Part II Logic and Set Theory

András Zsák

Lent 2024

Contents

1 Propositional Logic

Definition. The language of propositional logic consists of a set P of *primitive propositions* and the set $L = L(P)$ of *propositions* (or *compound propositions*), which is defined inductively as follows.

(i) $P \subset L$,

(ii) $\bot \in L$ (the symbol '$\bot$' is read 'false' or 'bottom'),

(iii) if $p, q \in L$ then $(p \Rightarrow q) \in L$.

Examples. We often use $P = \{p_1, p_2, p_3, \dots\}$. In other words, we need a countable infinite set of primitive propositions. The following are then examples of compound propositions.

$$(p_1 \Rightarrow p_2)\ , \quad ((p_1 \Rightarrow \bot) \Rightarrow p_3)\ , \quad ((p_1 \Rightarrow p_2) \Rightarrow (p_1 \Rightarrow p_3))$$

or for any $p \in L$,

$$((p \Rightarrow \bot) \Rightarrow \bot)$$

is also in L.

Remarks. 1. 'L defined inductively' means, more precisely, that $L = \bigcup_{n\in\mathbb{N}} L_n$, where

$$L_1 = P \cup \{\bot\}$$

and for $n \in \mathbb{N}$,

$$L_{n+1} = L_n \cup \{(p \Rightarrow q) : p, q \in L_n\} \ .$$

2. A proposition is a finite string of symbols from the alphabet $P \cup \{\bot,\ \Rightarrow,\ (,\ )\}$. It is easy to check that L is the smallest (with respect to inclusion) subset of the set Σ of all finite strings of symbols from the alphabet $P \cup \{\bot,\ \Rightarrow,\ (,\ )\}$ satisfying clauses (i)-(iii) above. Note that $L \neq \Sigma$. For example, the string $\Rightarrow p_1)($ is in Σ but is not a proposition.

3. Every proposition is built uniquely using clauses (i)-(iii) above, *i.e.*, for every $p \in L$, either $p \in P$ or $p = \bot$ or p is $(q \Rightarrow r)$ for unique $q, r \in L$.

4. How about other logical connectives like $\wedge$ and $\vee$? We will introduce these a little later.

Semantic Entailment

Definition. A *valuation* on L is a function $v\colon L \to \{0,1\}$ such that:

(i) $v(\bot) = 0$,

(ii) $v(p \Rightarrow q) = \begin{cases} 0 & \text{if } v(p) = 1, v(q) = 0, \\ 1 & \text{otherwise} \end{cases}$

for all $p, q \in L$.

Example. If $v(p_1) = 1$ and $v(p_2) = 0$, then $v\big((\bot \Rightarrow p_1) \Rightarrow (p_1 \Rightarrow p_2)\big) = 0$.

Proposition 1.

(i) If v and v' are valuations with $v\restriction_P = v'\restriction_P$, then $v = v'$.

(ii) For any function $w\colon P \to \{0,1\}$, there exist a valuation v on L such that $v\restriction_P = w$.

Remark. The proposition says that a valuation is determined by its values on P, and any values will do.

Proof. (i) Let $L' = \{p \in L : v(p) = v'(p)\}$. By assumption, $L_1 \subset L'$. If $p, q \in L'$, then $v(p \Rightarrow q) = v'(p \Rightarrow q)$ by definition. It follows that $L_n \subset L'$ implies $L_{n+1} \subset L'$ for every $n \in \mathbb{N}$. By induction, $L_n \subset L'$ for all $n \in \mathbb{N}$, and thus $v = v'$.
(ii) Set $v(p) = w(p)$ for every $p \in P$, and set $v(\bot) = 0$. This defines v on L_1. Having defined v on L_n for some $n \in \mathbb{N}$, for $p \in L_{n+1} \setminus L_n$, write $p = (q \Rightarrow r)$ for unique $q, r \in L_n$, and define

$$v(p) = \begin{cases} 0 & \text{if } v(q) = 1, v(r) = 0, \\ 1 & \text{otherwise.} \end{cases}$$

This defines v on L_{n+1}. Thus, we obtain a valuation v on L with $v\restriction_P = w$. □

Definition. Say $t \in L$ is a *tautology* if $v(t) = 1$ for all valuations v.

Definition. At this point, we enrich our language by adding the symbols $\top$ ('true' or 'top'), $\wedge$ ('and'), $\vee$ ('or') and $\neg$ ('not') as abbrevations as follows.

$$\begin{aligned} \top &= (\bot \Rightarrow \bot) \\ \neg p &= (p \Rightarrow \bot) \\ (p \vee q) &= (\neg p \Rightarrow q) \\ (p \wedge q) &= \neg(p \Rightarrow \neg q) \end{aligned}$$

for any $p, q \in L$. We have $v(\top) = 1$ for any valuation v. Similarly, $v(\neg p)$, $v(p \vee q)$ and $v(p \wedge q)$ have the expected values.

Examples. The following three examples of tautologies will play a role later.

1. $(p \Rightarrow (q \Rightarrow p))$ ('a true statement is implied by anything').

We check:

$v(p)$	$v(q)$	$v(q \Rightarrow p)$	$v(p \Rightarrow (q \Rightarrow p))$
0	0	1	1
0	1	0	1
1	0	1	1
1	1	1	1

Note that an identically 1 column indicates a tautology.

2. $(\neg\neg p \Rightarrow p)$, *i.e.*, $(((p \Rightarrow \bot) \Rightarrow \bot) \Rightarrow p)$ ('the law of excluded middle').

This can also be written as $(\neg p \vee p)$.

$v(p)$	$v(p \Rightarrow \bot)$	$v((p \Rightarrow \bot) \Rightarrow \bot)$	$v(((p \Rightarrow \bot) \Rightarrow \bot) \Rightarrow p)$
0	1	0	1
1	0	1	1

so a tautology.

