Matematický ústav Slezské univerzity v Opavě

Logic and set theory

M. Marvan

LOGIC AND SET THEORY

This text is intended for students attending the lecture *Logic and Set Theory.* The lecture is designed as an introduction to foundations of modern mathematics. In accordance with the author's preferences, it deals with the Kelley–Morse theory with the axiom of choice and the axiom of regularity.

1. Introduction

Set theory and mathematical logic are two interrelated mathematical disciplines that lie at the foundations of mathematics.

Mathematical logic resulted from efforts to formalise the process of mathematical reasoning. It provides a language to formulate mathematical propositions and their proofs. It deals with general axiomatic theories, consistency, provability, theorem-proving methods, etc. It is the tool to establish truth in mathematics.

Logic has been the foundations of mathematics since its ancient beginnings, but established itself as a separate domain only in the 19th century. One of the most important contributions of that time was *Begriffslehre* (Conceptology, 1879) by Gottlob Frege (1848–1925), which was devoted to recognising and describing all logical principles of mathematical proofs, with the ultimate goal of making arithmetic a part of logic. Frege is also credited for the introduction of the quantifiers $\forall$, $\exists$.

Set theory is the science of sets—finite or infinite collections of arbitrary objects called elements. In its original formulation it was created by the German mathematician Georg Cantor (1845–1918). An important predecessor to Cantor was the Prague scholar Bernard Bolzano (1781–1848), the author of *Paradoxien des Unendlichen* (Paradoxes of the Infinite, 1851). The modern set theory, together with logic, makes it possible to build all other branches of mathematics essentially "from nothing."

1.1. Naive set theory

In the paper *Über eine Eigenschaft des Inbegriffes aller reellen algebraischen Zahlen* (*On one property of the set of all real algebraic numbers*, 1874), Cantor proves the existence of infinitely many transcendent numbers (in his time only two transcendent numbers were known, e and π). His paper can be considered the founding publication on set theory. However, many scholars considered it controversial, some even refused to consider it mathematics[1] and only few encountered Cantor's ideas with enthusiasm[2].

Cantor uses natural language to define all notions. His work contains no precise definition of a set. Nevertheless, Cantor elaborated the theory to a considerable depth; focusing on the notion of cardinality (massiveness, see below) of a set. Today, this informally constructed theory is called the *naive set theory.*

However, a paradox emerged once Cantor took the set $\mathfrak{M}$ of all possible cardinalities into consideration. As it turns out, the set $\mathfrak{M}$ would have an even greater cardinality, one not

[1] Leopold Kronecker even called Cantor a "corrupter of youth".

[2] David Hilbert, who described Cantor's work as "the finest product of mathematical genius and one of the supreme achievements of purely intellectual human activity," once claimed: "No one shall drive us from the paradise that Cantor created for us."

contained in $\mathfrak{M}$. Two years earlier, in 1887, Cesare Burali-Forti (1861–1931) discovered a similar paradox in Cantor's theory of ordinal numbers.

A significant step towards uncovering the nature of the problem was made by Frege. In the two-volume work *Grundgesetze der Arithmetik* (1893 and 1903), he introduced the set

$$M_\phi = \{X \mid \phi(X)\}$$

of all X, satisfying condition $\phi(X)$. By coincidence, just before the publication of Frege's second volume, Bertrand Russell (1872–1970) pointed out a paradox arising in the special case

$$N = \{X \mid X \notin X\}.$$

The *Russel's paradox* states that both options $N \in N$ and $N \notin N$ lead to a contradiction. Indeed, if $N \in N$, then $N \notin N$ by the definition of N. Conversely, if $N \notin N$, then $N \in N$ for the same reason.

Since naive attempts to build a set theory did not lead to success, it was eventually built as an axiomatic theory. The first such theory was proposed in the beginning of the 20th century by Ernst Zermelo (1871–1953). Later, Abraham Fraenkel (1891–1965) made several important corrections, which resulted in what is called the Zermelo–Fraenkel theory. The latter represents the standard theory of sets today.

1.2. Logical paradoxes

Examples of logical reasoning leading to paradoxical conclusions have been known since ancient times.

Example. Epimenides' paradox. Epimenides, himself a Cretan, said that the Cretans lie under all circumstances. The paradox arises when we try to determine whether Epimenides' statement is true. In a modern form, the paradox is contained in the statement: "This statement is false."

It is easy to spot a logical circle in Epimenides' paradox because the sentence speaks of itself.

Example. Barber's paradox. An army barber was ordered to shave all men who don't shave themselves. Should he shave himself? This is how Russell popularised his paradox mentioned above.

Example. Grelling–Nelson's paradox. Let's classify adjectives into heterological and homological. Let an adjective be *homological* (or self-descriptive), if it has the property it describes (for example, the adjectives *English, existing, polysyllabic, pronounceable, unhyphenated* are homological) and *heterological* (or non-self-descriptive) in the opposite case. The paradox becomes apparent if we try to classify the adjective *heterological.*

Example. Berry's paradox. Only a finite number of numbers can be determined in less than thirteen words. Therefore, there are numbers that cannot be determined in less than thirteen words. However, every nonempty subset of the set of natural numbers has a least element. Therefore, there is "the least number that cannot be determined in less than thirteen words." This is paradoxical because the number in question is determined by a sentence of twelve words.

The following example shows that a logical paradox can have profound consequences in the real world, too.

Example. An arbitrator's paradox. Two companies entered into a contract in which, among other things, they empowered an arbitrator to resolve all their possible disputes, instead of going to court. One of the companies then produced irrefutable evidence that the contract has been void ab initio (from the outset). If the arbitrator finds that the contract is void, he will lose the right to arbitrate the dispute. If, on the other hand, he accepts the validity of the contract and decides the dispute, then he cannot accept the validity of the contract in view of the evidence presented.

Logical paradoxes arise because some statements apply more broadly than originally intended and anticipated, leading to some form of logical circle. In linguistic paradoxes, we also observe mixing the language we speak about with the language we speak.

2. Axiomatic theory of sets and classes

Within naive set theory, Russel's paradox arises when we set

$$N = \{X \mid X \notin X\}.$$

In general, we can ask whether the collection of all sets X possessing some property $\phi(X)$,

$$M_\phi = \{X \mid \phi(X)\},$$

is correctly defined, and if so, whether M_ϕ is a set. The answer is given by axiomatic set theory.

The first ever Zermelo–Fraenkel theory (**ZF** for short) blocks Russell's paradox by allowing only the notation $N = \{X \in U \mid X \notin X\}$, where U is some (already known) set. In the framework of **ZF**, N *is not* a set, whereas the question of what N *is* remains unanswered. Also the collection of all sets is not a set in **ZF**. Neither is the collection of all groups, all topological spaces and similar structures. However, even such collections can be subject to mathematical considerations (this happens, in particular, in category theory). It is therefore desirable to give even these "non-set" collections a place in the constructed theory.

2.1. Classes

By itself, the collection $\{X \mid \phi(X)\}$ of all sets X with the property $\phi(X)$ does not lead to any paradoxes. Only, as we have already seen, it cannot always be a set. A collection like this one is called a *class*, and not all classes are sets. For instance, the class of all sets, the class of all groups and the class of all topological spaces are not sets.

If we axiomatise only sets, there is *terra incognita* beyond the territory occupied by sets, if we axiomatise classes, there will be something we can work with. Classes are axiomatised, e.g., in the Gödel–Bernays–von Neumann and the Kelley–Morse theories. The difference between these two is that the latter is slightly more daring (as specified below). Kelley published his theory of classes in an appendix to *General Topology* in 1957. With insignificant deviations, we shall construct just the Kelley–Morse theory in what follows.

The purpose of our exposition is to have the reader understand "what one can do with classes" and "what one can do with sets."

2.2. Formulas

Formulas allow us to write "meaningful" statements about sets (and classes), i.e., statements which, in principle, can be true or false.
The *language of class theory* uses an alphabet consisting of symbols specified as follows:

1. Symbol $\in$, read as "is an element of" or "belongs to."

2. Logical operators $\urcorner, \wedge, \vee, \Rightarrow, \Leftrightarrow$.
3. Quantifiers $\forall, \exists$.
4. Brackets (,).
5. Symbols of variables and constants, which are upper and lower case letters of the Latin alphabet, digits 0 to 9, and possibly other symbols and combinations thereof, excluding the symbols listed in items 1–4.

For instance, x, X, y_1, U_2 can be symbols used to denote classes (including sets). Superfluous brackets can be omitted. We shall gradually add more symbols, e.g., $=$, $\notin$, $\cap$, $\cup$, $\{$, $\}$, $\emptyset$, 0, 1, 2, 3, $\aleph$.

Constants, contrary to variables, have a unique, conventional value. For instance, $\emptyset$ (empty set) will be a constant.

Next, we shall inductively define class theory *formulas.* All formulas will be non-empty finite sequences of symbols of the class theory alphabet. Not all sequences will be formulas, because not all sequences make sense in class theory. For example, the sequence $x \in y$ makes sense (and can be either true or false) and will be a formula, while the sequence $\exists \in\in \forall$ will be not.

We shall also define which variables occurring in a formula are *free* and which are *bound.* Each variable will be either free or bound, not both.

The formulas themselves will be denoted by letters of the Greek alphabet. They will be defined inductively, using four rules $\mathbf{F}_1$ to $\mathbf{F}_4$, listed below. A formula will be any sequence of symbols that results from a finite number of steps, each consisting in the application of any one of the rules $\mathbf{F}_1$ through $\mathbf{F}_4$.

$\mathbf{F}_1$ Each sequence $X \in Y$, where X, Y are variables or constants, is a formula. Each of X, Y is free, if it is a variable.

A formula introduced by this rule is called an *atomic formula.* The name reflects the fact that neither part $X, \in$ or Y is a formula. There will be no other atomic formulas.

$\mathbf{F}_2$ If ϕ is a formula, then $\urcorner(\phi)$ is a formula. The bound and free variables of $\urcorner(\phi)$ are the same as those of ϕ.

Example. Formula $\urcorner(X \in Y)$ will be abbreviated as $X \notin Y$, which introduces the symbol $\notin$. If variables, X, Y remain free. Let us stress that $X \notin Y$ is *not* an atomic formula.

$\mathbf{F}_3$ If ϕ is a formula and x is its free variable, then $(\forall x)(\phi)$ and $(\exists x)(\phi)$ are formulas. The variable x is bound in both formulas $(\forall x)(\phi)$ and $(\exists x)(\phi)$; the other variables are free or bound depending on whether they were free or bound in the formula ϕ.

Formulas $(\forall x)$ and $(\exists x)$ are read "for every x" and "there exists x," respectively. The variable x ceases to be a free variable and becames a bound variable. Rule $\mathbf{F}_3$ is called *quantification.* It *quantifies* the variable x.

Every bound variable can be renamed, similarly to summation indices or integration variables in mathematical analysis.

Example. An example of a formula generated by rule $\mathbf{F}_3$ is

$$(\exists x)\,(x \in X).$$

Rule $\mathbf{F}_3$ has been applied to the atomic formula $x \in X$, where both variables x, X were free. In the resulting formula $(\exists x)\,(x \in X)$, the variable x is bound, while X remains free. The meaning of the formula $(\exists x)\,(x \in X)$ is that X is nonempty. Renaming x to y, we get the formula $(\exists y)\,(y \in X)$ of the same meaning.

2.1. Remarks. 1. Parentheses may be omitted if clarity is not compromised. For example, rule $\mathbf{F}_3$ requires that we write $(\exists x)((\exists X)(x \in X))$, but so many brackets do not add to clarity. The notation $(\exists x)(\exists X)(x \in X)$ is also clear enough, or even $\exists x \, \exists X \; x \in X$ with no parentheses at all. In this text we will try to keep a reasonable amount of parentheses to achieve good readability even without introducing explicit rules for reading.

2. It is acceptable to bind variables that do not appear in the formula at all, but such quantification is unnecessary.

Next we introduce the notion of *compatibility.* For compatible formulas, every variable can be either bound or free, but not both.

2.2. Definition. Formulas ϕ, ψ are said to be *compatible*, if each variable that occurs in both formulas is either bound or free in both formulas.

$\mathbf{F}_4$ If ϕ, ψ are compatible formulas, then

$$(\phi) \wedge (\psi), \quad (\phi) \vee (\psi), \quad (\phi) \Rightarrow (\psi), \quad (\phi) \Leftrightarrow (\psi)$$

are formulas; they're called *compound* formulas. The variable is free or bound in the compound formula if it is free or bound in at least one of the formulas ϕ, ψ.

Example. The formulas $\phi = (\exists x) \, x \in y$ and $\psi = x \in z$ are incompatible because x is bound in the former and free in the latter. The remaining variables y, z occur in only one of the formulas, and therefore have no effect on compatibility. Consequently,

$$\phi \wedge \psi = ((\exists x) \, x \in y) \wedge (x \in z)$$

is *not* a formula.

2.3. Proposition. *If a variable X occurs in a formula ϕ, then it is either free or bound, but not both.*

Proof. The statement holds for atomic formulas, since they contain only free variables. If the statement is true for a formula ϕ, it is also true for $\neg\phi$, because the free and bound variables are the same for both. If the statement is true for a formula ϕ with a free variable X, it is also true for both $(\forall x)(\phi)$ and $(\exists x)(\phi)$, because the free and bound variables are the same for both, except for X, which has changed from free to bound. If the statement holds for compatible formulas ϕ, ψ, there is no variable that is free in one and bound in the other, and therefore the statement holds for the compound formulas $\phi \wedge \psi$, $\phi \vee \psi$, $\phi \Rightarrow \psi$, $\phi \Leftrightarrow \psi$.

If two formulas are not compatible, then any common variable that is bound in only one of the formulas can be renamed, i.e., denoted with a different symbol so that the new symbol does not appear in the other formula. In this way, compatibility can always be enforced. We will show later that such renaming does not affect the truth of the formula.

Example. Consider again two incompatible formulas $\phi = (\exists x) \, x \in y$ and $\psi = x \in z$. If we replace x with u in the formula ϕ to obtain $\phi = (\exists u) \, u \in y$, the two formulas are compatible and

$$\phi \wedge \psi = ((\exists u) \, u \in y) \wedge (x \in z)$$

is a formula.

For clarity, sometimes a formula symbol is followed by the list of all its free variables enclosed in parentheses.

Example. The formula $(\exists x)\, x \in y$ can be denoted by ϕ or $\phi(y)$, but neither $\phi(x, y)$ nor $\phi(y, z)$.

2.3. True and false formulas

Let us turn to the question of the truthfulness of formulas. As in the classical logic, two truth values are possible: 1 (the formula is true, the formula holds) and 0 (the formula is false, the formula does not hold).

We will decide the truth of formulas based on the truth of their subformulas using rules $\mathbf{P}_2$ to $\mathbf{P}_4$, corresponding to the rules $\mathbf{F}_2$ to $\mathbf{F}_4$ for constructing formulas of class theory. The rules $\mathbf{P}_n$ are as follows.

$\mathbf{P}_2$ The formula $\neg\phi$ is true if and only if ϕ is false.

$\mathbf{P}_3$ The formula $(\forall x)\, \phi(x)$ is true if and only if $\phi(x)$ holds for all x; the formula $(\exists x)\, \phi(x)$ is true if and only if there exists x such that $\phi(x)$ holds.

$\mathbf{P}_4$ The truth of the formulas $\phi \wedge \psi$, $\phi \vee \psi$, $\phi \Rightarrow \psi$, $\phi \Leftrightarrow \psi$ depends on the truth of the formulas ϕ, ψ according to the well-known rules, given in the following table:

ϕ	ψ	$\phi \wedge \psi$	$\phi \vee \psi$	$\phi \Rightarrow \psi$	$\phi \Leftrightarrow \psi$
1	1	1	1	1	1
1	0	0	1	0	0
0	1	0	1	1	0
0	0	0	0	1	1

There is no rule $\mathbf{P}_1$ that would allow to assess the truth of *all* atomic formulas. If there was one, we could decide the truth of all possible formulas, and there would be no need to build an axiomatic theory. Actually, the truth of atomic formulas has to be derived from the axioms.

A formula without free variables has a uniquely determined truth value 1 or 0. A formula containing free variables has no determined truth value, until the free variables are assigned values (concrete classes). The truth value of the formula then depends on the values assigned to the free variables.

2.4. Logical equivalence

Note that $\phi \Leftrightarrow \psi$ is true exactly when ϕ and ψ have have the same truth values. The formulas ϕ and ψ are then interchangeable in terms of truth and falsity. We say that the formulas ϕ, ψ are *logically equivalent.*

We shall not introduce any symbol to denote logical equivalence (to avoid confusing the language we speak about with the language we speak). In particular, '$\Leftrightarrow$' alone does not denote logical equivalence. To express logical equivalence of ϕ and ψ, we have to say '$\phi \Leftrightarrow \psi$ is true' or '$\phi \Leftrightarrow \psi$ holds.'

2.5. Inference and truth

We have now gathered the basic requisites for building an axiomatic theory of classes. Next we shall systematically introduce one axiom after another, simultaneously introducing important auxiliary concepts. If we were to list all the axioms at once, without the auxiliary notions, the axioms would be very complicated and completely incomprehensible.

All axioms are considered to be true formulas. Other true formulas (valid propositions) can be derived from axioms by using the rules $\mathbf{P}_2$ to $\mathbf{P}_4$ and standard proving methods, which are assumed to be well known.

Example. We have already introduced the symbol $\notin$ in a preceding example. The formula $X \notin Y$ has the same meaning as $\urcorner(X \in Y)$. Whatever we substitute for X, Y, the formulas $X \notin Y$ and $\urcorner(X \in Y)$ will always have the same truth value. Consequently, $X \notin Y$ and $\urcorner(X \in Y)$ are logically equivalent.

Let us emphasise that $X \notin Y$ is not an atomic formula, although, at first glance, it appears to be. It is just a short notation for the non-atomic formula $\urcorner(X \in Y)$. The truth value of the formula $X \notin Y$ is clearly derivable from the truth value of the atomic formula $X \in Y$. We introduce other similar short notations below.

2.6. Sets

In Kelly–Morse axiomatic class theory, sets can and will be *defined*.

2.4. Definition. A class X is a *set* if $(\exists Y)(X \in Y)$ holds. A class that is not a set is said to be a *proper class*.

Otherwise said, a class is a set if and only if it is an element of another class. Proper classes cannot be elements anywhere.

2.7. Equality of classes

Let's introduce one more notion and one more symbol. We say that two classes X, Y are equal and write $X = Y$ if they have the same elements. A formal definition follows.

2.5. Definition. Classes X, Y are *equal*, if

$$(\forall U)\,(U \in X \Leftrightarrow U \in Y).$$

We write $X = Y$.

In the preceding text we have already used the symbol '=' several times to denote general equality or identity of formulas or classes. For classes, we have now introduced equality as the condition of having the same elements, which does not conflict with the earlier usage. For formulas we informally define equality as identity. Obviously, identical formulas are logically equivalent, but not vice versa.

Let us note that if U is a proper class, then neither $U \in X$ nor $U \in Y$ holds, so the equivalence $U \in X \Leftrightarrow U \in Y$ is true for all proper classes U. Hence, it is sufficient to consider the equivalence $U \in X \Leftrightarrow U \in Y$ only assuming U to be a set.

Actually, it will be seen below that "=" is a relation between classes that is reflexive, symmetric and transitive.

Exercise. Show that

$$(\forall X)\, X = X,$$
$$(\forall X)(\forall Y)\,(X = Y \Rightarrow Y = X),$$
$$(\forall X)(\forall Y)(\forall Z)\,((X = Y \wedge Y = Z) \Rightarrow X = Z).$$

Help: Let us prove the first statement. This is equivalent to

$$(\forall X)(\forall U)\,(U \in X) \Leftrightarrow (U \in X).$$

Since $(U \in X) \Leftrightarrow (U \in X)$ holds for any truth value of the atomic formula $U \in X$, the first statement is proved.

By the definition of equality of classes, $X = Y$ implies $U \in X \Leftrightarrow U \in Y$ for each U. However, it does not follow from anywhere that $X = Y$ also implies $X \in V \Leftrightarrow Y \in V$ for every V. This must be postulated, and of course, it is sufficient to do so for one implication.

A. Axiom of invariance.

$$(\forall X)(\forall Y)(\forall V)\,((X = Y \wedge X \in V) \Rightarrow Y \in V).$$

Otherwise said, if $X = Y$, then $X \in V$ implies $Y \in V$.

2.6. Proposition. *Let $X = Y$, let V be an arbitrary class. Then*

$$X \in V \Leftrightarrow Y \in V$$

holds. Otherwise said, $X \in V$ a $Y \in V$ are true or not simultaneously.

