Undergraduate logic sequence: the notes

November 21, 2014

Contents

Chapter 1

Zermelo–Fraenkel set theory

1.1 Historical context

In 19th century, mathematicians produced a great number of sophisticated theorems and proofs. With the increasing sophistication of their techniques, an important question appeared now and again: which theorems require a proof and which facts are self-evident to a degree that no sensible mathematical proof of them is possible? What are the proper boundaries of mathematical discourse? The contents of these questions is best illustrated on several contemporary examples.

Example 1.1.1. The parallel postulate of Euclidean geometry was a subject of study for centuries. The study of geometries that fail this postulate was considered a non-mathematical folly prior to early 19th century, and Gauss for example withheld his findings in this direction for fear of public reaction. The hyperbolic geometry was discovered only in 1830 by Lobachevsky and Bolyai. Non-Euclidean geometries proved to be an indispensable tool in general theory of relativity, for example.

Example 1.1.2. Jordan curve theorem asserts that every non-self-intersecting closed curve divides the Euclidean plane into two regions, one bounded and the other unbounded, and any path from the bounded to the unbounded region must intersect the curve. The proof was first presented in 1887. The statement sounds self-evident, and the initial proofs were found confusing and unsatisfactory. The consensus formed that even statements of this kind must be proved from some more elementary properties of the real line.

Example 1.1.3. Georg Cantor produced an exceptionally simple proof of existence of non-algebraic real numbers, i.e. real numbers which are not roots of any polynomial with integer coefficients (1874). Proving that specific real

numbers such as π or e are not algebraic is quite difficult, and the techniques for such proofs were under development at that time. On the other hand, Cantor only compared the cardinalities of the sets of algebraic numbers and real numbers, found that the first has smaller cardinality, and concluded that there must be real numbers that are not algebraic without ever producing a single definite example. Cantor's methodology–comparing cardinalities of different infinite sets–struck many people as non-mathematical.

As a result, the mathematical community in late 19th century experienced an almost universally acknowledged need for an axiomatic development of mathematics modeled after classical Euclid's axiomatic treatment of geometry. It was understood that the primitive concept will be that of a set (as opposed to a real number, for example), since the treatment of real numbers can be fairly easily reinterpreted as speaking about sets of a certain specific kind. The need for a careful choice of axioms was accentuated by several paradoxes, of which the simplest and most famous is the *Russell's paradox*: consider the "set" x of all sets z which are not elements of themselves. Consider the question whether $x \in x$ or not. If $x \in x$ then x does not satisfy the formula used to form x, and so $x \notin x$. On the other hand, if $x \notin x$ then x does satisfy the formula used to form x, and so $x \in x$. In both cases, a contradiction appears. Thus, the axiomatization must be formulated in a way that avoids this paradox.

Several attempts at a suitable axiomatization appeared before Zermelo produced his collection of axioms in 1908, now known as Zermelo set theory with choice (ZC). After a protracted discussion and two late additions, the axiomatization of set theory stabilized in the 1920's in the form now known as Zermelo–Fraenkel set theory with the Axiom of Choice (ZFC). This process finally placed mathematics on a strictly formal foundation. A mathematical statement is one that can be faithfully represented as a formula in the language of set theory. A correct mathematical argument is one that can be rewritten as a formal proof from the axioms of ZFC. Here (roughly), a formal proof of a formula ϕ from the axioms is a finite sequence of formulas ending with ϕ such that each formula on the sequence is either one of the axioms or follows from the previous formulas on the sequence using a fixed collection of formal derivation rules.

The existence of such a formal foundation does not mean that mathematicians actually bother to strictly conform to it. Russell's and Whitehead's Principia Mathematica [9] was a thorough attempt to rewrite many mathematical arguments in a formal way, using a theory different from ZFC. It showed among other things that a purely formal treatment is excessively tiresome and adds very little insight. Long, strictly formal proofs of mathematical theorems of any importance have been produced only after the advent of computers. Mathematicians still far prefer to verify their argument by social means, such as by presentations at seminars or conferences or in publications. The existence of a strictly formal proof is considered as an afterthought, and a mechanical consequence of the existence of a proof that conforms to the present socially defined standards of rigor. In this treatment, we will also produce non-formal rigorous proofs in ZFC with the hope that the reader can accept them and learn to

emulate them.

1.2 The language of the theory

Zermelo–Fraenkel set theory with the Axiom of Choice (ZFC) belongs to the class of theories known as first order theories. General first order theories will be investigated in Chapter 5. Here, we only look at the special case of ZFC. The language of ZFC consists of the following symbols:

- an infinite supply of variables;
- a complete supply of logical connectives. We will use implication $\rightarrow$, conjunction $\wedge$, disjunction $\vee$, equivalence $\leftrightarrow$, and negation $\neg$;
- quantifiers. We will use both universal quantifier $\forall$ (read "for all") and existential quantifier $\exists$ (read "there exists");
- equality $=$;
- special symbols. In the case of ZFC, there is only one special symbol, the binary relational symbol $\in$ (membership; read "belongs to", "is an element of").

The symbols of the language can be used in prescribed ways to form expressions–formulas. In the case of ZFC, if x, y are variables then $x = y$ and $x \in y$ are formulas; if ϕ, ψ are formulas then so are $\phi \wedge \psi$, $\neg\phi$, etc.; and if ϕ is a formula and x is a variable then $\forall x\ \phi$ and $\exists x\ \phi$ are formulas; in these formulas, ϕ is called the scope of the quantifier $\forall x$ or $\exists x$. Formulas are customarily denoted by Greek letters such as ϕ, ψ, θ. A variable x is free in a formula ϕ if it appears in ϕ outside of scope of any quantifier. Often, the free variables of a formula are listed in parentheses: $\phi(x)$, $\psi(x, y)$. A formula with no free variables is called a sentence.

Even quite short formulas in this rudimentary language tend to become entirely unreadable. To help understanding, mathematicians use a great number of shorthands, which are definitions of certain objects or relations among them. Among the most common shorthands in ZFC are the following:

- $\forall x \in y\ \phi$ is a shorthand for $\forall x\ x \in y \rightarrow \phi$, and $\exists x \in y\ \phi$ is a shorthand for $\exists x\ x \in y \wedge \phi$;
- $\exists! x\ \phi$ is short for "there exists exactly one", in other words for $\exists x(\phi\ \wedge \forall y \phi(y) \rightarrow y = x)$;
- $x \subseteq y$ (subset) is short for $\forall z\ z \in x \rightarrow z \in y$;
- 0 is the shorthand for the empty set (the unique set with no elements);
- $x \cup y$ and $x \cap y$ denote the union and intersection of sets x, y;

- $\mathcal{P}(x)$ denotes the powerset of x, the set of all its subsets.

After the development of functions, arithmetical operations, real numbers etc. more shorthands appear, including the familiar $\mathbb{R}$, $+$, $\sin x$, $\int f(x)dx$ and so on. Any formal proof in ZFC using these shorthands can be mechanically rewritten into a form which does not use them. Since the shorthands really do make proofs shorter and easier to understand, we will use them whenever convenient.

1.3 The most basic axioms

At the basis of any first order theory, there is a body of axioms known as the *logical axioms*. They record the behavior of the underlying logic and have nothing to do with the theory per se. The choice of logical axioms depends on the precise definition of the formal proof system one wants to use. They are typically statements like the following: $\forall x\ x = x$, $\forall x \forall y \forall z\ (x = y \wedge y = z) \to x = z$, or $\phi \to (\psi \to \phi)$ for any formulas ϕ, ψ. The possible choices for the system of the logical axioms are discussed in Chapter 5; we will not explain them here.

Move on to the axioms specific to ZFC set theory.

Definition 1.3.1. The *Empty Set Axiom* asserts $\exists x \forall y\ y \notin x$.

It would be just as good to assert the existence of any set, $\exists x\ x = x$. The existence of the empty set would then follow from Comprehension below. We do need to assert though that the universe of our theory contains some objects.

Definition 1.3.2. The *Extensionality Axiom* states that $\forall x \forall y\ (\forall z\ z \in x \leftrightarrow z \in y) \to x = y$.

In other words, two sets with the same elements are equal. Restated again, a set is determined by its elements. In particular, there can be only one set containing no elements and we will denote it by 0.

Definition 1.3.3. The *Pairing Axiom* says that $\forall x \forall y \exists z \forall u\ u \in z \leftrightarrow (u = x \vee u = y)$.

In other words, given x, y one can form the pair $\{x, y\}$. This is our first use of the *set builder notation*. Note that if $x = y$, we are getting a singleton set $\{x\}$. The pair $\{x, y\}$ is unordered: looking at it, we cannot tell whether x comes first and y second or vice versa. We will also use the ordered pair $\langle x, y\rangle$, which (a definition of Sierpinski) is the set $\{\{x\}, \{x, y\}\}$. This can be formed using the pairing axiom several times. Ordered triples would be defined as $\langle x, y, z\rangle = \langle\langle x, y\rangle, z\rangle$, similarly to ordered n-tuples for every natural number n.

Definition 1.3.4. The *Union Axiom* asserts that $\forall x \exists y \forall z\ (z \in y \leftrightarrow \exists u\ u \in x \wedge z \in u)$.

In other words, for every set x (note that elements of x are again sets as in our discourse everything is a set) one can form the union y of all elements of x. The notation commonly used is $y = \bigcup x$. Note that the union of x is uniquely given by this description by the Axiom of Extensionality.T

he union and pairing axioms make it possible to formulate several operations on sets. If x, y are sets then there is a set (denoted by $x \cup y$) containing exactly elements of x and elements of y: $x \cup y = \bigcup\{x, y\}$. Given a finite list $x_0, x_1, x_2, \ldots x_n$ of sets, we can form the set $\{x_0, x_1, x_2, \ldots x_n\}$, which is the unique set containing exactly the sets on our list.

Definition 1.3.5. The *Powerset Axiom* asserts that $\forall x \exists y \forall z\ z \in y \leftrightarrow z \subseteq y$.

In other words, for every set x there is a set consisting exactly of all subsets of x. This set is called the *powerset* of x and commonly denoted by $\mathcal{P}(x)$. Note that the powerset of x is uniquely given by this description by the Axiom of Extensionality.

Exercise 1.3.1. Does $0 = \{0\}$ hold? Why?

Exercise 1.3.2. Evaluate $\bigcup\{0, \{0\}, \{0, \{0\}\}\}$ in the set builder notation.

Exercise 1.3.3. Write a formula $\phi(x, y, z)$ of set theory which says the following: "z is an ordered pair whose first element is in x and the second element is in y".

Exercise 1.3.4. Evaluate $\mathcal{P}(\{0, \{0\}\})$ in the set builder notation.

1.4 Axiom schema of Comprehension

Also known as Separation or Collection. It is in fact an infinite collection of axioms, with one instance for each formula ϕ of set theory.

Definition 1.4.1. Let ϕ be a formula of set theory with $n+1$ free variables for some natural number n. The instance of the *Axiom schema of Comprehension* associated with ϕ is the following statement. $\forall x \forall u_0 \forall u_1 \ldots \forall u_{n-1} \exists y \forall z\ z \in y \leftrightarrow z \in x \wedge \phi(z, u_0, \ldots u_{n-1})$.

We will use this axiom schema tacitly whenever we define sets using the *set builder notation*: $y = \{z \in x \colon \phi(x, u_0, u_1, \ldots u_n)\}$. Comprehension makes it possible to form a great number of new sets. Given sets x, y, we can form the intersection $x \cap y = \{z \in x \colon z \in y\}$, the set difference $x \setminus y = \{z \in x \colon z \notin y\}$ as well as the symmetric difference $x \Delta y = (x \setminus y) \cup (y \setminus x)$. Given a set x of sets, we can form the intersection of all sets in x: $\bigcap x = \{z \in \bigcup x \colon \forall y \in x\ z \in y\}$.

We will use a similar process to form what we call classes:

Definition 1.4.2. A *class* is a collection C of sets such that there is a formula ϕ of $n+1$ variables, and sets $u_0, \ldots u_{n-1}$, such that $z \in C \leftrightarrow \phi(z, u_0, u_1, \ldots u_{n-1})$. A *proper class* is a class which is not a set.

The set builder notation: $C = \{z\colon \phi(z, u_0, u_1, \dots u_{n-1})\}$ is often used to denote classes. A class may not be a set since the axiom schema of comprehension cannot be a priori applied due to the lack of the ambient set x. On some intuitive level, classes may fail to be sets on the account of being "too large".

Example 1.4.3. The class $\{z\colon z \notin z\}$ is not a set.

Proof. This is the Russell's paradox. Suppose that there is a set x such that for every set z, $z \in x$ just in case $z \notin z$. Ask whether $x \in x$ or not. If $x \in x$ then by the definition of x, $x \notin x$ which is a contradiction. If $x \notin x$ then by the definition of x, $x \in x$ holds and we have a contradiction again. Both options lead to contradiction, proving that x does not exist. □

The universal class $\{z\colon z = z\}$ is often called V, the set theoretical universe. It is certainly not a set: if V were a set, then all classes would turn into sets by inserting the ambient set V into their definitions. However, we just produced a class which is not a set.

Historical debate. The formulation of the axiom schema of comprehension is motivated by the desire to avoid Russell's paradox. The use of the ambient set x makes it impossible to form sets such as $\{z : z \notin z\}$ since we are missing the ambient set: $y = \{z \in ?\colon z \notin z\}$. This trick circumvents all the known paradoxes, it comes naturally to all working mathematicians, and it does not present any extra difficulties in the development of mathematics in set theory.

There were other attempts to circumvent the paradoxes by limiting the syntactical nature of the formula ϕ used in the comprehension schema as opposed to requiring the existence of the ambient set x. One representative of these efforts is Quine's *New Foundations* (NF) axiom system [8]. Roughly stated, in NF the formula ϕ has to be checked for circular use of $\in$ relation between its variables before it can be used to form a set. This allows the existence of the universal set $\{z\colon z = z\}$, but it also makes the development of natural numbers and general practical use extremely cumbersome. This seems like a very poor trade. As a result, NF is not used in mathematics today.

There was an objection to possible use of *impredicative definitions* allowed by the present form of comprehension. Roughly stated, the objecting parties (including Russell and Poincaré) claimed that a set must not be defined by a formula which takes into account sets to which the defined set belongs (a formula ϕ defining a subset of some set x should not use the powerset of x as one of its parameters, for example). Such a definition would form, in their view, a *vicious circle*. It is challenging to make this objection precise. Mathematicians use impredicative definitions quite often and without care–for example the usual proof of completeness of the real numbers contains a vicious circle in this view. Attempts to build mathematics without impredicative definitions turned out to be awkward. The school of thought objecting to impredicative definitions in mathematics mostly fizzled out before 1950.

Exercise 1.4.1. An intersection of a class and a set is a set.

Exercise 1.4.2. If C is a nonempty class then $\bigcap C$ is a set.

1.5 Relations and functions

For the statement of the remaining axioms, it is useful to develop the standard parlance regarding relations and functions. This is possible to do using the axioms introduced so far.

Definition 1.5.1. The Cartesian product $x \times y$ is the set of all ordered pairs $\langle u, v\rangle$ such that $u \in x$ and $v \in y$.

Note that the Cartesian product is a subset of $\mathcal{PP}(x \cup y)$ obtained by an application of the comprehension schema.

Definition 1.5.2. A *relation* between x and y is any subset of $x \times y$. A *binary relation* on x is any subset of $x \times x$, a *ternary relation* on x is any subset of $x \times x \times x$ etc. A (partial) *function* from x to y is a relation $f \subset x \times y$ such that for every $u \in x$ there is at most one $v \in y$ such that $\langle u, v\rangle \in f$.

For a relation $R \subset x \times y$ write $\mathrm{dom}(R) = \{u \in x \colon \exists v\ \langle u, v\rangle \in R\}$ and $\mathrm{rng}(R) = \{v \in y \colon \exists u\ \langle u, v\rangle \in R\}$. A function $f \colon x \to y$ is *total* if $\mathrm{dom}(f) = x$.

The standard notation and terminology concerning functions and relations as used in this book and elsewhere includes the following:

- if f is a function and $\langle u, v\rangle \in f$ then we write $f(u) = v$ and say that the output of f at input u is v;
- a function $f : x \to y$ is a *surjection* if $\mathrm{rng}(f) = y$. It is an *injection* if $u \neq v \in \mathrm{dom}(f)$ implies $f(u) \neq f(v)$. f is a *bijection* if $\mathrm{dom}(f) = x$ and f is both an injection and surjection.
- if $a \subseteq \mathrm{dom}(f)$ then $f''a$, the *image* of a under f, is the set $\{f(x) \colon x \in a\}$. If b is a set then $f^{-1}b$, the *preimage* of b under f, is the set $\{x \in \mathrm{dom}(f) \colon f(x) \in b\}$;
- if f is a function and a is a set then $f \restriction a$, the *restriction of f to a*, is the function $\{\langle u, v\rangle \in f \colon u \in a\}$;
- if x, y are sets, then y^x is the collection of all functions f such that $\mathrm{dom}(f) = x$ and $\mathrm{rng}(f) \subset y$;
- if $R \subset x \times y$ is a relation then $R^{-1} \subset y \times x$, the *inverse* of R, is the set $\{\langle u, v\rangle : \langle v, u\rangle \in R\}$.

Note that a function f is an injection if and only if its inverse is a function. There are other ways to denote the image and preimage of a set under a function used in the literature; for example, the image of a under f may be denoted by $f(a)$ or $f[a]$ etc. Our usage follows the current set theoretic standard.

1.6 Axiom of Infinity

Definition 1.6.1. The *Axiom of Infinity* is the statement $\exists x\ 0 \in x \wedge \forall y \in x\ y \cup \{y\} \in x$.

A brief discussion reveals that the set x in question must be in some naive sense infinite: its elements are 0, $\{0\}$, $\{0, \{0\}\}$ and so on. One must keep in mind that the distinction between finite and infinite sets must be defined formally. This is done in Section 2.2 and indeed, every set x satisfying $0 \in x \wedge \forall y \in x\ y \cup \{y\} \in x$ must be in this formal sense infinite. A natural question occurs: why is the axiom of infinity stated in precisely this way? Of course, there are many formulations which turn out to be equivalent. The existing formulation makes the development of natural numbers in Section 2.1 particularly smooth.

Historical debate. As there are no collections in sensory experience that are infinite, there was a considerable discussion, mostly predating the axiomatic development of set theory, regarding the use of infinite sets in mathematics.

Aristotle discerned between two kinds of infinity: the potential infinity and the actual infinity. A potentially infinite sequence is a sequence with a rule of extending it for an arbitrary number of steps (such as counting the natural numbers $0, 1, 2\ldots$). An actual infinity then grasps the whole result of repeating all these steps and views it as a completed object (the set of natural numbers). Zeno's paradoxes (5th century BC) have long been regarded as a proof that the actual infinity is an inherently contradictory concept. Bernard Bolzano, a catholic philosopher, produced an argument that there are infinitely many distinct truths which must be all present in omniscient God's mind, and therefore God's mind must be infinite (1851). This was intended as a defense of the use of infinite sets in mathematics. Poincaré and Hermann Weyl can be listed as important opponents of the use of infinite sets among 19th–20th century mathematicians. *Finitism*, the rejection of the Axiom of Infinity, still has a small minority following among modern mathematicians. On a practical level, while a great deal of mathematics can be developed without the Axiom of Infinity, the formulations and proofs without the axiom become cumbersome and long.

1.7 Axiom of Choice

Definition 1.7.1. The *Axiom of Choice* (AC) is the following statement. For every set x consisting of nonempty sets, there is a function f with $\mathrm{dom}(f) = x$ and $\forall y \in x\ f(y) \in y$. The function f is referred to as the *selector*.

Historical debate. The Axiom of Choice is the only axiom of set theory which asserts an existence of a set (the selector) without providing a formulaic description of that set. The Axiom of Infinity is presently stated in such a way as well, but (unlike AC) it can be reformulated to provide a definition of a certain infinite set. Naturally, AC provoked the most heated discussion of all the axioms.

Zermelo used AC in 1908 to show that the set of real numbers can be well-ordered (see Section 3.2). This seemed counterintuitive, as the well-ordering of the reals is an extremely strong construction tool, and at the same time it is entirely unclear how one could construct such a well-ordering. A number of people (including Lebesgue, Borel, and Russell) voiced various objections to AC as the main tool in Zermelo's theorem. A typical objection (Lebesgue) claimed that a proof of an existence of an object with a certain property, without a construction or definition of such an object, is not permissible. In the end, certain consequences of the axiom proved indispensable to the development of certain theories, such as Lebesgue's own theory of measure. A repeated implicit use of certain consequences of AC in the work of its very opponents also strengthened the case for adoption of the axiom.

One reason for the acceptance of the axiom was the lack of a constructive alternative. A plausible and useful alternative appeared in the 1960's in the form of Axiom of Determinacy (AD), asserting the existence of winning strategies in certain infinite two-player games [7]. At that point, the Axiom of Choice was already part of the orthodoxy and so AD remained on the sidelines.

Pleasing consequences. The axiom of choice is helpful in the development of many mathematical theories. Typically, it allows proving general theorems about very large objects.

- (Algebra) Every vector space has a basis;
- (Dynamical systems) A continuous action of a compact semigroup has a fixed point;
- (Topology) Product of any family of compact spaces is compact;
- (Functional analysis) Hahn–Banach theorem.

Foul consequences. Some weak consequences of AC are necessary for the development of theory of integration. However, its full form makes a completely harmonious integration theory impossible to achieve. It produces many "paradoxical" (a better word would be"counterintuitive") examples which force integration to apply to fairly regular functions and sets only.

- There is a nonintegrable function $f : [0, 1] \to [0, 1]$;
- (Banach–Tarski paradox) there is a partition of the unit ball in $\mathbb{R}^3$ into several parts which can be reassembled by rigid motions to form two solid balls of unit radius.

The upshot. The axiom of choice is part of the mathematical orthodoxy today, and its suitability is not questioned or doubted by any significant number of mathematicians. A good mathematician notes its use though, and (mostly) does not use it when an alternative proof without AC is available. The proof without AC will invariably yield more information than the AC proof. Almost

every mathematical theorem asserting the existence of an object without (at least implicitly) providing its definition is a result of an application of the axiom of choice.

Definition 1.7.2. If x is a collection of nonempty sets, then $\prod x$, the *product* of x, is the collection of all selectors on x.

It is not difficult to see that $\prod x$ is a set. The Axiom of Choice asserts that the product of a collection of nonempty sets is nonempty. In the case that x consists of two sets only, this definition gives a nominally different set than Definition 1.5.1, but this will never cause any confusion.

1.8 Axiom schema of Replacement

As was the case with the axiom schema of comprehension, this is not a single axiom but a schema including infinitely many axioms, one for each formula of set theory defining a class function.

Definition 1.8.1. The *Axiom schema of Replacement* states the following. If f is a class function and x is a set, then $f''x$ is a set as well.

Replacement was a late contribution to the axiomatics of ZFC (1922). It is the only part of the axiomatics invented by Fraenkel. It is used almost exclusively for the internal needs of set theory; we will see that the development of ordinal numbers and well-orderings would be akwward without it. The only "mathematical" theorem for which it is known to be indispensable is the Borel determinacy theorem of Martin, ascertaining the existence of winning strategies in certain types of two player infinite games [6].

Exercise 1.8.1. Show that the Axiom schema of Replacement is equivalent to the statement "each class function with set domain is a set".

Exercise 1.8.2. The statement "the range of a set function is a set" can be proved without replacement. Use Comprehension to prove the following: $\forall f \forall x$ if f is a function then $\exists y \forall z\ z \in y \leftrightarrow \exists v \in x\ f(v) = z$.

Exercise 1.8.3. There is no class injection from a proper class into a set.

1.9 Axiom of Regularity

Also known as Foundation or Well-foundedness.

Definition 1.9.1. The *Axiom of Regularity* states $\forall x\ x = 0 \vee \exists y \in x \forall z \in x\ z \notin y$.

Restated, every nonempty set contains an $\in$-minimal element. This is the only axiom of set theory that explicitly limits the scope of the set-theoretic universe, ruling out the existence of sets such as the following:

Exercise 1.9.1. Use the Axiom of Regularity to show that there is no set x with $x \in x$, and there are no sets x, y such that $x \in y \in x$.

The motivation behind the adoption of this axiom lies in the fact that the development of common mathematical notions within set theory uses sets that always, and of necessity, satisfy regularity. The formal development of set theory is smoother with the axiom as well. The present form of the axiom is due to von Neumann [12]. Mathematical interest in the phenomena arising when the Axiom of Regularity is denied has been marginal [1].

Chapter 2

Basic notions

2.1 Von Neumann's natural numbers

The purpose of this section is to develop natural numbers in ZFC.

Definition 2.1.1. For a set x, write $s(x) = x \cup \{x\}$. A set y is *inductive* if $0 \in y$ and for all x, $x \in y$ implies $s(x) \in y$.

Definition 2.1.2. (Von Neumann) The $\subseteq$-smallest inductive set is denoted by ω. A set x is a *natural number* if $x \in \omega$.

With every definition of this sort, one has to make sure that it actually makes sense. This is the contents of the following theorem.

Theorem 2.1.3. *There is an $\subseteq$-smallest inductive set.*

Proof. Let w be the intersection of all inductive sets. Clearly, if there is an $\subseteq$-smallest inductive set then w must be it, so it is enough to verify that w is inductive.

First, note that w is in fact a set. Just let z be any inductive set as guaranteed by the Axiom of Infinity, and note that $w = \{x \in z : \forall y \text{ if } y \text{ is inductive then } x \in y\}$ which can be formed by the Axiom schema of Comprehension.

Second, note that w is an inductive set. For that, we have to verify that $0 \in w$ and for every $x \in w$, $s(x) \in w$ holds. As 0 belongs to every inductive set, $0 \in \omega$ by the definition in ω. Now suppose that $x \in \omega$; we must show that $s(x) \in \omega$. For every inductive set y, $x \in y$ holds by the definition of ω. As y is inductive, $s(x) \in y$ as well. We have just proved that $s(x)$ belongs to every inductive set, in other words $s(x) \in \omega$. This completes the proof. □

The main feature of ω is that we can use induction to prove various statements about natural numbers. This is the contents of the next theorem.

Theorem 2.1.4. (Induction) *Suppose that ϕ is a formula, $\phi(0)$ holds, and $\forall x \in \omega\ \phi(x) \to \phi(s(x))$ also holds. Then $\forall x \in \omega\ \phi(x)$ holds.*

Proof. Consider the set $y = \{x \in \omega : \phi(x)\}$. We will show that y is an inductive set. Then, since ω is the smallest inductive set, it follows that $y = \omega$, in other words $\forall x \in \omega\ \phi(x)$ as desired.

Indeed, $0 \in y$ as $\phi(0)$ holds. If $x \in y$ then $s(x) \in y$ as well by the assumptions on the formula ϕ. It follows that y is an inductive set as desired. □

We will use the standard terminology for induction: $\phi(0)$ is the *base step*, the implication $\phi(x) \to \phi(s(x))$ is the *induction step*, and the formulas $\phi(x)$ in the induction step is the *induction hypothesis*. The next step is to verify that $\in$ on ω is a linear ordering that emulates the properties of natural numbers. Firstly, define what is meant by a linear ordering here.

Definition 2.1.5. An *ordering* on a set x is a two place relation $\leq \subset x \times x$ such that

1. $u \leq u$ for every $u \in x$;
2. $u \leq v \leq w$ implies $u \leq w$.
3. $u \leq v$ and $v \leq u$ implies $u = v$.

A *linear ordering* is an ordering which satisfies in addition

4. for every $u, v \in x$, $u \leq v$ or $v \leq u$ holds.

A *strict ordering* on x is a two place relation $<$ such that

1'. for every $u \in x$, $u < u$ is false;

2'. $u < v < w$ implies $u < w$.

Clearly, a strict ordering on x is obtained from an ordering by removing the diagonal, i.e. the set $\{\langle u, u\rangle : u \in x\}$. On the other hand, an ordering can be obtained from any strict ordering by adding the diagonal. The two notions are clearly very close and we will sometimes confuse them.

Theorem 2.1.6. (Linear ordering of natural numbers)

1. *If $x \in \omega$ and $y \in x$ then $y \in \omega$;*
2. *the relation $\in$ is a strict linear ordering on ω.*

Proof. For (1), let $\phi(x)$ be the statement $\forall y \in x\ y \in \omega$, and by induction on x prove $\forall x \in \omega\ \phi(x)$. *Base step.* The statement $\phi(0)$ holds since its first universl quantifier ranges over the empty set. *Successor step.* Suppose that $\phi(x)$ holds. To prove $\phi(s(x))$, let $y \in s(x)$. Either $y \in x$, in which case $y \in \omega$ by the induction hypothesis. Or $y = x$, in which case $y \in \omega$ since $x \in \omega$. This proves (1).

To prove (2), we have to verify the transitivity and linearity of $\in$ on ω. We will start with transitivity. Let $\phi(x)$ be the statement $\forall y \in x\ \forall z \in y\ z \in x$,

and by induction on $x \in \omega$ prove $\forall x \in \omega\ \phi(x)$. *Base step.* The statement $\phi(0)$ holds as its first universal quantifier ranges over the empty set. *Induction step.* Suppose that $\phi(x)$ holds and work to verify $\phi(s(x))$. Let $y \in s(x)$ and $z \in y$. By the definition of $s(x)$, there are two cases. *Either* $y \in x$, then by the induction hypothesis $z \in x$, and as $x \subseteq s(x)$, $z \in s(x)$ holds. *Or*, $y = x$, then $z \in x$ and as $x \subseteq s(x)$, $z \in s(x)$ holds again. This confirms the induction step and proves the transitivity.

Next, we proceed to linearity. The following two preliminary claims will be useful:

Claim 2.1.7. *For every $y \in \omega$, $0 = y$ or $0 \in y$.*

Proof. Let $\psi(y)$ be the statement $0 = y \vee 0 \in y$, and by induction on $y \in \omega$ prove $\forall y \in \omega\ \psi(y)$. *Base step.* $\psi(0)$ holds as $y = 0$ is one of the disjuncts. *Induction step.* Suppose that $\psi(y)$ holds and work to verify $\psi(s(y))$. The induction hypothesis offers two cases. *Either*, $y = 0$, in which case $y = 0 \in s(y)$ by the definition of $s(y)$. *Or*, $0 \in y$ and then $0 \in s(y)$ since $y \subseteq s(y)$. In both cases, the induction step has been confirmed. □

Claim 2.1.8. *For every $y \in \omega$, for every $x \in y$ $s(x) \in s(y)$ holds.*

Proof. Let $\psi(y)$ be the statement $\forall x \in y\ s(x) \in s(y)$ and by induction on $y \in \omega$ prove the statement $\forall y \in \omega\ \phi(y)$. *Base step.* $\psi(0)$ is trivially true as its universal quantifier ranges over an empty set. *Induction step.* Assume $\psi(y)$ holds and work to verify $\psi(s(y))$. Let $x \in s(y)$ be any element. By the definition of $s(y)$, there are two cases. *Either*, $x \in y$, then by the induction hypothesis $s(x) \in s(y)$, and as $s(y) \subseteq s(s(y))$, $s(x) \in s(s(y))$ holds. *Or*, $x = y$, in which case $s(x) = s(y) \in s(s(y))$ by the definition of $s(s(y))$. In both cases, the induction step has been confirmed. □

Now, let $\phi(x)$ be the statement $\forall y \in \omega\ x = y \vee x \in y \vee y \in x$. By induction on $x \in \omega$ prove $\forall x \in \omega\ \phi(x)$. *Base step.* The statement $\phi(0)$ is the contents of Claim 2.1.7. *Induction step.* Suppose that $\phi(x)$ holds, and work to verify $\phi(s(x))$. Let $y \in \omega$ be arbitrary. The induction hypothesis yields a split into three cases. *Either*, $y \in x$ and then, as $x \subseteq s(x)$, $y \in s(x)$. *Or*, $y = x$ and then $y \in s(x)$ by the definition of $s(x)$. *Or*, $x \in y$, and then by Claim 2.1.8 $s(x) \in s(y)$ holds, which by the definition of $s(y)$ says that either $s(x) \in y$ or $s(x) = y$. In all cases, the induction step has been confirmed. The linearity of the $\in$ ordering on natural numbers has been verified. □

Thus, as a set, every natural number is exactly the set of all natural numbers smaller than it. Now, a rather routine induction arguments (see the exercises) show that for every $x \in \omega$, $s(x)$ is the smallest natural number larger than x, and for every nonzero natural number x there is a largest number y smaller than x and $x = s(x)$. In order to develop further concepts associated with the natural numbers, such as the arithmetic operations, one uses recursive definitions as captured in the following theorem.

Theorem 2.1.9. (Recursive definitions) *Suppose that F is a class function such that $F(x)$ is defined for every set x. Then there is a unique class function G such that* $\mathrm{dom}(G) = \omega$ *and for every $n \in \omega$, $G(n) = F(G \restriction n)$.*

Proof. First, prove that for every $m \in \omega$ there is a unique set function G_m such that $\mathrm{dom}(G_m) = m + 1$ and for every $n \in m + 1$, $G_m(n) = F(G_m \restriction n)$. The proof proceeds by induction on $m \in \omega$. The base step $m = 0$ is trivial: $G_0(0) = F(0)$. For the induction step, suppose that the unique function G_m with domain $m+1$ has been found. Let $G_{m+1} = G_m \cup \{\langle m+1, F(G_m)\rangle\}$. This is the unique function such that for every $n \in m + 2$, $G(n) = F(G \restriction n)$.