3. $\big((p \Rightarrow (q \Rightarrow r)) \Rightarrow ((p \Rightarrow q) \Rightarrow (p \Rightarrow r))\big)$

Suppose not a tautology. Then

$$v\big(p \Rightarrow (q \Rightarrow r)\big) = 1 \quad \text{and} \quad v\big((p \Rightarrow q) \Rightarrow (p \Rightarrow r)\big) = 0$$

for some valuation v. Then $v(p \Rightarrow q) = 1$ and $v(p \Rightarrow r) = 0$. Thus, $v(p) = 1$, $v(r) = 0$, and so $v(q) = 1$, $v(q \Rightarrow r) = 0$ and $v\big(p \Rightarrow (q \Rightarrow r)\big) = 0$, a contradiction.

Definition. For $S \subset L$ and $t \in L$, say S *entails* t (or *semantically entails* t), written $S \models t$, if for every valuation v,

$$v(s) = 1 \text{ for all } s \in S \quad \text{implies} \quad v(t) = 1\ ,$$

i.e., 'whenever all of S is true, t is true as well.'

Examples. $\{p,\ p \Rightarrow q\} \models q \qquad \{p \Rightarrow q,\ q \Rightarrow r\} \models (p \Rightarrow r)$

Note. t is a tautology if and only if $\emptyset \models t$, which we abbreviate to $\models t$.

Definition. For $t \in L$, if $v(t) = 1$, then say t is *true in* v or v is a *model of* t. For $S \subset L$, a valuation v is a *model of* S if $v(s) = 1$ for all $s \in S$.

Note. $S \models t$ says that t is true in every model of S.

Syntactic Entailment

A notion of proof consists of axioms and deduction rules. In Propositional Logic we adopt the following propositions as axioms.

(A1) $p \Rightarrow (q \Rightarrow p) \quad (p, q \in L)$

(A2) $\big(p \Rightarrow (q \Rightarrow r)\big) \Rightarrow \big((p \Rightarrow q) \Rightarrow (p \Rightarrow r)\big) \quad (p, q, r \in L)$

(A3) $\neg\neg p \Rightarrow p \quad (p \in L)$

These are more accurately called ‘axiom-schemes’, as each is an infinite collection of axioms.

Note. The axioms are all tautologies.

We will have only one deduction rule, called *modus ponens*: ‘from p and $p \Rightarrow q$, can deduce q’.

For $S \subset L$ and $t \in L$, a *proof of* t *from* S is a finite sequence $t_1, t_2, \ldots, t_n$ of propositions such that $t_n = t$ and for each i,

(i) either t_i is an axiom,

(ii) or t_i is a member of S (*premiss* or *hypothesis*)

(iii) or t_i follows from earlier lines by modus ponens (MP): there exist $j, k < i$ with $t_k = (t_j \Rightarrow t_i)$.

If there exists a proof of t from S, say S *proves* t, or *syntactically entails* t, written $S \vdash t$.

Say t is a *theorem* if $\emptyset \vdash t$, which we simply denote $\vdash t$.

Examples. 1. $\{p \Rightarrow q,\ q \Rightarrow r\} \vdash (p \Rightarrow r)$

$$\begin{array}{ll}
\big(p \Rightarrow (q \Rightarrow r)\big) \Rightarrow \big((p \Rightarrow q) \Rightarrow (p \Rightarrow r)\big) & \text{(A2)} \\
(q \Rightarrow r) \Rightarrow (p \Rightarrow (q \Rightarrow r)) & \text{(A1)} \\
q \Rightarrow r & \text{(premiss)} \\
p \Rightarrow (q \Rightarrow r) & \text{(MP)} \\
(p \Rightarrow q) \Rightarrow (p \Rightarrow r) & \text{(MP)} \\
p \Rightarrow q & \text{(premiss)} \\
p \Rightarrow r & \text{(MP)}
\end{array}$$

2. $\vdash (p \Rightarrow p)$

$$\begin{array}{ll}
p \Rightarrow \big((p \Rightarrow p) \Rightarrow p\big) & \text{(A1)} \\
\Big(p \Rightarrow \big((p \Rightarrow p) \Rightarrow p\big)\Big) \Rightarrow \Big(\big(p \Rightarrow (p \Rightarrow p)\big) \Rightarrow (p \Rightarrow p)\Big) & \text{(A2)} \\
\big(p \Rightarrow (p \Rightarrow p)\big) \Rightarrow (p \Rightarrow p) & \text{(MP)} \\
p \Rightarrow (p \Rightarrow p) & \text{(A1)} \\
p \Rightarrow p & \text{(MP)}
\end{array}$$

In showing $S \vdash t$, the following result is often helpful.

Proposition 2. (Deduction Theorem) Let $S \subset L$ and $p, q \in L$. Then we have $S \vdash (p \Rightarrow q)$ if and only if $S \cup \{p\} \vdash q$.

Remark. So '$\Rightarrow$' really does behave like implication in formal proofs.

Note. To show $\{p \Rightarrow q,\ q \Rightarrow r\} \vdash (p \Rightarrow r)$ (Example 1 above), it is enough to show that $\{p \Rightarrow q,\ q \Rightarrow r,\ p\} \vdash r$. This is much easier: write down all three premisses and apply modus ponens twice.

Proof. Assume that $S \vdash (p \Rightarrow q)$. Write down a proof of $p \Rightarrow q$ from S and add the following two lines to obtain a proof of q from $S \cup \{p\}$:

$$\begin{array}{ll}
p & \text{(premiss)} \\
q & \text{(MP)}
\end{array}$$

Conversely, assume that $S \cup \{p\} \vdash q$ and let $t_1, t_2, \ldots, t_n = q$ be a proof of q from $S \cup \{p\}$. We show that S proves $p \Rightarrow t_i$ for each i by induction. Then in particular, S proves $p \Rightarrow q$ and we are done.

Case 1: If t_i is an axiom or $t_i \in S$, then

$$\begin{array}{ll}
t_i \Rightarrow (p \Rightarrow t_i) & \text{(A1)} \\
t_i & \text{(axiom or premiss)} \\
p \Rightarrow t_i & \text{(MP)}
\end{array}$$

is a proof of $p \Rightarrow t_i$ from S.