Proof. By the axiom of invariance, $X \in V$ implies $Y \in V$ and $Y \in V$ implies $X \in V$.

2.7. Proposition. *Assume $X = Y$ to be two equal classes. Assume $U_1, \dots, U_n$ to be arbitrary classes. Then*

$$\phi(X, U_1, \dots, U_n) \Leftrightarrow \phi(Y, U_1, \dots, U_n)$$

holds, i.e., $\phi(X, U_1, \dots, U_n)$ a $\phi(Y, U_1, \dots, U_n)$ are true or not simultaneously.

We prove the assertion by induction on the length of the formula. The length of a formula is defined to be the number of symbols in a formula constructed according to the rules $\mathbf{F}_1$ to $\mathbf{F}_4$ including all parentheses. Obviously, atomic formulas are the shortest (each consisting of three symbols) and each rule $\mathbf{F}_2$ to $\mathbf{F}_4$ produces formulas longer than the components thereof. A formal proof follows.

Proof. 1. If $\phi(X, U)$ is the atomic formula $U \in X$, the statement follows from the definition of equality of classes.

2. If $\phi(X, U)$ is the atomic formula $X \in U$, the statement follows from Proposition 2.6.

For all other formulas, one proceeds by induction. Assume that the statement holds for all formulas shorter than $\phi(X, U_1, \dots, U_n)$.

3. If the formula $\phi(X, U_1, \dots, U_n)$ is of the form specified in rule $\mathbf{F}_2$, i.e., $\neg\psi(X, U_1, \dots, U_n)$, then $\psi(X, U_1, \dots, U_n) \Leftrightarrow \psi(Y, U_1, \dots, U_n)$ holds by induction, since $\psi(X, U_1, \dots, U_n)$ is shorter than $\phi(X, U_1, \dots, U_n)$. But then $\phi(X, U_1, \dots, U_n) \Leftrightarrow \phi(Y, U_1, \dots, U_n)$ holds as well.

4. If $\phi(X, U_1, \dots, U_n)$ is of the form specified in rule $\mathbf{F}_3$, that is, $(\forall U)\,\psi(U, X, U_1, \dots, U_n)$ or $(\exists U)\,\psi(U, X, U_1, \dots, U_n)$, then $\psi(U, X, U_1, \dots, U_n) \Leftrightarrow \psi(U, Y, U_1, \dots, U_n)$ holds by induction, since $\psi(U, X, U_1, \dots, U_n)$ is shorter than $\phi(X, U_1, \dots, U_n)$. Then, however, $\phi(X, U_1, \dots, U_n) \Leftrightarrow \phi(Y, U_1, \dots, U_n)$ holds as well.

5. If the formula $\phi(X, U_1, \dots, U_n)$ is of the form specified in rule $\mathbf{F}_4$, then we proceed analogously. Complete the details as an exercise.

$\mathbf{B}_\phi$. Specification axioms. Let $\phi(x, U_1, \dots, U_n)$ be a formula containing free variables x, $U_1, \dots, U_n$, let Z be a variable that does not appear in ϕ. Then

$$(\forall U_1) \cdots (\forall U_n)(\exists Z)(\forall x)\,(x \in Z \Leftrightarrow (x \text{ is a set} \wedge \phi(x, U_1, \dots, U_n)))$$

is an axiom. Here 'x is a set ' replaces $(\exists Y)(x \in Y)$ (according to Definition 2.4).

Otherwise said, for every formula ϕ there exists a class Z, the elements of which are precisely all sets x such that $\phi(x, U_1, \dots, U_n)$ holds.

Axioms $\mathbf{B}_\phi$ form what is called a *scheme of axioms*, meaning one axiom for every formula ϕ.

The class Z is called the *class specified by the formula* $\phi(x, U_1, \ldots, U_n)$ and is denoted

$$\{x \mid \phi(x, U_1, \ldots, U_n)\}.$$

From this immediately follows a rule for deciding the truth of atomic formulas with the right-hand side $Z = \{x \mid \phi(x, U_1, \ldots, U_n)\}$.

P$_1$ Let $Z = \{x \mid \phi(x, U_1, \ldots, U_n)\}$ be the class specified by $\phi(x, U_1, \ldots, U_n)$. Then, for each set x,

$$x \in Z \Leftrightarrow \phi(x, U_1, \ldots, U_n).$$

holds. Otherwise said, if x is a set, then $x \in Z$ is true if and only if $\phi(x, U_1, \ldots, U_n)$ is true.

The requirement that x is a set is important because it prevents Russell's paradox to appear.

In the introduction we mentioned two variants of class theory, Gödel–Bernays–von Neumann and Kelley–Morse. The former admits only sets and the latter admits general classes as variables admitted in formulas ϕ of the *specification scheme*. Accordingly, there are more Kelley–Morse classes than there are Gödel–Bernays–von Neumann classes.

2.8. The universal class

We will now introduce the class $\mathcal{U}$ by

$$\mathcal{U} = \{x \mid x = x\}.$$

Since $x = x$ is always true (obviously, we could use any universally valid formula), $\mathcal{U}$ is the class of all sets. It is also called the *universal class* or the *universe*. Thus, $x \in \mathcal{U}$ means just that x is a set.

Later we shall show that $\mathcal{U}$ is a proper class.

2.9. The empty class

The empty class $\emptyset$ is introduced by

$$\emptyset = \{x \in \mathcal{U} \mid x \neq x\}.$$

Since $x \not x$ does not hold for any set, there is no set that lies in the empty class, which explains its name.

We could also write $\emptyset = \{x \mid x \neq x\}$, which would mean the same by definition. The long form $\{x \in \mathcal{U} \mid x \neq x\}$ is a reminder that x needs to be a set. We shall use the long form henceforth.

2.10. The algebra of classes

The specification scheme makes it possible to introduce operations with classes that are a direct generalisation of the well-known set operations.

$$U \cap V = \{x \in \mathcal{U} \mid x \in U \wedge x \in V\},$$

$$U \cup V = \{x \in \mathcal{U} \mid x \in U \vee x \in V\},$$

$$U \setminus V = \{x \in \mathcal{U} \mid x \in U \wedge \neg(x \in V)\}.$$

We also have the unary operation of the class complement

$$\widetilde{U} = \{x \in \mathcal{U} \mid \neg(x \in U)\},$$

where there is no set analogy because the complement of a set is always a proper class.

Exercise. Show that

$$\begin{aligned} &U \cap \mathcal{U} = U, && U \cup \emptyset = U,\\ &U \cap \emptyset = \emptyset, && U \cup \mathcal{U} = \mathcal{U},\\ &U \cap U = U, && U \cup U = U,\\ &U \cap V = V \cap U, && U \cup V = V \cup U,\\ &(U \cap V) \cap W = U \cap (V \cap W), && (U \cup V) \cup W = U \cup (V \cup W),\\ &(U \cap V) \cup W = (U \cup W) \cap (V \cup W), && (U \cup V) \cap W = (U \cap W) \cup (V \cap W). \end{aligned}$$

Exercise. Show that

$$\begin{aligned} &\widetilde{\mathcal{U}} = \emptyset, && \tilde{\emptyset} = \mathcal{U},\\ &(U \cap V)^{\sim} = \widetilde{U} \cup \widetilde{V}, && (U \cup V)^{\sim} = \widetilde{U} \cap \widetilde{V},\\ &\widetilde{\widetilde{U}} = U. \end{aligned}$$

Inclusion also has a direct generalisation.

2.8. Definition. We say that class U is *subclass* of class V and write $U \subseteq V$, if the formula

$$(\forall x \in \mathcal{U})\,(x \in U \Rightarrow x \in V)$$

holds.

The relation $A \subseteq B$ is called *inclusion.* The *sharp inclusion* $A \subset B$ is introduced by the formula

$$A \subset B \Leftrightarrow (A \subseteq B) \wedge (A \neq B).$$

Thus, $A \subset B$ holds if and only if $A \subseteq B$ and $A \neq B$ hold.

To show that U is a subclass in V, it is sufficient to show that an arbitrary element of U is also an element of V.

2.9. Proposition. *Every class is a subclass in* $\mathcal{U}$.

Proof. Every set $x \in U$ is also an element of $\mathcal{U}$, since all sets are elements of $\mathcal{U}$.

The following exercise shows that $\subseteq$ has the properties of an ordering.

Exercise. Let U, V, W be classes. Show that

$$\begin{aligned} &U \subseteq U && \text{(reflexivity)},\\ &(U \subseteq V \wedge V \subseteq U) \Rightarrow U = V && \text{(antisymetry)},\\ &(U \subseteq V \wedge V \subseteq W) \Rightarrow U \subseteq W && \text{(transitivity)}. \end{aligned}$$

Exercise. Let U, V be classes. Show that

$$U \subseteq V \Leftrightarrow U \cup V = V, \qquad U \subseteq V \Leftrightarrow U \cap V = U.$$

If U is a class, we introduce the *restricted quantifiers* $(\exists X \in U)$ and $(\forall X \in U)$ by the requirement that

$$(\exists X \in U)\, \phi(X, V_1, \dots, V_n) \Leftrightarrow (\exists X)\, X \in U \wedge \phi(X, V_1, \dots, V_n),$$
$$(\forall X \in U)\, \phi(X, V_1, \dots, V_n) \Leftrightarrow (\forall X)\, X \in U \Rightarrow \phi(X, V_1, \dots, V_n)$$

holds for all formulas $\phi(X, V_1, \dots, V_n)$.

If U is a class, we introduce its *union* $\bigcup U$ and its *intersection* $\bigcap U$ by

$$\begin{aligned}\bigcup U &= \{x \in \mathcal{U} \mid (\exists X \in U)\, x \in X\}\\ &= \{x \in \mathcal{U} \mid (\exists X)\, (X \in U \wedge x \in X)\},\\ \bigcap U &= \{x \in \mathcal{U} \mid (\forall X \in U)\, x \in X\}\\ &= \{x \in \mathcal{U} \mid (\forall X)\, (X \in U \Rightarrow x \in X)\}.\end{aligned}$$

We see that $\bigcap U$ contains exactly those x that lie in all sets $X \in U$, while $\bigcup U$ contains exactly those x that lie in at least one set $X \in U$. Moreover, if for every set $X \in U$ there is a set $F(X) \in V$ (anticipating introduction of a mapping $F : U \to V$ few pages later), we can define

$$\bigcap_{X \in U} F(X) = \{Y \in V \mid (\forall X \in U) Y \in F(X)\}$$
$$\bigcup_{X \in U} F(X) = \{Y \in V \mid (\exists X \in U) Y \in F(X)\}.$$

For example,

$$\bigcup U = \bigcup_{X \in U} X, \qquad \bigcap U = \bigcap_{X \in U} X.$$

2.10. Proposition. *We have*

$$\bigcup \emptyset = \emptyset, \qquad \bigcap \emptyset = \mathcal{U}.$$

Proof. Concerning the first equality, the inclusion $\emptyset \subseteq \bigcup \emptyset$ is obvious. Let us prove the opposite inclusion $\bigcup \emptyset \subseteq \emptyset$. Let $x \in \bigcup \emptyset$ be arbitrary. Then, by the definition of $\bigcup$, there is a class X such that $X \in \emptyset \wedge x \in X$. But $X \in \emptyset$ never holds. This means that no x satisfies $x \in \bigcup \emptyset$, i.e., $\bigcup \emptyset \subseteq \emptyset$.

Concerning the second equality, the inclusion $\bigcap \emptyset \subseteq \mathcal{U}$ is obvious, since every class is a subclass in $\mathcal{U}$. To prove the opposite inclusion $\mathcal{U} \subseteq \bigcap \emptyset$, let $x \in \mathcal{U}$ be arbitrary. Then x is a set. Now,$x \in \bigcap \emptyset$ holds if and only if for every class X the implication $X \in \emptyset \Rightarrow x \in X$ holds. However, this implication does hold, since the assumption $X \in \emptyset$ is never true.

Exercise. Let $U \subseteq V$. Show that

$$\bigcup U \subseteq \bigcup V, \qquad \bigcap U \subseteq \bigcap V.$$

Exercise. Let $x \in X$. Show that

$$\bigcap X \subseteq x \subseteq \bigcup X.$$

2.11. The class of all subsets

The class $\wp U$ of all subsets of a class U is defined by

$$\wp U = \{Z \in \mathcal{U} \mid Z \subseteq U\}.$$

C. Two axioms about subsets.

$\mathbf{C}_1$ If X is a set and $Z \subseteq X$, then Z is a set.
$\mathbf{C}_2$ If X is a set, then $\wp X$ is a set.

It can be shown that axioms $\mathbf{C}_1$**,** $\mathbf{C}_2$ can be replaced by a single axiom, specified in the following theorem.

2.11. Proposition. *The pair of axioms* $\mathbf{C}_1$**,** $\mathbf{C}_2$ *is equivalent to*

$$\mathbf{C} \qquad (\forall X \in \mathcal{U})(\exists Y \in \mathcal{U})(\forall Z)\,(Z \subseteq X \Rightarrow Z \in Y).$$

Proof. Let axioms $\mathbf{C}_1$ and $\mathbf{C}_2$ hold. Let X be a set. Let $Y = \wp X$, which is a set according to $\mathbf{C}_2$. Each subclass of $Z \subset X$ is a set under $\mathbf{C}_1$. Therefore, $Z \in \wp X$. This proves formula $\mathbf{C}$.

Let $\mathbf{C}$ hold. Let X be a set. By $\mathbf{C}$, there exists a set Y such that the implication $Z \subseteq X \Rightarrow Z \in Y$ holds for every class Z. Let us prove $\mathbf{C}_1$. If $Z \subseteq X$, then $Z \in Y$, and hence Z is a set and $\mathbf{C}_1$ is proved. Let us prove $\mathbf{C}_2$. It is easy to see that $\wp X \subseteq Y$, whence $\wp X$ is a set by $\mathbf{C}_1$. because Y is a set. This proves $\mathbf{C}_2$.

2.12. Proposition. *The universe* $\mathcal{U}$ *is a proper class.*

Proof. Assume by contradiction that $\mathcal{U}$ is a set. By $\mathbf{C}_1$, the subclass $N = \{X \in \mathcal{U} \mid X \notin X\}$ is a set, which is a contradiction by Russel's argument. Thus, $\mathcal{U}$ is not a set.

2.12. The existence of sets

Note that we have not proved the existence of a set yet. All axioms mentioned so far allow for the possibility that no class is a set at all, i.e., the possibility that $\mathcal{U} = \emptyset$. In particular, even $\emptyset$ needs not be a set.

D. Axiom of existence of sets.

$$\emptyset \in \mathcal{U},$$

i.e., the empty class is a set.

Equivalently, it is sufficient to assume the existence of at least one set.

2.13. Proposition. *Axiom* $\mathbf{D}$ *and*

$$\mathcal{U} \neq \emptyset$$

are equivalent.

Proof. If $\mathcal{U}$ is non-empty, then there is at least one set. Denote it by U. Since $\emptyset \subseteq U$, $\emptyset$ is also a set according to $\mathbf{C}_1$.

2.14. Proposition.

$$\bigcap \mathcal{U} = \emptyset, \qquad \bigcup \mathcal{U} = \mathcal{U}.$$

Proof. Concerning the first equality, it is obvious that $\emptyset \subseteq \bigcap \mathcal{U}$. Let us prove the opposite inclusion $\bigcap \mathcal{U} \subseteq \emptyset$. Suppose there is an element $x \in \bigcap \mathcal{U}$. This implies that x belongs to all sets contained in $\mathcal{U}$, and therefore it also belongs to the empty set, which is a contradiction. Thus, no x belongs to $\bigcap \mathcal{U}$.

Concerning the second equality, it is obvious that $\bigcup \mathcal{U} \subseteq \mathcal{U}$. Let us prove the opposite inclusion $\mathcal{U} \subseteq \bigcup \mathcal{U}$. Let $x \in \mathcal{U}$ be an arbitrary set. Then there exists a class X such that $X \in \mathcal{U}$ and $x \in X$, for example, $X = \wp x$. Consequently, $x \in \bigcup \mathcal{U}$.

2.13. One-element set

If x is a set, define the class $\{x\}$ by

$$\{x\} = \{y \in \mathcal{U} \mid y = x\}.$$

Exercise. Decide whether
a) $\emptyset \in \{\emptyset\}$;
b) $\emptyset \subseteq \{\emptyset\}$.

2.15. Proposition. *We have*

$$(\forall x \in \mathcal{U})\, \{x\} \in \mathcal{U}.$$

Otherwise said, if x is a set, then $\{x\}$ is a set.

Proof. Let x be a set. Then $x \in \wp x$ because $x \subseteq x$. Therefore, $\{x\} \subseteq \wp x$, whence $\{x\}$ is a set by $\mathbf{C}_1$ and $\mathbf{C}_2$.

The set $\{x\}$ is called the *one-element set* with the element x.

Exercise. What is $\{X\}$ if X is a proper class?

We have an infinite sequence of sets $\emptyset$, $\{\emptyset\}$, $\{\{\emptyset\}\}$, $\{\{\{\emptyset\}\}\}$, $\{\{\{\{\emptyset\}\}\}\}$, etc.

Exercise. Prove by induction that any two are different.

Except $\emptyset$, so far we have been able to construct only one-element sets.

2.16. Definition. If x, y are different sets, $x \neq y$, we introduce the class

$$\{x, y\} = \{u \in \mathcal{U} \mid u = x \vee u = y\},$$

which is called the se *two-element set* containing elements x, y.

E. The axiom of pair.

$$(\forall x \in \mathcal{U})(\forall y \in \mathcal{U})(\exists u \in \mathcal{U})(\forall z \in \mathcal{U})\, (z \in u \Leftrightarrow (z = x \vee z = y))$$

If x, y are sets, then $\{x, y\}$ is a set.

F. The axiom of union.

$$(\forall S \in \mathcal{U}) \left(\bigcup S \in \mathcal{U}\right).$$

If S is a set, then $\bigcup S$ is a set.

By unifying a set of sets, we always get a set. This, of course, need not be true for union of a proper class of sets. The analogue for intersections (if S is a set, then $\bigcap S$ is a set) follows from axiom $\mathbf{C}_1$.

Exercise. Let X, Y be sets. Show that

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

We see that the union of two sets is a set (by the axiom of union).

Obviously, $\{x, y\} = \{x\} \cup \{y\}$. If x, y, z are sets, we can analogously introduce

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

And so on.

The reservoir of provably existing sets can be expanded again, for example by

$$\{\emptyset, \{\emptyset\}\},$$
$$\{\emptyset, \{\emptyset, \{\emptyset\}\}\},$$
$$\{\emptyset, \{\emptyset, \{\emptyset, \{\emptyset\}\}\}\},$$

etc. However, they are still sets with finitely many elements.

2.14. Cartesian product

The Cartesian product is another important set construction, that can be extend to classes without problems.

2.17. Definition. Let a, b be sets, then

$$[a, b] = \{\{a\}, \{a, b\}\}$$

is called the *ordered pair* $[a, b]$.

2.18. Proposition. *If a, b, c, d are sets, then $[a, b] = [c, d]$ if and only if $a = c \wedge b = d$.*

Proof. By analysing all possible cases, we easily establish that $\{\{a\}, \{a, b\}\} = \{\{c\}, \{c, d\}\}$ if and only if $a = c$ aand $b = d$.

The ordered pair $[a, b]$ is uniquely determined by specifying two elements and their order, unlike the set a, b, which is determined by the two elements regardless of order.

2.19. Definition. Given two classes A, B, we define the class

$$A \times B = \{x \in \mathcal{U} \mid (\exists a \in A)(\exists b \in B)\, x = [a, b]\}$$

of all ordered pairs $[a, b]$ of elements of classes $a \in A$, $b \in B$. We call it the *Cartesian product* of the classes A, B.

It is also common to write

$$A \times B = \{[a, b] \mid a \in A, b \in B\}.$$

Example. $\{\heartsuit, \diamondsuit\} \times \{\spadesuit, \clubsuit\} = \{[\heartsuit, \spadesuit], [\heartsuit, \clubsuit], [\diamondsuit, \spadesuit], [\diamondsuit, \clubsuit]\}$.

Exercise. 1. Let A, B be arbitrary classes, let $A' \subseteq A$, $B' \subseteq B$ be subclasses. Show that

$$A' \times B' = (A \times B') \cap (A' \times B).$$

2. Let A, B be arbitrary classes, let A', A'' be subclasses in A. Show that

$$(A' \cap A'') \times B = (A' \times B) \cap (A'' \times B),$$
$$(A' \cup A'') \times B = (A' \times B) \cup (A'' \times B)$$

(distributive laws). 3. Let U be a class. Show that

$$A \times \bigcup_{X \in U} X = \bigcup_{X \in U} (A \times X), \qquad A \times \bigcap_{X \in U} X = \bigcap_{X \in U} (A \times X)$$

(distributive laws for systems of sets).