Now, note that for natural numbers $m \in k$, it must be the case that $G_m \subset G_k$: $G_k \restriction m + 1$ satisfies that for every $n \in m + 1$, $G_k(n) = F(G_k \restriction n)$ and by the uniqueness of G_m, $G_k \restriction m + 1 = G_m$ must hold. Let G be the class defined by $\langle m, x\rangle \in G$ if and only if $m \in \omega$ and $G_m(m) = x$. This is the unique class function required in the theorem.

For the uniqueness of the function G, suppose that H is a class function with $\mathrm{dom}(H) = \omega$ and such that for every $n \in \omega$, $H(n) = F(H \restriction n)$. Suppose for contradiction that $H \neq G$. The set $x = \{n \in \omega : G(n) \neq H(n)\}$ is nonempty, and therefore contains a smallest element m. Then, $G \restriction m = H \restriction m$ and so $G(m) = F(G \restriction m) = F(H \restriction m) = H(m)$. This contradicts the assumption that $x \in m$. □

As an interesting application of recursive definitions, we will develop the notion of the transitive closure of a set.

Definition 2.1.10. A set x is *transitive* if for every $y \in x$ and every $z \in y$, $z \in x$ holds.

Thus, for example ω and every natural number is transitive by Theorem 2.1.6. An example of a nontransitive set is $\{\{0\}\}$. We will show that every set belongs to a transitive set.

Definition 2.1.11. Let x be a set. The *transitive closure* of x, $\mathtt{trcl}(x)$, is the inclusion-smallest transitive set containing x as an element.

Theorem 2.1.12. *For every set x, $\mathtt{trcl}(x)$ exists.*

Proof. Recursively define a function G with $\mathrm{dom}(G) = \omega$ so that $G(0) = x$ and $G(n + 1) = \bigcup G(n)$. Theorem 2.1.9 shows that there is a unique function G satisfying these demands. By Axiom of Replacement, $\mathrm{rng}(G)$ is a set. Let $y = \{x\} \cup \bigcup \mathrm{rng}(G)$. We claim that y is a transitive set and if z is a transitive set containing x as an element, $y \subseteq z$ holds.

For the transitivity of y, suppose that $u \in y$ and $v \in u$. Then $u = x$ or there must be $n \in \omega$ such that $u \in G(n)$. By the definition of the function G, $v \in G(0)$ or $v \in G(n + 1)$ must hold. Thus, $v \in y$ and the transitivity of y has been confirmed.

For the minimality of y, suppose for contradiction that z is a transitive set containing x as an element and $y \not\subseteq z$. Thus, the set $y \setminus z$ must be nonempty,

containing some element v. There must be $n \in \omega$ such that $v \in G(n)$; choose $v \in y \setminus z$ so that this number n is minimal possible. By the definition of $G(n)$, there is $u \in G(n-1)$ such that $v \in u$. By the minimal choice of the number n, $u \in z$. By the transitivity of the set z, $u \in z$ and $v \in u$ imply that $v \in z$. This contradicts the initial choice of the set z. The theorem follows. □

Corollary 2.1.13. (Axiom of Regularity for classes) *Let C be a nonempty class. There is an element $x \in C$ such that no elements of x belong to C.*

Proof. Let y be any element of C. Consider the nonempty set $C \cap \mathtt{trcl}(y)$. The fact that this is indeed a set and not just a class follows from Exercise 1.4.1. Use the Axiom of Regularity to find an $\in$-minimal element x of $C \cap \mathtt{trcl}(y)$. All elements of x belong to $\mathtt{trcl}(y)$, and so by the minimal choice of x, none of them can belong to C. Thus, the set x works as required. □

Exercise 2.1.1. Write down the first four natural numbers as sets using the set builder notation and the symbol 0.

Exercise 2.1.2. By induction on $x \in \omega$ show that if $x \neq 0$ is a natural number then it has a predecessor: a number y which is largest among all numbers smaller than x, and such that $x = s(y)$.

Exercise 2.1.3. Without using the Axiom of Regularity, show that every nonempty subset of ω has an $\in$-smallest element.

Exercise 2.1.4. Define addition of natural numbers using a recursive definition.

2.2 Finite and infinite sets

The purpose of this section is to develop the definition of finiteness for sets. One reasonable way to proceed is to define a set to be finite if it is in a bijection with some natural number. We will use a different definition which has the virtues of being more intelectually stimulating, very efficient in proofs, and independent of the development of ω:

Definition 2.2.1. (Tarski) A set x is *finite* if every nonempty set $a \subseteq \mathcal{P}(x)$ has a $\subseteq$-minimal element: a set $y \in a$ such that no $z \in a$ is a proper subset of y. A set is *infinite* if it is not finite.

With a somewhat slick definition of this sort, it is necessary to verify that it corresponds to the intuitive notion of finiteness. We first provide a basic example of a finite and infinite set:

Example 2.2.2. 0 is a finite set. ω is an infinite set.

Proof. To see that 0 is finite, if $a \subseteq \mathcal{P}(0)$ is a nonempty set, then either it contains 0 and then 0 is its $\subseteq$-minimal element, or it does not contain 0 and then $\{0\}$ is its $\subseteq$-minimal element.

To see that ω is infinite, for every $n \in \omega$ let $y_n = \{m \in \omega \colon n \in m\}$ and let $a = \{y_n \colon n \in \omega\}$. This is a subset of $\mathcal{P}(\omega)$; let us show that it has no $\subseteq$-minimal element. Suppose y_n was such a minimal element. Then $y_{n+1} \in a$ is its proper subset, contradicting the minimality of y_n. □

Theorem 2.2.3. *The class of finite sets is closed under the following operations:*

1. *taking a subset;*
2. *adding a single element to a set;*
3. *union;*
4. *surjective image;*
5. *powerset.*

Proof. For (1), suppose that x is finite and $y \subset x$; we must argue that y is finite. Let $a \subset \mathcal{P}(y)$ be a nonempty set; we must show that a has a $\subseteq$-minimal element. Since $y \subset x$, it is the case that $a \subset \mathcal{P}(x)$. As x is finite, a must have a $\subseteq$-minimal element as desired.

For (2), suppose that x is finite and i is any set; we must verify that the set $y = x \cup \{i\}$ is finite. Let $a \subseteq \mathcal{P}(y)$ be a nonempty set; we must produce a $\subseteq$-minimal element of a. Let $b = \{u \cap x \colon u \in a\}$. This is a nonempty subset of $\mathcal{P}(x)$; as x is assumed to be finite, the set b has a $\subseteq$-minimal element v. There are now two cases. Either $v \in a$, in which case v is a $\subseteq$-minimal element of a. Or, $v \notin a$, in which case $u = v \cup \{i\}$ is a $\subseteq$-minimal element of a. This completes the proof of (2).

For (3), assume for contradiction that x, y are finite and $x \cup y$ is not. Let $a = \{z \subset x : z \cup y \text{ is not finite}\}$. This is a nonempty subset of x containing at least x as an element. Since x is finite, the set a has an inclusion-minimal element, say u. The set u must be nonempty since $y \cup 0 = y$ is a finite set. Let $i \in u$ be an arbitrary element, and let $v = u \setminus \{i\}$. By the minimality of u, $y \cup v$ is finite. By (2) $y \cup v \cup \{i\}$ is finite as well. As $y \cup v \cup \{i\} = y \cup u$, this contradicts the assumption that $u \in a$.

For (4), assume for contradiction that x is a finite set, $f : x \to y$ is a surjection, and y is not finite. Let $a = \{z \subset x \colon f''z \text{ is not finite}\}$. This is a subset of $\mathcal{P}(x)$ which by our contradictory assumption contains x as an element and therefore it is nonempty. As x is finite, a contains a $\subseteq$-minimal element u. Let $i \in u$ be an arbitrary element, and let $v = u \setminus \{i\}$. Then $v \notin a$ by the minimal choice of u, and so $f''v$ is finite. However, $f''u = f''v \cup \{f(i)\}$, which is finite by (2), contradicting the assumption that $u \in a$.

For (5), assume for contradiction that x is finite and $\mathcal{P}(x)$ is not finite. Let $a = \{y \subseteq x : \mathcal{P}(y) \text{ is not finite}\}$. This is a nonempty set, containing at least x as an element. Let u be a $\subseteq$-minimal element of a. Pick an element $i \in u$ and consider the set $v = u \setminus \{i\}$. Then, $\mathcal{P}(u) = \mathcal{P}(v) \cup \{z \cup \{i\} : z \in \mathcal{P}(v)\}$. The first set in the union is finite by the minimality of u, and the second is a surjective image of the first, therefore finite as well. By the previous items, $\mathcal{P}(u)$ is finite, and his is a contradiction to the assumption that $u \in a$. □

Corollary 2.2.4. *An injective image of an infinite set is infinite.*

Proof. Let x be an infinite set and let f be an injection with $\mathrm{dom}(f) = x$. We have to argue that $y = \mathrm{rng}(f)$ is infinite. Note first that $g = f^{-1}$ is a surjective function from y to x. So, if y were finite, then $g''y = x$ would be finite as well by Theorem 2.2.3(4), contradicting the assumption that x is infinite. □

The final theorem of this section is a characterization theorem. It characterizes finiteness in terms of natural numbers, and provides an equivalent reformulation of Definition 2.2.1. This allows one to prove theorems about finite sets by induction on their size. Note that the treatment of finiteness up to this point did not use natural numbers at all.

Theorem 2.2.5. *Let x be a set.*

1. *x is finite if and only if it is in bijection with a natural number;*
2. *x is infinite if and only if it contains an injective image of ω.*

The proof of (2) uses the Axiom of Choice. It is known that without the Axiom of Choice, (2) cannot be proved.

Proof. Start with (1). For the right-to-left implication, first argue by induction that $\forall n \in \omega\ n$ is finite. The base step is verified in Example 2.2.2, and the induction step follows from Theorem 2.2.3(2). Then, argue that every bijective image of a natural number is finite by Theorem 2.2.3(3).

For the left-to-right implication, suppose that x is finite and for contradiction assume that it is not in bijection with any natural number. Let $a = \{y \subseteq x : y$ is not in a bijective image with a natural number$\}$. This is a nonempty set, containing at least x as an element. Let $u \in a$ be a $\subseteq$-minimal element of a. Pick an arbitrary element $i \in u$ and let $v = u \setminus \{i\}$. By the minimal choice of u, v is a bijective image of an element of ω, and then u is a bijective image of its successor. This contradicts the assumption that $u \in a$.

Now for (2). For the right-to-left implication, note that ω is infinite by Example 2.2.2. An injective image of ω is infinite by Corollary **??**, and every superset of an infinite set is infinite by Theorem 2.2.3(1). This concludes the proof of the right-to-left implication.

For the left-to-right implication, let x be an infinite set; we must produce an injection from ω to x. Use the Axiom of Choice to produce a selector function $H\colon \mathcal{P}(x)\setminus\{0\} \to x$. Now consider the recursive definition of a function $F : \omega \to x$ given by $F(n) = H(x \setminus F''n)$. Note that for every natural number $n \in \omega$ the set $F''n \subset x$ will be finite by (1). Since x is infinite, the set $x \setminus F''n$ will be nonempty, the value $F(n) = H(x \setminus F''n)$ will be defined and different from all values of $F(m)$ for $m \in n$. It is then clear that F will be the requested injection from ω to x. □

In the following exercises, use Tarski's definition of finiteness.

Exercise 2.2.1. Prove that a surjective image of a finite set is finite.

Exercise 2.2.2. Let x be a finite set and $\leq$ a linear ordering on x. Prove that x has a largest element in the sense of the ordering $\leq$.

Exercise 2.2.3. Prove that the product of two finite sets is finite.

Exercise 2.2.4. Prove without the Axiom of Choice that if x is a finite set consisting of nonempty sets, then x has a selector.

Exercise 2.2.5. Without the use of Axiom of Infinity show that existence of an infinite set is equivalent to an existence of an inductive set.

2.3 Cardinality

In this section, we will develop the basic features of the set-theoretic notion of size–cardinality.

Definition 2.3.1. Let x, y be sets. Say that x, y have the same *cardinality*, in symbols $|x| = |y|$, if there is a bijection $f : x \to y$. Say that $|x| \leq |y|$ if there is an injection from x to y.

Theorem 2.3.2. *Having the same cardinality is an equivalence relation and* $\leq$ *is a quasiorder.*

Theorem 2.3.3. (Schröder–Bernstein) *If* $|x| \leq |y|$ *and* $|y| \leq |x|$ *then* x, y *have the same cardinality.*

Proof. Let x, y be sets and $f : x \to y$ and $g : y \to x$ be injections; we must produce a bijection. Identifying y with $\mathrm{rng}(g)$, we may assume that $y \subset x$ and g is the identity on y. By induction on $n \in \omega$ define sets $x_n, y_n \subset x$ by letting $x_0 = x, y_0 = y$ and $x_{n+1} = f''x_n$, $y_{n+1} = f''y_n$. By induction on $n \in \omega$ prove that $x_0 \supseteq y_0 \supseteq x_1 \supseteq y_1 \supseteq x_2 \supseteq \ldots$ Let $x_\omega = \bigcap_n x_n$. Consider the function $h : x \to y$ defined by $h(z) = z$ if $z \in x_\omega$, $h(z) = f(z)$ if $z \in x_n \setminus y_n$ for some $n \in \omega$, and $h(z) = z$ if $z \in y_n \setminus x_{n+1}$. This is the desired bijection. To see this, note that $h \restriction x_\omega$ is a bijection from x_ω to itself, $h \restriction x_n \setminus y_n$ is a bijection from $x_n \setminus y_n$ to $x_{n+1} \setminus y_{n+1}$, and $h \restriction y_n \setminus x_{n+1}$ is a bijection from $y_n \setminus x_{n+1}$ to itself. □

Theorem 2.3.4. *Distinct natural numbers have distinct cardinalities.*

Proof. It will be enough to show that if x, y are finite sets and $y \subseteq x$ and $y \neq x$ then y, x have distinct cardinalities. Suppose for contradiction that this fails for some x, y. Let $a = \{z \subseteq x : |z| = |x|\}$. The set $a \subset \mathcal{P}(x)$ is certainly nonempty, containing at the very least the set x itself. Let $z \in a$ be a $\subseteq$-minimal element. Note that $z \neq x$ since $y \in a$ and y is a proper subset of x. Let $h : x \to z$ be a bijection, and let $u = h''z$. Then $u \subseteq z$ and $|u| = |z|$, since $h \restriction z : z \to u$ is a bijection. Moreover, $u \neq z$: if i is any element of the nonempty set $x \setminus z$, then $h(i)$ belongs to $z \setminus u$. Thus, u is a proper subset of z which has the same cardinality of z and so the same cardinality as x. This contradicts the minimal choice of the set z. □

This theorem completely determines the possible cardinalities of finite sets. Every finite set has the same cardinality as some natural number by Theorem 2.2.5, and distinct natural numbers have distinct cardinalities. Thus, the cardinalities of finite sets are linearly ordered. One can ask if this feature persists even for infinite cardinalities. The answer depends on the axiom of choice. Assuming the axiom of choice, we will show that the even infinite cardinalities are linearly ordered.

We will conclude this section by proving that there are many distinct infinite cardinalities.

Theorem 2.3.5. (Cantor) *For every set x, $|x| \leq |\mathcal{P}(x)|$ and $|x| \neq |\mathcal{P}(x)|$.*

Proof. Clearly $|x| \leq |\mathcal{P}(x)|$ since the function $f : x \mapsto \{x\}$ is an injection from x to $\mathcal{P}(x)$.

To show that $|x| \neq |\mathcal{P}(x)|$ suppose for contradiction that x is a set and $f : x \to \mathcal{P}(x)$ is any function. It will be enough to show that $\mathrm{rng}(f) \neq \mathcal{P}(x)$, ruling out the possibility that f is a bijection. Consider the set $y = \{z \in x : z \notin f(z)\}$; we will show that $y \notin \mathrm{rng}(f)$. For contradiction, assume that $y \in \mathrm{rng}(f)$ and fix $z \in x$ such that $y = f(z)$. Consider the question whether $z \in y$. If $z \in y$ then $z \notin f(z)$ by the definition of y, and then $z \notin y = f(z)$. If, on the other hand, $z \notin y$ then $z \in f(z)$ by the definition of y, and so $z \in y = f(z)$. In both cases, we have arrived at a contradiction. □

Thus, $\mathcal{P}(\omega)$ has strictly greater cardinality than ω, $\mathcal{PP}(\omega)$ has strictly greater cardinality than $\mathcal{P}(\omega)$ and so on. We have produced infinitely many infinite sets with pairwise distinct cardinalities.

Exercise 2.3.1. Prove that if $|x| \leq |y|$ then $|\mathcal{P}(x)| \leq |\mathcal{P}(y)|$.

Exercise 2.3.2. Prove that for every set x, $|\mathcal{P}(x)| = |2^x|$.

Exercise 2.3.3. Use the Axiom of Choice to prove that if y is a surjective image of x then $|y| \leq |x|$.

Exercise 2.3.4. Prove that whenever x is a set then there is a set y such that $|z| < |y|$ for every $z \in x$.

2.4 Countable and uncountable sets

The most important cardinality-related concept in mathematics is countability. We will use it in this section to provide the scandalously easy proof of the existence of transcendental real numbers discovered by Cantor.

Definition 2.4.1. A set x is *countable* if $|x| \leq |\omega|$. A set which is not countable is *uncountable*.

As a matter of terminology, some authors require countable sets to be infinite. By the following theorem, this restricts the definition to the collection of sets which have the same cardinality as ω.

Theorem 2.4.2. *1. If x is countable then either x is finite or $|x| = |\omega|$.*

2. A nonempty set is countable if and only if it is a surjective image of ω.

3. A surjective image of a countable set is countable.

4. (With the axiom of choice) The union of a countable collection of countable sets is again countable.

Proof. For (1), first argue that for every set $x \subset \omega$, either x is finite or $|x| = |\omega|$. This is easy to see though: if the set $x \subset \omega$ is infinite, then its increasing enumeration is a bijection between ω and x.

Now suppose that x is an arbitrary countable set, and choose an injection $f : x \to \omega$. Let $y = \mathrm{rng}(f)$, so $f : x \to y$ is a bijection. By the first paragraph, the set y is either finite or has the same cardinality as ω, and so the same has to be true about x. This completes the proof of (1).

For (2), if $f : \omega \to x$ is a surjection of ω onto any set x, then the function $g : x \to \omega$ defined by $g(z) = \min\{n \in \omega : f(n) = z\}$ is an injection of x to ω, confirming that x is countable. On the other hand, if x is countable, then either x is infinite and then x is in fact a bijective image of ω by (1), or x is finite and then it is a bijective image of some natural number n. Any extension of this bijection to a function defined on the whole ω will be a surjection of ω onto x.

For (3), let x be a countable nonempty set and $f : x \to y$ be a surjection. By (2), there is a surjection $g : \omega \to x$ and then $f \circ g$ will be a surjection of ω onto y, confirming the countability of x.

For (4), we will first show (without AC) a special case: the set $\omega \times \omega$ is countable. Indeed, one bijection between $\omega \times \omega$ and ω is the Cantor pairing function defined by $f(n, m) = \frac{1}{2}(n + m)(n + m + 1) + m$. Now, suppose that $b = \{a_i : i \in \omega\}$ is a countable set, all of whose elements are again countable. To show that $\bigcup b$ is countable, we will produce a surjection from $\omega \times \omega$ to $\bigcup b$, which in view of the special case and (3) will show that $\bigcup b$ is countable.

For every $i \in \omega$, let c_i be the set of all surjections from ω to a_i. Since each a_i is assumed to be countable, each set c_i is nonempty. Use the Axiom of Choice to find a selector h: a map with domain ω such that for every $i \in \omega$, $h(i) \in c_i$. Let $f : \omega \times \omega \to \bigcup b$ be the map defined by $f(i, j) = h(i)(j)$. This is the desired surjection. □

Item (4) in its generality cannot be proved without the axiom of choice. Lebesgue, an opponent of AC, used item (4) unwittingly to develop his theory of integration; any such a theory has to use some form of (4).

Theorem 2.4.3. *The following sets are countable:*

1. the set of integers;

2. if x is any countable set then the set $x^{<\omega}$ of all finite sequences of elements of x;

3. the set of rational numbers;

4. *the set of all open intervals with rational endpoints;*

5. *the set of all polynomials with integer coefficients;*

6. *the set of all algebraic numbers.*

Theorem 2.4.4. $|\mathbb{R}| = |\mathcal{P}(\omega)|$.

While we have not developed the real numbers $\mathbb{R}$ formally, any usual concept of real numbers will be sufficient to prove this theorem.

Proof. By the Schröder-Bernstein theorem, it is enough to provide an injection from $\mathbb{R}$ to $\mathcal{P}(\omega)$ as well as an injection from $\mathcal{P}(\omega)$ to $\mathbb{R}$.

To construct an injection from $\mathbb{R}$ to $\mathcal{P}(\omega)$, we will construct an injection from $\mathbb{R}$ to $\mathcal{P}(x)$ for some countable infinite set instead, and finish the argument by Theorem 2.4.2(1). Let x be the set of all open intervals with rational endpoints, so x is countable by Theorem 2.4.3(4). Let $f : \mathbb{R} \to \mathcal{P}(x)$ be the function defined by $f(r) = \{i \in x : r \in i\}$; we claim that this is an injection. Let $r \neq s$ be two distinct real numbers. Then, there is an open interval i with rational endpoints that separates r from s, i.e. $r \in i$ but $s \notin i$. Then $i \in f(r)$ and $i \notin f(s)$, and therefore $f(r)$ and $f(s)$ must be distinct.

To construct an injection from $\mathcal{P}(\omega)$ to $\mathbb{R}$, consider the function $g : \mathcal{P}(\omega) \to \mathbb{R}$ defined by the following formula: $g(y)$ is the unique element of the closed interval $[0, 1]$ whose ternary expansion consists of 0's and 2's only, and n-th digit of the ternary expansion of $g(y)$ is 2 if $n \in y$, and the n-th digit is 0 if $n \notin y$. It is easy to check that this is an injection. □

Corollary 2.4.5. (Cantor) *There is a real number which is not the root of a nonzero polynomial with integer coefficients.*

Proof. The set $\mathcal{P}(\omega)$ is uncountable by Theorem 2.3.5, and so is $\mathbb{R}$. On the other hand, the set of algebraic real numbers is countable. Thus, there must be a real number which is not algebraic. □

The presented proof is incomparably easier than any proof that a specific real number (say π or e) is not algebraic. Also, it does not use almost any knowledge about real numbers.

Exercise 2.4.1. Let x be a countable set. Show that any set consisting of pairwise disjoint subsets of x is countable.

Chapter 3

The transfinite

In this chapter, we will show that the processes of induction and recursion can be extended far beyond ω. It is exactly the use of this extended, transfinite induction and recursion what sets set theory apart from other field of mathematics. While the idea may sound far-fetched at first, it is very powerful and found many uses in mathematics: the equivalence of Zorn's lemma and the Axiom of Choice, the Cantor–Bendixson analysis of closed sets of reals, or the stratification of Borel sets of reals into a hierarchy can serve as good examples.

3.1 Ordinals

We will first define the von Neumann *ordinal numbers*. Ordinals are typically denoted by lower-case Greek letters such as $\alpha, \beta, \gamma \ldots$ The collection of ordinals is itself naturally linearly ordered: given two ordinals α, β then either α is an initial segment of β or vice versa, β is an initial segment of α.

Definition 3.1.1. A set x is an *ordinal number*, or ordinal for short, if it is transitive and linearly ordered by $\in$.

In particular, every natural number as well as ω is an ordinal. There are other ordinals as well, e.g. $\omega \cup \{\omega\}$. If we want to develop the theory of ordinals without the Axiom of Regularity, the definition needs to be amended: a set x is an ordinal if it is transitive, linearly ordered by $\in$, and every subset of x has an $\in$-minimal element (the last clause is automatic if the Axiom of Regularity is present). Every ordinal with the membership relation is a linear ordering as per the definition, and we will always view ordinals as linear orderings.

We will first record the most useful technical properties of ordinals.

Theorem 3.1.2. *Let α be an ordinal.*

1. *Every element of α is an ordinal;*
2. *every $\in$-initial segment is either an element of α or equal to α.*

Proof. For (1), let $\beta \in \alpha$. We have to verify that β is linearly ordered by $\in$ and transitive. For the linearity, observe that $\beta \subseteq \alpha$ by the transitivity of α, and as α is linearly ordered by $\in$, so is β. For the transitivity, suppose that $\gamma \in \beta$ and $\delta \in \gamma$; we must conclude that $\delta \in \beta$. By the transitivity of α, all β, γ, δ are in α. Since $\in$ is a linear ordering on α and $\delta \in \gamma \in \beta$, $\delta \in \beta$ follows as required.

For (2), let $x \subseteq \alpha$ be an $\in$-initial segment of α and $x \neq \alpha$; we must argue that $x \in \alpha$. Let $\beta \in \alpha$ be the $\in$-minimal element of the nonempty set $\alpha \setminus x$ obtained by the Axiom of Regularity. We will show that $\beta = x$, and that will complete the proof as $\beta \in \alpha$.

For the inclusion $\beta \subseteq x$, choose $\gamma \in \beta$ and argue that γ must be an element of x. If this failed, γ would be an element of $\alpha \setminus x$ $\in$-smaller than β, contradicting the minimal choice of β. For the inclusion $x \subseteq \beta$, choose $\gamma \in x$ and argue that γ must be an element of β. If this failed, then by the linearity of $\in$ on the ordinal α, it would have to be the case that $\beta = \gamma$ or $\beta \in \gamma$. The former case is impossible as $\beta \notin x$ and $\gamma \in x$. In the latter case, note that x is an $\in$-initial segment of α and so $\beta \in \gamma$ and $\gamma \in x$ implies $\beta \in x$, again contradicting the assumption that $\beta \notin x$. The proof is complete. □

Now, it is time to prove the more salient features of ordinal numbers.

Theorem 3.1.3. (Linear ordering) *The class of ordinals is linearly ordered by* $\in$.

Proof. It is clear that $\in$ is a transitive relation on ordinals. If $\gamma \in \beta \in \alpha$ are ordinals then, as α is a transitive set, $\gamma \in \alpha$ must hold, and the transitivity of $\in$ has been proved.

For the linearity, let α, β be ordinals; we have to argue that either $\alpha \in \beta$ or $\beta \in \alpha$ or $\alpha = \beta$ holds. To this end, consider the set $\gamma = \alpha \cap \beta$. As an intersection of two transitive sets, it is transitive and therefore an $\in$-initial segment of both α and β.

Note that it must be the case that either $\gamma = \alpha$ or $\gamma = \beta$ holds. If both of these equalities failed, then both $\gamma \in \alpha$ and $\gamma \in \beta$ must hold by Theorem 3.1.2. But then $\gamma \in \alpha \cap \beta$, so $\gamma \in \gamma$ by the definition of γ, and this contradicts the Axiom of Regularity.

Thus, γ is equal to one of the ordinals α, β; say that $\gamma = \alpha$. If $\gamma = \beta$ then we conclude that $\alpha = \beta$. If $\gamma \neq \beta$ then by Theorem 3.1.2(2), $\gamma \in \beta$ and we conclude that $\alpha \in \beta$. In both cases, the linearity of $\in$ is confirmed. □

Corollary 3.1.4. *The class of ordinals is not a set.*

Proof. Assume for contradiction that the class of ordinals is a set x. The set x is transitive, as every element of an ordinal is again an ordinal by Theorem 3.1.2(1). It is linearly ordered by $\in$ by Theorem 3.1.3. Therefore, x is an ordinal, and so $x \in x$, contradicting the Axiom of Regularity. □

Theorem 3.1.5. (Rigidity) *Whenever α, β are ordinals and $i : \alpha \to \beta$ is an isomorphism of linear orders then $\alpha = \beta$ and $i = \mathrm{id}$.*

Proof. Assume that α, β are ordinals and $i : \alpha \to \beta$ is an isomorphism. Suppose for contradiction that i is not the identity. Then, there must be an ordinal $\gamma \in \alpha$ such that $i(\gamma) \neq \gamma$. Use the Axiom of Regularity to choose the $\in$-least ordinal $\gamma \in \alpha$ such that $i(\gamma) \neq \gamma$. Since $i(\gamma) \in \beta$, $i(\gamma)$ is an ordinal by Theorem 3.1.2(1). By the linearity (Theorem 3.1.3), there are three possible cases: either $i(\gamma) = \gamma$, or $i(\gamma) \in \gamma$, or $\gamma \in i(\gamma)$. We will reach a contradiction in each case.

First, $i(\gamma) = \gamma$ is impossible as γ was chosen precisely so that $i(\gamma) \neq \gamma$. Second, assume that $i(\gamma) \in \gamma$. By the minimal choice of γ, the equality $i(i(\gamma)) = i(\gamma)$ must hold. This means that the distinct ordinals γ and $i(\gamma)$ are sent by the isomorphism i to the same value $i(\gamma)$, which is a contradiction. Third, assume that $\gamma \in i(\gamma)$. In this case, since β is a transitive set and it contains $i(\gamma)$, it must contain also its element γ. Let $\delta \in \alpha$ be an element such that $i(\delta) = \gamma$. Since i is an isomorphism of linear orders and $i(\delta) = \gamma \in i(\gamma)$, it must be the case that $\delta \in \gamma$. By the minimality choice of γ, $i(\delta) = \delta \neq \gamma$, a final contradiction. □

What kind of ordinals are there? One infinite ordinal is ω. If α is an ordinal, one can form its *successor*, the ordinal $\alpha \cup \{\alpha\}$. Putting together ω with all the ordinals obtained from ω by iterating the successor operation finitely many times, we obtain the first *limit* ordinal larger than ω. The process can then be repeated, yielding larger nad larger ordinals. In general, we define

Definition 3.1.6. An ordinal α is a *successor ordinal* if there is a largest ordinal β strictly smaller than α. In this case, write $\alpha = \beta + 1$. If α is not a successor ordinal, then it is a *limit ordinal*.

Exercise 3.1.1. Verify that every natural number as well as ω is an ordinal.

Exercise 3.1.2. For every ordinal α there is a limit ordinal β such that $\alpha \in \beta$.

Exercise 3.1.3. For every set x of ordinals there is an ordinal larger than all elements of x. *Hint.* Consider the union of x.

Exercise 3.1.4. Prove Corollary 3.1.4 without the use of the Axiom of Regularity. *Hint.* Apply a Russell's paradox type of reasoning to the "set" of all ordinals.

3.2 Transfinite induction and recursion

The ordinal numbers allow proofs by transfinite induction and definitions by transfinite recursion much like natural numbers allow proofs by induction and definitions by recursion.

Theorem 3.2.1. (Transfinite induction) *Suppose that ϕ is a formula of set theory with parameters. Suppose that $\phi(0)$ holds, and for every ordinal α, $(\forall \beta \in \alpha\ \phi(\beta)) \to \phi(\alpha)$ holds. Then, for every ordinal α, $\phi(\alpha)$ holds.*

In parallel with induction on natural numbers, we will refer to $\phi(0)$ as the *base of induction* and to $(\forall \beta \in \alpha\ \phi(\beta)) \to \phi(\alpha)$ as the *induction step*.

Proof. Suppose for contradiction that there is an ordinal, call it γ, such that $\phi(\gamma)$ fails. Consider the set $x = \{\alpha \in \gamma + 1 : \neg\phi(\alpha)\}$. This is a nonempty set of ordinals, containing at least γ itself. By the Axiom of Regularity, the set x has an $\in$-minimal element α. Then $\forall\beta \in \alpha\ \phi(\beta)$ holds and $\phi(\alpha)$ fails, contradicting the assumptions. □

As in the case of induction on natural numbers, we will refer to the implication $(\forall\beta \in \alpha\ \phi(\beta)) \to \phi(\alpha)$ as the *induction step*. In most transfinite induction arguments, the proof of induction step is divided into the successor case and the limit case according to whether α is a successor or a limit ordinal.

Theorem 3.2.2. (Transfinite recursive definitions) *Suppose that F is a class function such that $F(x)$ is defined for all x. Then there is a unique class function G whose domain is the class of all ordinals and for every ordinal α, $G(\alpha) = F(G \restriction \alpha)$.*

In other words, the equation $G(\alpha) = F(G \restriction \alpha)$ may serve as a valid definition of the class function G.

Proof. We will prove first that for every ordinal β, there is a unique function G_β such that

(*) $\mathrm{dom}(G) = \beta$ and for every ordinal $\alpha \in \beta$, $G(\alpha) = F(G \restriction \alpha)$.

If this fails for some ordinal, then there must be the least ordinal β for which it fails. There are two cases:
Case 1. β is a limit ordinal. In such a case, consider the set $\{G_\gamma : \gamma \in \beta\}$. These functions can indeed be collected into a set by the axiom schema of replacement. It is also the case that if $\gamma \in \beta$ and $\delta \in \gamma$, then $G_\delta = G_\gamma \restriction \delta$ by the uniqueness property of the function G_δ with respect to (*) at δ. Thus, $\bigcup_{\gamma\in\beta} G_\gamma$ is a function with domain β, and it is clearly the unique function satisfying (*). This is a contradiction to the choice of β.
Case 2. β is a successor ordinal, $\beta = \gamma + 1$. In such a case, there is a unique function G_β satisfying (*), namely the function $G_\gamma \cup \langle\gamma, F(G \restriction \gamma)\rangle$. This is again a contradiction to the choice of β.

Now, the function G is defined as follows: $G(\alpha) = x$ if for every $\beta > \alpha$, $G_\beta(\alpha) = x$. This is the only possibility given the uniqueness of the functions G_β, and at the same time this function G works. □

3.3 Applications with choice

In the way of applications of the transfinite recursion procedure, we will state and prove two equivalent restatements of the axiom of choice. The first one is the famous well-ordering principle of Zermelo [13].