Case 2: If $t_i = p$, then $S \vdash (p \Rightarrow t_i)$ since $\vdash (p \Rightarrow p)$ (Example 2 above).

Case 3: Finally, if there exist $j, k < i$ such that $t_k = (t_j \Rightarrow t_i)$, then by induction hypothesis, there are proofs of $p \Rightarrow t_j$ and $p \Rightarrow (t_j \Rightarrow t_i)$ from S. Adding the lines

$$\begin{array}{ll}
\big(p \Rightarrow (t_j \Rightarrow t_i)\big) \Rightarrow \big((p \Rightarrow t_j) \Rightarrow (p \Rightarrow t_i)\big) & \text{(A2)} \\
(p \Rightarrow t_j) \Rightarrow (p \Rightarrow t_i) & \text{(MP)} \\
p \Rightarrow t_i & \text{(MP)}
\end{array}$$

we obtain a proof of $p \Rightarrow t_i$ from S. □

Our aim is to prove the Completeness Theorem: $S \vdash t$ if and only if $S \models t$. This is made up of *soundness* (if $S \vdash t$ then $S \models t$) and *adequacy* (if $S \models t$ then $S \vdash t$).

Soundness says that our notion of proof is sound: it doesn't lead to absurd conclusions. Adequacy says that our notion of proof is sufficiently strong to prove from S every semantic consequence of S.

Proposition 3. (Soundness Theorem) Let $S \subset L$ and $t \in L$. Then $S \vdash t$ implies $S \models t$.

Proof. Let $t_1, t_2, \dots, t_n = t$ be a proof of t from S. Let v be a model of S. We show that $v(t_i) = 1$ for all i by induction. Then in particular, $v(t) = 1$, as required.

If t_i is an axiom, then $v(t_i) = 1$ since axioms are tautologies. If t_i is a premiss, then $v(t_i) = 1$ since v is a model of S. Finally, if there exist $j, k < i$ such that $t_k = (t_j \Rightarrow t_i)$, then $v(t_j) = v(t_j \Rightarrow t_i) = 1$ by induction hypothesis, and hence $v(t_i) = 1$. □

Definition. Let $S \subset L$. Say S is *inconsistent* if $S \vdash \perp$; otherwise S is *consistent.*

Note. A special case of adequacy: $S \models \perp$ implies $S \vdash \perp$, *i.e.*, if S has no model, then S is inconsistent. Equivalently, if S is consistent, then S has a model.

Theorem 4. (Model existence lemma) Let $S \subset L$. If S is consistent, then S has a model.

Idea. Note that if $S \vdash t$, then $S \models t$ (soundness), so $v(t) = 1$ for every model v of S. So we could try

$$v(t) = \begin{cases} 1 & \text{if } S \vdash t\ , \\ 0 & \text{otherwise.} \end{cases}$$

However, this doesn't work. It is possible to have $t \in L$ with $S \nvdash t$ and $S \nvdash \neg t$. So we could try to enlarge S by adding either t or $\neg t$ for every $t \in L$, while keeping S consistent.

Proof. We present the proof in case the set P of primitive propositions is countable. The general case will be presented in Chapter 3. Note that if P is countable, then so is $L_1 = P \cup \{\perp\}$. It follows by induction that L_n is countable for every $n \in \mathbb{N}$, and hence $L = \bigcup_n L_n$ is also countable. We enumerate L as $t_1, t_2, t_3, \dots$.

Next observe that if $S \subset L$ is consistent and $t \in L$, then either $S \cup \{t\}$ or $S \cup \{\neg t\}$ is consistent. Indeed, if $S \cup \{t\} \vdash \perp$ and $S \cup \{\neg t\} \vdash \perp$, then $S \vdash \neg t$ by the Deduction Theorem, which in turn implies that $S \vdash \perp$ by modus ponens – a contradiction.

We now start with a consistent $S \subset L$, set $S_0 = S$ and define $S_n \subset L$ for each $n \in \mathbb{N}$ inductively as follows. Having defined S_{n-1} and assuming S_{n-1} is consitent, we let S_n be either $S_{n-1} \cup \{t_n\}$ or $S_{n-1} \cup \{\neg t_n\}$ whichever is consistent. Finally, we let $\overline{S} = \bigcup_{n \geqslant 0} S_n$.

By construction, for each $t \in L$, either $t \in \overline{S}$ or $\neg t \in \overline{S}$. We note the following two properties of $\overline{S}$.

$\overline{S}$ is consistent: if $\overline{S} \vdash \perp$, then since proofs are finite, we have $S_n \vdash \perp$ for some n – a contradiction.

$\overline{S}$ is *deductively closed*: if $\overline{S} \vdash t$, then $t \in \overline{S}$. Indeed, if $t \notin \overline{S}$, then $\neg t \in \overline{S}$. It follows that $\overline{S}$ proves both t and $\neg t$, and hence $\overline{S} \vdash \perp$ contradicting that $\overline{S}$ is consistent.

We now define $v\colon L \to \{0,1\}$ by

$$v(t) = \begin{cases} 1 & \text{if } t \in \overline{S}\ , \\ 0 & \text{otherwise.} \end{cases}$$

We show that v is a valuation on L. It will then follow that v is a model of S completing the proof.

Firstly, $v(\perp) = 0$ since $\perp \notin \overline{S}$ as $\overline{S}$ is consistent. Next, we examine $v(p \Rightarrow q)$ for arbitrary $p, q \in L$.

Case 1: $v(p) = 1$ and $v(q) = 0$, ie $p \in \overline{S}$ and $q \notin \overline{S}$. We need to show that $v(p \Rightarrow q) = 0$. If not, then $(p \Rightarrow q) \in \overline{S}$. Then the sequence

$$\begin{array}{ll} p & \text{(premiss)} \\ p \Rightarrow q & \text{(premiss)} \\ q & \text{(MP)} \end{array}$$

is a proof of q from $\overline{S}$. Since $\overline{S}$ is deductively closed, it follows that $q \in \overline{S}$ – contradiction.