2.20. Remark. We introduced the ordered pair $[a, b]$ as the set $\{\{a\}, \{a, b\}\}$ as a part of our effort to present all mathematics in terms of sets and classes. This has some unwanted consequences. In particular, ordered pairs can be used as arguments of common set operations, such as intersection and union. However, doing so is not always reasonable. Usually, the results of set operations with ordered pairs have no practical meaning.

Example. Here are some examples of "nonsensical" set operations mentioned in the last remark (a, b, c are pairwise different elements).

1. $[a, b] \cap [a, c] = \{\{a\}, \{a, b\}\} \cap \{\{a\}, \{a, c\}\} = \{a\}$.
2. $[a, c] \cap [b, c] = \{\{a\}, \{a, c\}\} \cap \{\{b\}, \{b, c\}\} = \emptyset$.
3. $\bigcap [a, b] = \bigcap \{\{a\}, \{a, b\}\} = \{a\} \cap \{a, b\} = \{a\}$.

Exercise. Let A, U be arbitrary classes. Show that, in general,

$$A \times \bigcap U \neq \bigcap (A \times U),$$
$$A \times \bigcup U \neq \bigcup (A \times U),$$

i.e., this is not the right way of writing the distributive law for systems of sets.

Later we shall prove that the Cartesian product of sets is a set, but we cannot do it without another axiom.

2.15. Relations and mappings

Relations and mappings between classes are defined in the same way as between ordinary sets. For completeness, we give all basic definitions and propositions below, leaving proofs as exercises.

2.21. Definition. Let A, B be arbitrary classes. A *relation* (correspondence) between classes A, B is a subclass of the Cartesian product $A \times B$. If $\rho \subseteq A \times B$ is a relation and if $a \in A$, $b \in B$ are sets such that $[a, b] \in \rho$, then a is said to be in relation ρ with b, which we write as $a \, \rho \, b$. Relation *on* a class A is the special case when $A = B$.

Examples. 1. The empty set $\emptyset \subseteq A \times B$ is a relation between classes A, B. No two elements $a \in A$, $b \in B$ are in this relation.

2. The whole class $A \times B$ is also a relation between classes A, B. Every two elements of $a \in A$, $b \in B$ are in this relation.

3. *Identical relation* on a class A is the subclass $\mathrm{id}_A = \{ [a, a] \mid a \in A \}$. Elements $a, b \in A$ are in this relation if and only if $a = b$.

2.22. Definition. If ρ is a relation between classes A, B, then the relation $\rho^{-1} \subseteq B \times A$ between classes B, A defined by

$$\rho^{-1} := \{ [b, a] \mid a \, \rho \, b \}$$

is said to be *inverse* to relation ρ.

Remember: $a \, \rho \, b \Leftrightarrow b \, \rho^{-1} \, a$.

2.23. Definition. Let ρ be a relation between classes A, B, let σ be a relation between classes B, C. The relation $\sigma \circ \rho$ (read „σ after ρ") between classes A, C defined by

$$\sigma \circ \rho = \{(a, c) \mid (\exists b \in B)(a \, \rho \, b \wedge b \, \sigma \, c)\},$$

is called the *composition* of relations ρ and σ.

Exercise. Let ρ be a relation between classes A, B. Then
1. $\rho \circ \mathrm{id}_A = \rho$;
2. $\mathrm{id}_B \circ \rho = \rho$.

Let, moreover, σ be a relation between classes B, C. Then
3. $(\sigma \circ \rho)^{-1} = \rho^{-1} \circ \sigma^{-1}$.

Exercise. 1. Show that $(\rho^{-1})^{-1} = \rho$.
2. Assuming $\rho \subseteq \rho'$, $\sigma \subseteq \sigma'$, show that $\rho \circ \sigma \subseteq \rho' \circ \sigma'$.

2.16. Mappings

A mapping is a special case of relation between classes A, B.

Definition. A relation f between classes A, B is called a *mapping* (or a *map*) from class A into class B and denoted $f \subseteq A \times B$, if for every element $a \in A$ there is exactly one element $b \in B$ such that $[a, b] \in f$ holds.

Intuitively, to each element of class A we assign exactly one "value" from class B. The element b is then usually denoted by $f(a)$, sometimes also by f_a, and called the *value* of mapping f at the element a or the *image* of the element a under the mapping f.

By writing $f : A \to B$ we express that f is a mapping from class A to class B. Another possible notation is $A \xrightarrow{f} B$. Instead of $b = f(a)$ we can also write $f : a \mapsto b$ or $a \overset{f}{\longmapsto} b$. The classes A and B are called the *domain* and the *codomain*, respectively.

Introducing a special quantifier $\exists!$ with the meaning "there is exactly one," the condition of f being a mapping can be written as

$$(\forall a \in A)(\exists! \, b \in B) \, [a, b] \in f.$$

"There is exactly one" means "there is at least one, and if there are two, then they are equal," or

$$(\exists! \, x \in X) \, \phi(x) \Leftrightarrow ((\exists x \in X) \, \phi(x)) \wedge ((\forall x \in X)(\forall x' \in X) \, (\phi(x) \wedge \phi(x')) \Rightarrow x = x').$$

Example. The identical relation on a class A is a mapping and is called *identical mapping* and is denoted by $\mathrm{id}_A : A \to A$. We have $\mathrm{id}_A(a) = a$ for every $a \in A$.

If $A \subseteq X$ is a subclass, then $\iota_A = \{(a, a) \mid a \in A\}$ is a mapping that to every element $a \in A$ assigns the same element $a \in X$: $\iota_{AX}(a) = a$. It is called the *inclusion mapping*.

Exercise. Let $f : AB$, $g : BC$ be two maps. Then the mapping $g \circ f : A \to C$ defined by

$$(\forall a \in A)\,(g \circ f)(a) = g(f(a)).$$

is called the *composition* of mappings f, g.

Exercise. (1) Let $f : A \to B$ ba a mapping. Then

$$f \circ \mathrm{id}_A = \mathrm{id}_B \circ f = f.$$

(2) Let $f : A \to B$, $g : B \to C$, $h : C \to D$ be mappings. Then

$$h \circ (g \circ f) = (h \circ g) \circ f.$$

As follows from (2), notation $h \circ g \circ f$ is unambiguous even with omitted parentheses.

If a relation $f \subseteq A \times B$ is a mapping $f : A \to B$, then it does not necessarily mean that the opposite relation f^{-1} is a mapping as well. If an element $b \in B$ is the image of an element $a \in A$, then the element a is called a *preimage* of the element b under the mapping f. The class of all preimages of element $b \in B$ under mapping f is denoted by $f^{-1}\{b\}$. That is,

$$f(a) = b \Leftrightarrow a \in f^{-1}\{b\}.$$

While the image of a general element $a \in A$ always exists and is unique, the preimage of an element $b \in B$ may not exist in general and may not be unique.

Given a subclass $B' \subseteq B$, we define

$$f^{-1}B' = \{a \in A \mid f(a) \in B'\}.$$

A special case is $f^{-1}\{b\}$ from the previous paragraph.

2.24. Remark. Let us emphasise that the preimage $f^{-1}\{b\}$ of an element b can be a proper class. Let us give an example. If we assign to each set an empty set, we get a constant mapping $\mathcal{U} \to \{\emptyset\}$ and the preimage of the element $\emptyset$ is the whole class $\mathcal{U}$.

However, if preimages are not sets, then they cannot form a class. This is an annoying limitation because, intuitively, the collection of all preimages exists and sometimes we need to work with it. This is a weakness of our axiomatisation.

Exercise. Show that for $f \subseteq A \times B$ and $B', B'' \subseteq B$ we have

$$f^{-1}(B' \cap B'') = f^{-1}B' \cap f^{-1}B'',$$
$$f^{-1}(B' \cup B'') = f^{-1}B' \cup f^{-1}B''.$$

2.25. Definition. A mapping $f : A \to B$ is called *surjective* (a *surjection*) or mapping *onto* the class B, if every element of $b \in B$ has at least one preimage in A:

$$(\forall b \in B)(\exists a \in A)\, b = f(a).$$

A mapping $f : A \to B$ is called *injective* (*injection*) or *one-to-one* if each element of $b \in B$ has at most one preimage in A:

$$(\forall a \in A)(\forall a' \in A)\,(f(a) = f(a') \Rightarrow a = a').$$

A mapping $f : A \to B$ is called *bijective* (*bijection*), if it is both surjective and injective.

Exercise. Let $f : A \to B$ be a mapping. Prove that the following conditions are equivalent:
(1) The mapping f is bijectvive.
(2) There exists a mapping $g : B \to A$ such that

$$g \circ f = \mathrm{id}_A, \qquad f \circ g = \mathrm{id}_B\,.$$

(3) The inverse relation $f^{-1} \subseteq B \times A$ is a mapping.
The mapping g satisfying conditions (2) is bijective as well and coincides with the relation f^{-1} of (3).

The mapping $g = f^{-1}$ from the previous statement is called *inverse* to f. It can be defined by

$$f^{-1}(b) = a \Leftrightarrow b = f(a).$$

Exercise. Let $f : A \to B$, $g : B \to C$ be bijections. Prove that $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$.

Exercise. Let $f : A \to B$, $g : B \to C$ be two mappings. Prove the following statements.
(1) If the mapping $g \circ f$ is injective, then so is the mapping f.
(2) If the mapping $g \circ f$ is surjective, then so is the mapping g.
(3) If the mapping $g \circ f$ is bijective, then so are both mappings f and g.

Exercise. Let $f : A \to B$, $g : A \to B$ be two mappings. Prove the following statements.
(1) If $h : B \to C$ is an injective mapping such that $h \circ f = h \circ g$, then $f = g$. In other words, an injective mapping can be cancelled from the left.
(2) If $h : D \to A$ is a surjective mapping such that $f \circ h = g \circ h$, then $f = g$. In other words, a surjective mapping can be cancelled from the right.

Example. Let $A \subseteq X$ be a subclass, $\iota_A = \{(a, a) \mid a \in A\}$ the corresponding inclusion mapping.
If $f : X \to Y$ is a mapping, then the composition $f \circ \iota_{AX}$ is a mapping $A \to Y$, which is called the *restriction* of the mapping f to a subset of A and is denoted by $f|_A$:

$$f|_A : A \xrightarrow{\iota_{AX}} X \xrightarrow{f} Y.$$

Note that the contraction is given by the same formula $a \mapsto f(a)$ as f.
If the set Y is contained in another set Z, then there is a composition $\iota_{YZ} \circ f : X \to Z$. In this case, we say that f arises by extension of the codomain, but there is no special notation for it.

Let $f : X \to Y$ be a mapping, let $A \subseteq X$ be a subclass. Define

$$fA = \{y \in Y \mid (\exists a \in A)\, y = f(a)\},$$

which is often abbreviated as

$$fA = \{f(a) \mid a \in A\}.$$

The class fA is called the *image* of the subclass $A \subseteq X$ under the mapping f.

Exercises. 1. Show that ι_{AX} is injective.
2. Let $f : X \to Y$ be a map. Show that $\bar{f} : X \to fX$, $x \mapsto f(x)$, is a surjective mapping and $f = \iota_{fX,Y} \circ \bar{f}$.
3. Show that every mapping f can be decomposed as $g \circ h$, where g is injective and h is surjective.

Exercise. Show that in those cases when all preimages under a mapping $f : A \to B$ are sets and form a class, then there is a bijection between it and the class fA.

If A is a set, we would expect fA to be a set as well, but in fact this requires a special axiom.

G. Axiom of substitution. If $f : X \to Y$ is a mapping and X is a set, then fX is a set.

In Zermelo–Fraenkel's axiomatisation, this axiom is actually a schema of axioms that uses substitutions, hence the name.

Example. Let $\mathcal{U}_1$ denote the class of all one-element sets, which we define as sets of the form $\{a\}$, where a is a set.

Let us show that $\mathcal{U}_1$ is a proper class. Consider the mapping $s : \mathcal{U} \to \mathcal{U}_1$, $a \mapsto \{a\}$. The mapping s is obviously surjective and injective (why?), i.e., a bijection. Therefore there is an inverse mapping $t : \mathcal{U}_1 \to \mathcal{U}$, $\{a\} \mapsto a$, which is also a bijection. If $\mathcal{U}_1$ were a set, then necessarily $\mathcal{U} = t\mathcal{U}_1$ would be also a set, which we know is not.

2.26. Definition. Let I be a set and $F : I \to \mathcal{U}$ a mapping which to each element of $i \in I$ assigns a set $F_i \in \mathcal{U}$. In general, such a mapping $F : I \to \mathcal{U}$ is called a *system of sets.* (Alternatively one can write $\{F_i\}_{i \in I}$.) We define

$$\bigcup_{i \in I} F_i = \bigcup \{F_i \mid i \in I\} = \bigcup FI,$$

$$\bigcap_{i \in I} F_i = \bigcap \{F_i \mid i \in I\} = \bigcap FI.$$

The axiom of substitution together with the axiom of union yields the following proposition.

2.27. Proposition. *Let I be a set and $F : I \to \mathcal{U}$ a mapping. Then*

$$\bigcup_{i \in I} F_i$$

is a set.

Proof. The class $\bigcup FI$ is a set according to the axiom of union, because FI is a set of sets under the substitution axiom.

Now we can prove that the Cartesian product of two sets is a set.

2.28. Proposition. *Let A, B be sets. Then $A \times B$ is a set.*

Proof. We have

$$A \times B = A \times \bigcup_{b \in B} \{b\} = \bigcup_{b \in B} A \times \{b\},$$

where $A \times \{b\}$ are sets according to the axiom of substitution because $A \times \{b\}$ is the image of the set A under the mapping $a \mapsto [a, b]$.

2.29. Definition. Let $f : X \to Y$ and $g : Y \to X$ be mappings. If

$$g \circ f = \mathrm{id}_X,$$

then we say that the mapping g is a *left inverse* of the mapping f. If

$$f \circ g = \mathrm{id}_Y,$$

then we say that the mapping g is a *right inverse of* of the mapping f.

2.30. Proposition. *Let $f : A \to B$ be an injective mapping of sets, where $A \neq \emptyset$. Then f has a left inverse.*

Proof. Since A is nonempty, there exists an element $c \in A$. Let us construct $g : B \to A$. If $b \in fA$, then it has exactly one preimage, say a, in which case we set $g(b) = a$. If, on the other hand, $b \notin fA$, then we put $g(b) = c$.

Let us show that $g \circ f = \mathrm{id}_A$. For arbitrary $a \in A$ we have $(g \circ f)(a) = g(f(a)) = g(b) = a$, which proves that $g \circ f = \mathrm{id}_A$.

Exercise. Is the assumption $A \neq \emptyset$ necessary?

The dual statement that each surjective mapping has a right inverse does not follow from the axioms presented so far. Recall the usual proof in the framework of naive set theory. Let $f : X \to Y$ be a surjective map. Then every element $y \in Y$ has at least one preimage $x \in X$ such that $f(x) = y$. For each $y \in Y$, let us choose one of the preimages and denote it by $g(y)$. We get the mapping $g : Y \to X$. The construction implies that $f \circ g = \mathrm{id}_Y$.

Remember that so far we can construct classes only by specification, that is, by some formula defining the class, i.e., by an unambiguous prescription. However, we have not constructed the subclass $g \subseteq Y \times X$ by an explicit prescription. Instead, for each $y \in Y$ we arbitrarily chose one element from the non-empty set $f^{-1}\{y\}$.

AC. Axiom of choice. If X is a set of nonempty sets, then there exists a mapping $\tau : X \to \bigcup X$ such that $(\forall x \in X)\, \tau(x) \in x$.

The mapping τ is called the selection mapping because it selects one element from each set $x \in X$.

2.31. Proposition. *Every surjective mapping $f : A \to B$ has a right inversion.*

Proof. Let $f : A \to B$ be surjective. Then $\{f^{-1}\{y\} \mid y \in Y\}$ is a system of nonempty sets. By the axiom of choice, there is a selection mapping τ satisfying $\tau(f^{-1}\{y\}) \in f^{-1}\{y\}$. Now it is sufficient to put $g(y) = \tau(f^{-1}\{y\})$ for $g(y) \in f^{-1}\{y\}$ to hold. Then $f \circ g = \mathrm{id}_Y$.

The axiom of choice has been used unconsciously for centuries until it was "discovered" in 1904 by Ernst Zermelo (1871–1953). The axiom of choice allows for non-constructive proofs of existence, that is, proofs that provide no prescription for the construction of the object whose existence is being proved. A typical case is Brouwer's fixed point theorem (any continuous mapping of the unit disc into itself has a fixed point). There was initially some distrust associated with the use of the axiom of choice, because it implied some rather paradoxical results, such as the Banach–Tarski theorem, according to which a unit ball in three-dimensional Euclidean space can be decomposed into a finite number of parts (five), from which two unit balls can be assembled using only rotations and translations (the parts are unmeasurable and therefore this is not a contradiction).

Nowadays, almost no one would question the axiom of choice. However, the arbitrariness associated with choice mapping causes "constructions" using the axiom of choice to be ambiguous and unrepeatable. We give some convincing examples in the next section (§ 2.17). It is therefore interesting to see which statements are equivalent to the axiom of choice. The existence of a right inverse to every surjection is a case in point.

2.32. Proposition. *The existence of a right inverse to every surjection between sets implies the axiom of choice.*

Proof. Let X be a set of nonempty sets. Then $X \times \bigcup X$ is a set and the subclass

$$E = \left\{ [x, a] \in X \times \bigcup X \;\middle|\; a \in x \right\}$$

of all pairs $[x, a]$ such that $a \in x \in X$ is also a set. Consider mappings

$$p_1 : E \to X, \quad [x, a] \mapsto x,$$
$$p_2 : E \to X, \quad [x, a] \mapsto a.$$

The mapping p_1 je surjective, because every set $x \in X$ is nonempty by assumption.

Let $q : X \to E$ be the right inverse of p_1. Then $p_1(q(x)) = x$ for all $x \in X$ and $q(x) = [x, p_2(q(x))] \in E$, whereupon $p_2(q(x)) \in x$ for each $x \in X$. We see that $p_2 \circ q$ is the selection map sought.

A selection mapping for a *finite* number of objects exists even without the axiom of choice. For example, for one nonempty set x containing the element c, the selection mapping is $\{[x, c]\}$.

Bertrand Russel popularised the axiom of choice by saying that the axiom is necessary when selecting a set from an infinite collection of socks, but it is not necessary when selecting a set from an infinite collection of shoes. Indeed, shoes being left and right, one can always select the left one (or the right one) from a pair, while with socks there is no way to say which is which.

2.33. Definition. Let $\{F_i\}_{i \in I}$ be a system of sets. The Cartesian product of the system is $\{F_i\}_{i \in I}$ is defined to be

$$\prod_{i \in I} F_i = \left\{ f : I \to \bigcup_{i \in I} F_i \;\middle|\; \underset{i \in I}{\forall} f(i) \in F_i \right\}.$$

The elements of the Cartesian product are called I-tuples. A convenient notation for them is $[f_i]_{i \in I}$, where $f_i \in F_i$.

2.34. Proposition. *The axiom of choice is equivalent with the statement that the Cartesian product of a nonempty set of sets is nonempty.*

Proof. An element of the Cartesian product element is the same thing as a selection mapping.

In class theory one can formulate axioms stronger than the axiom of choice for sets. We list two, but we will not use them.

Axiom of global choice. Let $T = \mathcal{U} \setminus \{\emptyset\}$ be the class of all non-empty sets. Then there exists a mapping $\tau : T \to \mathcal{U}$ such that $(\forall x \in T)\, \tau(x) \in x$.

The global axiom of choice says that q selection mapping exists even for proper classes of non-empty sets. It says nothing new about *sets*, only about proper classes.

The axiom of global choice implies that even the Cartesian product of a proper class of nonempty sets is non-empty.

Axiom of limitation of size. For every proper class C there exists a bijection $C \to \mathcal{U}$.

The axiom of limitation of size says that all proper classes are “the same” in that in the sense that any statement about one of them can be transferred to any other. Although its formulation is very daring, it does not contradict the other axioms.