Definition 3.3.1. A *well-ordering* is a linear ordering $\leq$ on a set x which in addition satisfies that every nonempty subset $a \subset x$ has a $\leq$-least element, i.e. an element u such that the conjunction $v \in a$ and $v \leq u$ implies $v = u$.

It is clear that every ordinal is a well-ordering: every subset of an ordinal has an $\in$-minimal element by the Axiom of Regularity, and by the linearity this is in fact an $\in$-smallest element. The next theorem shows that up to isomorphism, the ordinals are the only well-orderings out there.

Theorem 3.3.2. *Every well-ordering is isomorphic to a unique ordinal.*

Proof. The uniqueness part follows from the rigidity of ordinals, Theorem 3.1.5. For the existence part, let $\leq$ be a well-ordering on a set x. By transfinite recursion define a class function G on the class of all ordinals by letting $G(\alpha)$ be the $\leq$-least element of the set $x \setminus \mathrm{rng}(G \restriction \alpha)$ if the latter set is nonempty, and $G(\alpha) = \texttt{trash}$ otherwise. We will show that there is an ordinal α such that $G(\alpha) = \texttt{trash}$, and for the least such ordinal α, the function $G \restriction \alpha : \alpha \to x$ is an isomorphism of linear orders.

Suppose for contradiction that there is no ordinal α such that $G(\alpha) =$trash. Then G is an injection from the proper class of all ordinals to the set x. Such an injection does not exist by Exercise 1.8.3. This contradiction proves the existence of an ordinal α such that $G(\alpha) = \texttt{trash}$.

Now let α be the smallest ordinal such that $G(\alpha) = \texttt{trash}$, and consider the function $G \restriction \alpha$. Its domain is equal to α. Its range must be equal to x as this is the only way how $G(\alpha) = \texttt{trash}$ can occur. To conclude the proof, it will be enough to show that $G \restriction \alpha$ preserves the ordering.

Suppose for contradiction that $G \restriction \alpha$ does not preserve the ordering. Then there must be ordinals $\gamma \in \beta \in \alpha$ such that $G(\beta) < G(\gamma)$. But then, $G(\beta) \notin \mathrm{rng}(G \restriction \beta) \supset \mathrm{rng}(G \restriction \gamma)$. Therefore, by the recursive definition of G at γ, the element $G(\beta) \in x$ or something even smaller than it should have been picked as the value of $G(\gamma)$. This is a contradiction. □

Note the use of the Axiom schema of Replacement in the above proof. The theorem cannot be proved without it. The development of ordinals is one of the reasons why Replacement was incorporated into ZFC.

Definition 3.3.3. The *well-ordering principle* is the statement "every set can be well-ordered".

Theorem 3.3.4. (Zermelo) *The following are equivalent on the basis of ZF axioms:*

1. *the Axiom of Choice;*
2. *the well-ordering principle.*

Proof. (1) implies (2) is the more difficult implication. Assume the Axiom of Choice. Let x be an arbitrary set. It is enough to show that there is a bijection between x and an ordinal. Let h be a selector function on $\mathcal{P}(x) \setminus \{0\}$ as guaranteed by the Axiom of Choice. By transifinite recursion define a class function G on ordinals by $G(\alpha) = h(x \setminus \mathrm{rng}(G \restriction \alpha))$ if the set $x \setminus \mathrm{rng}(G \restriction \alpha$ is nonempty, and $G(\alpha) = \texttt{trash}$ otherwise.

There must be an ordinal β such that $G(\beta) = \texttt{trash}$, otherwise G would be an injection from the proper class of all ordinals to the set x. Such injections do not exist though by the result of Exercise 1.8.3. Let β be the smallest ordinal such that $G \restriction \beta = \texttt{trash}$. We will show that $G \restriction \beta$ is a bijection between x and β. This will prove (2).

First of all, $G \restriction \beta$ is a function with domain β by its definition. Its range must be equal to x, since there is no other way how $G(\beta) = \texttt{trash}$ could occur. Finally, $G \restriction \beta$ is an injection. If this failed, there would have to be ordinals $\delta \in \gamma \in \beta$ such that $G(\delta) = G(\gamma) \in x$; however, this contradicts the recursive definition of the value $G(\gamma)$ which cannot belong to $\mathrm{rng}(G \restriction \gamma)$, and therefore cannot be equal to $G(\delta)$.

To prove that (2) implies (1), assume that the well-ordering principle holds. To verify the Axiom of Choice, let x be a collection of nonempty sets. To produce a selector on x, just use the well-ordering principle to find a well-ordering on $\bigcup x$, and let f be the function such that $\mathrm{dom}(f) = x$ and $f(y)$ is the $\leq$-least element of y, whenever $y \in x$. This proves (1). □

Now we come to another equivalent of the Axiom of Choice, the *Zorn's lemma*. It is the most commonly used form of the Axiom of Choice in mathematics, since its use does not require technical tools such as transfinite recursion. Every good Pole will tell you that Zorn's lemma was first discovered by Kuratowski in 1922 [5].

Definition 3.3.5. Let $\langle P, \leq \rangle$ be a partially ordered set.

1. an *upper bound* of a set $A \subset P$ is any element $p \in P$ such that for every $q \in A$, $q \leq p$ holds;
2. an element $p \in P$ is *maximal* if there is no element $q \in p$ strictly larger than p.

Zorn's lemma is the following statement. Whenever $\langle P, \leq \rangle$ is a nonempty partially ordered set such that every linearly ordered subset of P has an upper bound, then P has a maximal element.

Theorem 3.3.6. (Kuratowski) *The following are equivalent on the basis of axioms of ZF set theory:*

1. *Axiom of Choice;*
2. *Zorn's lemma.*

Proof. We will start with (1)→(2) implication. Let P be a partially ordered set. Let $\texttt{trash}$ be a set which is not an element of P. Use the axiom of choice to find a selector h on the set $\mathcal{P}(P) \setminus \{0\}$. By transfinite recursion define a class function G on ordinals by the equation $G(\alpha) = h(a_\alpha)$ where $a_\alpha = \{p \in P : \forall \beta \in \alpha\ G(\beta) < p\}$ if the set $a_\alpha \subset P$ is nonempty, and $G(\alpha) = \texttt{trash}$ otherwise.

As in the previous proofs, there must be an ordinal β such that $G(\beta) = \texttt{trash}$; otherwise, the function G would be an injection from the proper class

of all ordinals to the set P, an impossibility by Exercise 1.8.3. Let β be the $\in$-smallest ordinal such that $G(\beta) = \texttt{trash}$. We will prove that β is a successor ordinal, $\beta = \gamma + 1$ for some γ, and $G(\gamma)$ is a maximal element of P.

To this end, observe that the recursion formula implies that the map $G \restriction \beta$ is a strictly increasing function from β to P–every value of G is larger than all the previous values. As a result, the set $G''\beta \subset P$ is linearly ordered, and by the assumption on the partial ordering P, it has an upper bound p. Note that necessarily $p \in G''\beta$ must hold, because otherwise the set a_β is nonempty, containing at least p, and then $G(\beta)$ would not be equal to $\texttt{trash}$. The only way how $p \in G''\beta$ can occur is that there is a largest ordinal $\gamma \in \beta$, and $G(\gamma) = p$.

To show that p is maximal in P, suppose for contradiction that it is not and that there is a strictly larger element $r \in P$. Then, the set a_β is nonempty, containing at least r, and so $G(\beta)$ would not be equal to $\texttt{trash}$. This contradiction completes the proof of the (1)$\rightarrow$(2) implication.

For the implication (2)$\rightarrow$(1), assume that Zorn's lemma holds. Let x be a set of nonempty sets. To confirm the axiom of choice, we must produce a selector for x. Consider the partially ordered set P of all functions f such that $\mathrm{dom}(f) \subseteq x$, and for all $y \in \mathrm{dom}(f)$, $f(y) \in y$. The ordering on P is inclusion: $f \leq g$ if $f \subseteq g$. Every linearly ordered subset of P has an upper bound: if $a \subset P$ is a collection linearly ordered by inclusion, then $\bigcup a \in P$ is the upper bound. By an application of Zorn's lemma, the partially ordered set P must have a maximal element, call it h. We will show that h is a selector on x.

Indeed, suppose for contradiction that h is not a selector on x. The only way how that can happen is that $\mathrm{dom}(h) \neq x$. Let $y \in x$ be some set not in the domain of h. Let $z \in y$ be an arbitrary element. Consider the set $f = h \cup \{\langle y, z\rangle\}$. It is clear that f is an element of the partially ordered set P, $h \subset f$, and $h \neq f$. This contradicts the maximal choice of h and completes the proof of the theorem. □

Since Zorn's lemma is such a common presence in many mathematical arguments, at least one application of it is called for. Note the typical form of the argument: a complicated object is constructed. The partially ordered set to which Zorn's lemma is applied consists of approximations to such an object, and a maximal approximation (granted by Zorn's lemma) is the object that we want. We will provide one typical example of this procedure.

Definition 3.3.7. Let x be a set. A *filter* on x is a set $F \subset \mathcal{P}(x)$ which is closed under supersets ($\forall y \in F\ \forall z \subseteq x\ y \subseteq z \rightarrow z \in F$) and intersections ($\forall y, z \in F\ y \cap z \in F$), and does not contain an empty set. An *ideal* on x is a set $I \subset \mathcal{P}(x)$ which is closed under subsets and unions, and does not contain x.

It should be clear that the notions of filter and ideal are in a sense dual: if F is a filter on a set x, then $I = \{x \setminus y\colon y \in F\}$ is an ideal on x, and vice versa, if I is an ideal on a set x, then $F = \{x \setminus y\colon y \in I\}$ is a filter on x. A filter typically serves as a measure of largeness of a subset of x, while an ideal serves as a notion of smallness.

Example 3.3.8. The *Fréchet ideal* on an infinite set x is the collection of all finite sets x.

Example 3.3.9. The *density zero ideal* on ω is the set of all sets $a \subset \omega$ whose upper asymptotic density $\limsup_n \frac{|a \cap n|}{n}$ is equal to zero.

In many circumstances, one would like to use a filter on a set x which for every set $y \subset x$ decides whether y is large or small, as in the following definition:

Definition 3.3.10. A filter F on a set x is an *ultrafilter* if for every set $y \subseteq x$, $y \in F$ or $x \setminus y \in F$. The ideal dual to an ultrafilter is a *maximal ideal*.

The catch is, how do we find an ultrafilter? There is a rather obvious and useless type of ultrafilter, the *principal* kind. An ultrafilter F is principal if there is an element $i \in x$ such that $y \in F$ if and only if $i \in y$. Are there any nonprincipal ultrafilters? The axiom of choice yields a positive answer:

Theorem 3.3.11. *(AC) There is a nonprincipal ultrafilter on every infinite set.*

Proof. Let x be an infinite set. Let P be the poset of all filters on x which do not contain any finite sets. The ordering on P is inclusion. We will use Zorn's lemma to produce a maximal element in P. Then, we will show that this maximal element is a nonprincipal ultrafilter.

First, observe that P is a nonempty poset. For this, consider $F = \{y \subseteq x \colon x \setminus y \text{ is finite}\}$. It is easy to check that F is a filter. Since x is infinite, $0 \notin F$. Since the union of finite sets is finite, F is closed under intersections. As a subset of a finite set is finite again, F is closed under supersets. Lastly, since x is infinite, F contains no finite sets.

Second, observe that every linearly ordered set $a \subset P$ has an upper bound. This upper bound is $\bigcup a$. To verify that $\bigcup a$ is indeed an element of P,

- $\bigcup a$ contains no finite sets as no filters in a contain any finite sets;
- to check the closure of a under supersets, let $y \subseteq x$ be an element of $\bigcup a$ and $y \subseteq z$ be a subset of x. Choose $F \in a$ such that $y \in F$. Since F is a filter, $z \in F$ and so $z \in \bigcup a$;
- to check the closure of $\bigcup a$ under intersections, we will finally use linearity of a. Suppose that $y, z \in \bigcup a$ and $F, G \in a$ are such that $y \in F$ and $z \in G$. By linearity of a, either $F \subseteq G$ or $G \subseteq F$ holds. For definiteness, suppose $F \subseteq G$. Then $y \in G$, and since G is a filter closed under intersections, $y \cap z \in G$ and so $y \cap z \in \bigcup a$ as required.

Now, Zorn's lemma shows that the poset P has a maximal element F. Let $x = y \cup z$ be a partition; we will show that either $y \in F$ or $z \in F$.

Claim 3.3.12. *Either $\forall u \in F\ u \cap y$ is infinite, or $\forall u \in F\ u \cap z$ is infinite.*

Proof. If both of the disjuncts failed, then there would be sets $u_y, u_z \in F$ such that $u_y \cap y$ is finite and $u_z \cap z$ is finite. Consider the set $u = u_y \cap u_z$. Since p is closed under intersections, $u \in F$. Since $x = y \cap z$, it must be the case that $u \subset (u_y \cap y) \cup (u_z \cap z)$. This is a union of two finite sets, and therefore finite. This contradicts the assumption that elements of P contain no finite sets. □

Now, one of the disjuncts in the claim must hold; for definiteness assume that $\forall u \in F\ u \cap y$ is infinite. Consider $G = \{v \subseteq z \colon \exists u \in F\ u \cap y \subseteq v\}$. This is a filter containing no finite sets, containing F as a subset, and y as an element. By the maximality assumption, it must be the case that $F = G$. Thus, $y \in F$ as requested. □

Exercise 3.3.1. Let $\langle A, \leq \rangle$ be a well-ordering on a set A. Let $B \subset A$ be any set. Then B equipped by the ordering inherited from A, is again a well-ordering.

Exercise 3.3.2. Let $\leq$ be a linear ordering. The following are equivalent:

1. $\leq$ is a well-ordering;
2. there is no infinite strictly descending sequence $x_0 > x_1 > x_2 > \dots$ in $\leq$.

Exercise 3.3.3. Every filter on a set x can be extended to an ultrafilter.

Exercise 3.3.4. Let $\langle P, \leq \rangle$ be a partial ordering. Show that there is a set $A \subset P$ such that any two elements of A are incomparable in $\leq$ and for every $p \in P$ there is $q \in Q$ such that p, q are comparable.

3.4 Applications without choice

Not all applications of the transfinite induction and recursion involve the axiom of choice. Our first such application yields the cumulative hirearchy of the set-theretic universe.

Definition 3.4.1. If α is an ordinal, let V_α be the set defined by the following recursive formula: $V_{\alpha+1} = \mathcal{P}(V_\alpha)$ and $V_\alpha = \bigcup_{\beta \in \alpha} V_\beta$ if α is limit.

Theorem 3.4.2. *1. Each V_α is a transitive set;*

2. $\alpha \leq \beta$ implies $V_\alpha \subseteq V_\beta$;

3. for every set x there is an ordinal α such that $x \in V_\alpha$.

Proof. For (1), first observe

Claim 3.4.3. *The class of transitive sets is closed under powerset and arbitrary unions.*

Proof. Suppose that x is a transitive set; we must show that $\mathcal{P}(x)$ is transitive. Let $y \subset x$ be any element of $\mathcal{P}(x)$ and $z \in y$ be arbitrary; we must show that $z \in \mathcal{P}(x)$. Clearly, $z \in x$, and since x was transitive to begin with, also $z \subset x$ holds. Therefore, $z \in \mathcal{P}(x)$ as desired.

Suppose that x is a set of transitive sets; we must show that $\bigcup x$ is transitive as well. Let $y \in \bigcup x$ and $z \in y$ be arbitrary; we must show that $z \in \bigcup x$. There must be $u \in x$ such that $y \in u$. Since u is transitive and $z \in y \in u$, it is the case that $z \in u$. Then $z \in \bigcup x$ as desired. □

Now, (1) is proved by induction on α. For the successor step of the induction, suppose that V_α is transitive; we must conclude that $V_{\alpha+1}$ is transitive. Since $V_{\alpha+1} = \mathcal{P}(V_\alpha)$, this follows from the claim. For the limit step, suppose that V_α is limit and the sets V_β for $\beta \in \alpha$ are already known to be transitive. Since $V_\alpha = \bigcup_{\beta\in\alpha} V_\beta$, the transitivity of V_α follows from the claim again. This completes the proof of the first item.

For (2), first observe

Claim 3.4.4. *For every ordinal β, $V_\beta \subseteq V_{\beta+1}$.*

Proof. Let $x \in V_\beta$. Since V_β is transitive by (1), $x \subseteq V_\beta$. Therefore, $x \in \mathcal{P}(V_\beta) = V_{\beta+1}$. □

The argument for (2) now proceeds by induction on β. For the successor step of the induction, suppose that the statement holds for β. To verify it for $\beta + 1$, suppose that $\alpha \leq \beta + 1$ is an ordinal. There are two cases. Either $\alpha = \beta + 1$ in which case certainly $V_\alpha \subset V_{\beta+1}$. Or $\alpha \leq \beta$ in which case $V_\alpha \subseteq V_\beta$ by the induction hypothesis, and $V_\beta \subseteq V_{\beta+1}$ by the claim; together $V_\alpha \subseteq V_\beta$ as desired. For the limit step of the induction, if β is a limit ordinal then $V_\alpha \subseteq V_\beta$ for every ordinal α by the definition of V_β.

The proof of (3) uses the Axiom of Regularity. Let $V = \bigcup_\alpha V_\alpha$. Suppose that the complement of V is a nonempty class. By the Axiom of Regularity for classes (Corollary 2.1.13) applied to the complement of V, there is a set $x \notin V$ such that all its elements are in V. For every $y \in x$ let $\mathtt{rk}(y)$ be the least ordinal α such that $y \in V_\alpha$; this exists as $y \in V$ by the minimal choice of x. By the Axiom of Replacement, $\mathtt{rk}''x \subset \mathrm{ON}$ is a set. By Exercise 3.1.3, there is an ordinal α larger than all ordinals in $\mathtt{rk}''x$. It follows that $x \subseteq V_\alpha$, and so $x \in V_{\alpha+1}$ by the definition of $V_{\alpha+1}$. This contradicts the assumption that $x \notin V$. □

The theorem makes it possible to define, for every set x, the ordinal $\mathtt{rk}(x)$ to be the smallest α such that $x \in V_\alpha$. Note that this is always a successor ordinal. The rank can serve as a rough measure of complexity of mathematical considerations. The theory of finite sets (such as most of finite combinatorics or finite group theory) takes place inside the structure $\langle V_\omega, \in\rangle$. Most mathematical analysis can be interpreted as statements about $V_{\omega+1}$. On the other hand, classical set theory often studies phenomena occurring high in the cumulative hierarchy. The high and low stages of the hierarchy are tied together more closely than one might expect.

The following theorem is a typical application of the transfinite recursion to mathematical analysis. Recall that a *basic open* set of reals is an interval (p, r) with rational endpoints, not including the endpoints. An open set of reals is one which is obtained as a union of some collection of basic open sets, and a closed set is one whose complement is open. A point $x \in A \subset \mathbb{R}$ is *isolated* in the set A if there is an open interval which contains x and no other points of the set A.

Theorem 3.4.5. (Cantor–Bendixson) *Every closed set of reals can be written as a disjoint union of a countable set and a closed set without isolated points.*

In fact, the decomposition is unique, as we will show later.

Proof. Let $C \subset \mathbb{R}$ be a closed set of reals. Use transfinite recursion theorem to define sets C_α for every ordinal α by the following recursive formula: $C_0 = C$, $C_{\alpha+1} = C_\alpha \setminus \{\text{isolated points of } C_\alpha\}$, and $C_\alpha = \bigcap_{\beta\in\alpha} C_\beta$.

Claim 3.4.6. *For every ordinal α, the set C_α is closed, and if $\beta \in \alpha$ then $C_\alpha \subseteq C_\beta$.*

Proof. By transfinite induction on α. At limit stage α, the construction takes an intersection of a collection of closed sets, which then must be closed and smaller than all sets in the intersection. At the successor stage, $C_{\alpha+1} \subseteq C_\alpha$ certainly holds. To prove that $C_{\alpha+1}$ is closed, for every point $x \in C_\alpha \setminus C_{\alpha+1}$ pick an open neighborhood O_x containing only x and no other elements of C_α. Then $C_{\alpha+1} = C_\alpha \setminus \bigcup_x O_x$, and as a difference of a closed set and an open set, the set C_α is closed. □

Now, there must be an ordinal α on which the recursive construction stabilizes in the sense that $C_{\alpha+1} = C_\alpha$. If this were not the case, the map G defined by $G(\alpha) = C_\alpha$ would be an injection from the proper class of all ordinals to the set $\mathcal{P}(R)$, and this is impossible by the result of Exercise 1.8.3. Let α be the smallest ordinal such that $C_{\alpha+1} = C_\alpha$, and let $D = C \setminus C_\alpha$. We will show that $C = C_\alpha \cup D$ is the desired decomposition of C into a closed set without isolated points and a countable set.

First of all, it is clear that the set C_α has no isolated points as $C_{\alpha+1} = C_\alpha \setminus \{\text{isolated points of } C_\alpha\}$ by the recursive definition, and $C_{\alpha+1} = C_\alpha$. To show that the set D is countable, we will construct an injection F from the set D to the countable set of all basic open sets. To define $F(x)$, find the smallest ordinal $\beta_x \in \alpha$ such that $x \in C_{\beta_x} \setminus C_{\beta_x+1}$. Since x is an isolated point of C_{β_x}, there is a basic open set which contains only x and no other elements of the set C_{β_x}. Let $F(x)$ be any such basic open set.

We must verify that the function F is in fact an injection on D. Let $x \neq y$ be distinct points of the set D; we must show that $F(x) \neq F(y)$. By symmetry, we can assume that $\beta_x \leq \beta_y$. Now, $y \in C_{\beta_y}$ by the definition of the ordinal β_y, the set $F(x)$ does not contain any points of the set C_{β_x} except for x, and since $C_{\beta_y} \subseteq C_{\beta_x}$ by the claim, $y \notin F(x)$. On the other hand, $y \in F(y)$ by the choice of the value $F(y)$. It follows that $F(x) \neq F(y)$. □

Exercise 3.4.1. Let x be a set. The following are equivalent:

1. $x \in V_\omega$;
2. $\texttt{trcl}(x)$ is finite.

Exercise 3.4.2. Show that V_ω is a countable set.

Exercise 3.4.3. Show that for every ordinal α, there is a set $x \in V_{\alpha+1}$ which does not belong to V_α.

Exercise 3.4.4. Show that the first stage at which the Cantor–Bendixson analysis of a closed set stabilizes is countable.

Exercise 3.4.5. Show that for every countable ordinal α there is a closed set C of reals such that the Cantor–Bendixson analysis of C does not stabilize before α.

3.5 Cardinal numbers

The purpose of this section is to further develop the theory of cardinalities under the Axiom of Choice. In particular, we will identify a canonical representative for each cardinality, and show that cardinalities are linearly ordered.

Definition 3.5.1. A *cardinal number*, or cardinal for short, is an ordinal number which is not in a bijective correspondence with any ordinal number smaller than it.

In particular, every natural number as well as ω is a cardinal number. In set-theoretic literature, cardinals are typically denoted by lowercase Greek letters such as $\kappa, \lambda, \mu, \ldots$

Theorem 3.5.2. *(AC) Every set is a bijective image of a unique cardinal number.*

Proof. Let x be any set. Let a be the class of all ordinal numbers which are bijective images of x. Observe that a is nonempty: by Zermelo's well-ordering theorem, x can be well-ordered and the well-ordering on it is isomorphic to some ordinal. The isomorphism is then a bijective function between x and the ordinal.

Now, the class a must have an $\in$-least element. Review the definition of a to check that this minimum of a is a cardinal number. This shows that x is in bijective correspondence with some cardinal number. The uniqueness of this cardinal number follows easily: if κ, λ are cardinals such that $|\kappa| = |x| = |\lambda|$, then κ and λ are in a bijective correspondence. This excludes both $\kappa \in \lambda$ and $\lambda \in \kappa$ by the definition of a cardinal number, and by the linearity of ordering of the ordinal numbers (Theorem 3.1.3), $\kappa = \lambda$ is the only option left. □

Corollary 3.5.3. *(AC) Whenever x, y are sets, then either $|x| \leq |y|$ or $|y| \leq |x|$.*

Proof. Let κ, λ be cardinals such that $|\kappa| = |x|$ and $|\lambda| = |y|$. By the linearity of ordering of ordinal numbers–Theorem 3.1.3, either $\kappa \subseteq \lambda$ or $\lambda \subseteq \kappa$ holds. Then, either $|\kappa| \leq |\lambda|$ or $|\lambda| \leq |\kappa|$ holds, as the identity map will be the required injection map. Thus, either $|x| \leq |y|$ or $|y| \leq |x|$ holds as desired. □

Thus, under the axiom of choice, cardinalities are linearly ordered (even well-ordered), and the cardinal numbers are canonical representatives of cardinalities. There is an enormous supply of cardinal numbers, as described in the following theorem:

Theorem 3.5.4. *For every ordinal α there is a cardinal κ such that $\alpha \in \kappa$.*

Proof. There are two possible, quite different proofs. For the first proof, fix an ordinal α. By Theorem 2.3.5, $|\mathcal{P}(\alpha)| > |\alpha|$. By the Axiom of Choice, there is a cardinal number κ such that $|\kappa| = |\mathcal{P}(\alpha)|$. Since $|\alpha| < |\kappa|$, it must be the case that $\alpha \in \kappa$.

The second proof does not use the Axiom of Choice. Consider the class function F from $\mathcal{P}(\alpha \times \alpha)$ to ordinals which maps a set T to α if T is a well-ordering and α is the unique ordinal isomorphic to T, and $F(T) = 0$ if T is not a well-ordering. By the Axiom schema of Replacement, $\mathrm{rng}(F)$ is a set. By Exercise 3.1.3, there is an ordinal β larger than all elements of $\mathrm{rng}(F)$. Let κ be the cardinal such that $|\kappa| = |\beta|$, and argue that $\alpha \in \kappa$. If this failed, then there would have to be an injection from κ to α, also an injection from β to α, and so there would be a well-ordering on a subset of α of ordertype α, contradicting the definition of β. □

Thus, the infinite cardinal numbers can be enumerated by ordinals in an increasing order: $\omega = \omega_0, \omega_1, \omega_2, \ldots, \omega_\omega, \omega_{\omega+1}, \ldots \omega_\alpha \ldots$ Set theoretical literature often makes a conceptual distinction between a cardinal number and the cardinality which that cardinal number represents. The cardinalities are denoted by $\aleph$, pronounced "aleph", the first letter of the Hebrew alphabet. Thus, $\aleph_0$ is the cardinality of ω_0, $\aleph_1$ is the cardinality of ω_1, and $\aleph_\alpha$ is the cardinality of ω_α.

Finally, we come to the formulation of the question which was one of the driving forces behind the development of modern set theory from its beginnings.

Question 3.5.5. (Continuum Hypothesis, CH) Is $|\mathbb{R}| = \aleph_1$? (The continuum problem) Determine the ordinal α such that $|\mathbb{R}| = \aleph_\alpha$. (The generalized continuum problem) For every ordinal α, determine the ordinal β such that $|\mathcal{P}(\omega_\alpha)| = \aleph_\beta$.

It turns out that the continuum problem cannot be resolved in ZFC. There is a good amount of speculation, some primitive and some highly sophisticated, as to what the "right" answer to the continuum problem "should" be. The author recommends a healthy dose of scepticism towards such speculation.

Before we leave the subject of cardinal numbers, we will develop the notion of cofinality:

Definition 3.5.6. Let κ, α be limit ordinals. Say that $\mathtt{cof}(\kappa) = \alpha$, or *the cofinality of κ is equal to α*, if α is the smallest ordinal such that there is a cofinal subset of κ of ordertype α. The ordinal κ is *regular* if $\mathtt{cof}(\kappa) = \kappa$. An ordinal which is not regular is called *singular*.

It is fairly immediate to observe that cofinality of any limit ordinal must be regular, and every regular ordinal is a cardinal. Many cardinals are regular, as becomes obvious from the following theorem:

Theorem 3.5.7. *Every successor cardinal is regular.*

Proof. This theorem requires the axiom of choice for its proof; without the axiom of choice it may even happen that every limit ordinal has cofinality equal to ω. We will just show that ω_1 is regular.

Suppose for contradiction that ω_1 is singular. Then, its cofinality must be equal to $\omega = \omega_0$ and there has to be a function $f : \omega \to \omega_1$ whose range is cofinal in ω_1. Then, $\omega_1 = \bigcup_n f(n)$ is a countable union of countable sets. Such unions are countable by Theorem 2.4.2(4), contradicting the definition of ω_1 as the first uncountable cardinal. □

The theorem immediately suggests a question:

Question 3.5.8. Is there an uncountable limit regular cardinal?

The question was considered by Hausdorff in 1908 and later greatly expanded by Tarski. The question cannot be resolved in ZFC. Limit regular cardinals are called *weakly inaccessible*, and they are the beginning of a hierarchy of *large cardinals* which is one of the main tools of modern set theory.

Chapter 4

Descriptive set theory

The purpose of this chapter is to develop the basics of the theory of definable sets of reals and "similar" spaces. This allows a careful development of all subjects of mathematical analysis such as integration theory and functional analysis.

4.1 Rational and real numbers

Before everything else, we must develop the real numbers in ZFC. This is not difficult, but we will use the opportunity to state and prove several interesting results on the way.

To develop the rational numbers in set theory, consider the set $\mathbb{Z}\times\mathbb{Z}\setminus\{0\}$ and define an equivalence on it: $\langle p_0, q_0\rangle\ E\ \langle p_1, q_1\rangle$ if $p_0q_1 = p_1q_0$. It is not difficult to check that E is indeed an equivalence. Let $\mathbb{Q}$ be the set of all equivalence classes of the relation E. Define the ordering $\leq$ on $\mathbb{Q}$ by setting $\langle p_0, q_0\rangle \leq \langle p_1, q_1\rangle$ if $p_0q_1 \leq p_1q_0$. It is not difficult to verify that $\leq$ is indeed an ordering respecting the equivalence classes. The ordering is countable, dense in itself, and it has no endpoints. Our first result shows that these features of $\mathbb{Q}$ identify it up to isomorphism.

Theorem 4.1.1. *Every countable dense linear order without endpoints is isomorphic to* $\langle\mathbb{Q}, \leq\rangle$.

Proof. The trick used is known as a "back-and-forth argument". Suppose that $\langle P, \leq_P\rangle$ and $\langle R, \leq_R\rangle$ are two dense countable linear orders without endpoints. We must prove that they are isomorphic. Let $\langle p_n : n \in \omega\rangle$ and $\langle r_n : n \in \omega\rangle$ are enumerations of P and Q respectively. By recursion on $n \in \omega$, build partial functions $h_n : P \to R$ such that

- $0 = h_0 \subset h_1 \subset h_2 \subset$;
- all maps h_n are finite injections;
- $p_n \in \mathrm{dom}(h_{2n+1})$ and $r_n \in \mathrm{rng}(h_{2n+2})$ for every $n \in \omega$;

- the maps h_n preserve the ordering: whenever $x <_P y$ are elements of $\mathrm{dom}(h_n)$ then $h_n(x) <_R h_n(y)$.

Once the recursion is performed, let $h = \bigcup_n h_n$. This is a function from P to Q which preserves the ordering, and $\mathrm{dom}(h) = P$ and $\mathrm{rng}(h) = Q$. That is, h is the requested isomorphism of the orderings P and Q.

To perform the construction, suppose that h_{2n} has been found. In the construction of h_{2n+1}, it is just necessary to include p_n in the domain of h_{2n+1}. If $p_n \in \mathrm{dom}(h_{2n})$ then let $h_{2n+1} = h_{2n}$ and proceed with the next stage of the recursion. If $p_n \notin \mathrm{dom}(h_{2n})$, then the construction of h_{2n+1} divides into several cases according to how p_n relates to the finite set $\mathrm{dom}(h_{2n}) \subset P$: ??? □

Exercise 4.1.1. Show that any two countable linear dense orderings with endpoints are isomorphic.

Definition 4.1.2. A linear ordering $\langle P, \leq \rangle$ is *complete* if every bounded subset of P has a *supremum*. That is, whenever $A \subset P$ is a set such that the set $B = \{p \in P : \forall q \in A\ q \leq p\}$ is nonempty, then the set B has a $\leq$-smallest element.

Definition 4.1.3. Let $\langle P, \leq_P$ be a linear ordering. A *completion* of P is a order-preserving map $c : P \to R$ to a complete linear ordering $\langle R, \leq_R \rangle$ such that $c''P \subset R$ is dense.

Theorem 4.1.4. *Every linear ordering has a completion. The completion is unique up to isomorphism.*

Proof. For simplicity of notation, we will consider only the case of dense linear ordering $\langle P, \leq_P \rangle$. First, construct some completion of P. Call a pair $\langle A, B \rangle$ a *Dedekind cut* if $A \cup B = P$, $A \cap B = 0$, for every $p \in A$ and every $q \in B$ $p <_P q$, and A does not have a largest element. Let R be the set of all Dedekind cuts, and define $\langle A_0, B_0 \rangle \leq_R \langle A_1, B_1 \rangle$ if $A_0 \subseteq A_1$.