Case 2: $v(q) = 1$, ie $q \in \overline{S}$. We need to show that $v(p \Rightarrow q) = 1$, *i.e.*, that $(p \Rightarrow q) \in \overline{S}$. The sequence

$$\begin{array}{ll} q & \text{(premiss)} \\ q \Rightarrow (p \Rightarrow q) & \text{(A1)} \\ p \Rightarrow q & \text{(MP)} \end{array}$$

is a proof of $(p \Rightarrow q)$ from $\overline{S}$. Since $\overline{S}$ is deductively closed, it follows that $(p \Rightarrow q) \in \overline{S}$.

Case 3: $v(p) = 0$, ie $p \notin \overline{S}$. We need to show that $v(p \Rightarrow q) = 1$, *i.e.*, that $(p \Rightarrow q) \in \overline{S}$. As in previous cases, since $\overline{S}$ is deductively closed, it is enough to show that $\overline{S} \vdash (p \Rightarrow q)$, which is equivalent to showing that $\overline{S} \cup \{p\} \vdash q$ by the Deduction Theorem. Note that $\neg p \in \overline{S}$. The following then is a proof of q from $\overline{S} \cup \{p\}$.

$$\begin{array}{ll} p & \text{(premiss)} \\ \neg p & \text{(premiss)} \\ \perp & \text{(MP)} \\ \perp \Rightarrow (\neg q \Rightarrow \perp) & \text{(A1)} \\ \neg\neg q & \text{(MP)} \\ \neg\neg q \Rightarrow q & \text{(A3)} \\ q & \text{(MP).} \end{array}$$

□

Corollary 5. (Adequacy Theorem) Let $S \subset L$ and $t \in L$. If $S \models t$, then $S \vdash t$.

Proof. If $S \models t$, then $S \cup \{\neg t\} \models \perp$. Hence by Theorem 4, we have $S \cup \{\neg t\} \vdash \perp$, which in turn implies $S \vdash \neg\neg t$ by the Deduction Theorem. We now obtain a proof of t from S by adding the following lines to a proof of $\neg\neg t$ from S.

$$\begin{array}{ll} \neg\neg t \Rightarrow t & \text{(A3)} \\ t & \text{(MP).} \end{array}$$

□

Theorem 6. (Completeness Theorem) Let $S \subset L$ and $t \in L$. Then $S \models t$ if and only if $S \vdash t$.

Proof. 'If' is soundness (Proposition 3).

'Only if' is adequacy (Corollary 5). □

Corollary 7. (Compactness Theorem) Let $S \subset L$ and $t \in L$. If $S \models t$, then there is a finite subset S' of S such that $S' \models t$.

Proof. Trivial if '$\models$' is replaced with '$\vdash$' since proofs are finite. □

Note. This is highly non-trivial without completeness. A special case of Corollary 7 is the following.

Corollary. Let $S \subset L$. If every finite subset of S has a model, then S had a model.

Proof. If S does not have a model, then $S \models \perp$. By Corollary 7 there is a finite subset S' of S such that $S' \models \perp$. □

Remark. Sometimes Corollary 7 is called the Compactness Theorem. It implies Corollary 7. Indeed, assume that $S \models t$. Then $S \cup \{\neg t\} \models \perp$. By Corollary 7 there is a finite subset S' of S such that $S' \cup \{\neg t\} \models \perp$, and thus $S' \models t$.

Corollary 8. (Decidability Theorem) Let $S \subset L$ be a finite set and $t \in L$. Then there is an algorithm that determines in finite time whether $S \vdash t$ or not.

Proof. Trivial if '$\vdash$' is replaced with '$\models$' by simply writing out a truth table S for the 2^n possible values of the primitive propositions appearing in members of S, where n is the number of such propositions. □

Remark. If $S \vdash t$, then a proof can be found in finite time: by writing out all proofs from S, we will eventually arrive at t. However, this algorithm does not terminate if $S \nvdash t$.

2 Well-Orderings and Ordinals

A *linear order* or *total order* on a set X is a relation $<$ on X that is

(i) *irreflexive*: $\neg(x < x)$ for all $x \in X$.

(ii) *transitive*: $\big((x < y) \wedge (y < z)\big) \Rightarrow (x < z)$ for all $x, y, z \in X$.

(iii) *trichotomous*: $(x < y) \vee (x = y) \vee (y < x)$ for all $x, y \in X$.

We will say 'X is linearly ordered by $<$' or simply 'X is a linearly ordered set'.

Note. In (iii) exactly one of the three possibilities hold. *E.g.,* if $x < y$ and $y < x$, then $x < x$ by (ii), which contradicts (i).

Examples. $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{R}$ in the usual order. Note that $\mathbb{N} = \{1, 2, 3, \dots\}$.

Note. For a set X of size at least 2, the relation on the power set $\mathbb{P}X$ of X (the set of all subsets of X) defined by $a < b$ if $a \subset b$ and $a \neq b$ is not trichotomous. Note that $a \subset b$ means: $(x \in a) \Rightarrow (x \in b)$ for all $x \in X$, which includes the case $a = b$.

Notation. We write '$x > y$' for '$y < x$' and '$x \leqslant y$' for '$x < y$ or $x = y$'. Then $\leqslant$ is

(i) *reflexive*: $x \leqslant x$ for all $x \in X$

(ii) *antisymmetric*: $\big((x \leqslant y) \wedge (y \leqslant x)\big) \Rightarrow (x = y)$ for all $x, y \in X$

(iii) *transitive*: $\big((x \leqslant y) \wedge (y \leqslant z)\big) \Rightarrow (x \leqslant z)$ for all $x, y, z \in X$

(iv) *trichotomous*: $(x \leqslant y) \vee (y \leqslant x)$ for all $x, y \in X$.

Note. If X is linearly ordered by $<$, then any subset Y of X is linearly ordered by $<$ (or, more precisely, by the restriction of $<$ to Y).

Definition. A *well-ordering* of a set X is a linear order $<$ on X such that every non-empty subset of X has a least element:

$$(\forall\, S \subset X)\ \big(S \neq \emptyset \Rightarrow (\exists\, x \in S)(\forall\, y \in S)(x \leqslant y)\big)$$

Note that the least element is unique by antisymmetry.