2.17. Equivalence relation

2.35. Definition. Let $\rho \subseteq A \times A$ be a relation on a class A. The relation ρ is called
– *reflexive*, if $a \,\rho\, a$ for all $a \in A$ platí ;
– *symmetric*, if $a \,\rho\, b \Rightarrow b \,\rho\, a$;
– *transitive*, if $(a \,\rho\, b \wedge b \,\rho\, c) \Rightarrow a \,\rho\, c$.

Examples. 1. Every identical relation id_A is reflexive, symmetric and transitive.

2. Relation "$\in$" on the universal class $\mathcal{U}$ is neither reflexive, nor symmetric, nor transitive. 3. Relation "$\subseteq$" on the universal class $\mathcal{U}$ is reflexive and transitive, but not symmetric.

Exercise. We can draw the graph of a relation on a set by representing the elements of the set A by points in the plane and drawing an arrow between two elements if they are in a relation. How to tell whether the relation represented by the graph is reflexive, symmetric, transitive?

Exercise. Find an error in the following "proof" of the false claim that every symmetric and transitive relation ρ is reflexive: "If $a \ \rho \ b$, then $b \ \rho \ a$ by symmetry, and $a \ \rho \ a$ by transitivity."

2.36. Definition. A reflexive, symmetric and transitive relation on a class is called an *equivalence relation* (or simply *equivalence* if it cannot be confused with the logical equivalence).

Examples. 1. The identical relation id_A is an equivalence on the class A.

2. On the universal class $\mathcal{U}$ we can introduce the equivalence relation $\sim$ by the following rule: $A \sim B$ if and only if there exists a bijection $A \to B$.

2.37. Definition. Let ρ be an equivalence on a class A. For $a \in A$, denote

$$[a]_\rho = \{x \in A \mid a \ \rho \ x\}.$$

The class $[a]_\rho$ is called the *equivalence class* with respect to the equivalence ρ.

2.38. Proposition. Let ρ be an equivalence on a class A. For any $a \in A$ the statements

$$a \in [a]_\rho,$$
$$a \ \rho \ b \Leftrightarrow [a]_\rho = [b]_\rho,$$
$$[a]_\rho \cap [b]_\rho \neq \emptyset \Rightarrow [a]_\rho = [b]_\rho.$$

hold.

2.39. Definition. For every $a \in A$, let the class $[a]_\rho$ be a set. The class

$$A/\rho = \{B \in \mathcal{U} \mid (\exists a \in A)\, B = [a]_\rho\} = \{[a]_\rho \mid a \in A\}$$

is called the *partition* or the *quotient* of the class A by the equivalence ρ.

Thus, an equivalence induces a partition only if each equivalence class is a set. When this is the case, then $a \mapsto [a]_\rho$ is a mapping $A \to A/\rho$ and

$$\bigcup_{a \in A} [a]_\rho = A.$$

If A is a set, then every equivalence ρ on A induces a partition. Moreover, A/ρ is a set as well, because A/ρ is the image of the set A under $A \to \mathcal{U}$, $a \mapsto [a]_\rho$.

Exercise. Find all partitions of the set $A = \{1, 2, 3\}$ (there are five of them).

Exercise. Let ρ be an equivalence on a class A, let σ be an equivalence on a class B. Define the relation $\gamma = \rho \times \sigma$ on the class $C = A \times B$ by

$$(a_1, b_1) \ \gamma \ (a_2, b_2) \ \Leftrightarrow \ (a_1 \ \rho \ a_2) \wedge (b_1 \ \sigma \ b_2).$$

Show that γ is an equivalence relation. Show that the equivalence classes of γ are just the classes $U \times V$, where U is an equivalence class of ρ and V is an equivalence class of σ.

According to the axiom of choice, a set can be formed by selecting one element from each equivalence class. The selected element is called a *representative* of the equivalence class. Any element can be chosen as a representative of the equivalence class it belongs to. However, when speaking about representatives, we assume that no two of them belong to one and the same equivalence class.

Examples. 1. In this example we use the sets $\mathbb{R}$ and $\mathbb{Q}$ of real and rational numbers, respectively. Everything you need to know about them you have learned in mathematical analysis. On the set $\mathbb{R}$ we introduce the relation $\sim_{\mathbb{Q}}$ by

$$x \sim_{\mathbb{Q}} y \Leftrightarrow y - x \in \mathbb{Q}.$$

It is easy to see that $\sim_{\mathbb{Q}}$ is an equivalence relation on $\mathbb{R}$. By the axiom of choice, there is a set $U \subseteq \mathbb{R}$ that has exactly one common element with every equivalence class $[x]$ (omitting the subscript $_{\sim_{\mathbb{Q}}}$. Otherwise said, U consists of representatives of the equivalence classes. Then, for every $x \in \mathbb{R}$ there is a unique decomposition $x = u + q$ such that $u \in U$ and $q \in \mathbb{R}$. This yields a bijection between $\mathbb{R}$ and the Cartesian product $U \times \mathbb{Q}$.

2. Choosing an arbitrarily small interval $[a, b] \subset \mathbb{R}$, $b > a$, observe that every equivalence class can be assigned a representative belonging to $[a, b]$ by applying an appropriate rational shift. Therefore, it can be assumed that $U \subseteq [a, b]$ with no loss in generality. In this case, U is called a *Vitali set.* Giuseppe Vitali (1875–1932) employed this set in his proof of existence of Lebesgue unmeasurable sets.

Assume $U \subseteq [a, b]$ to be measurable. Let, furthermore, $\mathbb{Q}_c = \{q \in \mathbb{Q} \mid -c \leq q \leq c\} = \mathbb{Q} \cap [-c, c]$. Consider the set

$$U + \mathbb{Q}_c = \{u + q \mid u \in U \wedge q \in \mathbb{Q}_c\}.$$

Let us prove that

$$[a, b] \subseteq U + \mathbb{Q}_{b-a} \subseteq [2a - b,\ 2b - a].$$

The second inclusion being obvious, we prove the first one. Let $r \in [a, b]$ and let u be the representative of $[r]$. Then $r - u$ is a rational number from the interval $[a - b,\ b - a]$, which means that $r - u$ belongs to $\mathbb{Q}_{b-a}$.

Then, by properties of the Lebesgue measure,

$$b - a = \ell([a, b]) \leq \ell(U + \mathbb{Q}_{b-a}) \leq \ell([2a - b,\ 2b - a]) = 3b - 3a$$

whereas

$$\ell(U + \mathbb{Q}_{b-a}) = \ell\Big(\bigcup_{q \in \mathbb{Q}_{b-a}} (U + q)\Big) = \sum_{q \in \mathbb{Q}_{b-a}} \ell(U + q) = \sum_{q \in \mathbb{Q}_{b-a}} \ell(U)$$

is an infinite sum of constants $\ell(U)$. The sum is either zero if $\ell(U) = 0$ or infinity if $\ell(U) > 0$, but in neither case it lies between $b - a$ and $3b - 3a$, which is a contradiction.

2.18. Von Neumann's construction of natural numbers

The axioms presented so far are not strong enough to prove the existence of an infinite set, such as the set of all natural numbers. Moreover, if our goal is to build mathematics "from nothing," we should construct the natural numbers using only tools provided by the set theory.

Von Neumann proposed to *define* a natural number as the set of all preceding natural numbers. Starting from $0 = \emptyset$, we get successively

$$\begin{aligned}
&0 = \emptyset,\\
&1 = \{0\} = \{\emptyset\}\\
&2 = \{0, 1\} = \{\emptyset, \{\emptyset\}\}\\
&3 = \{0, 1, 2\} = \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}\\
&4 = \{0, 1, 2, 3\} = \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}, \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}\}
\end{aligned}$$

etc. We see that $1 = 0 \cup \{0\}$, $2 = 1 \cup \{1\}$, $3 = 2 \cup \{2\}$, etc.

2.40. Definition. A set N is said to be *inductive*, if
(i) $\emptyset \in N$;
(ii) $n \in N$ implies $n \cup \{n\} \in N$.
Accordingly,

$$\mathcal{I} = \{N \in \mathcal{U} \mid (\emptyset \in N) \wedge (\forall n)\,(n \in N \Rightarrow n \cup \{n\} \in N)\}$$

is the *class of all inductive sets.*

H. Axiom of infinity.

$$\mathcal{I} \neq \emptyset$$

(the class of all inductive sets is not empty) i.e,

$$(\exists N \in \mathcal{U})\,((\emptyset \in N) \wedge (\forall n)\,(n \in N \Rightarrow n \cup \{n\} \in N))$$

(there exists an inductive set).

The set of natural numbers, denoted by $\mathbb{N}$, is defined as the intersection of all inductive sets. Thus,

$$\mathbb{N} = \bigcap_{N \in \mathcal{I}} N = \bigcap \mathcal{I},$$

which is a set because $\mathbb{N}$ is a subclass of at least one inductive set N that exists according to the axiom of infinity.

We have listed the first few natural numbers above, others can easily be supplied. If n is a natural number, then $n \cup \{n\}$ is called the *successor* of n and is denoted by σn or $n+1$ or $n^{\bullet}$.

2.19. The principle of mathematical induction

The principle of mathematical induction allows us to prove statements that depend on a natural number, i.e., assertions of the form

$$(\forall n \in \mathbb{N})\, \phi(n).$$

2.41. Proposition. *Let $K \subseteq \mathbb{N}$ be a set such that $0 \in K$ and the implication $n \in K \Rightarrow n+1 \in K$ hold. Then $K = \mathbb{N}$.*

Proof. Since K is inductive, K contains the intersection of all inductive sets, i.e., $\mathbb{N} \subseteq K$. Since $K \subseteq \mathbb{N}$ by assumption, we have $K = \mathbb{N}$.

2.42. Corollary. *Let $\phi(n)$ be a formula such that $\phi(0)$ holds and the implication $\phi(n) \Rightarrow \phi(n+1)$ holds as well. Then $\phi(n)$ holds for all $n \in \mathbb{N}$.*

Proof. Put $K = \{n \in \mathbb{N} \mid \phi(n)\}$.

The principle of mathematical induction also makes it possible to construct sets or even classes X_n dependent on a natural number n. To do this, it is sufficient to prove formulas of the form

$$(\forall n \in \mathbb{N})(\exists X_n)\, \phi(n, X_n).$$

In this case we say that classes X_n are defined recursively.

Example. *The muddy children puzzle.* Imagine a community of logically thinking children. The guardian tells them that at least one of them has a muddy face. Every child can look around and see all other children's faces, but no child can determine the state of its own face itself (there are no mirrors and children are not allowed to speak). Once in a while, the guardian asks children to step forward if they know for sure that their own face is muddy. Can the children determine the state of their face under such circumstances?

2.20. The Peano axioms

Peano's axioms provide the basis for deriving the properties of natural numbers. They are named after Giuseppe Peano (1858–1932).

2.43. Definition. A set N, a mapping $\sigma : N \to N$ and an element $0 \in N$ are said to satisfy Peano's axioms if

1. There is no $m \in N$ such that $\sigma m = 0$
2. The mapping $\sigma : N \to N$ is injective.
3. If $A \subseteq N$ satisfies $0 \in A$ and $a \in A \Rightarrow \sigma a \in A$, then $A = N$.

In this section we show that natural numbers and $\sigma a = a \cup \{a\}$ satisfy Peano's axioms. Let us emphasise that by natural numbers we always mean von Neumann natural numbers.

Proof of the first Peano axiom is easy. Consider the equality $m \cup \{m\} = \emptyset$. The left-hand side contains m as an element, but the right-hand side does not.

The third Peano axiom is the principle of mathematical induction and holds for natural numbers in direct consequence of Proposition 2.41.

Before proving the second Peano axiom, we establish another remarkable property of natural numbers.

2.44. Definition. A set A is said to be *transitive* if every element $a \in A$ is also a subset of A:

$$(\forall a \in A)(a \subseteq A).$$

2.45. Proposition. *Every natural number is transitive.*

Proof. Let $S \subseteq \mathbb{N}$ be the set of all transitive elements of $\mathbb{N}$. Let us prove that $S = \mathbb{N}$ by induction. The initial step: $\emptyset \in S$ because $\emptyset$ has no elements.

The induction step: Assume n to be transitive and show that $\sigma n = n \cup \{n\}$ is also transitive. Let m be an arbitrary element of $n \cup \{n\}$. Then either $m \in n$ or $m = n$. If $m \in n$, then $m \subseteq n \subseteq n \cup \{n\}$ by transitivity. If $m = n$, then $m = n \subseteq n \cup \{n\}$ obviously.

To simplify the proof of the second Peano axiom, we introduce yet another axiom now. We call it an auxiliary axiom, since we shall replace it by a stronger axiom later.

Auxiliary axiom. No set is a subset of its own element, i.e.,

$$(\forall X \in \mathcal{U})(\forall x \in X)\,\urcorner(X \subseteq x).$$

Without this axiom, there can be vicious circles in the determination of sets by their elements. For if a set M is a subset of its own element $m \in M$, then $m \in M \subseteq m$, whence $m \in m$. However, if sets are determined by their elements, then m takes part in its own determination. To understand why this might be a problem, consider two sets m, n such that $m = \{m\}$ and $n = \{n\}$. Then we cannot determine whether $m = n$ until we know whether $m = n$.

Needless to say, there is no known example of a set m satisfying $m = \{m\}$, except anomalies like $m = \{\{\{\cdots\}\}\}$ (infinitely many nested brackets with nothing but brackets inside).

2.46. Lemma. *Let m, n be two sets, at least one of them transitive. Then $n \in m$ and $m \in n$ cannot hold simultaneously.*

Proof. Let $n \in m$ and $m \in n$ and, say, n be transitive. Then $m \subseteq n$ by the transitivity of n, which contradicts the auxiliary axiom in view of $n \in m$.

2.47. Proposition. *Let $m, n \in \mathbb{N}$ satisfy $\sigma n = \sigma m$. Then $n = m$.*

Proof. Assume that $n \cup \{n\} = m \cup \{m\}$. The left-hand side contains n as an element and this must be also true for the right-hand side. Hence, $n \in m$ or $n = m$. The right-hand side contains m as an element and this must be also true for the left-hand side. Hence, $m \in n$ or $m = n$. By the preceding lemma, we cannot have both $n \in m$ and $m \in n$. Hence $n = m$.

This finishes the proof of the Peano axioms.

2.21. The arithmetic of natural numbers

In this section we shall introduce addition and multiplication of natural numbers and prove their algebraic properties. A mapping $\mathbb{N} \times \mathbb{N} \to \mathbb{N}$ is called a binary operation on the set $\mathbb{N}$.

2.48. Proposition. *There exists exactly one mapping $\mathbb{N} \times \mathbb{N} \xrightarrow{+} \mathbb{N}$ satisfying*

$$n + 0 = n, \quad n + \sigma m = \sigma(n + m),$$

exactly one mapping $\mathbb{N} \times \mathbb{N} \xrightarrow{\cdot} \mathbb{N}$ satisfying

$$n \cdot 0 = 0, \quad n \cdot \sigma m = n + (n \cdot m)$$

and exactly one mapping $\mathbb{N} \times \mathbb{N} \xrightarrow{\wedge} \mathbb{N}$ satisfying

$$n^\wedge 0 = 1, \quad n^\wedge \sigma m = n \cdot (n^\wedge m).$$

Proof. For every $n \in \mathbb{N}$, we define the mapping $\alpha_n : \mathbb{N} \to \mathbb{N}$ by the recursive formula

$$\alpha_n(0) = n, \quad \alpha_n(\sigma m) = \sigma\alpha_n(m),$$

the mapping $\beta_n : \mathbb{N} \to \mathbb{N}$ by the recursive formula

$$\beta_n(0) = 0, \quad \beta_n(\sigma m) = \alpha_n(\beta_n(m))$$

and the mapping $\gamma_n : \mathbb{N} \to \mathbb{N}$ by the recursive formula

$$\gamma_n(0) = 1, \quad \gamma_n(\sigma m) = \beta_n(\gamma_n(m)).$$

In this way, mappings α_n, β_n and γ_n are correctly defined on the whole $\mathbb{N}$. According to Proposition 2.41 and the first two Peano axioms they are defined uniquely. We set $n + m = \alpha_n(m)$, $n \cdot m = \beta_n(m)$, $n{}^\wedge m = \gamma_n(m)$.

Now $\sigma n = n + 1$ since

$$n + 1 = \alpha_n(1) = \alpha_n(\sigma 0) = \sigma\alpha_n(0) = \sigma n.$$

The binary operations $+$, $\cdot$, ${}^\wedge$ just defined have all expected properties of the addition, multiplication and exponentiation operations of natural numbers. In particular, the following formulas of natural number arithmetic apply (we turn to the usual notation a^b for exponentiation $a{}^\wedge b$).

$$\begin{array}{lll}
a + 0 = a, & a \cdot 1 = a, & \\
a + b = b + a, & a \cdot b = b \cdot a, & \\
a + (b + c) = (a + b) + c, & a \cdot (b \cdot c) = (a \cdot b) \cdot c, & \\
a \cdot (b + c) = (a \cdot b) + (a \cdot c), & a^{b+c} = a^b \cdot a^c, & \\
(a \cdot b)^c = a^c \cdot b^c, & (a^b)^c = a^{b \cdot c}, & \\
a^0 = 1, & a^1 = a, & 1^a = 1.
\end{array}$$

Proofs go by induction.

We also have the cancellation laws

$$\begin{aligned}
&a + c = b + c \Rightarrow a = b,\\
&c \neq 0 \wedge a \cdot c = b \cdot c \Rightarrow a = b,\\
&c \neq 0 \wedge a^c = b^c \Rightarrow a = b,\\
&c \neq 0 \wedge c \neq 1 \wedge c^a = c^b \Rightarrow a = b.
\end{aligned}$$

2.49. Remark. Let us emphasise that

$$0^0 = 1$$

whereas $0^a = 0$ *does not hold* for $a = 0$. This simplifies many formulas. For example, one can write

$$a_n x^n + \cdots + a_1 x + a_0 = \sum_0^n a_n x^n$$

for all x, including $x = 0$, if and only if $0^0 = 1$.

For indeterminate forms 0^0 see Remark 2.50 below.

2.22. Integer, rational, real and complex numbers

Now we introduce the sets of integer, rational, real and complex numbers. It will be clear from the construction that each forms a set. No detailed exposition arithmetic operations will be provided; they are introduced in a familiar way.

Well known is the construction of the set of integers as consisting of the pairs $[+, n]$, where $n \in \mathbb{N}$, and the pairs $[-, n]$, where $n \in \mathbb{N}\setminus\{0\}$ ($+$ and $-$ are symbols different from the symbols denoting natural numbers). Otherwise said, $\mathbb{Z}$ can be defined as the union $\big(\{+\} \times \mathbb{N}\big) \cup \big(\{-\} \times (\mathbb{N} \setminus \{0\})\big)$. Obviously, $\mathbb{Z}$ is a set.

Alternatively, $\mathbb{Z}$ can be introduced as $(\{+, -\} \times \mathbb{N})/\rho$, where ρ is the equivalence relation whose unique class containing more than one element is $\{[+, 0], [-, 0]\}$ (we identify $+0$ and -0).

Construction of rational numbers as pairs p/q of incommensurable numbers, where $p \in \mathbb{Z}$ and $q \in \mathbb{N}\setminus\{0\}$, is also well known. Thus, the class $\mathbb{Q}$ of rational numbers can be constructed as the factor set of the product $\mathbb{Z}\times(\mathbb{N}\setminus\{0\})$ under the equivalence $[p_1, q_1]\ \rho\ [p_2, q_2] \Leftrightarrow p_1q_2 = p_2q_1$. Obviously, $\mathbb{Q}$ is a set.

The construction of real numbers as Dedekind cuts is standard. Here we give a different definition using binary expansions. Namely, a real number

$$a = 0, a_1a_2a_3 \ldots = \sum_{i\in\mathbb{N}} a_i 2^{-i}$$

from the interval $\mathbb{I} = [0, 1)$ can be identified with a mapping $\mathbb{N} \longrightarrow \{0, 1\} = 2$, and is therefore an element of the set $2^{\mathbb{N}}$ of all mappings $\mathbb{N} \longrightarrow 2$. However, since

$$\begin{aligned} &0, 111 \ldots = 1, 000 \ldots, \\ &0, 0111 \ldots = 0, 1000 \ldots \end{aligned}$$

etc., one must eliminate sequences that starting at some point consist of only 1s; these will be called *redundant* sequences. Every sequence $0, a_1 \ldots a_n0111 \ldots$ can be replaced by the sequence $0, a_1 \ldots a_n1000 \ldots$, which, starting from the same point, consists of 0s. The 0 immediately preceding the 1s turns into one 1 immediately preceding the 0s. Thus, the semi-closed interval $\mathbb{I}$ can be identified with a subset of the set $2^{\mathbb{N}}$, from which we omit the redundant sequences in the above sense. The set of real numbers is then identifiable with the product $\mathbb{Z} \times \mathbb{I}$.