Claim 4.1.5. $\langle R, \leq_R \rangle$ *is a complete linear ordering.*

Proof. It is immediate that $\leq_R$ is an ordering. The first challenge is its linearity. Suppose that $\langle A_0, B_0 \rangle$ and $\langle A_1, B_1 \rangle$ are Dedekind cuts. We must show that either $A_0 \subseteq A_1$ or $A_1 \subseteq A_0$ holds. If $A_0 = A_1$ then this is clear. Otherwise, one of the sets $A_1 \setminus A_0$ or the set $A_0 \setminus A_1$ must be nonempty. Suppose for definiteness it is the set $A_1 \setminus A_0$, and choose an element $q \in A_1$ which is not in A_0. As $\langle A_0, B_0 \rangle$ is a Dedekind cut, it must be the case that $q \in B_0$ and all elements of A_0 are $<_P$-smaller than q. As $\langle A_1, B_1$ is a Dedekind cut, every element $p <_P q$ must belong to A_1. Therefore, $A_0 \subseteq A_1$. This confirms the linearity of $\leq_R$.

Now, we have to prove that $\leq_R$ is complete. Suppose that $S \subset R$ is a bounded set. Its supremum is defined as the pair $\langle A, B \rangle$ where $A = \bigcup \{A' : \exists B'\ \langle A', B' \rangle \in S\}$ and $B = \bigcap \{B' : \exists A'\ \langle A', B' \rangle \in S\}$. □

Now, we have to produce an order-preserving map $c : P \to R$ such that $c''P \subset R$ is dense. Just let $c(p) = \langle A, B\rangle$ where $A = \{q \in P : q <_P p\}$ and $B = \{q \in P : p \leq_P q\}$. ???

Thus, the map $c : P \to R$ is a completion of the ordering P. The final task is to show that any other completion of P is isomorphic to R. ??? □

Now it makes sense to define $\langle \mathbb{R}, \leq\rangle$ as the completion of $\langle \mathbb{Q}, \leq\rangle$, which is unique up to isomorphism. This is again a linear ordering which has some uniqueness features.

Theorem 4.1.6. *Every linear ordering which is separable, dense with no endpoints, and complete, is isomorphic to* $\langle \mathbb{R}, \leq\rangle$.

At this point, it is possible to introduce a problem which, together with the Continuum Hypothesis, shaped modern set theory. Say that a linear ordering $\langle P, \leq\rangle$ satisfies the *countable chain condition* if every collection of pairwise disjoint open intervals in P is countable. Note that every separable linear ordering P has the countable chain condition: if $D \subset P$ is a countable dense set and A is a collection of pairwise disjoint open intervals of P, for every $I \in A$ use the density of the set D to pick a point $f(I) \in D \cap I$. The function f is then an injection from A to D, showing that A is countable.

Question 4.1.7. (Suslin's problem) Suppose that a linear ordering is separable, dense with no endpoints, complete, and has the countable chain condition. Is it necessarily isomorphic to $\langle R, \leq\rangle$?

It turns out that the answer to the Suslin's problem cannot be decided within ZFC set theory.

4.2 Topological spaces

Many objects in mathematics are equipped with a structure that makes it possible to speak about continuous functions from one object to another–a topology.

Definition 4.2.1. A *topological space* is a pair $\langle X, T\rangle$ where X is a nonempty set and $T \subset \mathcal{P}(X)$ is a collection of subsets of X containing 0 and X and closed under finite intersections and arbitrary unions. The collection T is the *topology* and its elements are referred to as the *open sets*.

Definition 4.2.2. Suppose that $\langle X, T\rangle$ and $\langle Y, U\rangle$ are two topological spaces. A map $f : X \to Y$ is *continuous* if the f-preimages of open subsets of Y are open in X. The map f is a *homeomorphism* if it is a bijection and both f and f^{-1} are continuous maps.

Before we pass to examples, it is useful to note that most topologies are generated from collections of sets called subbases in the following way:

Definition 4.2.3. Let X be a set and $S \subset \mathcal{P}(X)$ be any set. The *topology generated by* S is the set $T = \{O \subset X : O = \bigcup B$ for some set B consisting of finite intersections of elements of $S\} \cup \{0, X\}$. The set S is a *subbasis* of T.

Proposition 4.2.4. *Whenever X is a set and $S \subset \mathcal{P}(X)$, the collection T above is in fact a topology on X.*

Proof. Clearly, $0, X \in T$ by the definition of T. We have to prove that T is closed under arbitrary unions and finite intersections.

The closure under arbitrary unions is immediate. If $U \subset T$ is any set, we must show that $\bigcup U \in T$. Let $B = \{P \subset X : P$ is an intersection of finitely many elements of S such that for some $O \in U$, $P \subset O\}$ It is not difficult to check that $\bigcup B = \bigcup U$ and so $\bigcup U \in T$ as required.

Now, we must show that T is closed under finite intersections. If $U \subset T$ is a finite set, we must show that $\bigcap U \in T$. Let $B = \{P \subset X : P$ is an intersection of finitely many elements of S such that $P \subset \bigcap U\}$. We will show that $\bigcup B = \bigcap U$; this will prove that $\bigcap U \in T$ as required. For the $\bigcup B \subseteq \bigcap U$ inclusion, note that B by definition consists of sets which are subsets of $\bigcap U$. For the $\bigcap U \subseteq \bigcup B$ inclusion, let $x \in \bigcap U$ be an arbitrary point. Since $U \subset T$, for every set $O \in U$ there is a set $P_O \subset O$ which is an intersection of finitely many elements of S and contains the points x. Since U is finite, the set $\bigcap_{O \in U} P_O$ is an intersection of finitely many elements of S, it is in B, and it contains the point x. Ergo, $x \in \bigcup B$. □

Example 4.2.5. The *discrete topology* on a set X is $T = \mathcal{P}(X)$. In other words, every set is open in the discrete topology.

Example 4.2.6. If $\langle L, \leq \rangle$ is a linear ordering, the *order topology* is generated by the subbasis consisting of all sets of the form (p, q) where $p < q$ are elements of L and (p, q) is the *open interval* $\{r : p < r < q\}$.

Example 4.2.7. The *Cantor space* is the set $2^\omega = \{f : \mathrm{dom}(f) = \omega, \mathrm{rng}(f) \subseteq \{0, 1\}\}$, equipped with the topology generated by the subbasis consisting of all sets of the form $\{f \in 2^\omega : f(n) = b\}$ where $n \in \omega$ and $b \in \{0, 1\}$.

Example 4.2.8. The *Baire space* is the set $\omega^\omega = \{f : \mathrm{dom}(f) = \omega, \mathrm{rng}(f) \subseteq \omega\}$, equipped with the topology generated by the subbasis consisting of all sets of the form $\{f \in \omega^\omega : f(n) = m\}$ where $n, m \in \omega$.

Example 4.2.9. The *Stone-Čech compactification of* ω is the following space denoted by $\beta\omega$: its underlying set is the set of all ultrafilters on ω, and the topology is generated by the subbasis consisting of all sets of the form $\{u : a \in u\}$ where $a \subset \omega$ is an arbitrary set.

Other examples of topological spaces are obtained by applying certain operations to preexisting spaces.

Example 4.2.10. Suppose that $\langle X, T \rangle$ is a topological space and $Y \subset X$. The *inherited topology* $T \restriction Y$ is the collection $\{A \cap Y : A \in T\}$.

In this way, we consider for example intervals $[0, 1]$ or $(0, 1) \subset \mathbb{R}$ with the inherited topology as topological spaces.

Example 4.2.11. Suppose that $\langle X_0, T_0\rangle$ and $\langle X_1, T_1\rangle$ are topological spaces. The *product space* is $\langle X_0, \times X_1, U\rangle$ where U is the topology on $X_0 \times X_1$ generated by the subbasis consisting of all sets of the form $O \times P$ where $O \in T_0$ and $P \in T_1$.

In this way, we consider for example the Euclidean spaces $\mathbb{R}$, $\mathbb{R} \times \mathbb{R}$, $\mathbb{R}^n$ for natural number $n \in \omega$ with the product topology. These spaces are pairwise nonhomeomorphic–the proof of this statement was the beginning of the field of *dimension theory.*

Example 4.2.12. Suppose that I is a set and $\langle X_i, T_i\rangle$ for $i \in I$ are topological spaces. The *product space* is the pair $\langle \prod_i X_i, U\rangle$ where $\prod_i X_i = \{f : \mathrm{dom}(f) = I, \forall i \in I f(i) \in X_i\}$ and U is generated by the subbasis consisiting of all sets of the form $\{f \in \prod_i X_i : f(j) \in O\}$ where $j \in I$ is an index and $O \in T_j$ is an open subset of X_j.

The most notorious space obtained in this way is the *Hilbert cube* $[0,1]^\omega$, the product of countably many copies of the interval $[0,1]$.

The following notions are ubiquitous in the treatment of topological spaces:

Definition 4.2.13. Let $\langle X, T\rangle$ be a topological space. A set $D \subset X$ is *dense* in the space if every nonempty open set $O \in T$ contains an element of D.

Definition 4.2.14. A topological space $\langle X, T\rangle$ is *separable* if it contains a countable dense set.

Exercise 4.2.1. Let $\langle X, T\rangle$ be a topological space and let $B \subset X$ be a set. Prove that there is the inclusion-smallest closed set $C \subset X$ which contains B as a subset. C is referred to as the *closure* of the set B, often denoted by $\bar{B}$.

Exercise 4.2.2. Let $\langle X, S\rangle$, $\langle Y, T\rangle$ be topological spaces. Consider the space $X \times Y$ with the product topology. Prove that the *projection function* $f : X \times Y \to X$ given by $f(x,y) = x$ is continuous.

Exercise 4.2.3. Let $\langle X, T\rangle$ be a topological space. Consider the space $X \times X$ with the product topology. Show that the function $f : X \to X \times X$ given by $f(x) = \langle x, x\rangle$ is continuous.

Exercise 4.2.4. Let X_0, X_1, Y_0, Y_1 be topological spaces with their topologies and let $f_0 : X_0 \to Y_0$ and $f_1 : X_1 \to Y_1$ be continuous functions. Conclude that the function $g : X_0 \times X_1 \to Y_0 \times Y_1$ given by $g(x_0, x_1) = (f_0(x_0), f_1(x_1))$ is continuous.

Exercise 4.2.5. Let $\langle X, S\rangle$ and $\langle Y, T\rangle$ be topological spaces, and $f : X \to Y$ be a continuous function. Then f viewed as a subset of $X \times Y$ is a closed subset of $X \times Y$.

Exercise 4.2.6. Let X, Y, Z be topological spaces, and f, g be continuous functions from X, Y respectively to Z. The set $C = \{\langle x, y\rangle \in X \times Y : f(x) = g(y)\}$ is closed. Similarly, if X_n for $n \in \omega$ are topological spaces and $f_n : X_n \to Z$ are continuous functions then the set $C = \{u \in \prod_n X_n : \forall n, m\ f_n(u(n)) = f_m(u(m))\}$ is closed in $\prod_n X_n$.

4.3 Polish spaces

Topological spaces defined in the previous section are quite abstract entities. There are many topological spaces with rather unusual properties. Fortunately, most topological spaces occurring in mathematical analysis are of a much more specific and concrete kind. Their topologies are in a natural sense generated from a notion of distance on the underlying set.

Definition 4.3.1. A *metric* on a set X is a function $d : X^2 \to \mathbb{R}$ such that

1. for every $x, y \in X$, $d(x, y) \geq 0$ and $d(x, y) = 0 \leftrightarrow x = y$;
2. $d(x, y) = d(y, x)$l
3. (the *triangle inequality*) for every $x, y, z \in X$, $d(x, z) \leq d(x, y) + d(y, z)$.

A pair $\langle X, d\rangle$ where d is a metric on X is a *metric space*.

Example 4.3.2. The *discrete metric* on any set X, assigning any two distinct points distance 1, is a metric. The*Euclidean metric* on $\mathbb{R}^n$ is a metric for every n. The *Manhattan metric* is a different metric on $\mathbb{R}^n$, defined by $d(x, y) = \sum_{i\in n} |x(i) - y(i)|$. The unit sphere S^2 in $\mathbb{R}^3$ can be equipped with at least two natural metrics: the metric inherited from the Euclidean metric on $\mathbb{R}^3$, or the Riemann surface metric defined by $d(x, y) =$the length of the shorter portion of the large circle connecting x and y.

Definition 4.3.3. If $\langle X, d\rangle$ is a metric space, $x \in X$ is a point and $\varepsilon > 0$ is a real number, the *open ball* $B(x, \varepsilon)$ is the set $\{y \in X : d(x, y) < \varepsilon\}$ and a *closed ball* $\bar{B}(x, \varepsilon)$ is the set $\{y \in X : d(x, y) \leq \varepsilon\}$. then the *topology generated by* d on the set X is the topology generated by the open balls $B(x, \varepsilon)$ for $x \in X$ and real $\varepsilon > 0$. A topology on the set X is *metrizable* if there is a metric which generates it.

We will often face the following challenge: given a metric d and a topology T on the same set X, decide whether d generates T or not. It turns out that there is a simple criterion for that.

Lemma 4.3.4. *Let X be a set, d be a metric on X and T be a topology on X. Then d generates T if and only if both of the following hold:*

1. *every open ball of the metric d is open in the topology T;*
2. *for every open set $O \in T$ and every $x \in O$ there is a real number $\varepsilon > 0$ such that $B(x, \varepsilon) \subset O$.*

Proof. Suppose on one hand that d generates T; we must prove (1) and (2). For (1), the open balls of the metric d are open in T by the definitions. For (2), suppose that $O \in T$ and $x \in O$; we must find a real number $\varepsilon > 0$ such that $X(x, \varepsilon) \subset O$. Since O is an open set in the topology generated by d, there must be finitely many open balls $B(y_i, \varepsilon_i)$ for $i \in n$ such that $\bigcap_i B(y_i, \varepsilon_i) \subset O$ and

$x \in \bigcap_i B(y_i, \varepsilon_i) \subset O$. Find a real number $\varepsilon > 0$ so small that $d(x, y_i) < \varepsilon_i - \varepsilon$ for every $i \in n$. Then, the triangle inequality shows that $B(x, \varepsilon) \subset B(y_i, \varepsilon_i)$ for every $i \in n$. In other words, $B(x, \varepsilon) \subset \bigcap_{i \in n} B(y_i, \varepsilon_i) \subset O$ as required.

Now suppose that (1) and (2) hold; we must prove that d generates T. Certainly all open balls of the metric are in T by (1). It will be enough to show that every open set $O \in T$ is a union of some collection of metric open balls. Let A be the set of all metric open balls which are subsets of O and argue that $O = \bigcup A$. Certainly, $\bigcup A \subseteq O$ since every set in the collection A is a subset of O. For the opposite inclusion $O \subseteq \bigcup A$, let $x \in O$ be an arbitrary point. Use (2) to find a real number $\varepsilon > 0$ such that $B(x, \varepsilon) \subset O$, and then observe that $B(x, \varepsilon) \in A$ and so $B(x, \varepsilon) \subset \bigcup A$ and $x \in \bigcup A$ as required. □

Among all possible metrics, there is a strongly preferred kind which enables many arguments from abstract analysis.

Definition 4.3.5. Let $\langle X, d \rangle$ be a metric space and let $\langle x_n : n \in \omega \rangle$ be a sequence of elements of X

1. A *limit* of the sequence is a point $y \in X$ such that $\lim_n d(x_n, y) = 0$.
2. the sequence is *Cauchy* if for every real number $\varepsilon > 0$ there is a number $n_\varepsilon \in \omega$ such that for every $n, m \in \omega$ greater than n_ε it is the case that $d(x_n, x_m) < \varepsilon$.

The metric d is *complete* if every Cauchy sequence has a limit.

Definition 4.3.6. A *Polish space* is a topological space $\langle X, T \rangle$ which is separable and completely metrizable.

Example 4.3.7. The Euclidean spaces are Polish as their topology is generated by the Euclidean metric.

Example 4.3.8. The Baire space is Polish. We will consider a *least difference metric* on ω^ω. If $x \neq y \in \omega^\omega$ are two distinct points, just let $\Delta(x, y) = \min\{n \in \omega : x(n) \neq y(n)\}$ and $d(x, y) = 2^{-\Delta(x,y)}$. It is not difficult to verify that d is a complete metric generating the topology of the Baire space.

There is an important point to note here. A Polish space is a topological space. By definition, there must be a complete metric generating its topology. However, there may not be any "canonical" choice of the metric. For example, in the case of the Euclidean spaces, both the Euclidean metric and the Manhattan metric generate the same topology. In the case of the Baire space, the definition of the least difference metric includes the choice of the constant 2. If the constant 2 is replaced by any other real number > 1, then the resulting metric generates the same topology and there is no clear reason for preferring one of these metrics over another. In more complicated spaces, the choice of the metric becomes more obscure still. Thus, the topology is the key feature of the Polish space, as opposed to the metric.

Most Polish spaces in mathematical analysis are obtained by various operations from simpler ones. In these notes, we will discuss only two operations for brevity.

Proposition 4.3.9. *Let $\langle X, T\rangle$ be a Polish space and $C \subset X$ a closed set. Then C with the inherited topology is a Polish space again.*

Proof. Let d be a complete metric on X. Let $d \restriction C$ be the metric d restricted to the points in the set C. It will be enough to show that the metric $d \restriction C$ on the set C is complete and it generates the inherited topology on the set C. □

Example 4.3.10. The two-dimensional sphere $S^2 \subset \mathbb{R}^3$ is a closed subset of $\mathbb{R}^3$ and therefore it is a Polish space with the inherited topology. Similarly for all other closed surfaces in $\mathbb{R}^3$.

Example 4.3.11. The *middle third Cantor set* is the closed set $C \subset \mathbb{R}$ defined as follows. By recursion on $n \in \omega$ define sets $C_n \subset [0,1]$ which are finite unions of closed intervals. The recursive specifications are $C_0 = [0,1]$, and C_{n+1} is obtained from C_n by removing the middle third of every interval which appears in C_n. Let $C = \bigcap_n C_n$. The middle third Cantor set is a closed subset of $\mathbb{R}$ and therefore Polish in the inherited topology.

Exercise 4.3.1. Show that the closed ball $\bar{B}(x,\varepsilon)$ is a closed set, and in fact it is the closure of the open ball $B(x,\varepsilon)$.

Exercise 4.3.2. Show that the Euclidean and Manhattan metric on a Euclidean space generate the same topology.

Exercise 4.3.3. Show that the Euclidean metric on $\mathbb{R}$ generates the order topology on $\mathbb{R}$.

Exercise 4.3.4. Every sequence in a metric space has at most one limit.

Exercise 4.3.5. If a sequence has a limit, then it is Cauchy.

Exercise 4.3.6. Let $\langle X_n, T_n\rangle$ be Polish spaces for every $n \in \omega$ such that the sets X_n are pairwise disjoint. Consider the space $X = \bigcup_n X_n$ equipped with the topology $T = \bigcup_n T_n$. Show that $\langle X, T\rangle$ is Polish.

4.4 Universality theorems

In every class of objects defined in mathematics, one seeks extremal cases: the most complicated or the simplest objects in that class. In the class of Polish spaces, we have the following information.

Theorem 4.4.1. *For every uncountable Polish space X there is a closed set $C \subset X$ which is homeomorphic to the Cantor space.*

In other words, the Cantor space is the simplest uncountable Polish space, as any other Polish space contains a closed copy of it.

Proof. Let X be an uncountable Polish space, let d be a complete metric for it, and let $D \subset X$ be a countable dense set. Write $P = \bigcup\{B(x,\varepsilon)\colon x \in D, \varepsilon \in \mathbb{Q}, B(x,\varepsilon) \text{ is countable}\}$. The set $P \subset X$ is open and (as a countable union of countable sets) it is countable.

By tree induction on $t \in 2^{<\omega}$ build open sets $O_t \subset X$ so that

- $O_0 = X$, $t \subset s$ implies $\bar{O}_t \subset O_s$;
- the diameter of O_t is not greater than $2^{-|t|}$;
- $O_{t^\frown 0} \cap O_{t^\frown 1} = 0$.

To perform the induction, suppose that O_t has been constructed. Pick distinct points $y_0 \neq y_1$ in the uncountable set $O_t \setminus P$ and choose a rational number $\varepsilon > 0$ which is smaller than $d(y_0,y_1)/4$ and such that both balls $B(y_0,3\varepsilon)$ and $B(y_1,3\varepsilon)$ are subsets of O_t. Let $x_0, x_1 \in D$ be points such that both distances $d(x_0,y_0)$ and $d(x_1,y_1)$ are smaller than ε, and let $O_{t^\frown 0} = B(x_0,\varepsilon)$ and $O_{t^\frown 1} = B(x_1,\varepsilon)$. The two open balls are certainly disjoint by the triangle inequality. They are also both uncountable: if one of them, say $B(x_0,\varepsilon)$, was countable, then it would be a subset of the set P, contradicting the fact that it contains the point y_0 which is not in the set P. Finally, the closures of both open balls $B(x_0,\varepsilon), B(x_1,\varepsilon)$ are subsets of O_t by the triangle inequality. This concludes the induction step and the inductive construction.

Now, we will show that for every point $z \in 2^\omega$, the intersection $\bigcap_n O_{z\restriction n}$ contains a single element $h(z) \in X$, and the map $h : 2^\omega \to X$ is a homeomorphism of the Cantor space with $\mathrm{rng}(h)$, and $\mathrm{rng}(h)$ is closed. ???

To show that $\mathrm{rng}(h)$ is closed (i.e. $X \setminus \mathrm{rng}(h)$ is open), let $x \notin \mathrm{rng}(h)$ be an arbitrary point; we must produce a neighborhood of x which is disjoint from $\mathrm{rng}(h)$. For every $\varepsilon > 0$ let $O_\varepsilon = \{y \in X\colon d(x,y) > \varepsilon\}$; thus $O_\varepsilon \subset X$ is an open set. Since $X = \{x\} \cup \bigcup_{\varepsilon>0} O_\varepsilon$, and $x \notin \mathrm{rng}(h)$, it is the case that $2^\omega = \bigcup_\varepsilon h^{-1}O_\varepsilon$. Since the function h is continuous, the sets $h^{-1}O_\varepsilon$ are open. As 2^ω is compact, the open cover $2^\omega = \bigcup_\varepsilon h^{-1}O_\varepsilon$ has a finite subcover, and so there is a positive rational $\varepsilon > 0$ such that $X = h^{-1}O_\varepsilon$. It follows that $B(x,\varepsilon)$ is an open neighborhood of x disjoint from $\mathrm{rng}(h)$ as desired. □

Theorem 4.4.2. *For every compact Polish space X there is a closed subset of $[0,1]^\omega$ homeomorphic to X.*

In other words, the infinite-dimensional Hilbert cube is the most complicated compact Polish space; it contains a closed copy of every other compact Polish space.

Proof. ??? □

Theorem 4.4.3. *Every Polish space is a continuous image of the Baire space ω^ω.*

Proof. Let $\langle X, T\rangle$ be a Polish space, and let d be a complete metric on X generating the topology T. By recursion on $n \in \omega$ build open balls B_t for all $t \in \omega^n$ so that

- $B_0 = X$;
- if $t \subset s$ then the closure of B_t is a subset of B_s;
- $B_t = \bigcup_m B_{t^\frown m}$;
- for every $n > 0$ and every $t \in \omega^n$, the diameter of B_t is $\leq 2^{-n}$.

Suppose for the moment that this construction has been performed. For every $y \in \omega^\omega$ define $f(y)$ to be the unique point in $\bigcap_n B_{y\restriction n}$. We will show that f is a correctly defined continuous function from ω^ω onto X.

First of all, we must prove that for every $y \in \omega^\omega$ the set $\bigcap_n B_{y\restriction n}$ contains exactly one point. There cannot be more than one point in this intersection: if $x \neq y$ were distinct point in it, there would be $n \in \omega$ such that $d(x, y) > 2^{-n}$ and then both x, y cannot fit into the set $B_{y\restriction n+1}$ by ??? above. On the othe hand, if ???

Second, we must show that the function f is continuous.

Third, the function f is onto. Let $x \in X$ be any point; we must produce $y \in \omega^\omega$ such that $x = f(y)$. By induction on $n \in \omega$ we can build sequences $t_n \in \omega^n$ so that $0 = t_0 \subset t_1 \subset t_2 \subset \dots$ and $x \in B_{t_n}$–this is possible by ??? above. Then, let $y = \bigcup_n t_n \in \omega^\omega$ and observe that $x \in \bigcap_n B_{t_n} = \bigcap_n B_{y\restriction n}$ and so necessarily $x = f(y)$.

All that remains to be done is to show that the inductive construction can be done. Suppose that B_t has been constructed. Fix a countable dense set $D \subset X$, and let $\{B_{t^\frown m} : m \in \omega\}$ be an enumeration of the countable set $C = \{B(x, \varepsilon) : x \in D \cap B_t,\ \varepsilon > 0$ is a rational number less than $2^{-|t|+1}$, and $\bar{B}(x, \varepsilon) \subset B_t\}$. It is necessary to verify that the induction hypotheses are satisfied. Only the third item may be problematic. To show that $B_t \subseteq \bigcup_m B_{t^\frown m}$, let $x \in B_t$ be an arbitrary point. Let $\delta > 0$ be a rational number such that $B(x, \delta) \subseteq B_t$. Let $z \in B(x, \delta/4)$ be any element of the set D, and consider the ball $B(z, \varepsilon/2)$. It is not difficult to verify that $B(z, \varepsilon/2) \in C$ and $x \in B(z, \varepsilon/2)$. Thus, $x \in \bigcup_m B_{t^\frown m}$ as desired. □

4.5 Borel sets

Open sets should be viewed as the simplest subsets of topological spaces. We will now develope the notion of a Borel subset of a topological space. Borel sets are more complicated than open, but they still possess many regularity features. The development of most of mathematical analysis (such as Lebesgue measure or Baire category) is impossible without the notion of Borel set. Intuitively, Borel sets are those sets which can be obtained from open sets by a repeated operations of countable union, countable intersection and complement.

Definition 4.5.1. Let X be a set. A set $B \subset \mathcal{P}(X)$ is a *σ-algebra of sets* if it contains $0, X \in B$ and B is closed under countable union, countable intersection, and complement.

For example, $\mathcal{P}(X)$ is a σ-algebra of sets. However, we will be interested in algebras that contain much fewer sets than the full powerset.

Definition 4.5.2. Let $\langle X, T\rangle$ be a topological space. The algebra of *Borel sets* is the inclusion-smallest σ-algebra of subsets of X containing the open sets.

A part of this definition is the statement that among the σ-algebras of subsets of X containing all open sets there indeed is an inclusion-smallest one. To prove this, let $A = \{C : C \text{ is a } \sigma\text{-algebra of subsets of } X \text{ which contains all open sets}\}$ and let $B = \bigcap A$. It will be enough to show that B is a σ-algebra of sets and it contains all open sets; then, it is clearly inclusion-smallest such by virtue of its definition. To see that B is a σ-algebra of sets, note that $0, X$ belong to every $C \in A$ and so they belong to B. We must show that B is closed under complements and countable unions and intersections; it will be enough to check the case of countable unions since the other cases are similar. Suppose that sets $D_n \subset X$ for $n \in \omega$ are in B. To show that $\bigcup_n D_n \in B$, note that for every σ-algebra $C \in A$ and for every $n \in \omega$, $D_n \in C$. Since C is a σ-algebra of sets, $\bigcup_n D_n \in C$. This means that for every $C \in A$, $\bigcup_n D_n \in C$, and so $\bigcup_n D_n \in B$.

The class of Borel sets allows a fine layering into a *Borel hierarchy* defined by transfinite recursion.

Definition 4.5.3. Let X be a Polish space. By transfinite recursion on $\alpha > 0$ define collections $\mathbf{\Sigma}^0_\alpha$ and $\mathbf{\Pi}^0_\alpha$ of subsets of X by the following demands:

1. $\mathbf{\Sigma}^0_1$ is the collection of all open subsets of X, $\mathbf{\Pi}^0_1$ is the collection of all closed subsets of X;
2. $\mathbf{\Sigma}^0_\alpha$ is the collection of all countable unions of sets in $\bigcup_{\beta\in\alpha} \mathbf{\Pi}^0_\alpha$, and $\mathbf{\Pi}^0_\alpha$ is the collection of all complements of sets in $\mathbf{\Pi}^0_\alpha$.

A Venn's diagram type of reasoning immediately shows that the sets in $\mathbf{\Pi}^0_\alpha$ are countable intersections of sets in $\bigcup_{\beta\in\alpha} \mathbf{\Sigma}^0_\alpha$.

Minor typographical points: the indexation of the Borel hierarchy begins with subscript 1 (as opposed to 0) for historical reasons. The role of the superscript 0 is not within the scope of this textbook; still, the superscript must not be omitted. The Greek letters are boldface. Lightface hierarchies exist as well, but again fall out of the scope of this textbook. The class $\mathbf{\Sigma}^0_2$ is often denoted by F_σ and the class $\mathbf{\Pi}^0_2$ is often denoted by G_δ. (F stands for French "fermé", or closed, while G stands for German "Gebiet", or region.) The following theorem captures the main features of the Borel hierarchy.

Theorem 4.5.4. *Let X be a Polish space.*

1. *Whenever $\beta \in \alpha$ are nonzero ordinals, then both $\mathbf{\Sigma}^0_\beta$ and $\mathbf{\Pi}^0_\beta$ are subsets of both $\mathbf{\Sigma}^0_\alpha$ and $\mathbf{\Pi}^0_\alpha$;*

2. *The construction stabilizes at* $\alpha = \omega_1$ *and* $\mathbf{\Sigma}^0_{\omega_1} = \mathbf{\Pi}^0_{\omega_1} = \bigcup_{\alpha\in\omega_1} \mathbf{\Sigma}^0_\alpha$ *is exactly the* σ*-algebra of Borel sets.*

Proof. For (1), the case of $1 = \beta \in \alpha = 2$ is handled separately. It is clear from the definitions that every closed set is F_σ and every open set is G_δ. We must show that every open set is F_σ; Venn's diagram reasoning then shows that every closed set is G_δ, proving the case $1 = \beta \in \alpha = 2$. Let d be a complete metric generating the topology of the space X. Since every open set is a union of countably many open d-balls, it is enough to show that every open ball is F_σ. Let $B(x, \varepsilon)$ be an open ball for some $x \in X$ and a real number $\varepsilon > 0$. Clearly, $B(x, \varepsilon) = \bigcup\{\bar{B}(x, \delta) : \delta > 0 \text{ is a rational number smaller than } \varepsilon\}$, where $\bar{B}(x, \delta)$ is the closed ball around x of radius δ. The right hand side of the equality is a countable union of closed sets, proving the case $1 = \beta \in \beta = 2$. To conclude the proof of (1), the case of $1 \in \beta \in \alpha$ follows immediately from the definitions for $\mathbf{\Sigma}^0_\alpha$ sets. On the $\mathbf{\Pi}$-side, use the fact that the sets in $\mathbf{\Pi}^0_\alpha$ are exactly all the countable intersections of sets in $\bigcup_{\beta\in\alpha} \mathbf{\Sigma}^0_\alpha$.

For (2), I will first show that every stage of the hierarchy consists of Borel sets only. This is proved by induction on α. For $\alpha = 1$, the open sets are Borel by definition, and the closed sets are Borel because they are complements of open (and therefore Borel) sets and the algebra of Borel sets is closed under complements. If $\alpha > 1$ is an ordinal and the sets in all classes $\mathbf{\Sigma}^0_\beta$ and $\mathbf{\Pi}^0_\beta$ for $\beta \in \alpha$ are already known to be Borel, then also sets in the classes $\mathbf{\Sigma}^0_\alpha$ and $\mathbf{\Pi}^0_\alpha$ must be Borel, since they are open as countable unions or intersections of some sets in $\bigcup_{\beta\in\alpha}(\mathbf{\Sigma}^0_\beta \cup \mathbf{\Pi}^0_\beta)$, these sets are Borel by the induction hypothesis, and the algebra of Borel sets is closed under countable unions and intersections.

Now, if we show that $\mathcal{C} = \bigcup_{\alpha\in\omega_1} \mathbf{\Sigma}^0_\alpha$ is a σ-algebra of sets, then (3) will follow by the minimality of the algebra of Borel sets, as the previous paragraph shows that $\mathcal{C} \subseteq \mathcal{B}$. To prove that $\mathcal{C}$ is a σ-algebra, verify the required closure properties one by one. For the closure under complement, suppose that $A \in \mathcal{C}$. Then there is $\alpha \in \omega_1$ such that $A \in \mathbf{\Sigma}^0_\alpha$, so $X \setminus A \in \mathbf{\Pi}^0_\alpha$ by (2), $\mathbf{\Pi}^0_\alpha \subseteq \mathbf{\Sigma}^0_{\alpha+1}$ by (1), and so $X \setminus A \in \mathbf{\Sigma}^0_{\alpha+1} \subseteq \mathcal{C}$ as required. For the closure under countable unions, suppose that A_n for $n \in \omega$ are sets in $\mathcal{C}$. There are ordinals $\alpha_n \in \omega_1$ such that $A_n \in \mathbf{\Pi}^0_{\alpha_n}$. Let $\alpha = \bigcup_n \alpha_n$. The set α, as a union of ordinals is again an ordinal by ???. As a countable union of countable sets, it is a countable ordinal by ??? By (1), all sets A_n for $n \in \omega$ are in $\mathbf{\Pi}^0_\alpha$ and so the union $\bigcup_n A_n$ is in $\mathbf{\Pi}^0_{\alpha+1}$. The closure under countable intersections is proved in a similar fashion. □

A basic feature of the Borel hierarchy is its invariance under continuous preimages.