We will say 'X is well-ordered by $<$' or simply 'X is a well-ordered set'.

Examples. $\mathbb{N}$ is well-ordered by the usual order.

$\mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{R}$ are not (*e.g.,* they have no least element).

The subset $[0, \infty)$ of $\mathbb{R}$ does have a least element, but it is not well-ordered: *e.g.,* the subset $(0, \infty)$ has no least element.

Note. A subset of a well-ordered set is well-ordered. We will see that $\mathbb{Q}$ has a rich collection of well-ordered subsets.

Definition. Two linearly ordered sets X and Y are *order-isomorphic* if there is a bijection $f\colon X \to Y$ that is order-preserving: $x < y$ implies $f(x) < f(y)$.

We say f is an *order-isomorphism.* Note that f^{-1} is also an order-isomorphism, and thus $x < y \iff f(x) < f(y)$.

Note. If the linearly ordered sets X and Y are order-isomorphic, and X is well-ordered, then so is Y.

Examples. $\mathbb{N}$ and $\mathbb{Q}$ are not order-isomorphic.

$\mathbb{Q}$ and $\mathbb{Q} \setminus \{0\}$ are order-isomorphic.

$A = \{\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \dots\}$ is order-isomorphic to $\mathbb{N}$ ($n \mapsto \frac{n}{n+1}$).

$B = A \cup \{1\}$ is well-ordered, not order-isomorphic to $\mathbb{N}$.

$C = A \cup \{2\}$ is order-isomorphic to B.

$D = A \cup (A + 1)$ is well-ordered, not order-isomorphic to A or B.

Definition. A subset I of a linearly ordered set X is an *initial segment* of X if $x \in I$ and $y < x$ implies $y \in I$.

Examples. $\{1, 2, 3, 4\}$ is an initial segment of $\mathbb{N}$, $\{1, 2, 3, 5\}$ is not.

The real interval $(0, 1)$ is an initial segment of $(0, \infty)$.

In general, for every $x \in X$, the subset $I_x = \{y \in X : y < x\}$ is a proper initial segment of X by transitivity. Not every proper initial segment of X is of this form in general (*e.g.,* the subset $(-\infty, 1]$ of $\mathbb{R}$).

Note. If I is a proper initial segment of a well-ordered set X, then $I = I_x$ for x the least element of $X \setminus I$. Indeed, if $y \in I_x$, then $y < x$, so $y \in I$ by choice of x. If $y \in I$ and $x \leqslant y$, then $x \in I$ as I is an initial segment – contradiction, so $y \in I_x$.

Lemma 1. Let X and Y be well-ordered sets, I be an initial segment of Y and $f\colon X \to I$ be an order-isomorphism. Then $f(x)$ is the least element of $Y \setminus \{f(y) : y < x\}$ for every $x \in X$.

Proof. The set $A = Y \setminus \{f(y) : y < x\}$ is not empty since $f(x) \in A$. Let a be the least element of A. Then $a \leqslant f(x)$ and $f(x) \in I$, so $a \in I$. Thus $a = f(z)$ for some $z \in X$. We need to show that $z = x$.

Since $f(z) = a \leqslant f(x)$, it follows that $z \leqslant x$ as f is order-preserving.

If $z < x$, then $f(z) \notin A$ by definition of A, which contradicts the choice of a. Thus, $z = x$ as required. □

Proposition 2. (Proof by induction) Let X be a well-ordered set and $S \subset X$ such that for every $x \in X$ the following holds: if $y \in S$ for all $y < x$, then $x \in S$. Then $S = X$.

Note. Assume that S is defined in terms of a property p: $S = \{x \in X : p(x)\}$. Then Proposition 2 says:

$$(\forall\, x)\Big(\big[(\forall\, y < x)p(y)\big] \Rightarrow p(x)\Big) \implies (\forall\, x)p(x) \ .$$

The ‘base case’ ($p(x)$ for the least element x of X) is included in the assumption ‘$\big[(\forall\, y < x)p(y)\big] \Rightarrow p(x)$’.

Proof. Is $S \neq X$, then $X \setminus S$ has a least element x, say. For $y < x$, we have $y \in S$ by choice of x. It follows from the assumption on S that $x \in S$ – contradiction. □

Remark. We next show an example of how induction is used.

Proposition 3. Let X and Y be well-ordered sets that are order-isomorphic. Then there is a unique order-isomorphism from X to Y.

Note. This is false in general for linearly ordered sets. *E.g.,* from $\mathbb{Z} \to \mathbb{Z}$ we have $n \mapsto n$ or $n \mapsto n + 17$; from $[0, \infty) \to [0, \infty)$ we have $x \mapsto x$ or $x \mapsto x^2$.

Proof. Let $f, g\colon X \to Y$ be order-isomorphisms. We show $(\forall\, x)(f(x) = g(x))$ by induction.

Let $x \in X$ and assume that $f(y) = g(y)$ for all $y < x$ (the induction hypothesis). We need to show that $f(x) = g(x)$. By Lemma 1, $f(x)$ is the least element of $A = Y \setminus \{f(y) : y < x\}$, and $g(x)$ is the least element of $B = Y \setminus \{g(y) : y < x\}$. By the induction hypothesis, $A = B$, and hence $f(x) = g(x)$. □

Remark. Induction allows us to prove things. We will also need a tool to construct things: recursion. We first recall that a function from a set X to a set Y is a subset f of $X \times Y$ such that

(i) for all $x \in X$ there exists $y \in Y$ with $(x, y) \in f$;

(ii) for all $x \in X$ and $y, z \in Y$, if $(x, y) \in f$ and $(x, z) \in f$, then $y = z$.

We of course write '$y = f(x)$' for '$(x, y) \in f$' and say that 'f maps x to y'.

Note that $f \in \mathbb{P}(X \times Y)$.