2.50. Remark. In mathematical analysis, 0^0 is considered an indeterminate form. However, it only means that the limit of type 0^0 can take arbitrary values, that is, if $\lim_{x\to c} f(x) = 0$ and $\lim_{x\to c} g(x) = 0$, then the limit

$$\lim_{x\to c} f(x)^{g(x)}$$

depends on the choice of the functions $f(x)$ a $g(x)$.

This in no way conflicts with the identification $0^0 = 1$ made above. However, this does mean that the function 0^x is discontinuous at zero (whereas x^0 is continuous at zero and both cannot be continuous at zero at the same time).

However, the above said does not prevent other authors from always considering 0^0 as an indeterminate form. This is fine, as long as necessary precautions are obeyed, one example of which we pointed out in Remark 2.49.

Concerning the complex numbers, the identification $\mathbb{C} \ni x + \mathrm{i}y \longleftrightarrow [x, y] \in \mathbb{R} \times \mathbb{R}$ does the work.

3. Cardinalities

A very important equivalence relation is when $A \sim B$ if and only if there is a bijection $A \to B$. This is sufficiently reasonable only for sets. Finite sets are equivalent if and only if they have the same number of elements (a form of the pidgeon-hole principle). Equivalence is a generalisation to general sets of the "equality of the number of elements."

Exercise. Show that the relation $\sim$ just defined is reflexive, symmetric and transitive.

Exercise. Show that $A \times A \sim A^2$ for any set A, where A^2 is the set of all mappings from the two-element set $2 = \{0, 1\}$ to A.

A very important example of equivalence is provided by the following proposition.

3.1. Proposition. *If X is a set, then*

$$\wp X \sim 2^X,$$

i.e., there is a bijection between the set $\wp X$ of all subsets in X and the set of all mappings from X to the two-element set $2 = \{0, 1\}$.

Proof. Consider the mapping $\wp X \to 2^X$ that assigns to a subset $A \subseteq X$ the mapping $f_A : X \to 2^X$ defined by the rule

$$f_A(x) = \begin{cases} 1, & \text{if } x \in A, \\ 0, & \text{if } x \notin A. \end{cases}$$

This is a bijection with the inverse $2^X \to \wp X$ defined as the mapping which to a mapping $f : X \to 2^X$ assigns the preimage $f^{-1}\{1\}$.

Equivalent sets are said to have the same *cardinality.* We shall postulate the existence of a class $\mathcal{C}$ and a surjective mapping $\# : \mathcal{U} \to \mathcal{C}$ satisfying

$$\#A = \#B \Leftrightarrow A \sim B.$$

Elements of $\mathcal{C}$ will be called *cardinal numbers* and particular symbols of constants will be assigned to some of them.

It would be natural to define cardinalities as representatives of the equivalence classes $\mathcal{U}/\!\sim$ with respect to the equivalence $\sim$. However, the equivalence classes are proper classes in general, and cannot form a class the cardinalities could be elements of.

We had to postulate the existence of the class $\mathcal{C}$ of cardinal numbers since we cannot *define* it yet. This shortcoming will be corrected later. It will be proved later that cardinal numbers form a proper class.

For each $n \in \mathbb{N}$, sets of n elements are the sets equivalent to the set $n = \{0, \ldots, n-1\}$. We call them also *n-element sets.* The class of n-element sets will be denoted by $\mathcal{U}_n$. For the corresponding cardinal number we simply choose $n \in \mathbb{N}$, that is, $\#n = n$. The set $\mathbb{N}$ can be identified with the set of *finite cardinal numbers.* Cardinal numbers not in $\mathbb{N}$ will be called *infinite cardinal numbers.*

Sets in

$$\bigcup_{n \in \mathbb{N}} \mathcal{U}_n$$

are called *finite sets*.

The sets equivalent to the set $\mathbb{N}$ are called *countable sets*. The corresponding cardinal number is called *countable* and denoted by $\aleph_0$ ($\aleph$ is the first letter of of the Hebrew alphabet and reads *aleph*). Thus, we have $\#\mathbb{N} = \aleph_0$. We shall see infinitely moany alephs later (a proper class of them). Each infinite cardinal number will be some aleph.

Example. You have already been introduced to countable sets in the lecture on mathematical analysis. You should be able to verify that

$$\mathbb{N} \sim \mathbb{Z}, \quad \mathbb{N} \sim \mathbb{N} \times \mathbb{N}, \quad \mathbb{N} \sim \mathbb{Q}$$

(by finding the appropriate bijections). Hint: Arrange the numbers in a sequence.

The result shows that $\#\mathbb{Z} = \#(\mathbb{N} \times \mathbb{N}) = \#\mathbb{Q} = \aleph_0$.

3.1. The arithmetic of cardinal numbers

Now, we shall introduce addition, multiplication and exponentiation of cardinal numbers, which will be a generalisation of of similar arithmetic operations with natural numbers. To add cardinal numbers, we need to introduce a disjoint union of sets.

3.2. Definition. Let A, B be two sets. The set $(\{0\} \times A) \cup (\{1\} \times B)$ is called the *disjoint union* of the sets A, B and is denoted by $A \sqcup B$.

It is easily seen that the sets $\{0\} \times A$ and $\{1\} \times B$ have no common elements (why?), which is the point of this construction. Later we shall generalise it to systems of sets.

3.3. Proposition. *Let A, B be two disjoint sets. Then $A \cup B \sim A \sqcup B$.*

Proof. Define the mapping $h : A \cup B \to A \sqcup B$ by

$$h(x) = \begin{cases} [0, x], & \text{když } x \in A, \\ [1, x], & \text{když } x \in B. \end{cases}$$

As an exercise, show that h is a bijection.

3.4. Proposition. *Let $A \sim A'$ be two equivalent sets and $B \sim B'$ two other equivalent sets. Then*

1° $\quad A \sqcup B \sim A' \sqcup B'$,

2° $\quad A \times B \sim A' \times B'$,

3° $\quad B^A \sim {B'}^{A'}$

are equivalent sets.

Proof. Let $g_A : A \to A'$ and $g_B : B \to B'$ be bijections.

1° : Introducing

$$g_A \sqcup g_B : A \sqcup B \to A' \sqcup B'$$

by

$$(g_A \sqcup f_B)(x) = \begin{cases} [0, g_A(a)], & \text{if } x = [0, a] \in \{0\} \times A, \\ [1, g_B(b)], & \text{if } x = [1, b] \in \{1\} \times B, \end{cases}$$

it is easily seen that $g_A \sqcup g_B$ is a bijection with the inversion ${g_A}^{-1} \sqcup {g_B}^{-1}$. Thus, $A \sqcup B$ a $A' \sqcup B'$ are equivalent.

2° : Introducing

$$g_A \times g_B : A \times B \to A' \times B'$$

by

$$(g_A \times g_B)(a, b) = [g_A(a), g_B(b)],$$

it is easily seen that $g_A \times g_B$ is a bijection with the inversion ${g_A}^{-1} \times {g_B}^{-1}$. Thus, $A \times B$ a $A' \times B'$ are equivalent.

3° . The mapping $B^A \to {B'}^{A'}$, $f \mapsto g_B \circ f \circ g_A^{-1}$, has the inverse mapping ${B'}^{A'} \to B^A$, $h \mapsto g_B^{-1} \circ h \circ g_A$, which shows that both mappings are bijective. Thus, B^A a ${B'}^{A'}$ are equivalent.

As a consequence of the last statement, we can introduce addition, multiplication and exponentiation of cardinal numbers by

1° $\qquad \#A + \#B = \#(A \sqcup B),$

2° $\qquad \#A \times \#B = \#(A \times B),$

3° $\qquad (\#B)^{(\#A)} = \#(B^A).$

In other words, we arbitrarily choose sets to represent the given cardinalities, perform the corresponding set operation on them and determine the cardinality of the result.

Exercise. Show that $1 + 1 = 2$, $1 \times 2 = 2$, $1^2 = 1$, $2^1 = 2$.

Exercise. Show that $\aleph_0 + 1 = \aleph_0 + 2 = \cdots = \aleph_0 + \aleph_0 = \aleph_0$.
Hint: Arrange in a sequence.

Exercise. Show that ${\aleph_0}^2 = {\aleph_0}^3 = \cdots = \aleph_0$. (Warning: ${\aleph_0}^{\aleph_0} = \aleph_0$ does not hold!)
Hint: Arrange in a sequence.

For operations with cardinal numbers, the usual laws of arithmetic apply.

3.5. Proposition. *Let* $\mathfrak{a}, \mathfrak{b}, \mathfrak{c}$ *be arbitrary cardinal numbers. Then*

$$\mathfrak{a} + \mathfrak{b} = \mathfrak{b} + \mathfrak{a},$$
$$0 + \mathfrak{a} = \mathfrak{a},$$
$$(\mathfrak{a} + \mathfrak{b}) + \mathfrak{c} = \mathfrak{a} + (\mathfrak{b} + \mathfrak{c}),$$
$$\mathfrak{a} \times \mathfrak{b} = \mathfrak{b} \times \mathfrak{a},$$
$$1 \times \mathfrak{a} = \mathfrak{a},$$
$$(\mathfrak{a} \times \mathfrak{b}) \times \mathfrak{c} = \mathfrak{a} \times (\mathfrak{b} \times \mathfrak{c}),$$
$$\mathfrak{a} \times (\mathfrak{b} + \mathfrak{c}) = \mathfrak{a} \times \mathfrak{b} + \mathfrak{a} \times \mathfrak{c},$$
$$0 \times \mathfrak{a} = 0,$$
$$\mathfrak{a}^{\mathfrak{b}+\mathfrak{c}} = \mathfrak{a}^{\mathfrak{b}} \times \mathfrak{a}^{\mathfrak{c}},$$
$$\mathfrak{a}^{\mathfrak{b}\times\mathfrak{c}} = (\mathfrak{a}^{\mathfrak{b}})^{\mathfrak{c}},$$
$$(\mathfrak{a} \times \mathfrak{b})^{\mathfrak{c}} = \mathfrak{a}^{\mathfrak{c}} \times \mathfrak{b}^{\mathfrak{c}},$$
$$\mathfrak{a}^0 = 1,$$
$$\mathfrak{a}^1 = \mathfrak{a}.$$

Proof. Let $\mathfrak{a} = \#A$, $\mathfrak{b} = \#B$, $\mathfrak{c} = \#C$. As an exercise, find appropriate bijections.

Exercise. Let $\mathfrak{m}$ be a cardinal number. Show that $1^{\mathfrak{m}} = 1$, $\mathfrak{m} + \mathfrak{m} = 2 \times \mathfrak{m}$, $\mathfrak{m}^2 = \mathfrak{m} \times \mathfrak{m}$.
And if $\mathfrak{m} \neq 0$, then $0^{\mathfrak{m}} = 0$.

Operations with cardinal numbers have analogies for systems, possibly infinite, of cardinal numbers.

3.6. Definition. The *disjoint union* of a system of sets $\{F_i\}_{i\in I}$ is defined to be the union

$$\bigsqcup_{i\in I} F_i = \bigcup_{i\in I} (\{i\} \times F_i).$$

As with the binary operation $\sqcup$, the sets $\{i\} \times F_i$ have no common elements, and again this is the point of this construction.

3.7. Proposition. *Let $\{F_i\}_{i\in I}$ a system of pairwise disjoint sets, i.e., such that $F_i \cap F_j = \emptyset$, whenever $i \neq j \in I$. Then $\bigcup_{i\in I} F_i \sim \bigsqcup_{i\in I} F_i$.*

Proof. Considering a mapping $h : \bigcup_{i\in I} F_i \to \bigsqcup_{i\in I} F_i$ defined by

$$h(x) = [i, x], \quad \text{if } x \in F_i,$$

show as an exercise that h is a bijection.

3.8. Proposition. *Consider two systems of equivalent sets $F_i \sim F_i'$, $i \in I$. Then*

$$1^\circ \qquad \bigsqcup_{i\in I} F_i \sim \bigsqcup_{i\in I} F_i',$$

$$2^\circ \qquad \prod_{i\in I} F_i \sim \prod_{i\in I} F_i'.$$

Proof. Let $g_i : F_i \to F_i'$ be bijections

1° : Consider the mapping

$$\bigsqcup_{i\in I} g_i : \bigsqcup_{i\in I} F_i \to \bigsqcup_{i\in I} F_i'$$

defined by

$$\left(\bigsqcup_{i\in I} g_i\right)(x) = [i, g_i(f)], \quad \text{if } x = [i, f] \in \{i\} \times F_i$$

It is easy to see that $\bigsqcup_{i\in I} g_i$ is a bijection with the inversion $\bigsqcup_{i\in I} g_i^{-1}$. Hence the statement.

2° : Consider the mapping

$$\prod_{i\in I} g_i : \prod_{i\in I} F_i \to \prod_{i\in I} F_i'$$

defined by

$$\left(\prod_{i\in I} g_i\right)([f_i]_{i\in I}) = [g_i(f_i)]_{i\in I}.$$

It is easy to see that $\prod_{i\in I} g_i$ is a bijection with the inversion $\prod_{i\in I} g_i^{-1}$. Hence the statement.

Considering cardinal numbers $\mathfrak{f}_i = \#F_i$, we can define the sum and the product of a system of cardinal numbers by

$$\sum_{i\in I} \mathfrak{f}_i = \#\bigcup_{i\in I} F_i,$$

$$\prod_{i\in I} \mathfrak{f}_i = \#\prod_{i\in I} F_i.$$

These are direct generalizations of the binary operations '+' and '×' defined above. (these are obtained when considering two-element set systems).

In analogy with Theorem 3.5, we have

$$\mathfrak{a} \times \sum_{i\in I} \mathfrak{b}_i = \sum_{i\in I} \mathfrak{a} \times \mathfrak{b}_i,$$

$$\mathfrak{a}^{\sum_{i\in I} \mathfrak{b}_i} = \prod_{i\in I} \mathfrak{a}^{\mathfrak{b}_i},$$

$$\left(\prod_{i\in I} \mathfrak{a}_i\right)^{\mathfrak{b}} = \prod_{i\in I} \mathfrak{a}_i^{\mathfrak{b}}$$

(prove them as an exercise).

The analogy of the commutative and the associative law is that no ordering of the index set occurs in the definition.

Exercise. Show that

$$\sum_{n\in\mathbb{N}} 1 = \sum_{n\in\mathbb{N}} 2 = \cdots = \sum_{n\in\mathbb{N}} \aleph_0 = \aleph_0$$

and, in general,

$$\sum_{n\in\mathbb{N}} \mathfrak{m} = \aleph_0 \times \mathfrak{m}.$$

3.2. Ordering of cardinal numbers

We say that a cardinal number $\mathfrak{a}$ is less than or equal to a cardinal number $\mathfrak{b}$, and we write $\mathfrak{a} \le \mathfrak{b}$, if there is an injective mapping $A \to B$ between sets A, B that represent $\mathfrak{a}$ and $\mathfrak{b}$, that is, $\mathfrak{a} = \#A$ and $\mathfrak{b} = \#B$. The injective mapping does or does not exist independently of the choice of the representatives A, B (why?). We get the relation $\le$ between cardinal numbers (notwithstanding the fact that we cannot define the class of cardinal numbers yet).

The relation $\le$ is obviously reflexive (via identical maps) and transitive (through the composition of injective maps). Cantor–Bernstein's theorem says that relation $\le$ is also antisymmetric. This is a non-trivial statement. If you have forgotten what antisymmetry is, see Definition of 4.1.

As usual, the sharp inequality $\mathfrak{a} < \mathfrak{b}$ means $\mathfrak{a} \le \mathfrak{b}$ and $\mathfrak{a} \ne \mathfrak{b}$.

3.9. The Cantor–Bernstein theorem. *If $\mathfrak{a} \le \mathfrak{b}$ and $\mathfrak{b} \le \mathfrak{a}$, then $\mathfrak{a} = \mathfrak{b}$.*

Proof. Let $\mathfrak{a} = \#A$ a $\mathfrak{b} = \#B$. Then there exist injective mappings $f : A \to B$ and $g : B \to A$. Let us construct a bijective mapping $A \to B$. Denote

$$\begin{aligned} A_0 &= A, & B_0 &= B, \\ A_1 &= gB_0, & B_1 &= fA_0, \\ A_2 &= gB_1, & B_2 &= fA_1, \\ A_3 &= gB_2, & B_3 &= fA_2, \\ A_4 &= gB_3, & B_4 &= fA_3, \\ &\vdots \end{aligned}$$

and $A_{i+1} = gB_i$, $B_{i+1} = fA_i$ for every natural number i. We have

$$\begin{aligned} &A_0 \supseteq A_1 \supseteq A_2 \supseteq \cdots, \\ &B_0 \supseteq B_1 \supseteq B_2 \supseteq \cdots, \end{aligned}$$

Then g induces a bijection $B_i \to A_{i+1}$ and f induces a bijection $A_i \to B_{i+1}$. Consequently, $g|_{B_{i-1}\setminus B_i}$ is a bijection

$$B_{i-1} \setminus B_i \to A_i \setminus A_{i+1}$$

and similarly $f|_{A_{i-1}\setminus A_i}$ is a bijection

$$A_{i-1} \setminus A_i \to B_i \setminus B_{i+1}.$$

By composing, we get bijections

$$A_0 \setminus A_1 \to B_1 \setminus B_2 \to A_2 \setminus A_3 \to B_3 \setminus B_4 \to \cdots$$

via the mapping f and bijections

$$B_0 \setminus B_1 \to A_1 \setminus A_2 \to B_2 \setminus B_3 \to A_3 \setminus A_4 \to \cdots$$

via the mapping g. These yields bijections between the unions

$$\bigcup_{i\in\mathbb{N}} A_{2i} \setminus A_{2i+1} \leftrightarrow \bigcup_{i\in\mathbb{N}} B_{2i+1} \setminus B_{2i+2}$$

as well as between the unions

$$\bigcup_{i\in\mathbb{N}} B_{2i} \setminus B_{2i+1} \leftrightarrow \bigcup_{i\in\mathbb{N}} A_{2i+1} \setminus A_{2i+2}.$$

Altogether, this yields the bijection

$$\bigcup_{i\in\mathbb{N}} B_i \setminus B_{i+1} \leftrightarrow \bigcup_{i\in\mathbb{N}} A_i \setminus A_{i+1}.$$

The subsets $\bigcup_{i\in\mathbb{N}} B_i \setminus B_{i+1}$ and $\bigcup_{i\in\mathbb{N}} A_i \setminus A_{i+1}$ are equal to

$$\bigcup_{i\in\mathbb{N}} A_i \setminus A_{i+1} = A \setminus \bigcap_{i\in\mathbb{N}} A_i, \quad \text{and} \quad \bigcup_{i\in\mathbb{N}} B_i \setminus B_{i+1} = B \setminus \bigcap_{i\in\mathbb{N}} B_i,$$

respectively. Here the inclusion $\subseteq$ follows from $A_i \setminus A_{i+1} \subseteq A \setminus \bigcap_{i\in\mathbb{N}} A_i$ for all i; similarly for B. The opposite inclusion $\supseteq$ can be proved as follows. For every element $a \in A \setminus \bigcap_{i\in\mathbb{N}} A_i$ there exists the minimal i such that $a \notin A_{i+1}$, whence $a \in A_i \setminus A_{i+1}$.

Thus, we have found a bijection

$$A \setminus \bigcap_{i\in\mathbb{N}} A_i \leftrightarrow B \setminus \bigcap_{i\in\mathbb{N}} B_i.$$

To construct a bijection between $\bigcap_{i\in\mathbb{N}} A_i$ and $\bigcap_{i\in\mathbb{N}} B_i$, we can proceed as follows.

$$\begin{aligned}\bigcap_{i\in\mathbb{N}} A_i = \bigcap_{i\in\mathbb{N}} A_{2i} = \bigcap_{i\in\mathbb{N}} (g\circ f)^i A_i \overset{f}{\longleftrightarrow} \bigcap_{i\in\mathbb{N}} f(g\circ f)^i A_i \\ = \bigcap_{i\in\mathbb{N}} (f\circ g)^i f A_i = \bigcap_{i\in\mathbb{N}} (f\circ g)^i B_{i+1} = \bigcap_{i\in\mathbb{N}} B_{2i+1} = \bigcap_{i\in\mathbb{N}} B_i.\end{aligned}$$

This finishes the proof.

The Cantor–Bernstein's theorem is very useful in determining cardinality of particular sets. Proving two inequalities (finding two injective mappings) is very often much easier than finding a bijection. Examples will be given below (see the discussion of the cardinality of the continuum).