Theorem 4.5.5. *Let* $\alpha > 0$ *be a countable ordinal. Continuous preimages of* $\mathbf{\Sigma}^0_\alpha$*, resp.* $\mathbf{\Pi}^0_\alpha$ *sets are again* $\mathbf{\Sigma}^0_\alpha$*, resp.* $\mathbf{\Pi}^0_\alpha$*.*

Proof. Let X, Y be Polish spaces and $f : X \to Y$ be a continuous function. The statement of the theorem for f is proved by transfinite induction on α. For the base of the induction, f-preimages of $\mathbf{\Sigma}^0_1$ (open) sets are $\mathbf{\Sigma}^0_1$ (open) by the

definition of continuity. The rest of the induction process proceeds uneventfully using the fact that preimages respect complementation and arbitrary unions: for every set $A \subset Y$, $f^{-1}(Y \setminus A) = X \setminus f^{-1}A$, and for sets $A_n \subset Y$, $f^{-1}(\bigcup_n A_n) = \bigcup_n f^{-1}A_n$. □

A fairly common task in descriptive set theory is the following. Given a Polish space X and its subset $B \subset X$ (typically defined in mathematical analysis), decide whether B is a Borel set, and if it is, identify the smallest ordinal α such that $B \in \mathbf{\Sigma}^0_\alpha$ or $B \in \mathbf{\Pi}^0_\alpha$. This may be quite difficult in many instances. Here, we will limit ourselves to only very basic examples.

Example 4.5.6. Every countable set is F_σ and therefore Borel. This follows from the fact that every singleton in a Polish space is a closed set, and so the countable set is a countable union of singleton closed sets.

Example 4.5.7. The set $\mathbb{Q} \subset \mathbb{R}$ of rational numbers is F_σ, but it is not G_δ. To see why it cannot be written as a countable intersection of open sets, note that if O_n for $n \in \omega$ are open sets containing $\mathbb{Q}$, then they are in fact dense open, and therefore their intersection is uncountable by the Baire Category Theorem.

Example 4.5.8. The set $B = \{x \in \mathbb{R}^\omega : x \text{ converges}\} \subset \mathbb{R}^\omega$ is $\mathbf{\Pi}^0_3$. To see this, note that x converges if and only if it is Cauchy: For every rational $\varepsilon > 0$ there is $k \in \omega$ such that for every $n, m > k$, $|x(n) - x(m)| \leq \varepsilon$. Given n, m, ε, the set $C(n, m, \varepsilon) = \{x \in \mathbb{R}^\omega : |x(n) - x(m)| \leq \varepsilon\}$ is closed in $\mathbb{R}^\omega$. The set B is then written as $B = \bigcap_\varepsilon \bigcup_k \bigcap_{n,m>k} C(n, m, \varepsilon)$. Note how the quantifiers turn into countable unions and intersections.

Exercise 4.5.1. Suppose that B, C are Borel subsets of the respective Polish spaces X, Y. Then $B \times C$ is a Borel subset of the product space $X \times Y$.

Exercise 4.5.2. The Borel sets form the smallest class of sets containing the closed sets and closed under countable unions and countable intersections. In other words, one does not need the complement operation to build the Borel sets from closed sets.

Exercise 4.5.3. Suppose that X, Y are Polish spaces, $\alpha \in \omega_1$ is a countable ordinal, and $B \subset X \times Y$ is a $\mathbf{\Pi}^0_\alpha$ set. Then, for every $x \in X$, the set $\{y \in Y : \langle x, y\rangle \in B\}$ is a $\mathbf{\Pi}^0_\alpha$ as well. Similarly for $\mathbf{\Sigma}^0_\alpha$ sets.

Exercise 4.5.4. The set $\{x \in 2^\omega : \sum\{\frac{1}{n+1} : x(n) = 1\} < \infty\}$ is an F_σ subset of 2^ω.

4.6 Analytic sets

In the previous section, we showed that the collection of Borel sets is closed under several operations, among them the continuous preimages. The closure of Borel sets under continuous images leads to a much larger class of sets, identified by the following definition.

Definition 4.6.1. Let $\langle X, T\rangle$ be a Polish space. A set $A \subset X$ is *analytic* if there is a continuous function $f : \omega^\omega \to X$ such that $A = \mathrm{rng}(f)$.

The terminology should not be confused with the notion of analytic function in complex analysis. The class of analytic functions is often denoted by $\mathbf{\Sigma}^1_1$. A complement of an analytic set is *coanalytic*, and the class of coanalytic sets is often denoted by $\mathbf{\Pi}^1_1$.

The original notation for analytic sets introduced by Lusin was A-sets (as opposed to B-sets, which denoted Borel sets). One of Lusin students, Alexandroff (later an important contributor to the field of topology), assumed that the A stands for his last name, and when Lusin introduced the term "analytic", his feelings were severely hurt. The perceived injustice blew entirely out of proportion and eventually lead to a workplace trial (a common tool of bolshevik terror in the Soviet Union of 1930's) of Lusin for imaginary counterrevolutionary crimes. Lusin narrowly escaped labor camp or worse.

The main properties of the class of analytic sets are captured in the following theorem.

Theorem 4.6.2. *The class of analytic sets contains all closed sets and it is closed under countable unions and intersections.*

The class of analytic sets is *not* closed under complements. This is the main difference between analytic and Borel sets.

Proof. Every Polish space is a continuous image of the Baire space by Theorem 4.4.3, and therefore analytic. A closed subset of a Polish space with the inherited topology is Polish, and therefore analytic as well. For the purposes of the following two paragraphs, this means that a continuous image of a closed set is analytic: if $C \subset X$ is closed and $g : C \to Y$ is a continuous map, then there is a continuous surjection $h : \omega^\omega \to C$, and then $g''C = \mathrm{rng}(g \circ h)$ and so $g''C$ is analytic.

For the countable unions, suppose that X is a Polish space and $A_n \subset X$ are analytic sets for every $n \in \omega$; we must prove that $A = \bigcup_n A_n \subset X$ is an analytic set as well. Use the assumptions to find countably many pairwise disjoint copies Y_n of the Baire space and continuous functions g_n for $n \in \omega$ such that $A_n = \mathrm{rng}(g_n)$. Let Y be the union space $\bigcup_n Y_n$; it is Polish. Let $g : Y \to X$ be the function $g = \bigcup_n g_n$; it is a continuous function and $A = \mathrm{rng}(g)$. Thus, the set A is analytic by the first paragraph of the present proof.

For the countable intersections, suppose that X is a Polish space and $A_n \subset X$ are analytic sets for every $n \in \omega$; we must prove that $A = \bigcap_n A_n \subset X$ is an analytic set as well. Use the assumptions to find continuous funtions $g_n : \omega^\omega \to X$ such that $A_n = \mathrm{rng}(g_n)$ for every $n \in \omega$. Consider the space $Y = (\omega^\omega)^\omega$, the set $C = \{y \in Y : \forall m \in \omega\ f_m(y(m)) = f_0(y(0)\}$ and let $g : C \to X$ be the function defined by $g(y) = g_0(y(0))$. The set $C \subset Y$ is closed by Exercise 4.2.6; the function g is continuous. In view of the first paragraph, it is enough to observe that $A = g''C$. For the left-to-right inclusion, for any $x \in A$, for every $n \in \omega$ pick a point $y(n) \in \omega^\omega$ such that $x = g_n(y_n)$, and observe

that $y = \langle y_n : n \in \omega \rangle \in C$ and $x = g(y)$; therefore $x \in g''C$. For the right-to-left inclusion, if $x \in g''C$ then there is a point $y \in C$ such that $g(y) = x$. By the definition of the set C, this means that for every $n \in \omega$, $x = g_n(y_n)$; thus for every $n \in \omega$ $x \in \mathrm{rng}(g_n) = A_n$, and $x \in A = \bigcap_n A_n$. □

Corollary 4.6.3. *All Borel sets are analytic.*

Proof. In view of Theorem 4.6.2, it is enough to observe that all Borel sets can be obtained from closed sets by a repeated application of the operations of countable unions and intersections. This is the content of Exercise 4.5.2. □

Theorem 4.6.4. *The class of analytic sets is closed under Borel map images and preimages.*

Proof. Suppose that X, Y are Polish spaces, $f : X \to Y$ is a Borel function and $A \subset X, B \subset Y$ are analytic sets; we must show that $f''A \subset Y$ and $f^{-1}B \subset X$ are analytic sets as well. Let $h : \omega^\omega \to X \times Y$ be a continuous map such that $f = \mathrm{rng}(h)$. Here, the function f is understood as a subset of $X \times Y$. Let $h_0 : \omega^\omega \to X$ be the function defined as $h_0(v) =$the first coordinate of $h(v)$. Let $h_1 : \omega^\omega \to X$ be the function defined as $h_1(v) =$the second coordinate of $h(v)$. Thus, both functions h_0, h_1 are continuous.

To show that $f''A$ is an analytic set, let $g : \omega^\omega \to X$ be a continuous map such that $A = \mathrm{rng}(g)$. Consider the set $C \subset \omega^\omega \times \omega^\omega$ defined by $C = \{\langle u, v\rangle : g(u) = h_0(v)\}$. The set C is closed by Exercise 4.2.6 and therefore Polish. A review of the definitions reveals that $f''A = h_1''C$. Therefore, the set $f''A$ is analytic.

To show that $f^{-1}B$ is an analytic set, let $k : \omega^\omega \to Y$ be a continuous map such that $B = \mathrm{rng}(k)$. Let $D \subset \omega^\omega \times \omega^\omega$ be the set defined by $D = \{\langle u, v\rangle : k(u) = h_1(v)\}$. The set D is closed, $f^{-1}B = h_0''D$ and so the set $f^{-1}B$ is analytic. □

Exercise 4.6.1. Suppose that B, C are analytic subsets of the respective Polish spaces X, Y. Then $B \times C$ is an analytic subset of the product space $X \times Y$.

Exercise 4.6.2. Let X, Y be Polish spaces and $A \subset X \times Y$ be an analytic set. The vertical section $A_x = \{y \in Y : \langle x, y\rangle \in A\}$ is an analytic subset of Y for every $x \in X$.

4.7 Lebesgue's mistake

In 1915, Lebesgue wrote a paper containing a wrong assertion: continuous images of Borel sets are Borel. Suslin, a student of Lusin in Moscow, noticed the error and proved several theorems about it. The distinction between analytic and Borel sets turns out to be a crucial issue in descriptive set theory and mathematical analysis, and we will prove several theorems about it. First, we will prove Lebesgue wrong: there are analytic sets that are not Borel.

Theorem 4.7.1. *For every Polish space X, there is a universal analytic set $A \subset \omega^\omega \times X$.*

Here, a set $A \subset \omega^\omega \times X$ is *universal analytic* if it is analytic and for every analytic set $B \subset X$ there is $y \in \omega^\omega$ such that $B = \{x \in X : \langle y, x\rangle \in A\}$.

Proof. Let $C \subset \omega^\omega \times (\omega^\omega \times X)$ be a universal closed set for $\omega^\omega \times X$. Let $A \subset \omega^\omega \times X$ be the projection of C into the first and third coordinates. The set A is a continuous image of the closed set C under a continuous function, therefore it is analytic. We will show that A is in fact a universal analytic set.

Suppose that $B \subset X$ is an analytic set; we must find $y \in \omega^\omega$ such that $B = \{x \in X : \langle x, y\rangle \in A\}$. Let $f : \omega^\omega \to X$ be a continuous function such that $B = \mathrm{rng}(f)$. Since f is a continuous function, it is a closed subset of $\omega^\omega \times X$, and by the universality of the closed set C, there must be a point $y \in \omega^\omega$ such that $h = C_y$. We claim that $B = A_y$; this will complete the proof.

To see that $B \subseteq A_y$, let $x \in B$ be arbitrary; we must show that $\langle y, x\rangle \in A$. Use the fact that $B = \mathrm{rng}(h)$ to find a point $z \in \omega^\omega$ such that $f(z) = x$. Then, $\langle z, x\rangle \in h$, so $\langle y, z, x\rangle \in C$ and so $\langle y, x\rangle \in A$ by the definition of the set A.

To see that $A_y \subseteq B$, let $x \in A_y$ be arbitrary; we must show that $x \in \mathrm{rng}(h)$. Since $\langle y, x\rangle \in A$, there must be a point $z \in \omega^\omega$ such that $\langle y, z, x\rangle \in C$ by the definition of the set A. Since $h = C_y$, this means that $h(z) = x$ and so $x \in \mathrm{rng}(h) = B$ as desired. □

Corollary 4.7.2. *There is a subset of ω^ω which is analytic and not Borel.*

Proof. Let $A \subset \omega^\omega \times \omega^\omega$ be a universal analytic set for ω^ω. Let $B = \{x \in \omega^\omega : \langle x, x\rangle \in A\}$; we will show that this subset of ω^ω is analytic and not Borel.

First of all, the set B is analytic. The function $f : \omega^\omega \to \omega^\omega \times \omega^\omega$ defined by $f(x) = \langle x, x\rangle$ is continuous and $B = f^{-1}A$; thus, the analyticity of B follows from Theorem 4.6.2 (2).

Now, we will show that the complement of B is not analytic. Suppose for contradiction that it is. Then, as $A \subset \omega^\omega \times \omega^\omega$ is a universal analytic set, there would have to be an index $x \in \omega^\omega$ such that $\omega^\omega \setminus B = A_x$. Now, just like in the argument for Russell's paradox, $x \in B$ if and only if $\langle x, x\rangle \in A$ (this is by the definition of the set B) and $\langle x, x\rangle \in A$ if and only if $x \notin B$ (since $A_x = \omega^\omega \setminus B$). Putting the two equivalences together we see that $x \in B \leftrightarrow x \notin B$, which is a contradiction.

Now, it follows immediately that the set B is not Borel. If it were, its complement would be Borel and therefore analytic by Corollary 4.6.3. However, we have just proved that this is not the case. □

The set produced in Corollary 4.7.2 seem to have little mathematical relevance. In fact, there are many analytic non-Borel sets at the heart of mathematical analysis. Here, we will provide one example with proof and several examples without proof.

Let $X = \mathcal{P}(\omega^{<\omega})$ be a Polish space with the standard topology ??? A set $T \subset \omega^{<\omega}$ is a *tree* if it is closed under initial segment: if $s \subset t$ and $t \in T$ then

$s \in T$. An *infinite branch* through a tree T is an infinite subset of T linearly ordered by initial segment.

Theorem 4.7.3. *The set $A = \{T \in X : T$ is a tree with an infinite branch$\} \subset X$ is analytic and not Borel.*

Proof. To see that the set A is analytic ???

To show that the set A is not Borel, we will prove that A is in fact *complete analytic*: for every analytic set $B \subset \omega^\omega$ there is a continuous function $f : \omega^\omega \to X$ such that $B = f^{-1}A$. This means that A cannot be Borel: if A were Borel, then (as the class of Borel sets is closed under continuous preimages by Theorem ???) every analytic subset of ω^ω would be Borel. This contradicts Corollary 4.7.2 though.

Thus, suppose that $B \subset \omega^\omega$ is an analytic set, and let $h : \omega^\omega \to \omega^\omega$ be a continuous function such that $B = \mathrm{rng}(h)$. Define the function $f : \omega^\omega \to X$ by $f(y) = \{t \in \omega^{<\omega} : h \cap ([t] \times [y \restriction |t|]) \neq 0$. We claim that this function works. This is the content of the following two claims.

Claim 4.7.4. *The function f is continuous and its values are always trees.*

Proof. For the continuity of f, for every $t \in \omega^{<\omega}$ and $y \in \omega^\omega$ the status of $t \in f(y)$ depends only on a finite initial segment of y, namely $y \restriction |t|$.

To see that the values of f are always trees, let $y \in \omega^\omega$ be an arbitrary point. To see the closure of $f(y)$ under initial segment, suppose that $s \subset t$ are finite sequences of natural numbers and $t \in f(y)$; we must prove $s \in f(y)$. Since $t \in f(y)$, we have $h \cap ([t] \times [y \restriction |t|]) \neq 0$. At the same time, $[t] \subset [s]$ and $[y \restriction |t|] \subset [y \restriction |s|]$. We conclude that $h \cap ([s] \times [y \restriction |s|]) \neq 0$ and so $s \in f(y)$. □

Claim 4.7.5. *For every $y \in \omega^\omega$, $y \in B \leftrightarrow f(y) \in A$.*

Proof. For the left-to-right implication, use the assumption that $B = \mathrm{rng}(h)$ to find a point $z \in \omega^\omega$ such that $h(z) = y$; we claim that z is an infinite path through the tree $f(y)$ and so $f(y) \in A$. To see this, note that for every number $n \in \omega$, the pair $\langle z, y\rangle$ belongs to both h and $[z \restriction n] \times [y \restriction n]$ and so $z \restriction n \in f(y)$ by the definition of the function f.

For the right-to-left implication, suppose that $f(y) \in A$ and pick an infinite branch z through the tree $f(y)$. Naturally identify z with an element of ω^ω. We will show that $h(z) = y$ and therefore $y \in \mathrm{rng}(h) = B$. To see this, for every number $n \in \omega$ note that $z \restriction n \in f(y)$ and so there are points $z_n, y_n \in \omega^\omega$ such that $h(z_n) = y_n$ and $z_n \in [z \restriction n]$ and $y \in [y \restriction n]$ by the definition of the function f. Now $\lim_n z_n = z$, the function h is continuous and so $h(z) = \lim_n h(z_n) = \lim_n y_n = y$ as desired. □

This completes the proof of the theorem. □

Exercise 4.7.1. Let X be an uncountable Polish space. Show that there is no universal Borel set $B \subset \omega^\omega \times X$, i.e. a Borel set such that for every Borel set $C \subset X$ there is a point $y \in \omega^\omega$ such that $C = \{x \in X : \langle y, x\rangle \in B\}$.

4.8 Suslin's fix

Theorem 4.8.1. (Lusin separation theorem) *Let X be a Polish space and $A_0, A_1 \subset X$ be disjoint analytic sets. Then there are disjoint Borel sets $B_0, B_1 \subset X$ such that $A_0 \subset B_0$ and $A_1 \subset B_1$.*

Proof. For contradiction assume that the requested Borel sets cannot be found. Let $g_0 : \omega^\omega \to X$ and $g_1 : \omega^\omega \to X$ be continuous functions such that $A_0 = \mathrm{rng}(g_0)$ and $A_1 = \mathrm{rng}(g_1)$. By induction on $n \in \omega$ build sequences t_0^n and t_1^n so that

- t_0^n, t_1^n is a sequence of length n consisting of natural numbers;
- $m < n$ implies $t_0^m \subset t_0^n$ and $t_1^m \subset t_1^m$;
- there are no disjoint Borel sets $B_0^n, B_1^n \subset X$ such that $g_0''[t_0^n] \subset B_0^n$ and $g_1''[t_1^n] \subset B_1^n$.

For the base of the induction, the sequences $t_0^0 = t_0^1 = 0$ work as in the third item by our initial assumption. To perform the induction step, suppose that the sequences t_0^n and t_1^n have been found. For contradiction assume that for any two numbers $k, l \in \omega$ the sequences $(t_0^n)^\frown k$ and $(t_1^n)^\frown l$ cannot be used in the next stage of the induction. This means that the third item must fail for them, and there are disjoint Borel sets $B_0^{kl}, B_1^{kl} \subset X$ such that $g_0''[(t_0^n)^\frown k] \subset B_0^{kl}$ and $g_1''[(t_1^n)^\frown l] \subset B_1^{kl}$. But then, the sets $B_0^n = \bigcup_k \bigcap_l B_0^{kl}$ and $B_1^n = \bigcup_l \bigcap_k B_1^{kl}$ are Borel and contradict the third item of the induction hypothesis at n.

Thus, the induction can be performed. Let $y_0 = \bigcup_n t_0^n$ and $y_1 = \bigcup_n t_1^n \in \omega^\omega$. The points $g_0(y_0)$ and $g_1(y_1) \in X$ belong to the respective sets A_0 and A_1. As $A_0 \cap A_1 = 0$, the two points are distinct, separated by some disjoint open sets $O_0, O_1 \subset X$. Since the functions g_0, g_1 are continuous, there must be a number $n \in \omega$ such that $g_0''[t_0^n] \subset O_0$ and $g_1''[t_1^n] \subset O_1$. However, this contradicts the third item of the induction hypothesis at n.

This contradiction proves the theorem. □

Corollary 4.8.2. (Suslin) *Let $B \subset X$ be a subset of a Polish space. The following are equivalent:*

1. *B is Borel;*
2. *both B and $X \setminus B$ are analytic sets.*

Proof. On one hand, if the set B is Borel, then so is its complement, and then both B and $X \setminus B$ as Borel sets are analytic.

On the other hand, if both sets B and $X \setminus B$ are analytic, then they are disjoint analytic sets that can be separated by Borel sets as per Theorem 4.8.1. However, the two sets are also complementary, so the only sets separating them are B and $X \setminus B$ themselves. It follows that B is Borel. □

Chapter 5

First order logic

In this chapter, we will develop the basic theory of first order logic. The first order logic is a formal calculus that mathematicians use to form grammatically correct mathematical expressions and formal derivations of certain expressions from others.

The first order logic is only one of a large family of formal logics. Typically, a formal logic consists of syntax (description of how expressions in its language can be formed), a *formal deduction system* (description of how some expressions can be derived from others), and *semantics* (description of how the formal logic expressions speak about some underlying structures). The most desirable features of a formal logic are soundness and completeness, which say that the formal deduction system proves exactly those expressions which are true of all possible underlying structures. The claim to fame of first order logic resides in the fact that most trained mathematicians nowadays tend to formulate their ideas in it or in a language that is easily equivalent to it. Many other formal logics (modal logic, intuitionist logic etc.) have been developed and play an important role in more specific context, such as ???.

5.1 Propositional logic

To illustrate the concerns of first order logic on a simple example, we will consider the case of classical propositional logic. As is the case for most logics, there are two faces of propositional logic, the syntactical and the semantical, and then there is a completeness theorem tying these two faces together.

To describe the syntax of propositional logic, its language consists of atomic propositions, logical connectives, and parentheses. Atomic propositions are just pairwise distinct symbols such as $A, B, C \ldots$; there must be at least one, there may be finitely or infinitely many of them. The set of logical connectives must be adequate (capable of describing any boolean combination). Common choices are $\neg, \wedge, lor$ (this is often used with Gentzen natural deduction system), $\neg, \rightarrow$ (this is used with Hilbert deduction system, and it is our choice in this book),

and | (Sheffer stroke or NAND, popular in computer science since this single connective is complete; it has a deduction system of its own). The parenthetisation can be handled in a number of satisfactory ways, and we will not be particularly careful about it.

The language of propositional logic can be used to form formulas. Every atomic proposition is a formula; if ϕ, ψ are formulas then $\neg(\phi)$ and $\phi \to \psi$ are formulas; and every formula is obtained by a repeated application of these two rules. We will often prove various proposition by induction on complexity of formulas.

Part of the syntactical face of propositional logic is a choice of formal deduction system. Every deduction system has logical axioms and rules of inference. In this book, we will use Hilbert deduction system. The axioms of this system are described by the following list. If ϕ, ψ, χ are formulas, then the following are axioms of Hilbert deduction system:

A1. $\phi \to \phi$

A2. $\phi \to (\psi \to \phi)$

A3. $(\phi \to (\psi \to \chi)) \to ((\phi \to \psi) \to (\phi \to \chi))$

A4. $(\neg\psi \to \neg\phi) \to (\phi \to \psi)$.

The only rule of inference is modus ponens: from ϕ and $\phi \to \psi$ we are allowed to infer ψ. A formal proof from a set Γ of formulas is a finite sequence of formulas $\phi_0, \phi_1, \ldots \phi_n$ such that each of the formulas is either a logical axiom, an element of Γ, or a formula derived by modus ponens from the previous formulas. We write $\Gamma \vdash \phi$ (and read Γ proves ϕ) if there is a formal proof from Γ in which ϕ appears. ϕ is a theorem of propositional logic if $0 \vdash \phi$.

The semantics of propositional logic uses truth assignments. An atomic truth assignment is any map V from the set of atomic propositions to a two element set $\{T, F\}$. A truth assignment is a function V from the set of all formulas to $\{0, 1\}$ such that

- whenever ϕ is a formula and $V(\phi) = 0$ then $V(\neg\phi) = 1$. If $V(\phi) = 1$ then $V(\neg\phi) = 0$;
- whenever ϕ, ψ are formulas then $V(\phi \to \psi) = 0$ if and only if $V(\phi) = 1$ and $V(\psi) = 0$.

We write $\Gamma \models \phi$ (and read ϕ is a tautological consequence of Γ) if for every truth assignment V, if $V(\psi) = T$ for every formula $\psi \in \Gamma$ then $V(\phi) = T$. ϕ is a tautology if $0 \models \phi$.

The completeness theorem for every type of logic will assert something to the effect that relations $\vdash$ and $\models$ are the same. In the case of propositional logic, this is indeed true:

Theorem 5.1.1. (Completeness theorem for propositional logic) *Whenever Γ is a set of formulas and ϕ is a formula, then $\Gamma \vdash \phi$ if and only if $\Gamma \models \phi$.*

The proof of the completeness theorem will be preceded by a number of lemmas, each of which is interesting in its own right.

Lemma 5.1.2. (Deduction) *Suppose that* Γ *is a set of formulas and* ϕ, ψ *are formulas.* $\Gamma \vdash \phi \to \psi$ *if and only if* $\Gamma, \phi \vdash \psi$.

Proof. The left-to-right implication is an immediate application of modus ponens. The right-to-left implication is more difficult. Suppose that $\Gamma, \phi \vdash \psi$, and let $\langle \theta_i : i \leq n \rangle$ be the formal proof of ψ from Γ, ϕ. We will rewrite it to get a formal proof of $\phi \to \psi$ from Γ. Each formula θ_i will be replaced by several formulas according to the following cases. In each case, a formula of the form $\phi \to \theta_i$ will appear in the rewritten proof.
Case 1. If θ_i is a formula in Γ or a logical axiom, replace θ_i with the statements $\theta_i \to (\phi \to \theta_i)$ (logical axiom), θ_i, and $\phi \to \theta_i$ (modus ponens).
Case 2. If $\theta_i = \phi$ then replace θ_i by $\phi \to \phi$ (logical axiom).
If θ_i is obtained by modus ponens from some previous formulas θ_j and $\theta_k = \theta_j \to \theta_i$ for some $j, k < i$, then replace θ_i with $(\phi \to (\theta_j \to \theta_i)) \to (\phi \to \theta_j) \to (\phi \to \theta_i))$ (logical axiom), $(\phi \to \theta_j) \to (\phi \to \theta_i)$ (modus ponens), and $\phi \to \theta_i$ (modus ponens).

This completes the argument. □

Definition 5.1.3. A set Γ of formulas is *contradictory* or *inconsistent* if there is a formula ϕ such that $\Gamma \vdash \phi$ and $\Gamma \vdash \neg\phi$. Otherwise, Γ is consistent.

Lemma 5.1.4. *Let* Γ *be an inconsistent theory. Then for every formula* ϕ, $\Gamma \vdash \phi$.

Proof. Fix a formula θ such that Γ proves both θ and $\neg\theta$. Concatenate the two formal proofs and adjoin the following formulas: $\neg\theta \to (\neg\phi \to \neg\theta)$ (axiom) $\neg\phi \to \neg\theta$ (modus ponens) $(\neg\phi \to \neg\theta) \to (\theta \to \phi)$ (axiom) $\theta \to \phi$ (modus ponens) ϕ (modus ponens). □

Lemma 5.1.5. (Proof by contradiction) *If* Γ *is a set of formulas and* ϕ *is a sentence,* $\Gamma \vdash \phi$ *if and only if* $\Gamma, \neg\phi$ *is contradictory.*

Proof. For the right-to-left implication, suppose that $\Gamma, \neg\phi$ is contradictory. By Lemma 5.1.4, $\Gamma, \neg\phi \vdash \neg(\phi \to (\phi \to \phi))$. By Lemma 5.1.2, $\Gamma \vdash \neg\phi \to \neg(\phi \to (\phi \to \phi))$. Adjoin to this formal proof the following formulas. $\neg\phi \to \neg(\phi \to (\phi \to \phi)) \to ((\phi \to (\phi \to \phi)) \to \phi)$ (axiom) $(\phi \to (\phi \to \phi)) \to \phi$ (modus ponens) $\phi \to (\phi \to \phi)$ (axiom) ϕ (modus ponens). This demonstrates that $\Gamma \vdash \phi$.

For the left-to-right implication of the lemma, if $\Gamma \vdash \phi$ then also $\Gamma, \neg\phi \vdash \phi$ and so $\Gamma, \neg\phi$ is contradictory, as it proves both ϕ and $\neg\phi$. □

Lemma 5.1.6. (Proof by cases) *If* Γ *is a set of formulas and* ϕ, ψ *are sentences, if both* $\Gamma, \phi \vdash \psi$ *and* $\Gamma, \neg\phi \vdash \psi$ *hold, then* $\Gamma \vdash \psi$ *holds.*

Proof. Assume that both $\Gamma, \phi \vdash \psi$ and $\Gamma, \neg\phi \vdash \psi$ hold. By Lemma 5.1.5, it is enough to show that $\Gamma, \neg\psi$ is contradictory. It is clear that $\Gamma, \neg\psi, \neg\phi$ is contradictory, since it proves both ψ (as $\Gamma, \phi \vdash \psi$) and $\neg\psi$ (assumption). By Lemma 5.1.5, $\Gamma, \neg\psi \vdash \phi$. Now, as $\Gamma, \phi \vdash \psi$, Lemma 5.1.2 shows that $\Gamma \vdash \phi \to \psi$. By modus ponens $\Gamma, \neg\psi \vdash \psi$, and so $\Gamma, \neg\psi$ is contradictory as desired. □

Definition 5.1.7. A theory Γ is *complete* if for every formula ϕ, either $\phi \in \Gamma$ ot $\neg\phi \in \Gamma$ holds.

Lemma 5.1.8. (Lindenbaum's theorem) *Every consistent theory can be extended into a complete consistent theory.*

Proof. We will treat only the case where there are only countably many atomic propositions. In such a case, there are only countably many formulas, and we can list them as $\langle \phi_n : n \in \omega \rangle$.

Let Γ be a consistent theory. By induction on $n \in \omega$ build theories Γ_n such that

- $\Gamma = \Gamma_0 \subseteq \Gamma_1 \subseteq \Gamma_2 \subseteq \dots$ and each theory Γ_n is consistent;
- $\phi_n \in \Gamma_{n+1}$ or $\neg\phi_n \in \Gamma_{n+1}$ holds.

The construction of Γ_{n+1} from Γ_n uses the proof by cases lemma. We claim that for at least one of $\Gamma_n \cup \{\phi_n\}$, $\Gamma_n \cup \{\neg\phi_n\}$ is a consistent theory, which can then serve as Γ_{n+1}. Suppose for contradiction that both of these theories are inconsistent. By Lemma 5.1.4, for any fixed formula θ they both prove both θ and $\neg\theta$. Bby Lemma 5.1.6, Γ_n proves both θ and $\neg\theta$. This means that Γ_n is inconsistent, contradicting the induction hypothesis.

After the induction has been performed, let $\Delta = \bigcup_n \Gamma_n$. This is certainly a complete theory by the second item of the induction hypothesis. It is also consistent: any putative proof of inconsistency from Δ uses only finitely many formulas from Δ, which then must all be included in some Γ_n for some $n \in \omega$. This contradicts the consistency of the theory Γ_n. □

Complete consistent theories have one key feature: if a formula is provable from such a theory then it belongs to it, as its negation cannot be provable by consistency and so does not belong to Γ.

Definition 5.1.9. A truth assignment V is a *model* of a theory Γ if $V(\phi) = 1$ for every $\phi \in \Gamma$.

Lemma 5.1.10. *A theory Γ is consistent if and only if it has a model.*

Proof. For the right-to-left direction, suppose that V is a model of Γ. To show that Γ is consistent, we will argue that every formula ϕ which occurs on a formal proof from Γ satisfies $V(\phi) = 1$. In such a case Γ cannot be inconsistent, since a formula and its negation have opposite truth values in V. So, let $\phi_0, \phi_1, \dots \phi_n$ be a formal proof from Γ and by induction on $i \leq n$ proof that $V(\phi_i) = 1$. At stage i of the induction, there are several cases. Either $\phi_i \in \Gamma$ and then

$V(\phi_i) = 1$ by the assumptions. Or, ϕ_i is an axiom of logic, in which case we easily check that all axioms of logic are tautologies and $V(\phi_i) = 1$ again. Or, ϕ_i is obtained via modus ponens from some ϕ_j and $\phi_k = \phi_j \to \phi_i$ for some $j, k < i$. In this case, as $V(\phi_j) = V(\phi_k) = 1$ by the inductive assumption, $V(\phi_i) = 1$ as desired again. This completes the proof of the right-to-left direction.

For the left-to-right direction, assume that Γ is a a consistent theory. Expand Γ to a complete consistent theory and by a slight abuse of notation call this possibly larger theory Γ again. Let V be the function from the set of all formulas to $\{0, 1\}$ defined by $V(\phi) = 1$ if and only if $\phi \in \Gamma$. We claim that V is a model of Γ; for this, it is just enough to confirm that V is indeed a truth assignment. The verification of the requisite truth assignment properties breaks into cases.