For $Z \subset X$, the restriction of f to Z is $f{\restriction_Z}= \{(x, y) \in f : x \in Z\}$ which is a function $Z \to Y$. Note that $f{\restriction_Z}$ is a subset of $Z \times Y$, and so in particular $f{\restriction_Z}$ is also an element of $\mathbb{P}(X \times Y)$.

Theorem 4. (Definition by recursion) Let X be a well-ordered set and Y an arbitrary set. Then for every function $G\colon \mathbb{P}(X \times Y) \to Y$ there is a unique function $f\colon X \to Y$ such that $f(x) = G(f{\restriction_{I_x}})$ for all $x \in X$.

Proof. Say h is an *attempt* if h is a function $I \to Y$, where I, the domain of h denoted $\operatorname{dom}(h)$, is an initial segment of X such that $h(x) = G(h{\restriction_{I_x}})$ for every $x \in \operatorname{dom}(h)$ (note that $x \in \operatorname{dom}(h)$ implies $I_x \subset \operatorname{dom}(h)$). We need to show that there is a unique attempt whose domain is X.

We first show that if h, h' are attempts then $h(x) = h'(x)$ for all $x \in \operatorname{dom}(h) \cap \operatorname{dom}(h')$. We prove this by induction. Fix $x \in \operatorname{dom}(h) \cap \operatorname{dom}(h')$ and assume that $h(y) = h'(y)$ for all $y < x$ (induction hypothesis). Then $h{\restriction_{I_x}}= h'{\restriction_{I_x}}$, and thus $h(x) = G(h{\restriction_{I_x}}) = G(h'{\restriction_{I_x}}) = h'(x)$.

We complete the proof by setting $f = \bigcup\{h : h \text{ is an attempt}\}$. Then f is a function since for any $x \in \operatorname{dom}(f)$, the value $h(x)$, where h is an attempt defined at x, is independent of h by what we showed above.

The domain of f is the union $\bigcup\{\mathrm{dom}(h) : h \text{ is an attempt}\}$, and thus $\mathrm{dom}(f)$ is an initial segment of X.

For any $x \in \mathrm{dom}(f)$, there is an attempt h such that $f(x) = h(x)$. It follows that $f(y) = h(y)$ for all $y < x$, and hence $f(x) = h(x) = G(h\restriction_{I_x}) = G(f\restriction_{I_x})$. Thus, f is an attempt.

Finally, assume that $\mathrm{dom}(f) \neq X$. Then $\mathrm{dom}(f) = I_x$ for some $x \in X$. In particular, there is no attempt defined at x. However, $f \cup \{(x, G(f)\}$ is an attempt defined at x. Thus, f is defined on the whole of X. □

Proposition 5. (Subset collapse) Let Y be a well-ordered set and $X \subset Y$. Then X is order-isomorphic to a unique initial segment of Y.

Proof. For uniqueness, assume that f is an order-isomorphism from X to an initial segment of Y. By Lemma 1, we have

$$f(x) = \min\big(Y \setminus \{f(y) : y \in X,\ y < x\}\big) \ .$$

It follows by induction that f is uniquely determined.

For existence, we may assume $Y \neq \emptyset$. Fix $y_0 \in Y$ and define $f \colon X \to Y$ by recursion as follows:

$$f(x) = \begin{cases} \min\big(Y \setminus \{f(y) : y \in X,\ y < x\}\big) & \text{if this exists,} \\ y_0 & \text{otherwise.} \end{cases}$$

We first show that the 'otherwise' clause never arises by showing that $f(x) \leqslant x$ for all $x \in X$. Indeed, fix $x \in X$ and assume that $f(y) \leqslant y$ holds for all $y \in X$ with $y < x$. Then $x \in Y \setminus \{f(y) : y \in X,\ y < x\}$, and hence $f(x) \leqslant x$. The claim follows by induction.

Given $y < x$ in X, since

$$f(x) \in Y \setminus \{f(z) : z \in X,\ z < x\} \subset Y \setminus \{f(z) : z \in X,\ z < y\} \ ,$$

it follows that $f(y) < f(x)$. Thus, f is order-preserving.

Finally, assume that $a \in Y \setminus \mathrm{im}(f)$. We show by induction that $f(x) < a$ for all $x \in X$, which shows that $\mathrm{im}(f)$ is an initial segment of Y. Fix $x \in X$ and assume that $f(y) < a$ for all $y \in X$ with $y < x$. Then $a \in Y \setminus \{f(y) : y \in X,\ y < x\}$, and thus $f(x) < a$, as required. □

Remark. It follows from Proposition 5 that a well-ordered set X is not order-isomorphic to any proper initial segment of X.

Notation. For well-ordered sets X, Y, we write $X \leqslant Y$ if X is order-isomorphic to an initial segment of Y.

Theorem 6. Let X, Y be well-ordered sets. Then either $X \leqslant Y$ or $Y \leqslant X$.

Proof. Assume that $Y \nleqslant X$. Then in particular, $Y \neq \emptyset$. Fix $y_0 \in Y$ and define $f \colon X \to Y$ by recursion as follows.

$$f(x) = \begin{cases} \min\big(Y \setminus \{f(y) : y < x\}\big) & \text{if this exists,} \\ y_0 & \text{otherwise.} \end{cases}$$

If the 'otherwise' clause ever arises, then let x be the least element of X for which this happens. Then $f(I_x) = Y$ and for $y < x$ the 'otherwise' clause does not arise in the definition of $f(y)$. It follows as in the proof of Proposition 5 that f is an order-isomorphism from I_x to Y contradicting $Y \not\leqslant X$. Hence the 'otherwise' clause never arises, and then it follows again as in the proof of Proposition 5 that f is an order-isomorphism from X to an initial segment of Y. □

Proposition 7. Let X, Y be well-ordered sets. If $X \leqslant Y$ and $Y \leqslant X$, then X and Y are order-isomorphic.