Also useful is the following proposition.

3.10. Proposition. *Let the cardinalities* $\mathfrak{a}, \mathfrak{a}', \mathfrak{b}, \mathfrak{b}'$ *satisfy* $\mathfrak{a} \le \mathfrak{a}'$ *and* $\mathfrak{b} \le \mathfrak{b}'$. *Then*

1° $\quad \mathfrak{a} + \mathfrak{b} \le \mathfrak{a}' + \mathfrak{b}'$,

2° $\quad \mathfrak{a} \times \mathfrak{b} \le \mathfrak{a}' \times \mathfrak{b}'$,

3° $\quad \mathfrak{a} \ne 0 \Rightarrow \mathfrak{a}^{\mathfrak{b}} \le \mathfrak{a}'^{\mathfrak{b}'}$.

Proof. Let $\mathfrak{a} = \#A$, $\mathfrak{a}' = \#A'$, $\mathfrak{b} = \#B$, $\mathfrak{b}' = \#B'$. By assumption, there exist injective mappings $f : A \to A'$, $g : B \to B'$. It is easy to see that the mappings $f \sqcup g$ a $f \times g$ constructed in the proof of Proposition 3.4, part 1° and 2°, are injective (carry out in detail yourself).

In case 3° one has to construct an injective mapping $A^B \to A'^{B'}$. Let $h : B \to A$ be arbitrary. By the injectivity of $g : B \to B'$, there exists a mapping $\bar{h} : B' \to A$ such that $\bar{h} \circ g = h$ (for this we need $A \ne \emptyset$). Then the composition $h' = f \circ \bar{h}$ is the mapping $B' \to A'$ sought. Observe that $h' \circ g = f \circ \bar{h} \circ g = f \circ h$.

To prove the injectivity of the assignment $h \mapsto h'$, consider two mappings $h_1, h_2 : B \to A$ that are assigned one and the same mapping $h_1' = h_2' : B' \to A'$. Then $h_1' \circ g = h_2' \circ g$, whence $f \circ h_1 = f \circ h_2$. Since f is injective, we can cancel it from the left, obtaining $h_1 = h_2$.

The classical Dirichlet principle states that for any natural number n there is no injective mapping $n + 1 \to n$. We prove it as part of a more general statement.

3.11. Dirichlet's principle. *We have*

$$0 < 1 < 2 < 3 < \cdots < \aleph_0.$$

Proof. Recall that $n+1 = \{0, 1, 2, \ldots, n\}$. Non-sharp inequalities $\le$ follow from the existence of injective insertions

$$\emptyset \to \{0\} \to \{0,1\} \to \{0,1,2\} \to \cdots \to \mathbb{N}.$$

Inequalities $n+1 \neq n$ directly follow from the classical Dirichlet principle. The classical Dirichlet's principle will be proved by induction. For $n \in \mathbb{N}$, let D_n denote the satement: "there is no injective mapping $n+1 \to n$." Statement D_0 is obvious, since $0 = \emptyset$, while $1 \neq \emptyset$.

Let D_n hold for some n, let us prove the validity of D_{n+1}. Assume by contradiction the existence of an injective mapping $f : n+2 \to n+1$. Recall that $n+2 = \{0, 1, 2, \dots, n+1\}$. By injectivity, the element $f(n+1) \in n+1$ is not an image $f(k)$ of any other element $k \in n+2$ than $n+1$. Consider the restriction $f|_{n+1} : n+1 \to n+1 \setminus \{f(n+1)\}$, which is well defined and also injective. Obviously, $n+1 \setminus \{f(n+1)\} \sim n$, so we got an injective mapping $n+1 \to n$, in contradiction with D_n.

It remains to prove the inequality $n < \aleph_0$ for all $n \in \mathbb{N}$. Assume the existence of an injective mapping $f : \mathbb{N} \to n$. Then the restriction $f|_{n+1}$ is also injective, which contradicts the D_n above.

3.12. Proposition. *If $\mathfrak{a} < \mathfrak{b} < \mathfrak{c}$, then $\mathfrak{a} < \mathfrak{c}$.*

Proof. Let $\mathfrak{a} < \mathfrak{b} < \mathfrak{c}$. Then $\mathfrak{a} \le \mathfrak{c}$, since composition of injective mappings is injective. Assume that $\mathfrak{a} = \mathfrak{c}$. Then $\mathfrak{a} < \mathfrak{b} < \mathfrak{a}$. Consequently, $\mathfrak{a} = \mathfrak{b}$ by the Cantor–Bernstein theorem, which is a contradiction.

3.13. Cantor's theorem. *For every cardinal number $\mathfrak{a}$ we have $\mathfrak{a} < 2^{\mathfrak{a}}$.*

Proof. Let $\mathfrak{a} = \#A$. Then $a \mapsto \{a\}$ is an injective mapping $A \to \wp A$. Hence, $\mathfrak{a} \le 2^{\mathfrak{a}}$.

Assuming that $A \sim \wp A$, we derive a contradiction. Let $F : A \to \wp A$ be a bijection. Denote

$$X = \{a \in A \mid a \notin F(a)\}.$$

Since F is surjective, there exists $x \in A$ such that $F(x) = X$. Then exactly one of the following two possibilities will happen.

1) $x \in X$, i.e., $x \in F(x)$, but then $x \notin X$ by the definition of X, which is a contradiction;

2) $x \notin X$, i.e., $x \notin F(x)$, but then $x \in X$ by the definition of X, which is a contradiction.

This shows that $A \not\sim \wp A$, whence $\mathfrak{a} \neq 2^{\mathfrak{a}}$.

3.14. Důsledek. *For every cardinal number there is a larger cardinal number.*

Note that we do not yet know whether there exist incomparable cardinalities $\mathfrak{a}, \mathfrak{b}$, i.e., ones for which none of $\mathfrak{a} < \mathfrak{b}$, $\mathfrak{a} = \mathfrak{b}$, $\mathfrak{a} > \mathfrak{b}$ holds. The answer is negative (they do not exist), but it requires more effort. We will postpone it to one of the next chapters along with the so-called laws of absorption, which allow us to determine sums and products of arbitrary cardinalities without tedious calculations. In the meantime, we shall discuss other cardinality laws occurring in mathematical analysis.

3.3. The cardinality of the continuum

The cardinality $2^{\aleph_0}$, of which we know so far only that it is uncountable ($> \aleph_0$ by Cantor's theorem), is called *the cardinality of the continuum*. It will be denoted $\mathfrak{c}$. The name comes from the name "continuum" for the set of real numbers.

3.15. Proposition. *We have $\mathfrak{c} = \#\mathbb{R}$.*

Proof. 1) To prove that $\mathfrak{c} \le \#\mathbb{R}$, we construct an injective mapping $2^{\mathbb{N}} \to \mathbb{R}$. An element of $2^{\mathbb{N}}$ is a mapping $\mathbb{N} \to 2$, which is the same thing as a sequence $[a_n]_{n \in \mathbb{N}}$ of zeroes and ones. To such a sequence we can assign the real number with the decimal expansion

$$\sum_{n\in\mathbb{N}} a_n 10^n = 0, a_0 a_1 a_2 \dots$$

The assignment is injective because different sequences correspond to different real numbers (problematic merges such as $1 = 0.999\dots$ do not occur because we chose a number system with base $10 \neq 2$).

2) To prove that $\#\mathbb{R} \leq \mathfrak{c}$, we construct an injective mapping $\mathbb{R} \to 2^{\mathbb{Q}}$ (the rational numbers are also countably many). Let the assignment be

$$r \mapsto \mathbb{Q}_{<r} = \{q \in \mathbb{Q} \mid q < r\}.$$

To prove injctivity, we assume that $r_1 < r_2$ are different real numbers. Then there exists a rational number q such that $r_1 < q < r_2$. The number q belongs to $\mathbb{Q}_{<r_2}$ and does not belong to $\mathbb{Q}_{<r_1}$. Hence, $\mathbb{Q}_{<r_2} \neq \mathbb{Q}_{<r_1}$, which finishes the proof.

It is easy to see that the intervals $\mathrm{I}(a,b)$, $\mathrm{I}(a,b]$, $\mathrm{I}[a,b)$, $\mathrm{I}[a,b]$, where $a < b$ are real numbers (we use the common notation preceded by the letter I for clarity), also have the cardinality of the continuum. It is enough to notice that the above proof of equality $\mathfrak{c} = \#\mathbb{R}$ passes for intervals if replacing the injective map $2^{\mathbb{N}} \to \mathbb{R}$ with an injective map $2^{\mathbb{N}} \to \mathrm{I}$, where I is one of the intervals $\mathrm{I}(a,b)$. $\mathrm{I}(a,b]$, $\mathrm{I}[a,b)$, $\mathrm{I}[a,b]$.

The proof of the Cantor–Bernstain theorem is constructive enough to provide an explicit bijection between any pair of sets $2^{\mathbb{N}}$, $\mathbb{R}$, $\mathrm{I}(a,b)$, $\mathrm{I}(a,b]$, $\mathrm{I}[a,b)$, $\mathrm{I}[a,b]$. However, it is obvious that the constructions would be rather complicated. In the following set of exercises we shall show how to construct such bijections more easily.

Exercise. Find a bijection between $\mathrm{I}(0,1)$ and $\mathbb{R}$.
Hint: a composition of tan and a linear function.

Exercise. Find a bijection between $\mathrm{I}(0,1)$ a $\mathrm{I}(0,1]$.
Hint: Decompose every interval into the union of countably many one-point sets and countably many open intervals, e.g.,

$$\mathrm{I}(0,1) = \bigcup_{i\in\mathbb{N}\setminus\{0\}} \left\{\frac{1}{2^n}\right\} \quad \cup \quad \bigcup_{i\in\mathbb{N}} \mathrm{I}\left(\frac{1}{2^{n+1}}, \frac{1}{2^n}\right),$$

$$\mathrm{I}(0,1] \;=\; \bigcup_{i\in\mathbb{N}} \left\{\frac{1}{2^n}\right\} \quad \cup \quad \bigcup_{i\in\mathbb{N}} \mathrm{I}\left(\frac{1}{2^{n+1}}, \frac{1}{2^n}\right).$$

Exercise. Find a bijection between $\mathrm{I}(0,1]$ a $\mathrm{I}[0,1]$.

Exercise. Show that

$$\mathfrak{c} = \mathfrak{c}+1 = \mathfrak{c}+2 = \cdots = \mathfrak{c}+\aleph_0 = \mathfrak{c}+\mathfrak{c} = \mathfrak{c}+\mathfrak{c}+\mathfrak{c} = \cdots = \aleph_0 \times \mathfrak{c}$$
$$= \mathfrak{c}\times\mathfrak{c} = \mathfrak{c}\times\mathfrak{c}\times\mathfrak{c} = \cdots = \mathfrak{c}^{\aleph_0}.$$

Exercise. Show that $\#\mathbb{C} = \mathfrak{c}$.

Exercise. Show that the set of all continuous functions $\mathbb{R} \to \mathbb{R}$ has the cardinality of the continuum.
Hint: A continuous function is determined by its values at all rational numbers.

Exercise. Show that the set of all open subsets in $\mathbb{R}^n$ has the cardinality of the continuum.
Hint: An open subsets in $\mathbb{R}^n$ is determined by its rational points (points with rational coordinates).

Example. **Cantor's discontinuum** In the first part of the proof of Theorem 3.15 we constructed an injective mapping $2^{\mathbb{N}} \to \mathbb{R}$. To a sequence $[a_n]_{n\in\mathbb{N}}$ of zeros and ones, we assigned a real number with decimal expansion

$$\sum_{n\in\mathbb{N}} a_n 10^n = 0, a_0 a_1 a_2 \ldots$$

The image of this mapping is a subset in $\mathbb{R}$, which is known as the Cantor discontinuum. It can be described as the set of all real numbers from the interval $\mathrm{I}[0,1]$ whose decimal expansion contains only zeros and ones (including those with infinitely many consecutive ones). We obtain the same set by excluding the open interval $\mathrm{I}(\frac{1}{10}, \frac{9}{10})$ (the middle eight tenths) from the interval $\mathrm{I}[0,1]$, obtaining the union of two closed intervals $\mathrm{I}[0, \frac{1}{2}] \cup \mathrm{I}[\frac{9}{10}, 1]$, which can be subject to an analogous procedure and so on and so forth. When repeating indefinitely the procedure of removing the middle eight tenths, we arrive at Cantor's discontinuum. What the first part of the proof of Theorem 3.15 actually means is that the Cantor discontinuum has the cardinality of the continuum.

Exercise. Prove that the overall length of the removed intervals is equal to 1.

3.4. Cantor's proof of the existence of transcendent numbers

In Cantor's time, only particular transcendent numbers were known. Moreover, the set of transcendent numbers is not fully explored till today. Cantor proved with surprising ease that there are infinitely many transcendent numbers.

Let us recall the definitions. An algebraic number is a complex number that is the root of a nonzero polynomial with rational coefficients. A transcendent number is a complex number that is not algebraic.

3.16. Proposition. *The cardinality of the set of algebraic numbers is* $\aleph_0$.

Proof. Let $\mathfrak{A}$ be the set of all algebraic numbers.

1) $\aleph_0 \leq \#\mathfrak{A}$, since every rational number is algebraic (why?).

2) Let us show that $\#\mathfrak{A} \leq \aleph_0$. The set of algebraic numbers can be expressed as the union

$$\mathfrak{A} = \bigcup_{n\in\mathbb{N}} \mathfrak{A}_n,$$

where n is the degree of the respective minimal polynomial (the polynomial of the minimal degree among all monic polynomials with rational coefficients possessing the root a). The cardinality of the set $\mathfrak{A}_n$ is $n \times \#\mathbb{Q}^n = n \times {\aleph_0}^n = \aleph_0$ (why?). Thus, we get

$$\#\mathfrak{A} \leq \# \bigcup_{n\in\mathbb{N}} \mathfrak{A}_n \leq \sum_{n\in\mathbb{N}} \#\mathfrak{A}_n = \sum_{n\in\mathbb{N}} \aleph_0 = \aleph_0 \times \aleph_0 = \aleph_0.$$

Consequently, $\#\mathfrak{A} = \aleph_0$.

3.17. Corollary. *The cardinality of the set of all transcendent numbers is* $\mathfrak{c}$.

3.5. Other cardinalities of mathematical analysis

We have already mentioned that there are infinitely many (even a proper class) of infinite cardinalities. Besides alephs $\aleph$, which are rather difficult concepts, there are infinitely many cardinalities, which are called $\beth$-numbers ($\beth$ is the second letter of the Hebrew alphabet). We define $\beth_n$ for all natural numbers $n \in \mathbb{N}$ by recursion

$$\beth_0 = \aleph_0, \quad \beth_{n+1} = 2^{\beth_n}.$$

Thus,

$$\beth_1 = 2^{\aleph_0} = \mathfrak{c}, \quad \beth_2 = 2^{2^{\aleph_0}} = 2^{\mathfrak{c}}, \quad \beth_3 = 2^{2^{2^{\aleph_0}}} = 2^{2^{\mathfrak{c}}}, \quad \ldots$$

Obviously from Cantor's theorem, $\beth_0 < \beth_1 < \beth_2 < \ldots$ We also have

$$\beth_0 = \#\mathbb{N}, \quad \beth_1 = \#\wp\mathbb{N}, \quad \beth_2 = \#\wp\wp\mathbb{N}, \quad \beth_3 = \#\wp\wp\wp\mathbb{N}, \quad \ldots$$

$\beth$-numbers are abundant in mathematical analysis.

Exercise. Show that the following sets have the cardinality of $\beth_2$.

1) $\wp\mathbb{R}$ of all subsets of the set of real numbers;
2) $\mathbb{R}^{\mathbb{R}}$ of all real functions of one real variable;
3) $\sum_{n\in N} \mathbb{R}^{\mathbb{R}^n}$ of all real functions of finitely many real variables.

3.18. Remark. How the arithmetic of cardinal numbers differs from the arithmetic of natural numbers? First of all, one can cancel neither in sums nor in products. For example, $1 + \aleph_0 = 0+\aleph_0$, but $1 \neq 0$. Similarly, $1\times\aleph_0 = 2\times\aleph_0$, but $1 \neq 2$. Consequently, no reasonable subtraction and division can be introduced.

What the two arithmetics do have in common is that there exist no divisors of zero. If $\mathfrak{a} \times \mathfrak{b} = 0$, then $\mathfrak{a} = 0$ or $\mathfrak{b} = 0$, because the Cartesian product of nonempty sets is nonempty.

4. Well-ordered sets and ordinal numbers

In the previous pages we reached some basic understanding of the cardinal numbers, even though quite leaky one for now. We have a relatively complete knowledge about countable sets, where the powerful method of mathematical induction works. The method can be generalised to so-called *transfinite induction*, which is applicable to so-called *well-ordered sets*.

While cardinal numbers generalise natural numbers in the sense of quantity, ordinal numbers generalise natural numbers in the sense of ordering. Both continue beyond finite numbers, but in different ways. To a single infinite cardinal number there correspond infinitely many ordinal numbers. For example, while there is only one countable cardinality $\aleph_0$, there will be uncountably many countable ordinal numbers (exactly $\aleph_1$). And since there are never enough surprises, we shall establish a bijection between cardinal numbers and ordinal numbers.

4.1. Ordering

To start with, we recall general ordering relations.

4.1. Definition. A relation $\leq$ on a class A is said to be *antisymmetric*, if $(a \leq b \wedge b \leq a) \Rightarrow a = b$ holds for all $a, b \in A$.

A reflexive, antisymmetric and transitive relation is called an *ordering*.

If $\leq$ is an ordering, then $\geq$ defined by $a \geq b \Leftrightarrow b \leq a$ is also an ordering.

If $\leq$ is an ordering on class A, then the relation $<$ defined by by the prescription $a < b \Leftrightarrow (a \leq b \wedge a \neq b)$ is called *is a sharp ordering*. Similarly, $>$.

A class along with a specified ordering is called an *ordered class*. A set together with a specified ordering is called an *ordered set*.

Every subclass $B \subseteq A$ of an ordered class A is an ordered class, if we define the *induced ordering* on B by the formula $\leq_B = \leq_A \cap (B \times B)$, that is, if we require $a \leq_B b \Leftrightarrow a \leq_A b$ for all $a, b \in B$. Similarly for the sharp ordering.

Example. The inclusion $\subseteq$ is an arrangement on the universal class $\mathcal{U}$ and all its subclasses. A sharp inclusion $\subset$ is a sharp ordering on the universal class $\mathcal{U}$ and all its subclasses.

Example. The sets $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}$ are ordered “by size” in the usual way.

Example. Empty relation $\emptyset \times \emptyset$ on the empty set is both an ordering and a sharp ordering.

4.2. Definition. Let A be an ordered class. An element $m \in A$ is called the *greatest element* resp. *least element*, if

$$(\forall a \in A)\, a \leq m \quad \text{resp.} \quad (\forall a \in A)\, m \leq a.$$

An element $m \in A$ is called a *maximal element* resp. a *minimal element*, if

$$(\forall a \in A)\, (m \leq a \Rightarrow m = a) \quad \text{resp.} \quad (\forall a \in A)\, (a \leq m \Rightarrow m = a).$$

Example. The universal class $\mathcal{U}$ has the least element $\emptyset$ and no greatest element.

Exercise. An ordered class has no more than one greatest and no more than one least element.

Exercise. The greatest element, if existing, is the only maximal element. The least element, if existing, is the only minimal element.

Exercise. Give an example of an ordered set that has a single maximal element and no greatest element.

Hint. The set is necessarily infinite.

4.3. Definition. Let X be an ordered class, $A \subseteq X$ its subclass. An element $m \in X$ is called an *upper bound* resp. a *lower bound* of the subclass A, if

$$(\forall a \in A)\, a \leq m \quad \text{resp.} \quad (\forall a \in A)\, m \leq a.$$

If there is a least upper bound, it is called the *supremum* and is denoted by $\sup A$. If there is a greatest bottom bound, it is called the *infimum* and is denoted $\inf A$.

Exercise. Every set $A \subseteq \mathcal{U}$ (i.e. a set of sets) has both a supremum and an infimum with respect to the ordering by inclusion. Namely, $\sup A = \bigcup A$ and $\inf A = \bigcap A$.

4.2. Chains

4.4. Definition. Elements a, b of an ordered set are called *comparable* if

$$(a \leq b) \vee (b \leq a).$$

An ordering $\leq$ on a class A is said to be *total* (or *complete*) if every two elements of A are comparable. A class together with a total ordering is called a *totally* (or *completely ordered class.*

A totally ordered set is called a *chain.*

Example. The set of real numbers with the usual ordering is a chain.