- if $V(\phi) = 1$ then we should verify that $V(\neg\phi) = 0$. Since $\phi \in \Gamma$, $\neg\phi \notin \Gamma$ by the consistency of Γ, and so indeed $V(\neg\phi) = 0$.
- if $V(\phi) = 0$ then we should verify that $V(\neg\phi) = 1$. Since $\phi \notin \Gamma$, $\neg\phi \in \Gamma$ by the completeness of Γ, and so indeed $V(\neg\phi) = 1$.
- if $V(\psi) = 1$ and ϕ is a formula then it should be the case that $V(\phi \to \psi) = 1$. The following formulas are in Γ: ψ (assumption), $\psi \to (\phi \to \psi)$ (axiom of logic), $\phi \to \psi$ (modus ponens). So $V(\phi \to \psi) = 1$ as required.
- if $V(\phi) = 0$ and ψ is a formula then it should be the case that $V(\phi \to \psi) = 1$. Here, the following formulas belong to Γ: $\neg\phi$ (assumption plus the second item) $\neg\phi \to (\neg\psi \to \neg\phi)$ (axiom of logic) $\neg\psi \to \neg\phi$ (modus ponens) $(\neg\psi \to \neg\phi) \to (\phi \to \psi)$ (axiom of logic) $\phi \to \psi$ (modus ponens). Thus $V(\phi \to \psi) = 1$ as desired.
- if $V(\phi) = 1$ and $V(\psi) = 0$ then it should be the case that $V(\phi \to \psi) = 0$. Here, if $\phi \to \psi \in \Gamma$, then also $\phi \in \Gamma$ (assumption) and so $\psi \in \Gamma$ (modus ponens), contradicting the assumption. So $\phi \to \psi \notin \Gamma$ and $V(\phi \to \psi) = 0$ as desired.

□

The completeness theorem for propositional logic follows. If Γ is a theory and ϕ is a formula, then the following are equivalent:

- $\Gamma \models \phi$;
- $\Gamma, \neg\phi$ has no model;
- $\Gamma, \neg\phi$ is inconsistent;
- $\Gamma \vdash \phi$.

The equivalence of the first two items follows from the definition of a model. The second and third items are equivalent by Lemma 5.1.10, and the third and fourth item are equivalent by the lemma on proof by contradiction.

Exercise 5.1.1. Without the use of the completeness theorem, prove that for every formula ϕ, $\phi \vdash \neg\neg\phi$ and $\neg\neg\phi \vdash \phi$. *Hint.* Use proof by cases.

Exercise 5.1.2. (Compactness theorem for propositional logic) Let Γ be a theory. Γ has a model if and only if every finite subset of Γ has a model.

5.2 Syntax

The language of a first order logic consists of several types of symbols.

- variables. There are infinitely many of them;
- equality symbol. The interest in languages without equality symbol is limited;
- the universal quantifier $\forall$. One can equivalently use existential quantifier $\exists$ or both;
- logical connectives. Our choice is again $\neg, \rightarrow$;
- parentheses;
- special functional or relational symbols. Each symbol has a fixed arity. 0-ary functional symbols are called constants.

The language of first order logic can be used to form terms and formulas. A variable is a term; if a functional symbol f has arity n and $t_0, t_1, \ldots t_{n-1}$ are terms, then $f(t_0, t_1, \ldots t_{n-1})$ is a term; and all terms are obtained by repeated application of these two rules. If t, s are terms then $t = s$ is a formula; if R is a relational symbol of arity n and $t_0, t_1, \ldots t_{n-1}$ are terms, then $R(t_0, t_1, \ldots t_{n-1})$ is a formula; if ϕ, ψ are formulas then $(\phi) \rightarrow (\psi)$ and $\neg(\phi)$ are formulas; if ϕ is a formula and x is a variable then $\forall x\ (\phi)$ is a formula; and all formulas are obtained by a repeated application of the previous rules.

We will have to pay closer attention to variables in formulas. If ϕ is a formula containing as a subformula the expression $\forall x\ \psi$, then ψ is called the *scope* of the quantifier $\forall x$ and every occurence of x inside a scope of a quantifier $\forall x$ is called *bounded*. An occurence of x is *free* if it is not bounded. x is *free* in ϕ if it has a free occurence in ϕ. A sentence is a formula with no free variables. A list of free variables of a formula is often appended to it in parentheses: the expression $\phi(\vec{x})$ intends to say that ϕ is a formula, $\vec{x}$ is a finite list of variables which includes all free variables of ϕ.

The process of term substitution (plugging in) is common in first order logic. If t is a term then $\phi(t/x)$ denotes the formula obtained from ϕ by replacing all free occurences of x with t. Similar notation applies to plugging in a list of terms into a list of variables of the same length: $\phi(\vec{t}/\vec{x})$. A substitution is *proper* if no variables occurring in the substituted terms become bounded in ϕ. We will have no opportunity to consider any other substitutions besides proper ones.

We will use the Hilbert–Ackermann deduction system for first order logic; a close competitor is the Gentzen natural deduction system. The Hilbert–Ackermann deduction system has many logical axioms. The first group of axioms deals only with logical connectives.

A1. $\phi \to \phi$

A2. $\phi \to (\psi \to \phi)$

A3. $(\phi \to (\psi \to \chi)) \to ((\phi \to \psi) \to (\phi \to \chi))$

A4. $(\neg\psi \to \neg\phi) \to (\phi \to \psi)$.

The second group of axioms shows the interaction between the universal quantifier and other expressions.

A5. $(\forall x\ \phi) \to \phi(t/x)$ whenever t is a term that can be substituted properly to x in ϕ

A6. $(\forall x\ \phi \to \psi) \to ((\forall x\phi) \to (\forall x\psi))$

A7. $\phi \to \forall x\ \phi$ if x is not a free variable of ϕ.

The third group of axioms describes the behavior of equality.

A8. $x = x$ for every variable x;

A9. $(x = y) \to (\phi(x/z) \to \phi(y/z)$ if x, y can be substituted properly to x.

Finally, every formula obtained from the previously mentioned logical axiom by preceding it with any string of universal quantifications is again an axiom of logic.

Let Γ be a set of formulas. A *formal proof* from Γ is a finite sequence of formulas ϕ_m for $m < n$ such that every entry on this sequence is either a formula from Γ, an axiom, or else it is obtained from the previous formulas on the sequence via modus ponens. If ϕ is a formula, write $\Gamma \vdash \phi$ (Γ *proves* ϕ) if there is a formal proof from Γ which contains ϕ. We write $\Gamma \vdash \phi$ if there is a formal proof from Γ on which ϕ appears. ϕ is said to be a theorem of logic if $0 \vdash \phi$.

A first order theory is a set of sentences in a fixed language. There are many first order theories of interest to mathematicians, some of them simple, others very complicated. Given a theory, the most commonly asked question is whether it is consistent, and if so, if one can recognize the theorems (formally provable sentences) of it with a computer algorithm.

Example 5.2.1. The theory of dense linear order without endpoints has a language with a single binary relational symbol $\leq$ and the following axioms:

- $\forall x \forall y \forall z\ x \leq y \wedge y \leq z \to x \leq z$, $x \leq y \wedge y \leq x \to y = x$, $x \leq y \vee y \leq x$;

- $\forall x \forall y \; x < y \rightarrow \exists z \; x < z < y$;
- $\forall x \exists z \; z < x \wedge \exists z \; x < z$.

The theory of dense linear order without endpoints has the pleasing property of being complete–i.e. for every sentence in its language, it either proves the sentence or its negation. As a consequence, there is a computer algorithm which decides whether a given sentence is a theorem of the theory or not.

Example 5.2.2. The theory of groups has a language with a binary functional symbol for multiplication, a unary symbol for inverse, and a constant symbol for the unit. The axioms are

- $\forall x \forall y \forall z \; x(yz) = (xy)z$;
- $\forall x \; x1 = 1x = x$;
- $\forall x \; xx^{-1} = x^{-1}x = 1$.

Despite the terminology, one should not get the impression that mathematicians working in group theory just prove sentences of this first order formal theory. In fact, their work mostly concentrates on properties of groups that are not expressible in such a simple language.

Example 5.2.3. The theory of real closed fields is designed to capture the first order properties of the real line with addition and multiplication. It has constant symbols $0, 1$, binary relational symbol $\leq$, and binary functional symbols $+, \cdot$. The axioms say

- $+, \cdot$ form a field: i. e. $+$ is a commutative group operation with neutral element 0, $\cdot$ is a group operation on the nonzero elements with neutral element 1, and $\forall x \forall y \forall z \; (x + y)z = xy + xz$;
- $\leq$ is a linear ordering and it is a group ordering vis-a-vis addition: i.e. $\forall x, y \geq 0 \; x + y \geq 0$;
- every polynomial of odd degree has a root. This is a collection of infinitely many axioms, one for each odd number. For example, for cubic polynomials we have the sentence $\forall y_0, y_1, y_2, y_3$ if $y_3 \neq 0$ then there is x such that $y_3xxx + y_2xx + y_1x + y_0 = 0$.

A classical theorem of Tarski [11] shows that (among other things) the theory of real closed fields is complete. There is an algorithm which checks whether a given sentence is a theorem of the theory of real closed fields which runs in double exponential time in the length of the sentence [2], and this is best possible [3].

Example 5.2.4. The Peano Arithmetic is a first order theory which records our intuitions about natural numbers. It has functional special symbols for 0, successor, addition, multipication, and exponentiation, and a special relational symbol for the ordering. The axioms are

- $\leq$ is an ordering with least element 0, Sx (the successor of x) is the least element larger than x, and every element larger than zero has a predecessor;
- $\forall x \forall y\ S(x+y) = x + Sy$, $x + xy = x(Sy)$, and similar statement for exponentiation;
- the induction scheme: whenever $\phi(x)$ is a formula, the following is an axiom: $(\phi(0) \wedge (\forall x\ (\phi(x) \to \phi(Sx))) \to \forall x\ \phi(x)$.

There is no computer algorithm that correctly recognizes theorems of Peano Arithmetic. Ergo, this theory is much more complicated than the previous examples.

Example 5.2.5. Zermelo–Fraenkel set theory is a first order theory.

Thus, essentially all of modern mathematics can be formulated within the scope of a fixed first order theory. Still, it is interesting to study other theories as well– in a more restrictive context there may be more information available.

5.3 Semantics

Let $\mathfrak{L}$ be a language of first order logic. This is to say, $\mathfrak{L}$ specifies the special functional and relational symbols with their arities that we want to use. Let R_i, F_j be the relational and functional symbols of $\mathfrak{L}$ for indices i coming from some index sets I, J. An $\mathfrak{L}$-model (or $\mathfrak{L}$-structure) is a tuple $\mathfrak{M} = \langle M, R_i^{\mathfrak{M}} : i \in I, \dots F_j^{\mathfrak{M}} : j \in J \rangle$ where M is a set (the universe of the model $\mathfrak{M}$), for each $i \in I$ $R_i^{\mathfrak{M}}$ is a relation on M of the same arity as R_i (the realization of R_i in $\mathfrak{M}$), and for each $j \in J$ $F_j^{\mathfrak{M}}$ is a function on M of the same arity as F_j (the realization of F_j in $\mathfrak{M}$).

Given a term $t(\vec{x})$ and a list $\vec{m}$ of elements of the universe M of the same length as the list $\vec{x}$ of variables of the term t, we may substitute and get another element $t^{\mathfrak{M}}(\vec{m}/\vec{x})$ of the set M. This is defined by induction on the complexity of the term t as follows:

- if $t = x$ then $t(m/x) = m$;
- if $t = F_j(t_0, \dots)$ then $t^{\mathfrak{M}} = F_j^{\mathfrak{M}}(t_0^{\mathfrak{M}}(\vec{m}/\vec{x}) \dots)$.

Given a formula $\phi(\vec{x})$ and a list $\vec{m}$ of elements of the universe M of the same length as the list $\vec{x}$ of variables of the formula ϕ, we may consider the question whether $\mathfrak{M}$ *satisfies* the formula $\phi(\vec{m}/\vec{x})$, or written in symbols, whether $\mathfrak{M} \models \phi(\vec{m}/\vec{x})$. This is again defined by induction on the complexity of the formula ϕ:

- if ϕ is an atomic formula of the form $t_0 = t_1$ then $\mathfrak{M} \models \phi(\vec{m}/\vec{x})$ if $t_0^{\mathfrak{M}}(\vec{m}/\vec{x}) = t_1^{\mathfrak{M}}(\vec{m}/\vec{x})$;
- if ϕ is an atomic formula of the form $R_i(t_0, \dots)$ then $\mathfrak{M} \models \phi(\vec{m}/\vec{x})$ if $\langle t_0^{\mathfrak{M}}(\vec{m}/\vec{x}), \dots \rangle \in R_i^{\mathfrak{M}}$;

- if $\phi = \neg\psi$ then $\mathfrak{M} \models \phi$ if $\mathfrak{M} \models \psi$ fails. Similarly for the implication;
- if $\phi = \forall y\ \psi(y, \vec{x})$ then $\mathfrak{M} \models \phi$ if for every $n \in M$, $\mathfrak{M} \models \psi(n, \vec{m}/y, \vec{x})$.

If Γ is a theory the $\mathfrak{M}$ is a model of Γ if $\mathfrak{M} \models \phi$ for every $\phi \in \Gamma$. $\Gamma \models \phi$ denotes the situation that every model of Γ satisfies ϕ. The theory of the model $\mathfrak{M}$ is the set of all sentences that it satisfies. A sentence ϕ is *valid* if $0 \models \phi$.

The most immediate concerns at this stage are the following questions. Given a first order theory, is there a model of it? How many models? Given a model, can we decide which sentences in the appropriate first order language it satisfies? Questions such as these can be easy or difficult, and in most cases good answers are highly desirable.

Example 5.3.1. The theory of dense linear order without endpoints has exactly one countable model up to isomorphism, the rational numbers.

Example 5.3.2. Every group is a model of the theory of groups.Thus, the theory of groups has many different countable models, among them abelian groups (satisfying the sentence $\forall x \forall y\ xy = yx$) and nonabelian groups.

This together with the soundness of the proof system shows that the theory of groups does not prove the sentence $\forall x \forall y\ xy = yx$ nor its complement. One famous result says that there is an algorithm which decides which sentences $\mathbb{F}_n$ for $n \geq 2$ (the free groups on two generators) satisfy [4]. While these groups are pairwise nonisomorphic, they all satisfy the same sentences [10].

Example 5.3.3. Consider the structure $\mathfrak{M} = \langle \mathbb{R}, 0, 1, \leq, +, \cdot \rangle$. The theory of $\mathfrak{M}$ is axiomatized by the axioms of the theory of real closed fields.

Example 5.3.4. The model $\langle \mathbb{N}, 0, 1, S, +, \cdot \rangle$ is a model of Peano Arithmetic.

Despite the suggestive nature of the terminology, there are many other models of Peano Arithmetic. There is no computer algorithm which can decide whether a given sentence is satisfied by $\mathbb{N}$ or not.

5.4 Completeness theorem

Theorem 5.4.1. (Gödel's completeness theorem for first order logic) *A theory is consistent if and only if it has a model.*

As was the case in propositional logic, the proof is preceded by several syntactical lemmas of independent interest. The deduction theorem, the theorems on proof by contradiction and proof by cases transfer verbatim from the treatment of propositional logic.

Lemma 5.4.2. (Generalization rule) *Suppose that* Γ *is a theory and* x *is a variable that does not appear in any sentences of* Γ. *Then* $\Gamma \vdash \phi$ *implies* $\Gamma \vdash \forall x\ \phi$.

Proof. Let $\phi_i : i \in n$ be a formal proof of ϕ. We will rewrite each formula ϕ_i with several others among which $\forall x\ \phi_i$ occurs and so that the result is still a formal proof from Γ. This will complete the proof.

If ϕ_i is an axiom of logic then rewrite it with $\forall x\ \phi_i$, which is also an axiom of logic. If $\phi_i \in \Gamma$ then by assumption x does not appear in ϕ_i, and we can replace ϕ_i with ϕ_i (axiom of Γ), $\phi_i \to \forall x\ \phi_i$ (axiom of logic), $\forall x\ \phi_i$ (modus ponens). If ϕ_i is obtained from previous formulas ϕ_j and $\phi_k = \phi_j \to \phi_i$ by modus ponens, replace it with the sequence $\forall x\ (\phi_j \to \phi_i)$ (proved previously), $\forall x\ (\phi_j \to \phi_i) \to (\forall x \phi_j \to \forall x\ \phi_i)$ (axiom of logic) $\forall x \phi_j \to \forall x\ \phi_i$ (modus ponens), $\forall x\ \phi_j$ (proved previously) $\forall x\ \phi_i$ (modus ponens). This completes the rewriting process and the proof of the lemma. □

Lemma 5.4.3. (Change of variables) *Suppose that $\phi(y)$ is a formula and x is a variable that does not occur in ϕ. Then $\vdash \forall y\ \phi(y) \leftrightarrow \forall x\ \phi(x/y)$.*

Proof. For the left-to-right direction of the equivalence, $\forall y\ \phi(y) \to \phi(x/y)$ is an axiom of logic. Thus, $\forall y\ \phi(y) \vdash \phi(x/y)$. By the generalization rule, $\forall y\ \phi(y) \vdash \forall x\ \phi(x/y)$. The deduction lemma completes the proof of this direction.

For the other direction, let $\psi(x) = \phi(x/y)$. Then y can be properly substituted to x in ψ and $\psi(y/x) = \phi$. So, $\forall x\ \phi(x/y) \to \phi$ is an axiom of logic. Thus, $\forall x\ \phi(x/y) \vdash \phi$ and by the generalization rule, $\forall x\ \phi(x/y) \vdash \forall y\ \phi$. Now apply the deduction lemma again and complete the proof. □

Lemma 5.4.4. (Elimination of constants) *Suppose that Γ is a theory, c is a constant that does not appear in any sentence in Γ, and $\phi(x)$ is a formula such that $\Gamma \vdash \phi(c/x)$. Then $\Gamma \vdash \forall x\ \phi$.*

Proof. Let $\phi_i : i \in n$ be a formal proof of $\phi(c/x)$. Let y be a variable that does not appear in the proof. Directly verify that $\phi_i(y/c) : i \in n$ is a formal proof of $\phi(y/x)$. Let $\Gamma_0 \subset \Gamma$ be the set of sentences used in this proof. Then $\Gamma_0 \vdash \phi(y/x)$ and so by the Generalization Rule, $\Gamma_0 \vdash \forall y\ \phi(y/x)$ and $\Gamma \vdash \forall y\ \phi(y/x)$. The proof is completed by a reference to the Change of variables lemma. □

The most efficient proof of the completeness theorem is based on the following notion.

Definition 5.4.5. A theory Γ is *Henkin* if for every formula $\phi(x)$ there is a constant c such that the sentence $\neg\forall x\phi \to \neg\phi(c/x)$ appears in Γ.

The definition of Henkin property is often stated in the literature in an equivalent form using the existential quantifier.

Lemma 5.4.6. *Every consistent Henkin theory has a model.*

Proof. Let Γ be a consistent Henkin theory. Extend it if necessary to a complete consistent theory. This extension will be again Henkin. For constants c, d of the language of the theory Γ, write $c \equiv d$ if $\Gamma \vdash c = d$. It is not difficult to verify that $\equiv$ is an equivalence relation. The model $\mathfrak{M}$ of the theory Γ under construction will use as its universe M the set of all $\equiv$-classes. Below, for a

constant symbol c write $[c]_\equiv$ to denote the only equivalence class containing c. If $\vec{c}$ is a finite tuple of constant symbols with possible repetitions, let $[\vec{c}]_\equiv$ be the tuple of equivalence classes containing the respective symbols on the tuple $\vec{c}$.

To construct the realizations of the special relational symbols, let R_i be a relational symbol of arity n_i. Let $R_i^{\mathfrak{M}}$ be the set of all n_i-tuples $\vec{m}$ of elements of M such that for any n_i-tuple $\vec{c}$ of constant symbols such that $[\vec{c}]_\equiv = \vec{m}$ it is the case that $\Gamma \vdash R_i(\vec{c})$. Note that if $\vec{c}$ and $\vec{d}$ are n_i-tuples of constant symbols such that corresponding symbols on both tuples are equivalent, then $\Gamma \vdash R_i(\vec{c}) \leftrightarrow R_i(\vec{d})$ by the last logical axiom of equality.

To construct the realizations of the special functional symbols, let F_j be a relational symbol of arity n_j. Let $F_j^{\mathfrak{M}}$ be the function defined by $F_j^{\mathfrak{M}}(\vec{m}) = n$ if for any n_j-tuple $\vec{c}$ of constant symbols and a constant symbol d such that $[\vec{c}]_\equiv = \vec{m}$ and $[d]_\equiv = n$, it is the case that $\Gamma \vdash F_j(\vec{c}) = d$. Note that this is well defined. Whenever $\vec{c}$ is an n_j-tuple of constant symbols, then $\Gamma \vdash \neg\forall x\ \neg x = F_j(\vec{c})$ (why?). As the theory Γ is Henkin, there indeed is a constant symbol d such that $\Gamma \vdash d = F_j(\vec{c})$. If d, e are constant symbols such that $\Gamma \vdash d = F_j(\vec{c})$ and $e = F_j(\vec{c})$ then $d \equiv e$ by the first axiom of equality.

It is now necessary to prove that the model $\mathfrak{M} = \langle M, R_i^{\mathfrak{M}} : i \in I, F_j^{\mathfrak{M}} : j \in J\rangle$ is indeed a model of Γ. By induction on complexity of a formula $\phi(\vec{x})$ with some list $\vec{x}$ of all its free variables, we will prove that for every list $\vec{c}$ of functional symbols of the same length, $\mathfrak{M} \models \phi([\vec{c}]_\equiv/\vec{x})$ if and only if $\Gamma \vdash \phi(\vec{c}/\vec{x})$. This will complete the proof.

For atomic formulas ϕ this follows essentially directly from the definitions. If $\phi = \neg\psi$ and we know the result for ψ, this follows from the completeness of the theory Γ and the induction hypothesis. The implication is similar. The only challenging step is the universal quantification. So, suppose that $\phi(\vec{x}) = \forall y\ \psi(\vec{x}, y)$, we have handled the formula ψ successfully, and $\vec{c}$ is a sequence of constant symbols of the same length as $\vec{x}$. In this case, the following are equivalent:

- $\mathfrak{M} \models \phi([\vec{c}]_\equiv/\vec{x})$;
- for every $m \in M$, $\mathfrak{M} \models \psi([\vec{c}]_\equiv, m/y)$;
- for every constant symbol d, $\mathfrak{M} \models \psi([\vec{c}]_\equiv, [d]_\equiv/y)$;
- for every constant symbol d, $\psi(\vec{c}/\vec{x}, \dot{/}y) \in \Gamma$;
- $\forall y\ \psi(\vec{c}/\vec{x}, y) \in \Gamma$.

The equivalence of the first and second item is the definition of satisfaction for universal formulas, the equivalence of the second and third is the construction of the universe M (it consists solely of equivalence classes of constant symbols), the equivalence of third and fourth is the induction hypothesis, and the equivalence of fourth and fifth follows from the assumption that Γ is a complete Henkin theory. □

Lemma 5.4.7. *Every consistent theory can be extended to a complete consistent Henkin theory.*

Proof. We are going to handle only the case in which the underlying language $\mathfrak{L}$ has countably many special relational and functional symbols. Let $\mathfrak{L}'$ be a language obtained from $\mathfrak{L}$ by adding new constant symbols $\{c_n : n \in \omega\}$. Enumerate all sentences of the expanded language by $\{\phi_n : n \in \omega\}$. By induction on $n \in \omega$ build theories Γ_n in the expanded language so that

- $\Gamma = \Gamma_0 \subseteq \Gamma_1 \subseteq \ldots$, each theory is consistent and uses only finitely many of the new constant symbols;
- for every $n \in \omega$, the theory Γ_{2n+1} contains either ϕ_n or its negation;
- for every $n \in \omega$, if ϕ_n is a sentence of the form $\forall y\ \psi(y)$ then Γ_{2n+2} contains either ϕ_n or the sentence $\neg\psi(c/y)$ for some constant sumbol c.

Once the induction is performed, let $\Gamma' = \bigcup_n \Gamma_n$. This theory in the expanded language is consistent, since it is an increasing union of consistent theories. It is complete by the second inductive item, and it is Henkin by the third inductive item. This will complete the proof of the lemma.

To perform the induction, suppose that $n \in \omega$ is a number and the theory Γ_{2n} has been constructed. To find Γ_{2n+1}, use the lemma on proof by cases. If both theories Γ_{2n}, ϕ_n and $\Gamma_{2n}, \neg\phi_n$ were inconsistent, Γ_{2n} would be inconsistent as well, contradicting the induction hypothesis. So, one of Γ_{2n}, ϕ_n and $\Gamma_{2n}, \neg\phi_n$ is consistent, and this consistent choice will be our Γ_{2n+1}. Since Γ_{2n} contains only finitely many of the new constant symbols and ϕ_n does as well, also Γ_{2n+1} contains only finitely many new constant symbols.

Now suppose that $n \in \omega$ is a number and the theory Γ_{2n+1} has been obtained. To construct Γ_{2n+2}, if ϕ_n is not of the form $\forall y\ \psi(y)$ then let $\Gamma_{2n+2} = \Gamma_{2n+1}$. If $\phi_n = \forall y\ \psi(y)$, then choose a new constant symbol d which does not appear in Γ_{2n+1}. Observe that $\Gamma_{2n+1}, \neg\psi(d/y)$ is inconsistent if and only if $\Gamma \vdash \psi(d/y)$ if and only if $\Gamma \vdash \forall y\ \psi(y)$–the first equivalence is by the lemma on proof by contradiction, and the second equivalence is by the lemma on elimination of constants. Thus, there are two possibilities. Either, $\Gamma_{2n+1} \vdash \phi_n$–in this case, just let $\Gamma_{2n+2} = \Gamma_{2n+1}, \phi$ and proceed with the induction. Or, $\Gamma_{2n+1}, \neg\psi(d/y)$ is consistent–in this case let $\Gamma_{2n+2} = \Gamma_{2n+1}, \neg\psi(d/y)$ and proceed. The induction step has been completed. □

The completeness theorem has a long list of attractive corollaries. The first group of the corollaries is centered around the compactness theorem:

Corollary 5.4.8. (Compactness theorem for first order logic) *A theory Γ has a model if and only if every finite subset of Γ has a model.*

Proof. The completeness theorem shows that Γ has a model if and only if it is consistent. Since every formal proof from Γ uses only finitely many sentences in Γ, the theory Γ is consistent if and only if every finite subset of it is consistent. By the completeness theorem again, this latter statement is equivalent to the assertion that every finite subset of Γ has a model. □

Example 5.4.9. A construction of nonstandard model of Peano Arithmetic; i.e. a model which is not isomorphic to the "standard" model $\langle \mathbb{N}, 0, S, \leq, +, \cdot \rangle$. Add a constant symbol c to the language. Add the infinitely many statements $0 < c$, $S0 < c$, $SS0 < c$, ... to the theory. Every finite subset of the resulting theory has a model (the standard model with c realized as some large natural number), so the whole theory has a model $\mathfrak{M}$. The realization $c^{\mathfrak{M}}$ must be larger than all the "standard" natural numbers 0, $S0$, $SS0$, ... and so this model cannot be isomorphic to the standard model of Peano Arithmetic.

Example 5.4.10. Consider the language with no special symbols. I claim that there is no sentence ϕ in this language such that $\mathfrak{M} \models \phi$ just in case the universe of $\mathfrak{M}$ is finite. (In other words, finiteness/infiniteness is not expressible in this language.) Suppose for contradiction that ϕ is such a sentence. Let ψ_n is the sentence "there are at least n distinct objects", or $\exists x_0 \ldots \exists x_{n-1}\ x_0 \neq x_1 \wedge x_0 \neq x_2 \wedge \ldots x_{n-2} \neq x_{n-1}$. Consider the theory $\Gamma = \{\phi, \psi_n : n \in \omega\}$. Every finite subset of this theory has a model: just look at a sufficiently large finite set–it satisfies ϕ by the assumption on ϕ. Thus, Γ has a model. This model has to be an infinite model of ϕ, contradicting the properties of ϕ.

The second group of immediate corollaries to the completeness theorem ais centered around the notion of categoricity. It offers us a ready tool to show that various theories are complete.

Definition 5.4.11. Let $\mathfrak{M}$ and $\mathfrak{N}$ be models for the same language, with respective universes M, N. The models are *isomorphic* if there is a bijection $h : M \to N$ which transports the $\mathfrak{M}$ realizations to the $\mathfrak{N}$-realizations. A theory Γ is countably categorical if every two countable models of Γ are isomorphic.

Corollary 5.4.12. *If a countable theory Γ is countably categorical, it is complete.*

Proof. Suppose for contradiction that ϕ is a sentence such that Γ proves neither ϕ nor its negation. Then both theories Γ, ϕ and $\Gamma, \neg\phi$ are consistent and by the completeness theorem, they both must have countable models. These two models cannot be isomorphic, since one satisfies ϕ and the other does not. This contradicts our initial assumptions on Γ. □

Example 5.4.13. The theory of dense linear order without endpoints is complete. We showed that every two countable dense linear orders without endpoints are isomorphic. Thus, the theory is countably categorical, and therefore complete.

Exercise 5.4.1. Let Γ be a consistent theory in some language $\mathfrak{L}$. Let $\mathfrak{L}'$ be an expansion of this language by some new functional or relational symbols. Then Γ is still consistent in this new language.

Exercise 5.4.2. If a theory Γ has arbitrarily large finite models (i.e. for every $n \in \omega$ there is a finite model of Γ whose universe has size larger than n) then it has an infinite model.

Chapter 6

Model theory

Model theory is the branch of mathematics that compares and classifies models of various theories. Its goal is to improve the understanding of first order consequences of these theories, as well as the understanding of the complexity of objects that can be defined in various models.

6.1 Basic notions

Let $\mathfrak{L}$ be a language of first order logic, containing special relational symbols R_i of arity n_i for $i \in I$ and special functional symbols F_j of arity n_j for $j \in J$. Let $\mathfrak{M}$, $\mathfrak{N}$ two $\mathfrak{L}$-models with respective universes M, N.

Definition 6.1.1. The models $\mathfrak{M}$ and $\mathfrak{N}$ are *elementarily equivalent* if $Th(\mathfrak{M}) = Th(\mathfrak{N})$.

Clearly, if the models are isomorphic, then they are elementarily equivalent. The reverse implication does not hold though: the free groups on two and three generators respectively are elementarily equivalent, but they are not isomorphic.

Definition 6.1.2. $\mathfrak{M}$ is a *submodel* of $\mathfrak{N}$ if $M \subseteq N$ and $R_i^{\mathfrak{M}} = R_i^{\mathfrak{N}} \cap M^{n_i}$, and $F_j^{\mathfrak{M}} = F_j^{\mathfrak{N}} \restriction M^{n_j}$ for all $i \in I$ and all $j \in J$.

For example, if G is a subgroup of some group H with group operation $\cdot$, then $\langle G, \cdot \rangle$ is a submodel of $\langle H, \cdot \rangle$.

Definition 6.1.3. $\mathfrak{M}$ is an *elementary submodel* of $\mathfrak{N}$ if it is a submodel and for every formula $\phi(\vec{x})$ of the language with free variables $\vec{x}$, and every tuple $\vec{m}$ of elements of M of the same length as $\vec{x}$, $\mathfrak{M} \models \phi(\vec{m}/\vec{x})$ if and only if $\mathfrak{N} \models \phi(\vec{m}/vecx)$.

For example, $\langle \mathbb{Z}, \leq \rangle$ is a submodel of $\langle \mathbb{Q}, \leq \rangle$, but it is not an elementary submodel: the former satisfies $\forall x \neg 0 < x < 1$, while the latter satisfies the opposite. On the other hand, $\langle \mathbb{Q}, \leq \rangle$ is an elementary submodel of $\langle \mathbb{R}, \leq \rangle$. We will prove this later.

Definition 6.1.4. An injection $j : M \to N$ is an *elementary embedding* of $\vec{M}$ to $\vec{N}$ if for every formula $\phi(\vec{x})$ of the language with free variables $\vec{x}$, and every tuple $\vec{m}$ of elements of M of the same length as $\vec{x}$, $\mathfrak{M} \models \phi(\vec{m}/\vec{x})$ if and only if $\mathfrak{N} \models \phi(j\vec{m}/vecx)$.

It is customary in model theory to order models of a given complete theory by elementary embeddability. A prime model of a theory Γ is one which can be elementarily embedded into every other model of Γ, and ???

Definition 6.1.5. Let n_0 be a natural number. A set $A \subset M^n$ is definable (with parameters) if there is an $\mathfrak{L}$-formula $\phi(\vec{x}, \vec{y})$ with free variable lists $\vec{x}_0, \vec{x}_1$ of respective lengths n_0 and some n_1, and some n_1-tuple $\vec{m}_1$ of elements of M such that $A = \{\vec{m}_0 \in M^{n_0} : \mathfrak{M} \models \phi(\vec{m}_0, \vec{m}_1)\}$. The set A is definable without parameters if the formula ϕ can be chosen so that the variable list $\vec{x}_1$ is empty.

It is always of great interest to find a simple characterization of sets definable in a given model. For example, the famous Tarski theorem on real closed fields shows among other things that the only subsets of $\mathbb{R}$ definable in the model $\langle \mathbb{R}, 0, 1, \leq, +, \cdot \rangle$ are finite unions of open intervals and singletons. Therefore, sets such as $\mathbb{Z}$ are not definable. Also, all functions $f : \mathbb{R} \to \mathbb{R}$ definable in this model have polynomial rate of growth, i.e. there is a number n such that for all large enough real numbers r, $f(r) < r^n$. Thus, for example the function $f(x) = e^x$ is not definable in this model.

On the other hand, definable sets in complicated structures such as $\langle \mathbb{N}, 0, S, \leq, +, \cdot \rangle$ cannot be characterized in any useful way.

