad \delta_k g(a^n x) := \frac{g(a^n(x + h_k)) - g(a^n x)}{h_k}. \tag{J.12}$$

Then

$$\forall_{k,n \in \mathbb{N}} \quad |\delta_k g(a^n x)| \overset{(J.10)}{\leq} \frac{a^n\,|h_k|}{|h_k|} = a^n, \tag{J.13}$$

and, recalling $a \in \mathbb{N}$,

$$\underset{n>k\in\mathbb{N}}{\forall} \quad \delta_k g(a^n x) = \frac{g_{a^n}(x \pm \frac{1}{a^{k+1}}) - g_{a^n}(x)}{h_k} = \frac{g_{a^n}(x \pm \frac{a^{n-(k+1)}}{a^n}) - g_{a^n}(x)}{h_k} \overset{(J.9)}{=} 0. \quad (J.14)$$

Thus, for each $k \in \mathbb{N}$, we obtain

$$\begin{aligned} \delta_k f \quad &:= \quad \frac{f(x+h_k) - f(x)}{h_k} = \frac{\sum_{n=0}^{\infty} q^n \left(g(a^n(x+h_k)) - g(a^n x)\right)}{h_k} \\ &\overset{(J.14)}{=} \quad \sum_{n=0}^{k-1} q^n \, \delta_k g(a^n x) + q^k \, \delta_k g(a^k x) \end{aligned} \quad (J.15)$$

and estimate, recalling $aq > 2$,

$$\left|\sum_{n=0}^{k-1} q^n \, \delta_k g(a^n x)\right| \leq \sum_{n=0}^{k-1} q^n \, |\delta_k g(a^n x)| \overset{(J.13)}{\leq} \sum_{n=0}^{k-1} q^n \, a^n = \frac{1 - q^k a^k}{1 - qa} < \frac{q^k a^k}{qa - 1}. \quad (J.16)$$

We rewrite (J.16) as

$$\left|\sum_{n=0}^{k-1} q^n \, \delta_k g(a^n x)\right| \leq \eta \, q^k a^k, \quad \text{where} \quad \eta := \frac{1}{aq-1}, \quad 0 < \eta < 1. \quad (J.17)$$

According to (J.8), g_{a^k} is affine on intervals of length $\frac{1}{2a^k}$. Thus, since $a \geq 4$ implies

$$\underset{k\in\mathbb{N}}{\forall} \quad \frac{1}{a^{k+1}} \leq \frac{1}{4a^k}, \quad (J.18)$$

we can always choose the sign of h_k such that

$$|\delta_k g(a^k x)| = \frac{|a^k(x + h_k - x)|}{|h_k|} = a^k. \quad (J.19)$$

Using this choice for the h_k, we combine our estimates to obtain

$$\underset{k\in\mathbb{N}}{\forall} \quad |\delta_k f| \overset{(J.15)}{=} \left|\sum_{n=0}^{k-1} q^n \, \delta_k g(a^n x) + q^k \, \delta_k g(a^k x)\right| \overset{(J.17),(J.19)}{\geq} (1 - \eta) \, q^k a^k. \quad (J.20)$$

Thus, as $qa > 2$, $\lim_{k\to\infty} |\delta_k f| = \infty$, proving that f is not differentiable in x. As $x \in \mathbb{R}$ was arbitrary, f is nowhere differentiable.

References

[Bla84] A. Blass. *Existence of Bases Implies the Axiom of Choice*. Contemporary Mathematics **31** (1984), 31–33.

[EFT07] H.-D. Ebbinghaus, J. Flum, and W. Thomas. *Einführung in die mathematische Logik*, 5th ed. Spektrum Akademischer Verlag, Heidelberg, 2007 (German).

[EHH+95] H.-D. Ebbinghaus, H. Hermes, F. Hirzebruch, M. Koecher, K. Mainzer, J. Neukirch, A. Prestel, and R. Remmert. *Numbers*. Graduate Texts in Mathematics, Vol. 123, Springer-Verlag, New York, 1995, corrected 3rd printing.

[Jec73] T. Jech. *The Axiom of Choice*. North-Holland, Amsterdam, 1973.

[Kun80] Kenneth Kunen. *Set Theory*. Studies in Logic and the Foundations of Mathematics, Vol. 102, North-Holland, Amsterdam, 1980.

[Kun12] Kenneth Kunen. *The Foundations of Mathematics*. Studies in Logic, Vol. 19, College Publications, London, 2012.

[Lan65] Edmund Landau. *Grundlagen der Analysis*, 4th ed. American Mathematical Society, New York, 1965.

[Str08] Gernot Stroth. *Lineare Algebra*, 2nd ed. Berliner Studienreihe zur Mathematik, Vol. 7, Heldermann Verlag, Lemgo, Germany, 2008 (German).

[Wik15a] Wikipedia. *Coq — Wikipedia, The Free Encyclopedia*. 2015, `https://en.wikipedia.org/wiki/Coq` Online; accessed Sep-01-2015.

[Wik15b] Wikipedia. *HOL Light — Wikipedia, The Free Encyclopedia*. 2015, `https://en.wikipedia.org/wiki/HOL_Light` Online; accessed Sep-01-2015.

[Wik15c] Wikipedia. *Isabelle (proof assistant) — Wikipedia, The Free Encyclopedia*. 2015, `https://en.wikipedia.org/wiki/Isabelle_(proof_assistant)` Online; accessed Sep-01-2015.