A subset of a chain is a chain under the induced ordering (because it has no non-comparable elements). It is called a *subchain.*

Example. The intervals $\mathrm{I}(a,b)$, $\mathrm{I}(a,b]$, $\mathrm{I}[a,b)$, $\mathrm{I}[a,b]$ as well as the sets $\mathbb{N}, \mathbb{Z}, \mathbb{Q}$ are subchains in the chain of all real numbers.

Example. The ordering of the class $\mathcal{U}$ by inclusion is not complete. The sets $\{a\}$ and $\{b\}$, where $a \neq b$, are incomparable.

Let M be a chain, let $m, n \in M$ be two elements such that $m < n$ and there is no $p \in M$ such that $m < p < n$. Then we say that the element n is the *successor* of the element m and that the element m is the *predecessor* of the element n and we write $m <\circ\, n$ and $n \circ\!> m$.

Exercise. Show that every element $a \in M$ has at most one predecessor and at most one successor.

4.3. Well-ordering

4.5. Definition. We say that an ordered class T is *well ordered* if every non-empty subclass $A \subset T$ has the least element.

4.6. Proposition. *Every well-ordered class A is totally ordered.*

Proof. Suppose not, then there exist incomparable elements a, b in A. Then the subset $\{a, b\} \subseteq A$ is nonempty and has no least element, which is a contradiction.

4.7. Proposition. *Every finite chain, including the empty chain, is well ordered.*

Proof. By induction on the length n of a chain $B_n = \{b_0 < b_1 < \cdots < b_{n-1}\}$. For $n = 0$ there is no nonempty subset, i.e., nothing to be checked.

Let the statement hold for n. Consider an $(n+1)$-element chain $B_{n+1} = \{b_0 < b_1 < \cdots < b_{n-1} < b_n\}$ and its nonempty subset A. If $A \cap B_n$ is non-empty, then it has the least element, which is also the least element in A. If $A \cap B_n$ is empty, then $A = \{b_n\}$, which has the least element b_n.

Examples. 1. Sets $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ are not well ordered, as they contain no least element.

2. The semi-closed interval $\mathrm{I}[0,1)$ is not a well-ordered set because its non-empty subset $\mathrm{I}(0,1)$ does not contain the least element.

4.8. Proposition. *The set of all natural numbers is well ordered.*

Proof. Let $A \subseteq \mathbb{N}$ be a non-empty subset, let $a \in A$ be an arbitrary element. Elements larger than a do not affect the existence of the least element, elements smaller or equal to a form a non-empty finite chain that has a least element, say m. Then m is the least element of A.

4.9. Proposition. *Every subset A of a well-ordered set X is well ordered under the induced ordering.*

Proof. Let $B \subseteq A$ be a non-empty subset. Then $B \subseteq X$ has the least element, which is also the least element of B as the subset of A with the induced ordering.

Recall that an element b of the ordered set is a *successor* of the element a, if $a < b$ and there is no c such that $a < c < b$.

4.10. Proposition. *Every element of a well-ordered class, except the greatest, has a successor.*

Proof. Let X be a well-ordered class, Let $a \in X$ be an element that is not the greatest element in X. Let $B = \{x \in X \mid a < x\}$. If class B were empty, we would have $x \geq a$ for all $x \in X$ (since X is a chain) and a would be the greatest element in X, which, by assumption, is not. Thus, B is non-empty, hence has the least element. Let us denote it by b. Obviously, $a < b$. Let us show that b is the successor of a. Suppose not, that is, there exists an element c such that $a < c < b$. Then $c \in B$, and therefore necessarily $b \leq c$, which is a contradiction.

As we already know, every element of a chain has at most one successor. If existing, the successor of an element a will be denoted $a^\bullet$.

A well-ordered set can be represented by a diagram in which the elements are ordered from left to right from smaller to larger and each element is connected by a line to its successor.

Examples. The five-element chain has the diagram

$$\circ\!-\!\!\!-\!\circ\!-\!\!\!-\!\circ\!-\!\!\!-\!\circ\!-\!\!\!-\!\circ.$$

The set of natural numbers has the diagram

$$\circ\!-\!\!\!-\!\circ\!-\!\!\!-\!\circ\!-\!\!\!-\!\circ\!-\!\!\!-\,\cdots,$$

where the three dots represent the countable repetition of the pattern $\circ\!-\!\!\!-$.

Exercise. Show that a well-ordered set cannot contain an infinite strictly decreasing sequence of elements $\cdots < a_2 < a_1 < a_0$.

4.4. Isotone mappings and isomorphisms

4.11. Definition. A mapping $f : X \to Y$ between ordered classes is called *isotone*, if $a \leq b$ implies $f(a) \leq f(b)$ for all $a, b \in X$.

A mapping $f : X \to Y$ mezi uspořádanými třídami se nazývá *strictly isotone*, if $a < b$ implies $f(a) < f(b)$ for all $a, b \in X$.

Example. The identical mapping $\mathrm{id}_X : X \to X$ is both isotone and strictly isotone for arbitrary ordering on X. The constant mapping $X \to 1$ (all on one) is isotone. It is not strictly isotone, if X has more than one element.

Exercise. Prove that an injective isotone mapping is strictly isotone.

If a bijective mapping is isotone, it does not necessarily mean that the inverse mapping is also isotone. The simplest counterexample is a mapping of a two-element set consisting of two incomparable elements into a two-element chain.

4.12. Definition. A bijective mapping $f : X \to Y$ between ordered classes is called *isomorphism*, if both f i f^{-1} are isotone. Two ordered classes X, Y are said to be *isomorphic*, if there is an isomorphism between them. We write $X \cong Y$.

Exercise. Every isomorphism $X \to Y$ is a strictly isotone map.

4.13. Proposition. *Every bijective isotone mapping $X \to Y$, where X is totally ordered, is an isomorphism.*

Proof. Let $f(a) < f(b)$. Then either $a < b$ or $a = b$ or $b < a$., since X is totally ordered. The second possibility contradicts injectivity, the third one contradicts isotonicity, and we are left with $a < b$.

4.5. Ordinal numbers

Isomorphic ordered sets are said to have the same *ordinal type*. Ordinal numbers can be defined as ordinal types of well-ordered sets. This means that two ordinal numbers are equal if and only if they are ordinal types of isomorphic well-ordered sets. We introduce special symbols for ordinal types analogously to cardinal numbers, using isomorphisms of well-ordered sets instead of bijections between sets.

An isomorphism of well-ordered sets is a bijection. Consequently, isomorphic well-ordered sets have the same cardinality. Hence, every ordinal number has a certain cardinality, and multiple different ordinals can be of the same cardinality.

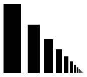
As with cardinal numbers, this kind of definition does not allow to introduce the class of ordinal numbers. This will be done later by using von Neumann's definition of ordinal numbers or ordinals, which is, however, not as illustrative as the definition using ordinal types.

Examples. 1. Every natural number n also denotes the ordinal type of a chain of n elements (we already know that it is well-ordered). Its diagram consists of n points connected by lines, e.g. $1, 2, 3$ have diagrams

$$\circ\,, \qquad \circ\!\!-\!\!\circ\,, \qquad \circ\!\!-\!\!\circ\!\!-\!\!\circ\,,$$

respectively.

2. Symbol ω denotes the ordinal type of the set of all natural numbers ordered by size (also a well-ordered set). We have given its diagram $\circ\!\!-\!\!\circ\!\!-\!\!\circ\!\!-\!\!\circ\!\!-\!\cdots$ above. Another suitable illustration of ω is the infinite row



of constituents (rods, columns, poles, bollards, milestones, pickets, or other objects) extending to the horizon.

Remark. It is instructive to view this `pdf` file at increasing resolution.

4.6. The arithmetic of ordinal numbers

Like cardinal numbers, ordinal numbers can be added, multiplied and exponentiated. We define the operations of addition and multiplication similarly to the operations of the same name with cardinal numbers, only we need to ensure that disjoint union and product of well-ordered sets are well ordered. None of the operations is commutative. Yet, exponentiation of ordinal numbers is completely different from exponentiation of cardinal numbers.

4.7. Ordinal addition

4.14. Definition. Consider two well-ordered sets A, B. The set $(\{0\} \times A) \cup (\{1\} \times B)$ ordered by

$$(i, a) \leq (j, b) \quad \Leftrightarrow \quad i < j \vee (i = j \wedge a \leq b)$$

is called the *disjoint union of well-ordered sets* or the *sum of well-ordered sets* A, B and is denoted by $A \sqcup B$ or $A + B$.

Elements of disjoint union are arranged

1. by the first components, assuming $0 < 1$;
2. or, in the case of equality of the first components, by the second components, using the ordering of the set where both elements are located.

Otherwise: Elements originating from A precede all elements originating from B, whereas elements originating from one and the same set are arranged in the same way as in “their” set.

Schematically,

$$A + B = \boxed{\text{A}}\!-\!\boxed{\text{B}}.$$

4.15. Definition. Let α and β be ordinal types of well-ordered sets A and B, respectively. We define the sum $\alpha + \beta$ as the ordinal type of the disjoint union $A \sqcup B$.

The definition is correct because if $A \cong A'$ and $B \cong B'$ are isomorphic well-ordered sets, then $A \sqcup B$, $A' \sqcup B'$ are isomorphic well-ordered sets. The proof of isomorphism is similar to the proof of equivalency of sets (if f, g are isomorphisms, then $f \sqcup g$ is an isomorphism). The proof of well-ordering can be found below even in a more general setting.

Example. What is $1 + 2$? We have

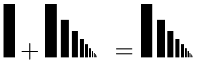
$$\circ \sqcup \circ\!\!-\!\!\circ = \boxed{\circ}\!-\!\boxed{\circ\!\!-\!\!\circ} = \circ\!\!-\!\!\circ\!\!-\!\!\circ,$$

and, therefore, $1 + 2 = 3$.

Example. What is $1+\omega$? We have

$$\circ \sqcup \circ - \circ - \cdots = \boxed{\circ} - \boxed{\circ - \circ - \cdots} = \circ - \circ - \cdots = \omega$$

or

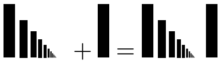


(one + row = row). Hence, $1+\omega=\omega$. Similarly $2+\omega=\omega$, etc.

Example. What is $\omega+1$? We have

$$\circ - \circ - \cdots \sqcup \circ = \boxed{\circ - \circ - \cdots} - \boxed{\circ} = \circ - \circ - \cdots \circ.$$

or

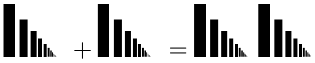


Thus, $\omega+1 \neq \omega$. Indeed, ω contains a unique element with no predecessor, while $\omega+1$ contains two such elements. Similarly $\omega+2 \neq \omega+1$, etc.

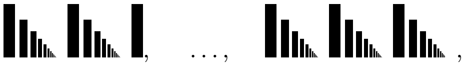
Example. What is $\omega+\omega$? We have

$$\begin{aligned} &\circ - \circ - \cdots \sqcup \circ - \circ - \cdots \\ &\quad = \boxed{\circ - \circ - \cdots} - \boxed{\circ - \circ - \cdots} \\ &\quad = \circ - \circ - \cdots \circ - \circ - \cdots \end{aligned}$$

or



(two rows). We see that $\omega+\omega$ is another ordinal number, different from those already constructed. Similarly $\omega+\omega+1$, …, $\omega+\omega+\omega$ are



respectively, i.e., are different new ordinal numbers.

The above examples show that, unlike cardinal addition, ordinal addition is not commutative. For example, $1+\omega=\omega\neq\omega+1$. Nonetheless, addition of ordinal numbers is associative.

Exercise. Prove associativity of addition.

We introduce a generalisation to systems. A well-ordered system of well-ordered sets is a system $\{A_i\}_{i\in I}$, where the individual summands A_i and the index set I are well ordered. The disjoint union

$$\bigsqcup_{i\in I} A_i = \bigcup_{i\in I}(\{i\}\times A_i)$$

will be ordered according to

$$(i,a_i)\le(j,a_j) \quad\Leftrightarrow\quad i<j \vee (i=j \wedge a_i \le a_j)$$

(here $a_i\in A_i$ a $a_j\in A_j$).

Again, elements of disjoint union are arranged

1. by the first components, assuming the ordering from the index set I;

2. or, in the case of equality of the first components, by the second components, using the ordering of the set where both elements are located.

Schematically,

$$\bigsqcup_{i\in I} A_i = \boxed{A_{i_1}}-\boxed{A_{i_2}}-\boxed{A_{i_3}}-,$$

with individual "blocks" ordered in the same way as their indices in the set I.

Clearly, the previously introduced disjoint union of two sets is a special case where the index set is the chain 0—1. The generalisation to an n-element chain is obvious — we get the sum of n sets.

4.16. Proposition. *The disjoint union of a well-ordered system of well-ordered sets is a well-ordered set.*

Proof. Let B be a non-empty subset in the disjoint union $\bigsqcup_{i\in I} A_i$. Let I_B be the set of all indices $i \in I$ such that $B \cap A_i \neq \emptyset$. Obviously, $I_B \neq \emptyset$ is a non-empty subset in a well-ordered set I, so it has the least element, which by definition is the least index of i such that $B \cap A_i \neq \emptyset$. But then $B \cap A_i$ is a nonempty subset of the well-ordered set A_i, so it has a least element that is also the least element of the whole subset B.

4.17. Definition. Let α_i be ordinal types of well-ordered sets A_i, $i \in I$, where I is well-ordered. We define the sum $\sum_{i\in I} \alpha_i$ as the ordinal type of a disjoint union $\bigsqcup_{i\in I} A_i$.

Example. We have

$$\sum_{i\in\omega} i = \omega,$$

as can be easily seen from the diagram

$$\boxed{\circ}-\boxed{\circ-\circ}-\boxed{\circ-\circ-\circ}-\cdots$$

4.8. Ordinal multiplication

4.18. Definition. Let A, B be two well-ordered sets. The set $A \times B$ with the ordering given by the formula

$$(a, b) \leq (a', b') \quad \Leftrightarrow \quad a < a' \vee (a = a' \wedge b \leq b')$$

is called the *lexicographic product* of well-ordered sets A, B and is denoted by $A \times B$.

In words: The elements of the lexicographic product are ordered

1. by the first components,
2. by the second components in the case of equal first components.

The origin of the attribute "lexicographic" can be found in the dictionary arrangement of two-letter words. This is illustrated in the examples below.

Otherwise said, A times B means that we replace the elements of the set A by copies of the set B; within the same copy we take the ordering from B, while between different copies we take the ordering from A. Schematically,

$$A \times B = \underbrace{\boxed{B}-\boxed{B}-\boxed{B}-\cdots}_{A}$$

Clearly, the product $A \times B$ is nothing but the disjoint sum

$$A \times B = \bigsqcup_A B$$

of a system consisting of copies of the set B, one for each element of A, and is therefore it is well ordered.

Example. Let us compute 3×3. We have

$$3 \times 3 = \left\{\begin{matrix} [0,0] & [0,1] & [0,2] \\ [1,0] & [1,1] & [1,2] \\ [2,0] & [2,1] & [2,2] \end{matrix}\right\}$$

with the lexicographical ordering

$$[0,0] < [0,1] < [0,2] < [1,0] < [1,1] < [1,2] < [2,0] < [2,1] < [2,2].$$

We see that it is nothing but the disjoint union

$$\boxed{[0,0]\text{—}[0,1]\text{—}[0,2]}\text{—}\boxed{[1,0]\text{—}[1,1]\text{—}[1,2]}\text{—}\boxed{[2,0]\text{—}[2,1]\text{—}[2,2]}.$$

Since the lexicographic product of two well-ordered sets is a special case of the sum of a system of sets, it is well-ordered.

Consider ordinal numbers α, β which are ordinal types of sets A, B, respectively. Define the lexicographic product $\alpha \times \beta$ as the ordinal type of the lexicographic product $A \times B$.

The meaning of lexicographic multiplication is "how many times what," that is, $\alpha \times \beta$ means "α-times β." Since lexicographic multiplication is generally non-commutative, the order matters.

4.19. Definition. Let A, B be two well-ordered sets. The set $A \times B$ with the ordering given by the formula

$$(a,b) \leq (a',b') \quad \Leftrightarrow \quad b < b' \vee (b = b' \wedge a \leq a')$$

is called the *anti-lexicographic product* of well-ordered sets A, B and is denoted by $A \cdot B$.

In words: The elements of the anti-lexicographic product are ordered

1. by the second components,
2. by the first components in the case of equal second components.

Obviously, the lexicographic product $A \times B$ is isomorphic to the anti-lexicographic product $B \cdot A$ via the bijection $(a,b) \longleftrightarrow (b,a)$.

Thus, the product $A \cdot B$ is identifiable the disjoint sum

$$A \cdot B = \bigsqcup_B A$$

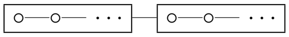
or, schematically,

$$A \cdot B = \underbrace{\boxed{A}\text{—}\boxed{A}\text{—}\boxed{A}\text{—}\cdots}_{B}$$

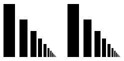
In terms of ordinal numbers, we get the reverse multiplication $\beta \cdot \alpha = \alpha \times \beta$, which will be read "$\alpha$ β-times."

In comparison to the lexicographic product, the anti-lexicographic product better combines with exponentiation, which is its "raison d'être," see below.

Example. What is $\omega \cdot 2 = \bigsqcup_2 \omega = 2 \times \omega$? We get

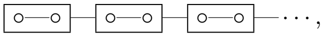


which is $\omega + \omega$ or

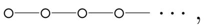


(two infinite rows one after another).

In contrast, $2 \cdot \omega = \bigsqcup_\omega 2 = \omega \times 2$ is

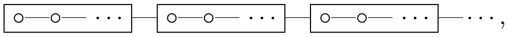


or

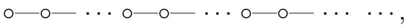


which is ω. We see that $\omega \cdot 2 \neq \omega = 2 \cdot \omega$.

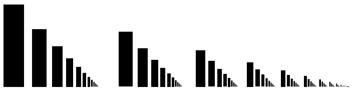
Example. What is $\omega \cdot \omega = \omega \times \omega = \bigsqcup_\omega \omega$? We get



or



or



(an infinite row of infinite rows, or a doubly infinite row).

We get the same result if we replace each rod with an infinite row in . Observe that ω elements have no predecessors.

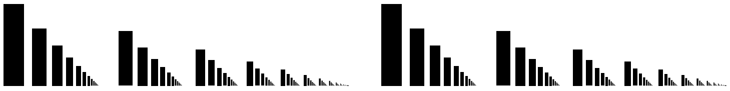
The product $\omega \cdot \omega$ can be written also as ω^2.

Exercise. What is $1 + \omega^2$? What is $\omega + \omega^2$?

Exercise. Show that each element of ω^2 is of the form $\omega \cdot a_1 + a_0$, where a_1, a_0 are natural numbers.

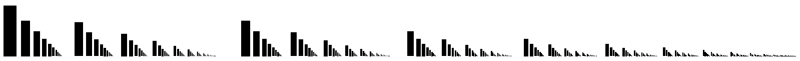
Hint. Cut the previous picture at some point.

Example. Twice the ordinal number ω^2 is $\omega^2 \cdot 2 = 2 \times \omega^2$, which is



(two infinite rows of infinite rows). How many elements have no predecessors? Similarly for higher multiples.

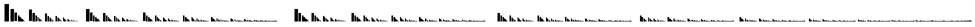
Example. Let us define ω^3 as ω times the ordinal number ω^2, i.e.,



(triply infinite row). We get the same result when we multiply ω^2 times ω. Note that ω^2 elements have no predecessors.

Exercise. Show that each element of ω^3 is of the form $\omega^2 \cdot a_2 + \omega \cdot a_1 + a_0$, where a_2, a_1, a_0 are natural numbers.

Example. Analogously, ω^4 is



(quadruply infinite row). In this case, ω^3 elements have no predecessor.

Exercise. Show that each element of ω^4 is of the form $\omega^3 \cdot a_3 + \omega^2 \cdot a_2 + \omega \cdot a_1 + a_0$, where a_3, a_2, a_1, a_0 are natural numbers.

Not all laws of usual arithmetic apply to ordinal numbers. We have already seen that the commutative law is generally invalid for addition and multiplication. The full list of valid identities follows.