6.2 Ultraproducts and nonstandard analysis

The purpose of this section is to build a solid logical foundation to *nonstandard analysis*. Nonstandard analysis is an attempt to formalize calculus with infinitesimals (infinitely small numbers), to make sense of the original, logically rather incoherent, language and argumentation of Newton. On our way to this goal, we have to introduce the important model-theoretic tool of *ultraproduct*.

Ultraproducts are a common way of producing complicated models of theories. Let $\mathfrak{L}$ be a first order language, and let $\mathfrak{M}_i$ for $i \in \omega$ be $\mathfrak{L}$-models with respective universes M_i. We want to define a product $\mathfrak{N}$ such that if ϕ is a sentence satisfied by all models $\mathfrak{M}_i$ then it is also satisfied by $\mathfrak{N}$–so for example the product of groups will be again a group, a product of linear orders will be again a linear order etc. For this, we need an important tool:

Definition 6.2.1. A filter on ω is a set $U \subset \mathcal{P}(\omega)$ such that

1. $0 \notin U, \omega \in U$;
2. $a, b \in U \to a \cap b \in U$ (closure under intersections);
3. $a \in U$ and $a \subset b$ implies $b \in U$ (closure under supersets).

An *ultrafilter* is a filter U on ω such that for every partition $\omega = a \cup b$, either $a \in U$ or $b \in U$.

Now, suppose that U is an ultrafilter on ω; we will form an ultraproduct $\mathfrak{N} = \prod_i^U \mathfrak{M}_i$, which will again be a $\mathfrak{L}$ model. To form the universe N of the model $\mathfrak{N}$, consider first the ordinary product $\Pi_i M_i$, which is the set of all functions u with domain ω such that for every $i \in \omega$, $f(i) \in M_i$. Consider the following relation E on $\Pi_i M_i$: $u\ E\ v$ if $\{i \in \omega : u(i) = v(i)\} \in U$.

Claim 6.2.2. *E is an equivalence relation.*

Let N, the universe of the model $\mathfrak{N}$, is the set of all E-equivalence classes of functions in $\prod_i M_i$. We must define the relatizations of special relational and functional symbols in the model $\mathfrak{N}$. Suppose that R is a special relational symbol of the language L of arity n. Define the realization $R^{\mathfrak{N}}$ to be the set of all n-tuples $[\vec{u}]_E$ such that the set $\{i \in \omega : \vec{u}(i) \in R^{\mathfrak{M}_i}\} \in U$. Suppose that F is a special functional symbol of the language $\mathfrak{L}$ of arity n. Define the realization $F^{\mathfrak{N}}$ to be the function which assigns to each n-tuple $\vec{u}]_E$ of elements of the set N the value $[v]_E$, where $v \in \prod_i M_i$ is the function defined by $v(i) = F^{\mathfrak{M}_i}(\vec{u}(i))$.

Theorem 6.2.3. (Łoś) *For every formula $\phi(\vec{x})$ of the language $\mathfrak{L}$ with n free variables, and every n-tuple $\vec{u}$ of functions in $\prod_i M_i$, the following are equivalent:*

1. *$\mathfrak{N} \models \phi([\vec{u}]_E/\vec{x})$;*
2. *the set $\{i \in \omega : \mathfrak{M}_i \models \phi(\vec{u}(i)/\vec{x})\}$ belongs to the ultrafilter U.*

In particular, if ϕ is a sentence satisfied by all models $\mathfrak{M}_i$, then it is also satisfied by the model $\mathfrak{N}$.

Proof. The proof goes by induction on the complexity of the formula ϕ. To make the induction go as smoothly as possible, we choose the language with logical connectives $\neg$ and $\wedge$ and the existential quantifier.

Suppose that the statement of the theorem has been proved for ϕ; we must verify it for $\neg\phi$. We will neglect the parameters of ϕ. The following chain of equivalences verifies the statement for $\neg\phi$. $\mathfrak{N} \models \neg\phi$ if and only if (by the definition of satisfaction relation) $\mathfrak{N} \models \phi$ fails if and only if (by the induction hypothesis) $\{i \in \omega : \mathfrak{M}_i \models \phi\} \notin U$ if and only if (as U is an ultrafilter) $\{i \in \omega : \mathfrak{M}_i \models \phi \text{ fails}\} \in U$ if and only if (by the definition of satisfaction relation) $\{i \in \omega : \mathfrak{M}_i \models \neg\phi\} \in U$.

Suppose that the statement of the theorem has been proved for ϕ and ψ; we must verify it for $\phi \wedge \psi$. Here, $\mathfrak{N} \models \phi \wedge \psi$ if and only if (by the definition of the satisfaction relation) $\mathfrak{N} \models \phi$ and $\mathfrak{N} \models \psi$ if and only if (by the induction hypothesis) $\{i \in \omega : \mathfrak{M}_i \models \phi\} \in U$ and $\{i \in \omega : \mathfrak{M}_i \models \psi\} \in U$ if and only if (as U is closed under intersections and supersets) $\{i \in \omega : \mathfrak{M}_i \models \phi$ and $\mathfrak{M}_i \models \psi\} \in U$ if and only if (by the definition of satisfaction relation) $\{i \in \omega : \mathfrak{M}_i \models \phi \wedge \psi\} \in U$.

???

□

As an important special case, consider the situation that the models $\mathfrak{M}_i$ are all equal to some model $\mathfrak{M}$ with universe M. Let $j : M \to N$ be the map defined by $j(m) = [c_m]_E$ where c_m is the map with domain ω such that for every $i \in m$, $c_m(i) = m$. The Łoś theorem then says precisely that the map is and elementary embedding. In this special case, the model $\mathfrak{N}$ is called an *ultrapower* of $\mathfrak{M}$.

One fairly well-known application of ultrapowers is found in the field of nonstandard analysis. The nonstandard analysis is an attempt to provide semantics to Newton's language of "infinitesimals" in the development of calculus and mathematical analysis.

Consider the model $\mathfrak{R} = \langle \mathbb{R}, \mathcal{P}(\mathbb{R}), \mathcal{PP}(\mathbb{R}), \ldots, \in \rangle$. Let U be a nonprincipal ultrafilter on natural numbers, and let $\mathfrak{R}^*$ be the ultrapower of $\mathfrak{R}$. The model $\mathfrak{R}^*$ is of the form $\langle \mathbb{R}^*, \ldots \varepsilon \langle$; the elements of $\mathbb{R}^*$ are often called *hyperreals*. The ultrapower elementary embedding is traditionally denoted by the star symbol: thus, if $r \in \mathbb{R}$ is a real, $r^* \in \mathbb{R}^*$ is its image among the hyperreals etc. The set $\mathbb{N}$ of all natural numbers is viewed naturally as a subset of the reals, and then $\mathbb{N}^*$ is its image under the ultrapower embedding.

Note that the hyperreals are elementarily equivalent to reals, and therefore their version of addition, multiplication, ordering etc. satisfy the same first order properties is those of the reals. However, the hyperreal line is in some sense richer than the real line, as is obvious from the following central definition and claim:

Definition 6.2.4. Let $\varepsilon > 0^*$ be a hyperreal. We call ε an *infinitesimal* if for every positive real number $r \in \mathbb{R}$, $\varepsilon < r^*$.

Claim 6.2.5. *Infinitesimals exist in* $\mathbb{R}^*$.

Proof. Consider the map $c : \omega \to \mathbb{R}$ defined by $c(n) = 1/n$. We will show that the equivalence class of this function in the ultrapower, $[c]_E$, is an infinitesimal. □

Now, the stage is set for finding equivalent restatments of limits, continuity, differentiability etc. using Neton's original language of infinitely small or ifnintely large quantities. We will prove only one illustrative theorem among many possibilities.

Definition 6.2.6. Hyperreals r, s are *infinitesimally close* if the difference $|r-s|$ is infinitesimal. A hyperreal r is *finite* if it is infinitesimally close to s^* for some real s. Otherwise, the hyperreal is *infinite*.

Theorem 6.2.7. *Let $s : \mathbb{N} \to \mathbb{R}$ be a sequence of real numbers and L a real number. Then the following are equivalent:*

1. $\lim s = L$;
2. *for every infinite hypernatural $n \in \mathbb{N}$, the value $s^*(n)$ is infinitesimally close to L^*.*

Theorem 6.2.8. *Let $f : \mathbb{R} \to \mathbb{R}$ be a function. The following are equivalent:*

1. *f is continuous;*
2. *for every real $r \in \mathbb{R}$, whenever a hyperreal s is infinitesimally close to r^*, the functional value $f^*(s)$ is infinitesimally close to $f^*(r^*)$.*

6.3 Quantifier elimination and the real closed fields

Let $\mathfrak{R}$ be the model $\langle \mathbb{R}, 0, 1, \leq, +, \cdot \rangle$. This is one of the more popular structures in mathematics. The purpose of this section is to state and outline the proof of a theorem of Tarski, which axiomatized the theory of $\mathfrak{R}$, showed that the theory is decidable, and characterized the sets definable in the structure. On the way to this goal, we will develop the powerful model theoretic concept of *quantifier elimination*.

Definition 6.3.1. A theory Γ has *quantifier elimination* if for every formula ϕ in the language of Γ (perhaps with some free variables) there is a formula ψ containing no quantifiers such that $\Gamma \vdash \phi \leftrightarrow \psi$.

Elimination of quantifiers typically offers a (highly desirable) algorithmic way of deciding which sentences are provable from Γ, and whether various formulas are satisfied in models of Γ. The question is, can we (efficiently) eliminate quantifiers from any formula? Which theories have quantifier elimination?

We prove several results on quantifier elimination, ordered by difficulty.

Theorem 6.3.2. *The theory of infinite set has quantifier elimination.*

As a motivational example, note that the theory of equality (without any non-logical axioms) does not have quantifier elimination, since the formula $\exists y\ y \neq x$ does not have a quantifier-free equivalent. There are essentially only two candidates for such an equivalent, $x = x$ and $x \neq x$. However, in the model with only one element m, the formula $x = x$ is satisfied at m while the formula $\exists y\ y \neq x$ is not, showing that $x = x$ and $\exists y\ y \neq x$ are not equivalent. In the model with at least two elements m, n, the formula $x \neq x$ is not satisfied at m while the formula $\exists y\ y \neq x$ is, showing that $x \neq x$ and $\exists y\ y \neq x$ are not equivalent.

Proof. Recall that the theory Γ of infinite set uses no special relational or functional symbols, and for each natural number n, it contains the statement $\exists x_0 \exists x_1 \ldots \exists x_n\ x_0 \neq x_1 \wedge x_0 \neq x_2 \wedge \ldots$ (there are at least $n+1$ many distinct elements).

Let $\vec{x}$ be a list of variables. A formula $\phi(\vec{x})$ is *complete* if it is a conjunction of atomic formulas $x = y$ or their negations where x, y range over all variables on the list $\vec{x}$. We will show that for every formula ψ of the language with equality, there is a disjunction θ of complete formulas such that $\Gamma \vdash \psi \leftrightarrow \theta$.

The proof proceeds by induction of complexity of the formula ψ. We will work with the language with logical connectives $\neg, \vee$ and the existential quantifier $\exists$. The atomic case is trivial, since every atomic formula is complete.

Finally, suppose that a formula $\phi(\vec{x}, y)$ is provably equivalent to some disjunction of complete formulas. We want to show that $\exists y\ \phi$ is also equivalent to disjunction of complete formulas. Since the existential quantifier distrubutes over disjunction ($\exists z\ \theta_0 \vee \theta_1$ is provably equivalent to $(\exists z\ \theta_0) \vee (\exists z\ \theta_1)$), it is enough to treat the case where ϕ is (equivalent to) a single complete formula. Let $\psi(\vec{x})$ be the formula obtained from $\phi(\vec{x}, y)$ by erasing all conjuncts that mention y. We claim that $\Gamma \vdash \exists y\ \phi(\vec{x}, y) \leftrightarrow \psi(\vec{x})$. This is proved in two distinct cases.

Case 1. Either there is a variable z in the list $\vec{x}$ such that ϕ contains $z = y$ as one of the conjuncts. In this case, $\exists y\ \phi(\vec{x}, y)$ is implied by $\psi(\vec{x})$ since the existential quantifier is witnessed by $z = y$. (*Example.* $\exists y\ y = x_0 \wedge x_0 \neq x_1$ is logically equivalent to $x_0 \neq x_1$.)

Case 2. Or, ϕ contains a conjunct of the form $z \neq y$ for every variable z on the list $\vec{x}$. In this case, $\phi(\vec{x}, y)$ is equivalent to the conjunction of $\psi(\vec{x})$ and the statement "y is not equal to anything on the list $\vec{x}$". Now, $\Gamma \vdash \psi(\vec{x}) \leftrightarrow \exists y\ \phi(\vec{x}, y)$, since the existence of y which is not equal to anything on the (finite) list $\vec{x}$ follows immediately from the axioms of the theory Γ. (*Example.* Γ proves that $\exists y\ y \neq x_0 \wedge y \neq x_1 \wedge x_0 \neq x_1$ is equivalent to $x_0 \neq x_1$.) □

Corollary 6.3.3. *Suppose that M is an infinite set. The sets definable in the model $\langle M, = \rangle$ are exactly the finite and cofinite subsets of M.*

Recall that a subset $N \subset M$ is *cofinite* if $M \setminus N$ is finite.

Proof. On one hand, every finite or cofinite set is clearly definable in the model. For example, the set $\{c_0, c_1, c_2\}$ is definable by the formula $\phi(x, y_0, y_1, y_2)$ equal to $x = y_0 \vee x = y_1 \vee x = y_2$ with the parameters c_0, c_1, c_2.

On the other hand, every definable set in the structure is either finite or cofinite. Since every definition can be replaced with an equivalent quantifier-free definition, it is enough to show that every set defined by a quantifier free formula is finite or cofinite. This is proved by induction on complexity of the defining quantifier-free formula ϕ. □

Theorem 6.3.4. *The theory of dense linear order without endpoints has quantifier elimination.*

As a motivational example, note that the theory of linear order (without the density axiom) does not have quantifier elimination. Consider the formula $\phi(x, y) = \exists z\ x < z < y$; it does not have a quantifier free equivalent. There are essentially only three options for the quantifier-free equivalent, $x < y$, $y < x$, and $y = x$, and neither of them is equivalent to $\phi(x, y)$. Note though that $x < y$ *is* equivalent to ϕ in dense linear orders.

Proof. The proof follows the lines of the argument for Theorem 6.3.2. Let Γ denote the first order theory of dense linear order without endpoints. We will use $x < y$ as the shorthand for $x \leq y \wedge x \neq y$. A formula $\phi(\vec{x})$ is called *complete* if it is a conjunction of atomic formulas or their negations and for every pair of variables x, y on the list $\vec{x}$, the conjuncts include $x = y$ or $x \neq y$, and they also

include $x < y$ or $x \not< y$. Note that for a given finite list of variables, there are onfly finitely many complete formulas up to logical equivalence. We will show that for every formula $\phi(\vec{x})$ there is a disjunction $\psi(vecx)$ of complete formulas such that $\Gamma \vdash \phi(\vec{x}) \leftrightarrow \psi(\vec{x})$. This will complete the proof. The argument proceeds by complexity of the formula ϕ. We will use the first order language that contains logical connectives $\neg, \vee$ and the existential quantifier $\exists$.

The case of atomic formulas, as well as the induction step for disjunction and negation are dealt with literally as in the previous proof. To perform the induction step for existential variables, assume that ϕ is a complete formula with variables $\vec{x}$ and y; we must show that $\exists y\ \phi(\vec{x}, y)$ is equivalent to a complete formula. Consider the formula θ that obtains from ϕ by erasing all conjuncts mentioning y; we will show that $\Gamma \vdash \exists y \phi(\vec{x}, y) \leftrightarrow \theta(\vec{x})$.

Case 1. Suppose that ϕ contains a conjunct of the form $x = y$ for some variable x on the list $\vec{x}$. In such a case $\exists y\ \phi(\vec{x}, y)$ is logically equivalent to $\theta(\vec{x})$ since satisfaction of the existential quantifier is witnessed by x. (*Example.* $\exists y\ x_0 = y < x_1$ is logically equivalent to $x_0 < x_1$.)

Case 2. Suppose that ϕ contains conjuncts of the form $x \neq y$ for every variable x on the list $\vec{x}$. Consider where y stands in the $<$-order of the other variables as specified by the formula ϕ. There are three distinct cases: either ϕ asserts that y is smaller than all variables on the list $\vec{x}$, or it is greater than all of them, or there are two variables x_0, x_1 on the list such that ϕ asserts that $x_0 < y < x_1$ and there is no variable on the list $\vec{x}$ strictly between x_0, x_1. Let us consider the third case. The dense liner order axiom then proves $x_0 < x_1 \rightarrow \exists y\ x_0 < y < x_1$ and therefore also $\exists y\ \phi(\vec{x}, y) \rightarrow \theta(\vec{x})$. (*Example.* The density of the ordering implies that $\exists y\ x_0 < y < x_1$ is equivalent to $x_0 < x_1$.) □

Corollary 6.3.5. *Let $\langle L, \leq \rangle$ be a dense linear order without endpoints. The sets definable in the model $\langle L, \leq \rangle$ are exactly the finite unions of open intervals and singletons.*

Proof. On one hand, a finite union of open intervals and singletons is clearly definable in the model. A set such as $(l_0, l_1) \cup (l_2, l_3) \cup \{l_4, l_5\}$ is definable via the formula $\phi(x, y_0, y_1, y_2, y_3, y_4, y_5) = (y_0 < x < y_1) \vee (y_2 < x < y_3) \vee x = y_4 \vee x = y_5$ with the parameters $l_0, l_1, l_2, l_3, l_4, l_5$.

On the other hand, every definable set is a finite union of open intervals and singletons. Since every formula is equivalent to a quantifier-free formula, it is enough to check that quantifier-free formulas can define only finite unions of open intervals and singletons. This is verified by induction on complexity of the quantifier-free formula ϕ. □

Theorem 6.3.6. *The theory of algebraically closed fields has quantifier elimination.*

Recall that the theory of fields has constant symbols $0, 1$ and binary functional symbols $+, \cdot$ and the following axioms:

- $+$ is a commutative group operation with 0 as the neutral element;

- $\cdot$ is a commutative group operation on nonzero elements, with 1 as the neutral element. Moreover, $\forall x\ x \cdot 0 = 0 \cdot x = 0$;
- (distributivity) $\forall x \forall y \forall z\ x(y+z) = xy + xz$ and $(x+y)z = xz + yz$.

The algebraically closed fields are obtained by adding axioms saying that every polynomial of degree larger than zero has roots. This is an infinite collection of axioms. For every natural number $n > 0$, there is a statement $\forall y_0 \forall y_1 \ldots \forall y_n\ y_n \neq 0 \to \exists x\ y_n x^n + y_{n-1}x^{n-1} + \cdots + y_0 = 0$.

As a motivational example, the theory of fields without the additional algebraic closure axioms does not have quantifier elimination. Consider the formula $\phi(x) = \exists y\ y \cdot y = x$; it does not have a quantifier-free equivalent in this theory. Suppose for contradiction that $\psi(x)$ is such a quantifier-free equivalent. ψ is just some boolean combination of statements of the form $p(x) = 0$ where p is a polynomial with integer coefficients. Consider the two fields $\mathbb{Q}$ and $\mathbb{R}$ with the usual addition and multiplication. Both fields evaluate the polynomials in the same way, and so $\mathbb{Q} \models \psi(2)$ if and only if $\mathbb{R} \models \psi(2)$. However, $\mathbb{R} \models \phi(2)$ while $\mathbb{Q} \models \neg\phi(2)$, since the square root of 2 is well-known to be irrational. This contradicts the equivalence of $\phi(x)$ and $\psi(x)$.

Proof. We will adopt the subtraction operation into the language to simplify the resulting expressions. The terms of the language are than just polynomials in several variables and integer coefficients, and every atomic formula can be rearranged into the form $p = 0$ where p is such a polynomial. The proof of quantifier elimination proceeds by induction on the complexity of formulas. As in the previous proofs, it is necessary to show how to eliminate the existential quantifier. There are several interesting special cases, which will be used to deal with the general case.

Claim 6.3.7. *If $p(x, \vec{y})$ is a polynomial with integer coefficients, then $\exists x\ p(x, \vec{y}) = 0$ is equivalent to a quantifier-free formula.*

Proof. In an algebraically closed field, the formula $\exists x\ p(x, \vec{y}) = 0$ is equivalent to the statement that p as a polynomial in x has nonzero degree or otherwise it is a zero polynomial. In other words, if $a_i : i \leq n$ are terms in the variables on the list $\vec{y}$ such that $p = \Sigma_{i \leq n} a_i x^i$, the formula $\exists x\ p(x, \vec{y}) = 0$ is equivalent to the formula $(a_1 \neq 0 \vee a_2 \neq 0 \vee \cdots \vee a_n \neq 0) \vee a_0 = 0$. □

Claim 6.3.8. *If $p(x, \vec{y})$ is a polynomial with integer coefficients, then $\exists x\ p(x, \vec{y}) \neq 0$ is equivalent to a quantifier-free formula.*

Proof. In every field, a polynomial with nonzero coefficients has at least one nonzero value. Thus, if $a_i : i \leq n$ are terms in the variables on the list $\vec{y}$ such that $p = \Sigma_{i \leq n} a_i x^i$, the formula $\exists x\ p(x, \vec{y}) \neq 0$ is equivalent to the formula $a_0 \neq 0 \vee a_1 \neq 0 \vee a_2 \neq 0 \vee \cdots \vee a_n \neq 0$. □

Claim 6.3.9. *If $p(x, \vec{y})$ and $q(x, \vec{y})$ are polynomials with integer coefficients, then $\exists x\ p(x, \vec{y}) = 0 \wedge q(x, \vec{y}) \neq 0$ is equivalent to a quantifier-free formula.*

Proof. In an algebraically closed field, the formula $\neg\exists x\ p(x,\vec{y}) = 0 \wedge q(x,\vec{y}) \neq 0$ (or "all roots of p are also roots of q) is equivalent to the statement that the polynomial p divides q^n where n is the degree of p: both polynomials factorize into linear factors, every linear factor of p must show up in q, and it can repeat at most n many times in the factorization of p. Thus it will be enough to show that the statement "p divides q" is equivalent to a quatifier-free formula.

This is essentially the long division algorithm. Divide q with p and consider the remainder, which is some polynomial r of degree less than the degree of p Let $a_i : i \leq m$ are terms in the variables on the list $\vec{y}$ such that $r = \Sigma_{i\leq m} a_i x^i$. Then "$p$ divides q" is equivalent to the quantifier-free formula $a_0 = 0 \wedge a_1 = 0 \wedge \cdots \wedge a_m = 0$. □

Claim 6.3.10. *If $p_i(x,\vec{y}) : i < n$ and $q_i(x,\vec{y}) : i < m$ are polynomials with integer coefficients, then $\phi = \exists x\ p_0 = 0 \wedge p_1 = 0 \wedge \cdots \wedge q_0 \neq 0 \wedge q_1 \neq 0 \wedge \ldots$ is equivalent to a quantifier-free formula.*

Proof. In every field, the condition $q_0 \neq 0 \wedge q_1 \neq 0 \wedge \ldots$ is equivalent to $q \neq 0$ where q is the polynomial which is the product of all polynomials on the list $q_0, q_1 \ldots$. Thus, it is enough to deal with the case where $m = 1$, i.e. there is only one q-polynomial.

The proof goes by induction on k, where k is the sum of the degrees of all polynomials on the list $p_i : i < n$. In the base case that $k = 0$, all the p-polynomials have degree zero, therefore do not mention x at all, and the formula ϕ is equivalent to $p_0 \neq 0 \wedge p_1 = 0 \wedge \cdots \wedge \exists x\ q \neq 0$, which is equivalent to a quantifier-free formula by Claim 6.3.8.

Now suppose that the induction hypothesis has been verified for some k, and argue that it holds at $k+1$. Suppose that $p_i(x,\vec{y}) : i < n$ and $q(x,\vec{y})$ are polynomials with integer coefficients such that the degrees of the polynomials p_i add up to $k+1$. We must verify that $\phi = \exists x\ p_0 = 0 \wedge p_1 = 0 \wedge \cdots \wedge q \neq 0$ is equivalent to a quantifier-free formula. If there is only one p-polynomial (i.e. $n = 1$), then this is the content of Claim 6.3.9. So suppose that $n > 1$, and (renumbering the polynomials if necessary) assume that the degree of p_0 is some d_0, the degree of p_1 is some d_1 with $d_1 \leq d_0$, and a_0, a_1 are the respective leading coefficients of the polynomials p_0, p_1. Then ϕ is equivalent to the formula $(a_1 = 0 \wedge \psi) \vee (a_1 \neq 0 \wedge \theta)$, where

- $\psi = \exists x\ p_0 = 0 \wedge \bar{p}_1 = 0 \wedge \cdots \wedge q \neq 0$ where $\bar{p}_1 = p_1 - a_1 x_1^d$. Observe that the degree of $\bar{p}_1$ is smaller than the degree of p_1;
- $\theta = \exists x\ \bar{p}_0 = 0 \wedge p_1 = 0 \wedge \cdots \wedge q \neq 0$ where $\bar{p}_0 = a_1 p_0 - a_0 x^{d_0 - d_1} p_1$. Observe that the degree of $\bar{p}_0$ is smaller than the degree of p_0.

The sum of degrees of polynomials mentioned in ψ or θ is in both cases at most k, and so by the induction hypothesis, both ψ, θ are equivalent to a quantifier-free formula. Ergo, ϕ is equivalent to a quantifier-free formula and the induction step has been performed. □

Now for the general case of eliminating the existential quantification, suppose that ψ is an arbitrary quantifier-free formula and x is a variable; we want to show that $\exists x\ \psi$ is equivalent to a quantifier-free formula. Rearranging ψ if necessary, we may assume that ψ is a disjunction $\theta_0 \vee \theta_1 \vee \dots$ where each θ_i is in turn a conjunction of atomic formulas or their negations. Then, $\exists x\ \psi$ is equivalent to $\exists x\ \theta_0 \vee \exists x\ \theta_1 \vee \dots$, and each formula $\exists x\ \theta_i$ is equivalent to a quantifier-free formula by Claim 6.3.10. This completes the proof of the theorem.

□

Corollary 6.3.11. *Let $\langle \mathbb{C}, 0, 1, +, \cdot \rangle$ be the field of complex numbers with addition and multiplication. The definable sets in this model are exactly the finite and cofinite sets.*

Proof. On one hand, every finite or cofinite set is clearly definable in the model. For example, the set $\{c_0, c_1, c_2\}$ is definable by the formula $\phi(x, y_0, y_1, y_2)$ equal to $x = y_0 \vee x = y_1 \vee x = y_2$ with the parameters c_0, c_1, c_2.

On the other hand, every definable set in the structure is either finite or cofinite. Since every definition can be replaced with an equivalent quantifier-free definition, it is enough to show that every set defined by a quantifier free formula is finite or cofinite. This is proved by induction on complexity of the defining quantifier-free formula ϕ. The important case is that of atomic formulas. An atomic formula $\phi(x, \vec{y})$ is (after perhaps some reorganization) just an equation $p(x) = 0$ where p is a polynomial in x with parameters that are some combination of the parameters on the list $\vec{y}$. A nonzero polynomial in a field has only finitely many roots, so the atomic formula defines a finite set. □

Theorem 6.3.12. (Tarski 1951) *The theory of real closed fields is complete and has quantifier elimination.*

Recall that the theory of real closed fields has constant symbols $0, 1$, binary functional symbols x, y, and a binary relational symbol $\leq$ and axioms as follows:

- $0, 1, +, \cdot$ form a field;
- $\leq$ is a linear order such that $\forall x \forall y\ (0 \leq x \wedge 0 \leq y) \to 0 \leq x + y$ (in other words, $+$ is an ordered group);
- every polynomial of odd degree has a root.

The intended model of the theory of real closed fields is $\mathfrak{R} = \langle \mathbb{R}, 0, 1, +, \leq, \cdot \rangle$.

The proof of the theorem is too long to include in these notes. We will only discuss two motivational examples of quantifier elimination in the structure $\mathfrak{R}$.

Example 6.3.13. The existential formula $\exists x\ ax^2 + bx + c = 0$ is equivalent to the quantifier-free formula $b^2 + 4ac \geq 0$.

Example 6.3.14. If $p(x)$ is a polynomial and $a < b$ are real numbers, the *Sturm's algorithm* provides an algorithmic way to decide whether $\exists x\ p(x) = 0 \wedge a \leq x \leq b$ holds. A more careful look at the algorithm will show that it in fact reduces this existential formula to a quantifier-free formula. There are many other root-finding algorithms.

Example 6.3.15. The ordering $\leq$ is deinable in the structure $\mathfrak{R}$ from the other functions: $x \leq y$ if and only if $\exists z\ z^2 + x = y$. However, without the symbol $\leq$, the quantifier elimination fails: the set $A = \{x : 0 \leq x\}$ is not definable without quantifiers from the remaining functions. To see this, suppose that $\phi(x, \vec{y})$ is a quantifier-free formula not mentioning $\leq$, and $\vec{r}$ is a sequence of real numbers of the same length as $\vec{y}$. I will produce a real number $s > 0$ such that $\mathfrak{R} \models \phi(s/x, \vec{r}/\vec{x}) \leftrightarrow \phi(-s/x, \vec{r}/\vec{y})$. This shows that $\phi(x, \vec{r}/\vec{y})$ does not define the set A in the model $\mathfrak{R}$.

The atomic subformulas in $\phi(x, \vec{r}/\vec{y})$ are of the form $p(x) = 0$ where p is some polynomial with real coefficients. Nonzero polynomials have only finitely many roots, so there is some real number $s > 0$ such that neither s nor $-s$ is a root of any nonzero polynomial mentioned in $\phi(x, \vec{r}/\vec{y})$. It is clear that the number s works as desired.

Example 6.3.16. The function $f(x) = e^x$ is not definable in the structure $\mathfrak{R}$. In fact, for every definable function g there is a number $n \in \omega$ and a real number $r \in \mathbb{R}$ such that for every $x > r$, $g(x) \leq x^n$. To see this, suppose that $g(x) = y$ is defined via some formula $\phi(x, y, \vec{z})$ and a string $\vec{r}$ of parameters of the same length as $\vec{z}$. By the quantifier elimination, we may assume that ϕ is quantifier free. For any real number s, the atomic formulas in $\phi(s/x, y, \vec{r}/\vec{z})$ are inequalities of the form $p(x) \geq 0$ where p is a polynomial with real coefficients. Let $h(s)$ be the largest real number which is a root of some nonzero polynomials mentioned in $\phi(s/x, y, \vec{r}/\vec{z})$. We will show that $g(s) \leq h(s)$ and h is bounded by a polynomial.

First of all, if $t, u > h(s)$ are real numbers, then $\mathfrak{R} \models \phi(s/x, t/y, \vec{r}/\vec{y}) \leftrightarrow \phi(s/x, u/y, \vec{r}/\vec{z})$, since no polynomial mentioned in $\phi(s/x, y, \vec{r}/\vec{z})$ changes sign past $h(s)$. This means that $g(s) \leq h(s)$.

Second, to bound the function h by a polynomial, we must use one of the theorems bounding roots of a polynomial. Theorem ??? of ??? states that if $p(y) = \Sigma_{i \leq n} a_i y^i$ is a polynomial with leading coefficient $a_n \neq 0$ then all of its complex roots have absolute value $\leq \frac{1}{|a_n|} \Sigma_{i<n} |a_i|$. Now note that the coefficients of the polynomials in the formula $\phi(s/x, y, \vec{r}/\vec{z})$ are themselves polynomials in s. This means that there is some real number s_0 and a constant $\varepsilon > 0$ such that the leading coefficients of these polynomials are in absolute value $> \varepsilon$ for all $s > s_0$. The function $h(s)$ for $s > s_0$ is then bounded by $1/\varepsilon$ times the sum of $1 + a^2$ for all coefficients a of the polynomials appearing in the formula ϕ.

Corollary 6.3.17. *Every subset of $\mathbb{R}$ definable in $\mathfrak{R}$ is a finite union of open intervals and singletons.*

Proof. The atomic formulas of the language of RCF can be written in the form of $p(\vec{x}) \geq 0$ or $p(\vec{x}) = 0$ for polynomials p of some variables $\vec{x}$. Polynomials are continuous functions, and number of roots is bounded by the degree of the polynomial. Therefore, the atomic formulas can define only a finite union of open intervals and singletons. A general quantifier-free formula is a boolean combination of atomic formulas, and so it also can only define a finite union of open intervals and singletons. □

The corollary is very attractive; it immediately leads to the following definition:

Definition 6.3.18. A model $\mathfrak{M}$ is *o-minimal* if its language contains a binary relation symbol $\leq$ such that $\leq^{\mathfrak{M}}$ is a linear ordering and every definable subset of the universe of $\mathfrak{M}$ is a finite union of open intervals in this ordering and singletons.

Which models are o-minimal? In particular, which relations or functions can be added to $\mathfrak{R}$ while preserving its o-minimality?

Theorem 6.3.19. (Wilkie 1996) *Let* $\mathfrak{E} = \langle \mathbb{R}, 0, 1, \leq, +, \cdot, e^x \rangle$. *The structure* $\mathfrak{E}$ *is o-minimal.*

The theory of the structure $\mathfrak{E}$ does not allow quantifier elimination. It is not known if the theory is decidable.

Chapter 7

The incompleteness phenomenon

The purpose of this chapter is to prove the famous first Gödel's incompleteness theorem.