Proof. Let $f\colon X \to Y$ and $g\colon Y \to X$ be order-isomorphisms to initial segments of Y and X, respectively. Then $g \circ f$ is an order-isomorphism from X to an initial segment of X. By uniqueness in Proposition 5, it follows that $g \circ f = \mathrm{Id}_X$, the identity on X. Similarly, $f \circ g = \mathrm{Id}_Y$. □

Remark. What the above shows is that '$\leqslant$' is a linear order (reflexive, antisymmetric, transitive and trichotomous) on the collection of well-ordered sets provided we identify order-isomorphic sets. (We haven't showed transitivity but that is straightforward.) It is natural to introduce the corresponding '$<$' sign as follows. For well-ordered sets X, Y, write $X < Y$ to mean '$X \leqslant Y$ and X not order-isomorphic to Y'. Equivalently, $X < Y$ if and only if X is order-isomorphic to a proper initial segment of Y. Then '$<$' is irreflexive, transitive and trichotomous. A natural question arises: Is the collection of all well-ordered sets a well-ordered set? We return to this question later in the chapter, but first show how to build new well-ordered sets from old ones.

There is always another one. Let X be a well-ordered set. Fix $z \notin X$ and define $X^+ = X \cup \{z\}$ well-ordered by extending the well-ordering $<$ on X by defining $x < z$ for all $x \in X$. Then X^+ is uniquely defined up to order-isomorphism (*i.e.*, it does not depend on the choice of z) and $X < X^+$.

Note. For any set X, there is always some z not in X, for example, because there is no surjection from X to the power set $\mathbb{P}X$ by Cantor's diagonal argument.

Upper bounds. We next want to show that if $\{X_i : i \in I\}$ is a set of well-ordered sets, then there is a well-ordered set X such that $X_i \leqslant X$ for all $i \in I$. For well-ordered sets $(X, <_X)$ and $(Y, <_Y)$, we say Y *extends* X if $X \subset Y$, $<_X$ is the restriction of $<_Y$ to X (formally, $<_X = <_Y \cap (X \times X)$) and X is an initial segment of Y. We say that the set $\{X_i : i \in I\}$ is *nested* if for all $i, j \in I$, either X_j extends X_i or X_i extends X_j.

Proposition 8. Let $\{X_i : i \in I\}$ be a nested set of well-ordered sets. Then there is a well-ordered set X such that $X_i \leqslant X$ for all $i \in I$.

Proof. Set $X = \bigcup_{i \in I} X_i$ and for $x, y \in X$ set $x < y$ if and only if there exists $i \in I$ with $x, y \in X_i$ and $x <_i y$, where $<_i$ denotes the well-ordering of X_i. From the assumption that the X_i are nested, it follows that $<$ is a well-defined linear order on X and each X_i is an initial segment of X.

Given a non-empty subset $S \subset X$, we have $S \cap X_i \neq \emptyset$ for some $i \in I$. Since X_i is well-ordered, $S \cap X_i$ has a least element x. Since X_i is an initial segment of X, it follows that x is a least element of S. □

Remark. The same result holds without the assumption that the X_i are nested (see Chapter 5).

Ordinals

Definitions. An *ordinal* is a well-ordered set with two ordinals regarded the same if they are order-isomorphic. The *order-type* of a well-ordered set X is the unique ordinal to which X is order-isomorphic.

Remark. The formal definition of ordinal will be given in Chapter 5. For now you can view the word ordinal as shorthand for identifying well-ordered sets that are order-isomorphic. The results in this chapter can be expressed purely in terms of well-ordered sets.

Examples. For $k \in \{0\} \cup \mathbb{N}$ we write k for the order-type of a well-ordered set of size k. We let ω denote the order-type of $\mathbb{N}$ (same as order-type of $\{0\} \cup \mathbb{N}$). Note that in $\mathbb{Q}$, the set $A = \{\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \dots\}$ also has order-type ω.

Definition. Let α, β be ordinals and X, Y be well-ordered sets with order-types α, β, respectively. We write $\alpha \leqslant \beta$ of $X \leqslant Y$ and $\alpha < \beta$ if $X < Y$. Similarly, we write α^+ for the order-type of X^+.

Note. The notions above are well-defined, *i.e.*, they don't depend on the choice of X and Y. For arbitrary ordinals α, β, we have either $\alpha \leqslant \beta$ or $\beta \leqslant \alpha$ (Theorem 6), and if $\alpha \leqslant \beta$ and $\beta \leqslant \alpha$, then $\alpha = \beta$ (Proposition 7).

Theorem 9. Let α be an ordinal. Then the ordinals strictly smaller than α form a well-ordered set of order-type α.

Proof. Fix a well-ordered set X with order-type α and form the set

$$X' = \{Y \subset X : Y \text{ is an initial segment of } X\} .$$

Then X' is linearly ordered by the relation $<$ by Propositions 5 and 7 and by Theorem 6. The map $x \mapsto I_x$ is an order-isomorphism from X to X', and thus X' is well-ordered with irder-type α. The set

$$\{\text{order-type}(Y) : Y \in X'\}$$

is then a well-ordered set consisting precisely of all ordinals strictly less than α. □

Note. It is natural to denote the set of ordinals $\{\beta : \beta < \alpha\}$ by I_α. It is an of a well-ordered set of order-type α.

Proposition 10. A non-empty set S of ordinals has a least element.

Proof. Fix $\alpha \in S$. If α is not a least element of S, then $S \cap I_\alpha \neq \emptyset$. Then $S \cap I_\alpha$ has a least element β by Theorem 9. Since I_α is an initial segment of ordinals, it follows that β is a least element of S. □

Theorem 11. (Burali-Forti paradox) The ordinals do not form a set.

Proof. Assume that the ordinals do form a set X. Then X is well-ordered by Proposition 10. Let α be the order-type of X. Then I_α is a proper initial segment of X that is order-isomorphic to X contradicting Proposition 5. □

Remark. Let $S = \{\alpha_i : i \in I\}$ be a set of ordinals. Applying Proposition 8 to the nested set $\{I_{\alpha_i} : i \in I\}$, we obtain an ordinal α that is an upper bound for S. A minimal element of the non-empty set $\{\beta \in I_\alpha \cup \{\alpha\} : \beta \text{ is an upper bound for } S\}$. of ordinals is a least upper bound for S denoted $\sup S$. It is easy to see that if $\alpha = \sup S$, then $I_\alpha = \bigcup_{i\in I} I_{\alpha_i}$.