4.20. Proposition. *Identities*

$$\begin{aligned}
&0 + \alpha = \alpha = \alpha + 0,\\
&(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma),\\
&1 \cdot \alpha = \alpha = \alpha \cdot 1,\\
&(\alpha \cdot \beta) \cdot \gamma = \alpha \cdot (\beta \cdot \gamma),\\
&\alpha \cdot (\beta + \gamma) = \alpha \cdot \beta + \alpha \cdot \gamma,\\
&0 \cdot \alpha = 0 = \alpha \cdot 0,
\end{aligned}$$

hold for arbitrary ordinal numbers α, β, γ.

Proof. A proof by transfinite induction will be given later.

4.21. Definition. Given a polynomial $p = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 \in \mathbb{N}[x]$ of a single indeterminate x with coefficients from $\mathbb{N}$, we can introduce the ordinal number

$$p(\omega) = \omega^n \cdot a_n + \omega^{n-1} \cdot a_{n-1} + \cdots + \omega \cdot a_1 + a_0.$$

The order of operations must be obeyed!

4.22. Proposition. *The numbers $p(\omega)$ exhaust the ordinal numbers expressible as finite sums and products of natural numbers and ω.*

4.9. Ordinal exponentiation

Unlike the sums, neither infinite products nor infinite powers carry a natural well-ordering. For example, $2^{\aleph_0}$, the set of countable sequences of numbers $0, 1$, does admit a lexicographic ordering, but we can easily find an infinite decreasing sequence

$$[1,0,0,0,0,\dots] \quad > \quad [0,1,0,0,0,\dots] \quad > \quad [0,0,1,0,0,\dots], \quad \dots$$

Infinite products and powers must be constructed in a different way.

4.23. Definition. Let α and β be ordinal types of well-ordered sets A and B, respectively. Then α^β is defined to be the ordinal type of

$$\bigcup_{k \in B} A^k,$$

where $\bigcup A^k$ is the plain, not disjoint, union of the sets A^k of type α^k. The finite powers

$$A^k = \underbrace{A \times \cdots \times A}_{k},$$

are embedded one in another in such a way that A^k is identified with $\{0\} \times A^k \subset A^{k+1}$, where 0 denotes the least element of A.

Example. Let us show that

$$2^\omega = \bigcup_{k\in\omega} 2^k = \omega.$$

Here 2^0 is the one-element set $1 = \{0\}$, embedded into $2^1 = \{[0], [1]\}$ as the subset $\{[0]\}$. Next, $2^1 = \{[0], [1]\}$ is embedded into 2^2 as $\{[0, 0], [0, 1]\}$, while 2^2 is embedded into 2^3 as $\{[0, 0, 0], [0, 0, 1], [0, 1, 0], [0, 1, 1]\}$, etc. We see that the union $\bigcup_{k\in\omega} 2^k$ can be identified with the set of all sequences $\dots, a_2, a_i, a_0$ infinite to the left, where finitely many elements a_i are 1, while the others are 0. The ordering remains lexicographic, so one decides according to the leftmost $a_i = 1$. We get

$$(\dots, 0, 0, 0) < (\dots, 0, 0, 1) < (\dots, 0, 1, 0) < (\dots, 0, 1, 1) <\cdots,$$

which is isomorphic to the chain ω, i.e.,

$$2^\omega = \omega.$$

Remark. Note that contrary to the cardinal number $2^{\aleph_0}$, the ordinal number 2^ω is *countable.*

Example. Let us show that

$$1^\omega = \bigcup_{k\in\omega} 1^k = 1.$$

Since $1 \subset 2$, we can use the procedure from the previous example, restricting ourselves to sequences composed of the unique element $0 \in 1$. However, there is only one such sequence, namely $(\dots, 0, 0, 0, 0)$.

Example. Let us compute ω^ω. Again, we can use the procedure from the previous example, but this time the the sequences $\dots, a_2, a_2, a_0$ are constructed from all natural numbers, while everything else remains the same, including the ordering. We get

$$\begin{aligned}
&(\dots, 0, 0, 0, 0) < (\dots, 0, 0, 0, 1) < (\dots, 0, 0, 0, 2) < \cdots\\
&< (\dots, 0, 0, 1, 0) < (\dots, 0, 0, 1, 1) < (\dots, 0, 0, 1, 2) < \cdots\\
&< (\dots, 0, 0, 2, 0) < (\dots, 0, 0, 2, 1) < (\dots, 0, 0, 2, 2) < \cdots\\
&< \cdots\\
&< (\dots, 0, 1, 2, 0) < (\dots, 0, 1, 2, 1) < (\dots, 0, 1, 2, 2) < \cdots,
\end{aligned}$$

which is a completely new ordinal number, because all ordinal numbers $p(\omega) = \omega^n \cdot c_n + \omega^{n-1} \cdot c_{n-1} + \cdots + \omega \cdot c_1 + c_0$ constructed so far correspond to finite sequences $a_n, \dots, a_2, a_2, a_0$.

The diagram of the set ω^ω is obtained when in the set ω we replace each constituent by the set ω (we get ω^2), then again replace each constituent with the set ω (resulting in ω^3), and so on ad infinitum. This produces a self-similar set, each part of which is similar to the whole, i.e., a fractal.

Example. A similar pattern can be used when constructing the number ω^{ω^ω}, only the sequences used will be indexed by the elements of ω^ω. Analogously $\omega^{\omega^{\omega^\omega}}$, etc.

Remark. All ordinal numbers ω^ω, ω^{ω^ω}, $\omega^{\omega^{\omega^\omega}}, \ldots,$ are *countable.*

Here we finish our excursion into the zoo of ordinal numbers. We see that the world of ordinal numbers is much more diverse than the world of cardinal numbers. Let us emphasise that all ordinal numbers constructed so far have a countable cardinality.

4.24. Proposition. *For arbitrarary ordinal numbers α, β, γ, we have*

$$\alpha^0 = 1,$$
$$\alpha^1 = \alpha.$$
$$\alpha^{\beta+\gamma} = \alpha^\beta \cdot \alpha^\gamma,$$
$$(\alpha^\beta)^\gamma = \alpha^{\beta\cdot\gamma}.$$

Proof. A proof by transfinite induction will be given later.

In terms of the lexicographic product, we would have $\alpha^{\beta+\gamma} = \alpha^\gamma \times \alpha^\beta$ and $(\alpha^\beta)^\gamma = \alpha^{\gamma\times\beta}$, the order of β and γ being reversed.

Exercise. Show that $(\beta+\gamma)\cdot\alpha$ differs from $\beta\cdot\alpha+\gamma\cdot\alpha$ in general. Thus only one of the distributive laws between multiplication and addition holds.
Hint: Choose $\alpha = 2$, $\beta = \omega$, $\gamma = 1$.

Exercise. Show that $(\alpha\cdot\beta)^\gamma$ can differ from both $\alpha^\gamma\cdot\beta^\gamma$ and $\beta^\gamma\cdot\alpha^\gamma$.
Hint: Choose $\alpha = \omega$, $\beta = \gamma = 2$.

4.10. Transfinite induction

The principle of transfinite induction enables us to prove statements of the form

$$(\forall x \in X)\, \phi(x),$$

where X is a well-ordered class. The induction step consists in proving that the validity of $\phi(x)$ for all $x < a$ implies the validity of $\phi(a)$.

4.25. Proposition. *Let $\phi(x)$ be a formula, let X be a well-ordered class. Let $(\forall x < a)\,\phi(x)$ imply $\phi(a)$, for all $a \in X$. Then $\phi(x)$ holds for all $x \in X$.*

Proof. Let $A \subseteq X$ denote the subclass of elements x such that $\phi(x)$ does not hold. Suppose that $A \neq \emptyset$. Let a be the least element in A, i.e., the least element for which ϕ does not hold. For all elements $x < a$ (if existing) then $\phi(x)$ does hold. By assumption, $\phi(a)$ holds as well, which contradicts $a \in A$. Therefore, $A = \emptyset$ and $\phi(x)$ holds for all $x \in X$.

It may seem that $\phi(x)$ need not hold for the least element a of the class X. The opposite is true, for the class $\{x \in X \mid x < a\}$ is empty in this case, and therefore the condition $(\forall x < a)\,\phi(x)$ holds.

The principle of transfinite induction also allows us to construct sets F_x dependent on an element of a well-ordered class X. To do this, it is sufficient to prove a formula of the form

$$(\forall x \in X)(\exists F_x \in \mathcal{U})\, \phi(x, F_x).$$

In this case we speak of a transfinite recursive construction of a class F_x.

5. Zermelo's theorem and its consequences

Zermelo's theorem says that every set can be well ordered. This is a very important theorem, the corollary of which is that transfinite induction is applicable to every set.

5.1. Zermelo's theorem. *There exists a well ordering on every set.*

Ernst Zermelo (1871–1953) proved this theorem first in the paper E. Zermelo, Beweis, daß jede Menge wohlgeordnet werden kann, *Math. Annalen* 59 (1904) 514–516.

Several proofs are known, all of which use the axiom of choice, which is necessary since the axiom is a consequence of Zermelo's theorem. In fact, given a well-ordered set, one can select the smallest element from each nonempty subset, which yields a selection mapping.

Exercise. Find a well-ordering of the set $\mathbb{Z}$ of integers.
Hint: Try the ordinal type $\omega + \omega$.

Exercise. Find a well-ordering of the set $\mathbb{Q}$ of rationals.
Hint: Try the ordinal type $\omega \times \omega$.

5.2. Remark. According to Zermelo's theorem, there is a well ordering of the set of real numbers. However, there is no known formula that would specify such an ordering. There is also no known formula that would specify a selection mapping $\wp\mathbb{R} \setminus \{\emptyset\} \to \mathbb{R}$, i.e., which would select of one element from each non-empty set of real numbers.

Before proving Zermelo's theorem, we present several its consequences, which are important in diverse branches of mathematics.

5.1. Hausdorff's principle

5.3. Hausdorff's maximality principle. *Every chain in an ordered set can be extended to a maximal chain.*

A maximal chain is a chain to which no element can be added (that is, it ceases to be a chain when any element is added). In the proof we use Zermelo's theorem.

Proof. Let $(A, \le)$ be any ordered set, B its subchain. By Zermelo's theorem, there exists a well-ordering on the set A, which we denote by $\preceq$. Using transfinite recursion, we successively add elements of A to B as long as the result remains a chain. More exactly, we define a map $C : A \to \wp A$ by

$$C(a) = \begin{cases} B & \text{if } a \text{ is the least element of } A \text{ with respect to } \le, \\ \{a\} \cup \bigcup\limits_{b \prec a} C(b), & \text{if } \{a\} \cup \bigcup\limits_{b \prec a} C(b) \text{ is a chain with respect to } \le, \\ \bigcup\limits_{b \prec a} C(b) & \text{otherwise.} \end{cases}$$

Then

$$C = \bigcup_{a \in A} C(b)$$

is the maximal chain sought, because it contains B, and if $c \in A$ is an element comparable with all elements of C, then it is already in C, since it had to be added during the transfinite recursion.

5.2. Zorn's lemma

Zorn's lemma is a tool for construction of maximal elements in various parts of mathematics. The lemma was published by the German mathematician Max Zorn in 1935. A slightly weaker version was published by the Polish mathematician Kazimierz Kuratowski in 1922.

5.4. Zorn's lemma. *Let A be an ordered set such that every chain in A has an upper bound. Then for every element $a \in A$ there is a maximal element $b \in A$, such that $a \leq b$.*

Zorn's lemma is an easy consequence of Hausdorff's principle.

Proof. According to the Hausdorff principle, the one-element chain $\{a\}$ can be extended to a maximal chain I that contains it. By assumption, I has an upper bound b. Obviously, b is the maximal element sought.

Zorn's lemma is often applied in situations where A is a system of sets (e.g., algebraic structures) for which the union of a chain of of nested sets is again an element of the system (in this form the lemma was proved by K. Kuratowski in 1922). It can thus be shown, for example, that every proper subalgebra can be embedded in a maximal proper subalgebra.

5.3. Initial segments and ordering

5.5. Definition. Let A be a chain. A subset $I \subseteq A$ is called an *initial segment* in A, if $b \in I$ whenever $b \in A$ and $b \leq a$ for some $a \in I$.

An initial segment I is said to be *proper* if $I \neq A$.

Example. Every chain is its own initial segment. The empty set is a proper initial segment of every nonempty chain.

Example. $\{0, 1, 2\}$ is an initial segment of $\mathbb{N}$.

Exercise. Prove that an initial segment $K \subseteq J$ of an initial segment $J \subseteq I$ is an initial segment $K \subseteq I$.

Exercise. 1. If $\{I_j\}_{j \in J}$ is a system of initial segments in a chain A, then the union $\bigcup_{j \in J} I_j$ is also an initial segment in A. Prove.

2. If $\{I_j\}_{j \in J}$ is a system of initial segments in a chain A, then the intersection $\bigcap_{j \in J} I_j$ is also an initial segment in A. Prove.

5.6. Proposition. *If I, I' are two initial segments in a chain A, then $I \subseteq I'$ or $I' \subseteq I$.*

Proof. Suppose, on the contrary, that both $I \backslash I'$ and $I' \backslash I$ are non-empty and contain elements i and i', respectively. Since A is a chain, i and i' are comparable and, therefore, $i \leq i'$ or $i' \leq i$. However, I, I' are initial segments, whence $i \in I'$ or $i' \in I$, which is impossible because of the way i and i' were introduced. Thus, at least one of $I \setminus I'$ and $I' \setminus I$ is empty, which means that $I \subseteq I'$ or $I' \subseteq I$.

From now on A will be a well-ordered set. For any element of $a \in A$, we introduce

$$A_a = \{x \in A \mid x < a\}.$$

The set A_a is obviously a proper initial segment in A.

5.7. Proposition. *Let $I \subset A$ be a proper initial segment. Then there exists an element $a \in A$ such that*

$$I = A_a.$$

Proof. If I is a proper initial segment, then the difference $A \setminus I$ is non-empty. Hence, $A \setminus I$ contains the smallest element. Let us denote it by a. Then $I = A_a$.

Example. The proper initial segments in ω coincide with the finite ordinal numbers.

5.8. Theorem (trichotomy). *Let A, B be well-ordered sets. Then one of the following three possibilities takes place:*

1. *A is isomorphic to B;*
2. *A is isomorphic to a proper initial segment in B;*
3. *a proper initial segment in A is isomorphic to B.*

The idea is to assign the smallest element that does not have an image to the smallest element that does not have a preimage. A formal proof can go as follows.

Proof. Consider the set $\mathcal{T}$ of all triples (I, ϕ, J) such that I is an initial segment in A, J is an initial segment in B, and $\phi : I \to J$ is an isomorphism. Let

$$\mathcal{I} = \bigcup_{(I,\phi,J)\in\mathcal{T}} I, \quad \text{and} \quad \mathcal{J} = \bigcup_{(I,\phi,J)\in\mathcal{T}} J.$$

be, respectively, the union of all initial segments I and J that correspond to triples $(I, \phi, J) \in \mathcal{T}$. Then, $\mathcal{I}$ is an initial segment in A, and $\mathcal{J}$ is an initial segment in B (see Exercise 1 before Proposition 6.6).

Let (I, ϕ, J), (I', ϕ', J') belong to $\mathcal{T}$. Let us prove that $\phi|_{I\cap I'} = \phi'|_{I\cap I'}$. Consider an arbitrary element $i_0 \in I \cap I'$. Denote $j_0 = \phi(i_0)$ and $j_0' = \phi'(i_0)$. Both j_0 and j_0' belong to B. Without loss in generality, $j_0 \leq j_0'$. If $j_0 < j_0'$, then we get a contradiction as follows.

Let $i_1 = \phi'^{-1}(j_0)$, $j_1 = \phi(i_1)$. Then $i_1 < i_0$, since $\phi'(i_1) = j_0 < j_0' = \phi'(i_0)$. Next, let $i_2 = \phi'^{-1}(j_1)$, $j_2 = \phi(i_2)$. Then $i_2 < i_1$, since $\phi'(i_2) = j_1 < j_0 = \phi'(i_1)$. This can be repeated infinitely, leading to an infinite strictly decreasing sequence $i_0 > i_1 > i_2 > \cdots$, which is impossible, since $\{i_0, i_1, i_2, \dots\}$ would be a nonempty subset with no least element.

$$\begin{tikzcd}
 & j_0' \\
i_0 \arrow[r] \arrow[ur] & j_0 \\
i_1 \arrow[r] \arrow[ur] & j_1 \\
i_2 \arrow[r] \arrow[ur] & j_2 \\
i_3 \arrow[r] \arrow[ur] & j_3 \\
 \vdots &
\end{tikzcd}$$

Thus, $j_0 = j_0'$, i.e., $\phi(i_0) = \phi'(i_0)$, which shows that $\phi|_{I\cap I'} = \phi'|_{I\cap I'}$. This means that every element of $i \in \mathcal{I}$ has exactly one image $\phi(i) \in \mathcal{J}$.

Let us prove that $\mathcal{I} = A$ or $\mathcal{J} = B$. Assume, on the contrary, that both $A \setminus \mathcal{I}$ and $B \setminus \mathcal{J}$ are non-empty. Let $a \in A \setminus \mathcal{I}$, $b \in B \setminus \mathcal{J}$ be the least elements. Denote $\mathcal{I}^\bullet = \mathcal{I} \cup \{a\}$ and $\mathcal{J}^\bullet = \mathcal{J} \cup \{b\}$ (the successors of $\mathcal{I}$ and $\mathcal{J}$ with respect to $\subseteq$). Then the isomorphism $\phi : \mathcal{I} \to \mathcal{J}$ can be extended to $\phi : \mathcal{I}^\bullet \to \mathcal{J}^\bullet$ by $\phi(a) = b$. This, however, contradicts the definition of $\mathcal{I}$ and $\mathcal{J}$. This proves that at least one of statements 1–3 holds.

5.9. Lemma. *Let A be a well-ordered set, let $\phi : A \to A$ be a strictly isotone map. Then $a \le \phi(a)$ for all $a \in A$.*

Proof. If not $a \le \phi(a)$, then $a > \phi(a)$. Denote $B = \{a \in A \mid a > \phi(a)\}$. Assume that A is not empty. Let $b \in B$ be the least element. Since $\phi(b) < b$ and ϕ is strictly isotone, we have $\phi(\phi(b)) < \phi(b)$, whence $\phi(b) \in B$. This contradicts the fact that b is the least element in B.

Example. The mapping $\phi : \mathbb{N} \to \mathbb{N}$, $\phi(n) = n + 1$, is strictly isotone and $n < \phi(n)$.

5.10. Lemma. *Let A be a well-ordered set. Then there is no strictly isotone mapping $\phi : A \to A_a$, where $a \in A$.*

Proof. Assume that a strictly isotone mapping $\phi : A \to A_a$ exists. Then $\phi(a) \in A_a$, whence $\phi(a) < a$. This contradicts the last lemma.

5.11. Corollary. *No well-ordered set A is isomorphic to a subset of a proper initial segment of A.*

Proof. Let $\phi : A \to \phi A$ be an isomorphism such that ϕA is a subset of a proper initial segment in A. By Proposition 6.7, there exists $a \in A$ such that $\phi A \subseteq A_a$. In particular, $\phi(a) \in A_a$. Hence, $\phi(a) < a$, which contradicts Lemma 6.9.

It easily follows that no two statements in Theorem 6.8 can be true simultaneously.

5.4. Ordering of ordinal numbers

5.12. Definition. Let α and β be ordinal types of well-ordered sets A and B, respectively. Let

$$\alpha = \beta \quad \text{or} \quad \alpha < \beta \quad \text{or} \quad \alpha > \beta$$

according to whether
A is isomorphic to B or
A is isomorphic to a proper initial segment in B or
B is isomorphic to a proper initial segment in A,
respectively.

5.13. Proposition. *The above definition introduces a total ordering between ordinal numbers.*

Proof. Reflexivity and transitivity are obvious. Antisymmetry then follows from Corollary 6.11. Theorem 6.8 implies that one of the possibilities $\alpha = \beta$, $\alpha < \beta$, $\alpha > \beta$ always occurs.

5.5. Comparability of cardinal numbers

It can be derived from Zermelo's theorem (which we have not proved yet) that every two cardinal numbers are comparable.

5.14. Proposition. *Let $\mathfrak{a}, \mathfrak{b}$ be cardinal numbers. Then $\mathfrak{a} = \mathfrak{b}$ or $\mathfrak{a} < \mathfrak{b}$ or $\mathfrak{a} > \mathfrak{b}$.*

Proof. Consider sets A and B of cardinalities $\mathfrak{a}$ and $\mathfrak{b}$, respectively. Both A, B can be well-ordered by Zermelo's theorem, and then by Theorem 6.8 there is an injective mapping $A \to B$ or $B \to A$.