7.1 Peano Arithmetic

Since the incompleteness theorem is most commonly stated for Peano Arithmetic, we will first take some time to describe this first order theory in some detail. Its language has a constant symbol 0, a unary functional symbol S (successor), binary functional symbols $+, \cdot$, and a binary relational symbol $\leq$. Its axioms are:

- $\leq$ is a linear ordering with 0 as the least element;
- for every x, $S(x)$ is the $\leq$-smallest element larger than x, and every nonzero x is $S(y)$ for some y;
- for all x, y, $x+0 = x$ and $x+Sy = S(x+y)$, $x \cdot 0 = 0$ and $x \cdot Sy = x \cdot y + x$;
- (the induction scheme) Whenever $\phi(\vec{x}, y)$ is a formula with all free variables listed, the following statement is an instance of the induction axiom scheme: $\forall \vec{x}\ (\phi(\vec{x}, 0) \wedge (\phi(\vec{x}, y) \rightarrow \phi(\vec{x}, Sy)) \rightarrow \forall y\ \phi(\vec{x}, y))$.

To illustrate the use of the induction scheme, we prove the following simple formal theorem of Peano Arithmetic.

Theorem 7.1.1. *PA proves the commutativity of addition,* $\forall y\ \forall x\ x+y = y+x$.

Proof. To prepare the ground, by induction on y prove the statement $\forall x\ \forall y\ x + Sy = Sx + y$. For the base step, $x + S0 = S(x+0)$ by the third group of axioms, $S(x+0) = Sx$ and $Sx = Sx + 0$ by the neutrality of 0, and so $x + S0 = Sx + 0$.

For the induction step, suppose that $x + Sy = Sx + y$ holds and work to prove $x + SSy = Sx + Sy$. To see how this is done, $x + SSy = S(x + Sy)$ by the third axiom group, $S(x + Sy) = S(Sx + y)$ by the induction hypothesis, and $S(Sx + y) = Sx + Sy$ by the third axiom group again.

Another useful preliminary fact is that $\forall x\ x + 0 = 0 + x$. This is proved by induction on x. The base step $0 + 0 = 0 + 0$ follows from the logical axioms of equality. For the induction step, the induction hypothesis $x + 0 = 0 + x$ must be shown to imply $Sx + 0 = 0 + Sx$. The following string of equalities proves exactly that: $0 + Sx = S(0 + x)$ by the third group of axioms of PA, $S(0+x) = S(x+0)$ by the induction hypothesis, $S(x+0) = Sx$ since $x+0 = x$ by the third group of axioms of PA, and $Sx = Sx + 0$ by the third group of axioms of PA again.

Finally, we are ready to prove the commutativity by induction on y. The base step is the statement $\forall x\ x + 0 = 0 + x$ proved in the previous paragraph. For the successor step, we must show that the induction hypothesis $x+y = y+x$ implies $x + Sy = Sy + x$. Indeed, $x + Sy = Sx + y$ by the first paragraph of this proof, $Sx + y = y + Sx$ by the induction hypothesis, and $y + Sx = Sy + x$ by the first paragraph of this proof again. □

7.2 Outline of proof

Theorem 7.2.1. (First Incompleteness Theorem) *Peano Arithmetic is not complete. There is a sentence ϕ of the language of Peano Arithmetic such that PA proves neither ϕ nor $\neg\phi$.*

We will present a slightly simplified proof of the incompleteness theorem. It consists of three parts.

Arithmetization of syntax. Plainly speaking, this says that the syntax of Peano Arithmetic can be encoded by natural numbers in a sensible way. We will produce injective maps $\phi \mapsto \widehat{\phi}$ and $t \mapsto \widehat{t}$ that send formulas and terms of the language of PA to natural numbers so that simple syntactical notions are definable in $\mathfrak{N}$. In particular, there are formulas

- $\texttt{Form}$ such that $\mathfrak{N} \models \texttt{Form}(n)$ just in case there is a formula ϕ such that $n = \widehat{\phi}$;
- $\texttt{Plug}$ such that $\mathfrak{N} \models \texttt{Plug}(k, l, m)$ just in case there is a formula ϕ with a single free variable x such that $\hat{\phi} = k$, and $m = \widehat{\phi(t/x)}$ where t is the numeral for l;
- $\texttt{Prov}$ such that $\mathfrak{N} \models \texttt{Prov}(n)$ just in case there is a sentence ϕ which is a theorem of PA and $n = \widehat{\phi}$.

In fact, essentially every imaginable syntactical notion will be definable using the coding in question. There are many equivalent ways to arithmetize syntax, but all of them require some tedious moves.

Diagonalization. This is the crux of the proof, a simple and confusing lemma with a simple and confusing proof.

Lemma 7.2.2. *For every formula θ of one free variable, there is a sentence ϕ such that $\mathfrak{N} \models \phi \leftrightarrow \theta(\widehat{\phi})$.*

A more precise form of the lemma makes the conclusion that PA$\vdash \phi \leftrightarrow \theta(\widehat{\phi})$. This is slightly more difficult to prove and we are not going to need it. In both cases, the arithmetization of syntax is necessary for the proof.

Proof. Let $\theta(x)$ be a formula of one free variable. Let y be a variable which does not appear in θ. Let $\psi(y)$ be the formula $\forall z\ \mathtt{Plug}(y, y, z) \rightarrow \theta(z/x)$. Let ϕ be the sentence $\psi(t/y)$ where t is the numeral for $\widehat{\psi}$. We claim that ϕ works as required. Observe the equivalence of the following items:

- $\mathfrak{N} \models \phi$;
- $\mathfrak{N} \models \psi(\hat{\psi})$;
- $\mathfrak{N} \models \theta(\widehat{\psi(t/y)})$ where t is the numeral for $\widehat{\psi}$;
- $\mathfrak{N} \models \theta(\widehat{\phi})$.

The first and second item are equivalent by the definition of ϕ. The second and third item are equivalent by the definition of ψ and `Plug`, and the third and fourth item are equivalent by the definition of ϕ again.

□

Final cinch. Once the diagonalization is proved, the incompleteness theorem is an easy corollary. Apply the diagonalization lemma with $\theta(x) = \neg\mathtt{Prov}(x)$. Find a sentence ϕ such that $\mathfrak{N} \models \phi \leftrightarrow \theta(\widehat{\phi})$. We claim that the sentence ϕ is not decidable in Peano Arithmetic:

- if PA$\vdash \phi$ then $\mathfrak{N} \models \phi$ and so $\mathfrak{N} \models \neg\mathtt{Prov}(\widehat{\phi})$, and therefore ϕ is not provable; this is a contradiction;
- if PA$\vdash \neg\phi$ then $\mathfrak{N} \models \neg\phi$, and so $\mathfrak{N} \models \mathtt{Prov}(\widehat{\phi})$, and so ϕ is provable in PA. This contradicts the consistency of PA.

7.3 Arithmetization of syntax

7.4 Other sentences unprovable in Peano Arithmetic

Gödel's incompleteness theorem provides a sentence unprovable in Peano Arithmetic. The sentence is in logical sense the simplest possible. However, in mathematical sense, it has the disadvantage of carrying no clear content. Over time, a number of mathematically meaningful sentences formalizable, but not provable, in Peano Arithmetic appeared.

Example 7.4.1. Ramsey's theorem. For every number $k \in \omega$, every $r \in \omega$ and every coloring $c : [\omega]^k \to r$ there is an infinite set $a \subset \omega$ such that all k-element subsets of a are colored with the same color. This theorem is not formalizable in the language of Peano Arithmetic due to the quantification over infinite objects.

We consider a finitization of this statement due to Paris and Harrington. For every $k, r \in \omega$ there is m such that for every coloring $c : [m]^k \to r$ there is a nonempty set $a \subset m$ such that $\min(a) < |a|$ and all k-element subsets of a are colored with the same color. This statement is formalizable, but not provable, in PA. The function $k, r \mapsto m$ grows very fast.

Example 7.4.2. Kruskal's tree theorem. A tree is a (finite) partially ordered set $\langle T, \leq \rangle$ such that for every $t \in T$, the set $\{s \in T : s \leq t\}$ is linearly ordered by $\leq$. For $t, s \in T$ write $\inf(t, s)$ for the $\leq$-largest element u such that $u \leq t$ and $u \leq s$. For trees T, S write $T \prec S$ if there is an injection $h : T \to S$ which preserves the ordering and infima.

Kruskal's tree theorem states that for every infinite sequence $\langle T_n : n \in \omega \rangle$ there are $n_0 < n_1$ such that $T_{n_0} \prec T_{n_1}$. This is not formalizable in Peano Arithmetic due to the quantification over infinite objects. We consider a *finitization* of this statement. For every $k \in \omega$ there is $m \in \omega$ such that for every sequence $\langle T_n : n < m \rangle$ in which every tree T_n has size at most $n + k$, there are $n_0 < n_1 < m$ such that $T_{n_0} \prec T_{n_1}$.

The finite version is formalizable, but not provable in Peano Arithmetic. The function $k \mapsto m$ grows extremely fast. Kruskal's theorem plays important role in computer science, proving termination of important algorithms for word problems.

Chapter 8

Computability

In this chapter, we formalize the notion of a "computable" function from natural numbers to natural numbers. There is a number of different approaches developed by separate research groups at about the same time in mid-1930's. They all lead to the same class of functions. This remarkable coincidence lead mathematicians to believe that this class of functions is truly the class of functions computable in an intuitive sense. This belief is encapsulated in a nonmathematical statement known as Church's thesis.

In the first three sections we develop three competing concepts of a computable function. In the fourth section, we show that these three concepts yield the same class of functions. The ultimate application of the concept of computability from mathematician's point of view is proving that certain naturally occurring problems are algorithmically unsolvable. In the last section of the chapter we will discuss some of these tough problems.

In several sections, we will speak about formal languages, and this is a suitable place to develop the appropriate notational conventions. An *alphabet* will always be just a finite nonempty set of symbols. A *word* in an alphabet Σ is just a finite sequence of symbols in Σ. One possible word is the empty word, denoted by 0. If $a \in \Sigma$ is a symbol and $n \in \omega$ is a natural number, a^n denotes the word consisting of n many a's. If v, w are words then vw denotes their concatenation. A *language* is a set of words in a fixed alphabet.

8.1 μ-recursive functions

Definition 8.1.1. Let $f : \omega^n \to \omega$ be a partial function. The symbol $f(x_i : i \in n) \uparrow$ denotes the fact that $f(x_i : i < n)$ is not defined. The function f is *total* if $f(x_i : i < n)$ is defined for each n-tuple $\langle x_i : i < n \rangle \in \omega^n$.

Definition 8.1.2. The class of partial μ-recursive functions is the smallest class containing

- the coordinate functions $f(x_i : i < n) = x_j$ for each $n > 0$ and $j < n$;

- the successor function $f(x) = x + 1$,

and closed under the following operations:

- composition: if f is a function of n variables and g_i for $i < n$ are all functions of m variables, obtain the function $h(g_0(x, y, z, \dots), g_1(x, y, z, \dots), \dots g_{n-1}(x, y, z \dots)$;
- primitive recursion: if f is a function of $n+2$ variables and g is a function of n variables, obtain the function h of $n+1$ variables given by $h(0, x_i : i < n) = g(x_i : i < n)$ and $h(m+1, x_i : i < n) = f(h(m, x_i : i < n), m, x_i : i < n)$;
- minimalization: if f is a function of $n+1$ variables then obtain a function μf of n variables, defined by $\mu f(x_i : i < n) = y$ if for every $z \leq y$ the functional value $f(x_i : i < n, z)$ is defined, if $z < y$ then this value is not zero, and if $z = y$ then this value is zero. If such y does not exist, then the value of $\mu f(x_i : i < n)$ is undefined.

Definition 8.1.3. The class of primitive recursive functions is the smallest class containing the coordinate functions and the successor function, and closed under the operation of composition and primitive recursion.

In particular, every primitive recursive function is *total*.

Example 8.1.4. Addition and multiplication are primitive recursive.

Proof. $x+y$ is defined by the recursive scheme $0+y = y$ and $(x+1)+y = (x+y)+1$. $x \cdot y$ is defined by the recursive scheme $0 \cdot y = 0$ and $(x+1) \cdot y = x \cdot y + y$. □

Example 8.1.5. The function $x \dot{-} y$, defined by $x \dot{-} y = 0$ if $x \leq y$ and $x \dot{-} y = x - y$ if $x > y$, is primitive recursive.

Proof. First, check that the function $g(x) = x \dot{-} 1$ is primitive recursive: $g(0) = 0$, $g(x+1) = x$. Then, define $x \dot{-} y$ by recursion on y: $x \dot{-} 0 = x$ and $x \dot{-} (y+1) = (x \dot{-} y) \dot{-} 1$. □

Example 8.1.6. The *Ackermann function* is total μ-recursive function which is not primitive recursive. It is uniquely given by the demands $A(0, n) = n + 1$, $A(m, 0) = A(m-1, 1)$, and $A(m, n) = A(m-1, A(m, n-1))$ if $m, n > 0$.

8.2 Turing machines

Another approach towards formalizing the notion of computability relies on modeling of computational devices. We will develop the simplest possibility, the deterministic finite automaton, as a baby case of the ultimate model, the Turing machine.

Remark. For Turing, the models were intended to model the work of secretaries in his office, as opposed to the (as yet nonexistent) computing devices. The (typically female) computing associates are the unsung heroes of applied mathematics before 1950. Armies of them were necessary to complete any significant job.

Definition 8.2.1. A *deterministic finite automaton* is a tuple $\langle \Sigma, S, A, s, T \rangle$ such that

- Σ is a finite nonempty set (the *alphabet*);
- S is a finite set (the set of *states*)
- $A \subset S$ is a set (the set of *accepting states*);
- $s \in S$ is the *starting state*;
- $T : S \times \Sigma \to S$ is a function.

Definition 8.2.2. If Σ is a finite set (an alphabet) then Σ^* is the set of all finite strings of elements of Σ (words). A *language* is a subset of Σ^*.

Definition 8.2.3. Let $\langle \Sigma, S, A, s, T \rangle$ be a finite automaton and $w \in \Sigma^*$ be a word of length n. A *computation* with input w is a sequence $\langle s_i : i \leq n \rangle$ of states such that $s_0 = s$ and for every $i < n$, $s_{i+1} = s_i = T(s_i, w(i))$. The automaton *accepts* the word w if $s_n \in A$; it *rejects* the word if $s_n \notin A$. A language L is *recognizable by a finite automaton* if there is an automaton such that for every word w, $w \in L$ if and only if the automaton accepts w.

Example 8.2.4. The language of all words of even length in a given alphabet is recognizable by finite automaton. Just let $S = \{s, t\}$, let the function T flip the state on any given input, and let $A = \{s\}$. Thus, for any given input word w, the computation on input w keeps oscillating between the states s, t. If it ends in the state s, the word has even length, otherwise the word has odd length.

Example 8.2.5. The language L of all words in the alphabet $\{a, b\}$ with equal number of occurences of letters a, b is not recognizable by finite automaton.

Proof. Suppose for contradiction that $\langle \Sigma, S, A, s, T \rangle$ is a finite automaton recognizing L. Let n be the size of the set S, and consider the word $w = a^{n+1}b^{n+1}$. In the computation on input w, the same state (call it t) must appear on two distinct positions $i < j < n + 1$. Let $m = j - i$ and consider the word $v = a^{n+1+m+1}b^{n+1}$. The computation on input v proceeds similarly as the computation on input w, with the difference that it traverses the cycle between the positions $i < j$ twice. Therefore, the computations on input v, w end in the same state. This is impossible, since $w \in L$ while $v \notin L$ and so w must be accepted while v must be rejected. □

The last example makes it clear that finite automaton is too weak a model for computation. The computing device must have an unlimited amount of memory for notes, otherwise the sheer amount of data may overwhelm it even in the case of very simple tasks.

Definition 8.2.6. A *Turing machine* is a tuple $\langle \Sigma, S, A, s, T \rangle$ such that

- Σ is a finite set of size at least two, with a designated "blank" symbol (the alphabet accepted by the machine);

- S is a finite set (the set of states);
- A is a subset of S (the set of accepting states);
- $s \in S$ is an element of S (the starting state);
- $T : S \times \sigma \to S \times \Sigma \times \{-1, 0, 1\}$ is a function (the action of the machine).

Intuitively speaking, the machine has a tape, which is a sequence of boxes indexed by (both positive and negative) integers. Each box can hold a single letter of the alphabet. The machine has a head that can read a single symbol on the tape. At a given stage of the computation, the machine reads the symbol in the location of its head, and depending on the state in which it is in, it moves to a different state, rewrites the symbol, and moves the head to the left or right on the tape (or the head stays in the same location). This intuition is formalized in the following definition.

Definition 8.2.7. Let $z : \mathbb{Z} \to \Omega$ be a function. A *run* of the machine on the input z is a sequence $\langle z_i, b_i, n_i : i \in \omega \rangle$ such that

- z_i is a function from $\mathbb{Z}$ to Σ, $s_i \in S$, and $n_i \in \mathbb{Z}$;
- $z_0 = z$, $s_0 = s$, $n_0 = 0$;
- if $T(s_i, z_i(n_i)) = (c, u, v)$ then $s_{i+1} = c$, $z_{i+1} = z_i$ except that the n_i-th entry of z_i is replaced with u, and $n_{i+1} = n_i + v$.

The machine *accepts* the input z if the run on the input z visits one of the accepting states, in other words *halts*. A language L is recognizable by a Turing machine if there is a Turing machine such that for every finite word w, the machine accepts w if and only if $w \in L$.

One of the most important differences between Turing machines and finite automatons is that computations of Turing machines may never halt; in such a case, the programmer never gets the information he most likely seeks.

There are many other computing devices that one can formalize. There may be multiple tapes, or FIFO or LIFO stacks present. These variations may make it easier to construct various machines, but they do not change the overall computational power of the device.

8.3 Post systems

Still another approach to computability was developed by Emil Leon Post in 1936. It is intended to model simple manipulations in algebra or calculus, but its computational power turns out to be equivalent to Turing machines. In this approach, the word, instead of serving as in input of a computational device, is obtained from a finite list of initial words (axioms) using a finite list of editing rules (productions).

Definition 8.3.1. Let Σ be an alphabet. A *production rule* is an expression of the form

$$g_0 S_0 g_1 S_1 \dots S_n g_{n+1} \to h_0 S_{i_0} h_1 S_{i_1} \dots S_{i_m} h_m$$

where

1. $g_0, g_1, \dots$ and $h_0, h_1, \dots$ are words (perhaps null words);
2. $i_0, i_1, \dots$ are numbers between 0 and n.

The production rule can be applied to a word w if w is of the form $g_0 v_0 g_1 v_1 \dots v_n g_{n+1}$ for some (perhaps null) words $v_0, v_1, \dots v_n$, and the application of the rule to the word w then results in a word $h_0 v_{i_0} h_1 v_{i_1} \dots v_{i_m} h_m$.

Example 8.3.2. The production rule $xSxyT \to xSSTxy$ can be applied to the word $xyxyxyx$ in two ways. In the first, we let $S = y$ and $T = xyx$ and produce $xyyxyxxy$. The second way obtains if we let $S = yxy$ and $T = x$ and produce $xyxyyxyxxy$.

Definition 8.3.3. A *Post system* is a pair $\langle A, P \rangle$ where A is a finite set of words (the *axioms*) and P is a finite set of production rules. The language *generated by* the Post system is the set of all words that can be obtained from some word in A by a finite succession of applications of the production rules in P. A language L in a finite alphabet Σ is Post-generable if there is a Post system in a possibly larger alphabet $\Delta \supset \Sigma$ such that the language K generated by it satisfies $K \cap \Sigma^* = L$.

Example 8.3.4. The language L consisting of all words in the language $\Sigma = \{a, b\}$ which have the same number of a's and b's is Post-generable.

Proof. Consider the Post system with just one axiom 0 and productions $ST \to SabT$ and $ST \to SbaT$. First of all, the word 0 is in the language L and the production rules applied to words in L lead again to words in L. Therefore, only words in L can be generated by the production rules in the system.

On the other hand, we can prove by induction on the length of the word $w \in L$ that w can be generated by repeated application of the production rules in the system. This is clear if the length of w is 0, since then $w = 0$ and w is the initial axiom. Suppose that the length of w is greater than 0 and for shorter words the induction hypothesis has been verified. The word w must contain either the group ab or the group ba, so it must be of the form $g_0 abg_1$ or $g_0 bag_1$ for some strings g_0, g_1. Now the word $v = g_0 g_1$ is in the language L, it is shorter than w, and so by the induction hypothesis it is obtained from 0 using the production rules in the system. Now, the word w is obtained from v using a single application of the production rules by the definition of v. □

Switching from generating languages to computing functions is easy.

Definition 8.3.5. A partial function $f : \omega^m \to \omega$ is *Post-computable* if the language L consisting of all expressions of the form $1^{n_0} : 1^{n_1} : \ldots 1^{n_{m-1}} : 1^{f(n_0,n_1,\ldots)}$ in the language $\{1, :\}$ is Post-generable.

Example 8.3.6. The function $f(n) = n^2$ is Post–computable.

Proof. The equality $(n+1)^2 = n^2+2n+1$ (itself a rewriting rule of sorts) plays a key role. Just let : be the only axiom of the Post system and $S : T \to S1 : TSS1$ be the only rewriting rule. It is easy to verify that the system produces the desired function. □

8.4 Putting it together

Theorem 8.4.1. *The following classes of functions are equal:*

1. *the class of μ-recursive functions;*
2. *the class of Turing-computable functions;*
3. *the class of Post-computable-functions;*
4. *the class of functions Σ_1-definable in $\mathfrak{N}$.*

To prove that every μ-recursive function is Σ_1, we will show that the basic functions are Σ_1 and that the generating operations applied to Σ_1 functions yield again Σ_1 functions.

The basic functions are easily Σ_1: for example, the function $f(x, y, z) = x$ is the set of all quadruples $\langle x, y, z, u\rangle$ such that $u = x$–so in fact it is definable by an atomic formula.

For the primitive recursion operation, suppose for definiteness that we are defining a function of two variables. Suppose that g, h are Σ_1 functions, g is a function of one variable and h is a function of three variables, and define f by the recursive scheme $f(0, y) = g(y)$ and $f(x + 1, y) = h(x, y, f(x, y))$. Then $f(x, y) = z$ is equivalent to the following formula $\phi(x, y, z)$: there is a code for a sequence s such that $s(0) = g(y)$ and $\forall u < x\ s(x + 1) = h(x, y, s(x))$ and $s(x) = z$. The formula ϕ is Σ_1 by the closure properties of Σ_1 properties in ???

For the search operation, suppose for definiteness that we are defining a function of one variable. Suppose that g is a Σ_1 function of two variables, and f is defined by the search operator: $f(y) = \mu g(x, y) = 0$. Then $f(y) = z$ is equivalent to the following formula $\phi(y, z)$: $\forall x < z\ \exists u\ u \neq 0 \wedge g(x, y) = u$ and $g(z, y) = 0$. The formula ϕ is Σ_1 by the closure properties of the class of Σ_1 formulas.

For composition, suppose for definiteness that we are composing functions of a single variable. Let g, h be Σ_1 functions, and let f be their composition: $f = g \circ h$. Then $f(x) = y$ is equivalent to the following formula $\phi(x, y)$: $\exists z\ h(x) = z \wedge g(z) = y$.

To prove that every Σ_1 function is μ-recursive, we will first show that

Claim 8.4.2. *The characteristic function of any Δ_0 formula is primitive-recursive.*

Here, the characteristic function of a Δ_0 formula $\phi(x, y)$ of say two free variables is the function $\chi_\phi : \omega^2 \to 2$ defined by $\chi_\phi(x, y) = 1 \leftrightarrow \phi(x, y)$ holds.

Proof. The proof proceeds by induction on the complexity of the Δ_0 formula ϕ. The atomic formulas are of the form $s \leq t$ for some terms s, t. The terms are primitive recursive functions of their variables, as they are built from the variables and 0 by adding one, addition, and multiplication. Then $\chi_{s \leq t}$ is equivalent to $(t \dot{-} s)$ which is primitive recursive by Example 8.1.5.

If ϕ, ψ are formulas whose characteristic functions are primitive-recursive, then also $\phi \wedge \psi$ has the same property, since its characteristic function is the product of χ_ϕ and χ_ψ. The negation is just as easy, since $\chi_{\neg\phi} = 1 \dot{-} \chi_\phi$.

Finally, consider the case of bounded quantifiers. Suppose that ϕ is a formula such that χ_ϕ is primitive recursive. Let x, y be variables such that y does not appear in ϕ. Then the characteristic function of $\forall x < y\ \phi$ is defined by primitive recursion on y as follows: $f(0) = 1$ and $f(y + 1) = f(y) \cdot \chi_\phi(y)$. If s is a term not mentioning x then the characteristic function of $\forall x < s\ \phi$ is defined as $f \circ s$. The case of a bounded existential quantifier is similar. □

Now suppose that f is a Σ_1 function; we must show that it is μ-recursive. For simplicity, assume that f is a function of a single variable. Thus, there is a Δ_0 formula $\phi(x, y, z)$ such that $f(x) = y$ is equivalent to $\exists z\ \phi(x, y, z)$. Use the μ-operator and the claim to show that the function $g(x)$ =the smallest code for a pair of natural number such that $1 \dot{-} \chi_\phi(x, s(0), s(1)) = 0$ is μ-recursive. Let h be the primitive recursive function such that for every code s for a pair of natural numbers, $g(s)$ =the first number coded by s. Then $f = h \circ g$ by the definitions and so f is μ-recursive.

To show that every μ-recursive function is Post-computable, we will show that the atomic functions are Post-computable, and the operations of primitive recursion, composition, and search appplied to Post-computable functions return Post-computable functions again.

For the atomic functions, adding 1 is computed by the system with the axiom : 1, and a single production rule $S : T \to S1 : T1$. The function $f(x, y, z) = y$ is computed by the system with a single axiom ::: and production rules $S : T : U : V \to S1 : T : U : V$, $S : T : U : V \to S : T1 : U : V1$ and $S : T : U : V \to S : T : U1 : V$. The other atomic functions are just as simple.

For the composition, suppose for simplicity that f, g are Post-computable functions of a single variable. We must produce a Post system which computes $f \circ g$. Suppose that f is computed by a Post system with axioms in the set A and productions in the set P; the function g is computed by a Post system with axioms in the set B and production rules in the set Q. The following system will the compute the composition $f \circ g$. The axioms are all words of the form be_0ae_1 where $b \in B$ and $a \in A$; the symbols e_0, e_1 are some fixed new symbols added to the alphabet. The production rules are of two kinds: the first kind includes the rules $q_0e_0Se_1 \to q_1e_0Se_1$ and $Se_0p_0e_1 \to Se_0p_1e_1$ whenever $q_0 \to q_1$ is a production rule in the set Q and $p_0 \to p_1$ is a production rule in the set P.

These rules allow a computation of the function g to the left of the e_0 symbol, and of the function f to the right of it. For the composition, we must check that the resulting output of g is equal to the input of f. This is achieved by the production rules $S1e_01Te_1 \to Se_0Te_1e_2$, $S1e_01Te_1e_2 \to Se_0Te_1e_2$, and finally $S : e_0 : Te_1e_2 \to S : T$. The repeated application of the first two productions in this group removes an equal number of 1's from the right and left side of the e_0 symbol, and the last production erases the scaffolding.

For the primitive recursion, assume for simplicity that g is a Post-computable function of two variables, and f is a function of a single variable defined by a primitive recursive scheme $f(0) = c$ and $f(n+1) = g(n, f(n))$, where c is some constant. We must show that f is Post-computable.

In view of the central theorem ???, we will make the following definitions.

Definition 8.4.3. A partial function $f : \omega^k \to \omega$ is *computable* if it is a Σ_1 subset of ω^{k+1}.

Definition 8.4.4. Let $k \in \omega$ be a natural number. A set $A \subset \omega^k$ is called *recursively enumerable*, or r.e., or computably enumerable, if it is Σ_1 definable in the structure $\mathfrak{N}$. The set A is called *recursive*, or computable, if both A and its complement are recursively enumerable.

8.5 Decidability

In this section, we want to use the preceding developments to provide some examples of algorithmically undecidable problems in mathematics, and discuss the methodology of proving that they in fact are algorithmically undecidable. The following definition is somewhat imprecise in its wording, but it is nevertheless extremely useful.

Definition 8.5.1. A problem is *algorithmically undecidable* if it is a yes-no question whose inputs can be coded efficiently with natural numbers such that the set of (codes for) inputs for which the answer is yes is not recursive.

Most interesting undecidable problems are such that the set of codes for inputs for which the answer is yes is recursively enumerable. In order to be able to identify any such problems, we must produce at least one recursively enumerable set which is not recursive. The initial example of such a set is obtained by an application of the diagonal method, and all the other ones refer to this initial example one way or another.

Definition 8.5.2. Let $k \in \omega$ be a natural number. A set $A \subset \omega \times \omega^k$ is *universal recursively enumerable set* in dimension k if

1. A is recursively enumerable;
2. for every recursively enumerable set $B \subset \omega^k$ there is $p \in \omega$ such that for every k-tuple $\vec{n} \in \omega^k$, $\langle p, \vec{n} \rangle \in A \leftrightarrow \vec{n} \in B$.

Theorem 8.5.3. *There is a universal recursively enumerable set in every dimension $k \in \omega$.*

Corollary 8.5.4. *There is a recursively enumerable set $C \subset \omega$ which is not recursive.*

Proof. Let $A \subset \omega \times \omega$ be a universal recursively enumerable set in dimension 1. Let $C = \{n \in \omega : \langle n, n\rangle \in A\}$. The set C is recursively enumerable by ???. We must show that C is not recursive, i.e. its complement is not recursively enumerable.

Suppose for contradiction that $\omega \setminus C$ is recursively enumerable.Then, by the universality of the set A, there must be a number $p \in \omega$ such that for every number $n \in \omega$, $\langle p, n\rangle \in A \leftrightarrow n \in \omega \setminus C$. Consider the question whether $p \in C$.

If $p \in C$, then by the definition of the set C, $\langle p, p\rangle \in A$, and so by the choice of the number p, $p \notin C$. This is a contradiction. On the other hand, if $p \notin C$, then by the definition of the set C again, $\langle p, p\rangle \notin A$ and so $p \in C$ by the choice of p. This is again a contradiction. No other cases are possible, so this completes the proof. □

There are many algorithmically undecidable problems in mathematics. Certain algorithmically undecidable problems are related to the notion of computation itself:

Example 8.5.5. (Halting problem) Decide whether a given Turing machine will terminate on blank input.

Example 8.5.6. (Busy beaver problem) Among the finitely many Turing machines on fixed number of states and fixed alphabet, find one which on blank input writes the longest sequence of nonblank symbols and halts.

A large class of undecidable problems comes from first order theories of various structures.

Example 8.5.7. (Tarski 1953) Theory of groups is undecidable. There is no algorithm deciding whether a sentence ϕ in the language of group theory is formally provable from axioms of group theory. By the completeness theorem, this is the same as to say that ϕ holds in all groups.

Example 8.5.8. Theory of finite groups is undecidable. There is no algorithm deciding whether a sentence in the language of group theory holds in all finite groups or not.

Example 8.5.9. (Robinson 1969) The theory of $\langle \mathbb{Q}, +, \cdot\rangle$ is undecidable.

Example 8.5.10. The theory of $\langle \mathbb{C}, +, \cdot, \exp\rangle$ is undecidable.

Other undecidable problems come from algebraic/combinatorial challenges.

Example 8.5.11. (Hilbert's 10th problem) (Matiyasevich) There is no algorithm deciding whether a given multivariate polynomial equation with integer coefficients has an integer solution.

Example 8.5.12. (Word problem ???)

Index

Bibliography

[1] Peter Aczel. *Non-well-founded sets*. CSLI Lecture Notes 14. Stanford University, Stanford, 1988.

[2] Michael Ben-Or, Dexter Kozen, and John Reif. The complexity of elementary algebra and geometry. *Journal of Computer and Systems Sciences*, 32:251264, 1986.

[3] James H. Davenport and Joos Heintz. Real quantifier elimination is doubly exponential. *J. Symb. Comput.*, 1988.

[4] Olga Kharlampovich and Alexei Myasnikov. Elementary theory of free non-abelian groups. *J. Algebra*, 302:451552, 2006.

[5] Casimir Kuratowski. Une méthode d'élimination des nombres transfinis des raisonnements mathématiques. *Fundamenta Mathematicae*, 3:76108, 1922.

[6] D. Anthony Martin. A purely inductive proof of Borel determinacy. In A. Nerode and R. A. Shore, editors, *Recursion theory*, number 42 in Proceedings of Symposia in Pure Mathematics, pages 303–308. American Mathematical Society, Providence, 1985.

[7] Jan Mycielski and H. Steinhaus. A mathematical axiom contradicting the axiom of choice. *Bulletin de l'Académie Polonaise des Sciences. Série des Sciences Mathématiques, Astronomiques et Physiques*, 10:13, 1962.

[8] W. V. Quine. *New Foundations for Mathematical Logic*, pages 80–101. Harvard Univ. Press, 1980.

[9] Bertrand Russel and Alfred Whitehead. *Principia Mathematica*. Cambridge University Press, Cambridge, 1910.

[10] Z. Sela. Diophantine geometry over groups. vi. The elementary theory of a free group. *Geom. Funct. Anal.*, 16:707730, 2006.

[11] Alfred Tarski. *A Decision Method for Elementary Algebra and Geometry*. Univ. of California Press, Los Angeles, 1951.

[12] J. von Neumann. Über die Definition durch transfinite Induktion und verwandte Fragen der allgemeinen Mengenlehre. *Mathematische Annalen*, 99:373391, 1928.

[13] Ernst Zermelo. Beweis, dass jede menge wohlgeordnet werden kann. *Math. Ann.*, 59:51–516, 1904.