Examples. We now list further examples of ordinals. We have already seen $0, 1, 2, \dots, \omega$. Note that $\omega = \sup\{0, 1, 2, \dots\}$. The next ordinal is ω^+ which we write as $\omega + 1$. For now this is just notation but it is consistent with ordinal addition defined later in this chapter. We then have $\omega + 1, \omega + 2, \dots$. It is natural to denote $\sup\{\omega, \omega + 1, \omega + 2, \dots\}$ by $\omega + \omega$ or by $\omega \cdot 2$. The latter notation is consistent with ordinal multiplication defined later in the chapter. We continue with a few more ordinals.

$0,\ 1,\ 2,\ 3, \dots\ \omega,\ \omega+1$ (officially ω^+), $\omega+2$ (officially ω^{++}), $\omega+3,\ \dots$

$\omega+\omega = \omega\cdot 2$ (officially $\sup\{\omega,\ \omega+1,\ \omega+2,\ \dots\}$), $\omega\cdot 2+1,\ \omega\cdot 2+2,\ \omega\cdot 2+3,$

$\dots\ \omega\cdot 3,\ \dots\ \omega\cdot 4\ \dots \omega\cdot 5,\ \omega\cdot\omega = \omega^2$ (officially $\sup\{\omega,\ \omega\cdot 2,\ \omega\cdot 3,\ \dots\}$), $\omega^2+1,$

$\omega^2+2,\ \dots\ \omega^2+\omega,\ \omega^2+\omega+1,\ \dots\ \omega^2+\omega\cdot 2,\ \dots\ \omega^2+\omega\cdot 3,\ \dots\ \omega^2+\omega^2 = \omega^2\cdot 2,$

$\omega^2\cdot 2+1,\ \dots\ \omega^2\cdot 2+\omega,\ \dots\ \omega^2\cdot 3,\ \dots\ \omega^2\cdot 4,\ \dots\ \omega^3,\ \dots\ \omega^3\cdot 2,\ \dots\ \omega^3\cdot 3,\ \dots$

$\omega^4,\ \dots\ \omega^5,\ \dots\ \omega^\omega$ (officially $\sup\{\omega,\ \omega^2,\ \omega^3,\ \dots\}$), $\dots\ \omega^\omega\cdot 2,\ \dots\ \omega^\omega\cdot 3,\ \dots$

$\omega^\omega\cdot\omega = \omega^{\omega+1},\ \dots\ \omega^{\omega+2},\ \dots\ \omega^{\omega+3},\ \dots\omega^{\omega\cdot 2},\ \dots\ \omega^{\omega\cdot 3},\ \dots\ \omega^{\omega\cdot 4},\ \dots\ \omega^{\omega^2} = \omega^{(\omega^2)},$

$\dots\ \omega^{\omega^2\cdot 2},\ \dots\ \omega^{\omega^2\cdot 3},\ \dots\ \omega^{\omega^3},\ \dots\ \omega^{\omega^4},\ \dots\ \omega^{\omega^\omega},\ \dots\ \omega^{\omega^{\omega^2}},\ \dots\ \omega^{\omega^{\omega^3}},\ \dots\ \omega^{\omega^{\omega^\omega}},$

$\dots\ \varepsilon_0 = \omega^{\omega^{\omega^{\omega^{\cdot^{\cdot^{\cdot}}}}}}$ (officially $\sup\{\omega,\ \omega^\omega,\ \omega^{\omega^\omega},\ \dots\}$), $\varepsilon_0+1,\ \varepsilon_0+2,\ \dots\ \varepsilon_0+\omega,$

$\dots\ \varepsilon_0+\varepsilon_0 = \varepsilon_0\cdot 2,\ \dots\ \varepsilon_0\cdot 3,\ \dots\ \varepsilon_0\cdot\omega,\ \dots\ \varepsilon_0\cdot\varepsilon_0 = \varepsilon_0^2,\ \dots\ \varepsilon_0^3,\ \dots\ \varepsilon_0^\omega,\ \dots\ \varepsilon_0^{\omega^2},$

$\dots\ \varepsilon_0^{\omega^3},\ \dots\ \varepsilon_0^{\omega^\omega},\ \dots\ \varepsilon_0^{\omega^{\omega^{\omega^{\cdot^{\cdot^{\cdot}}}}}} = \varepsilon_0^{\varepsilon_0},\ \dots\ \varepsilon_0^{\varepsilon_0^{\varepsilon_0}},\ \dots\ \varepsilon_0^{\varepsilon_0^{\varepsilon_0^{\varepsilon_0}}},\ \dots\ \varepsilon_1 = \varepsilon_0^{\varepsilon_0^{\varepsilon_0^{\varepsilon_0^{\cdot^{\cdot^{\cdot}}}}}},\ \dots\varepsilon_2,$

$\dots\ \varepsilon_3,\ \dots\varepsilon_\omega,\ \dots\ \varepsilon_{\varepsilon_0},\ \dots\ \varepsilon_{\varepsilon_{\varepsilon_0}},\ \dots\ \varepsilon_{\varepsilon_{\varepsilon_{\cdot_{\cdot_{\cdot}}}}},\ \dots$

Note. All the ordinals above are countable! By this we mean, of course, that they are order-types of countable well-ordered sets.

Questions. Does there exist an uncountable ordinal? *I.e.,* does there exist an uncountable well-ordered set? Can we well-order $\mathbb{R}$?

Theorem 12. There exist an uncountable ordinal.

Idea. If there is an uncountable ordinal, then there is a least one, α say. Then I_α is the set of countable ordinals, *i.e.*, the set of order-types of well-orderings of subsets of $\mathbb{N}$.

Proof. We can form the set

$$A = \{(M, R) \in \mathbb{PN} \times \mathbb{P}(\mathbb{N} \times \mathbb{N}) : \ R \text{ is a well-ordering of } X\}$$

of well-orderings of subsets of $\mathbb{N}$. Then the set

$$B = \{\text{order-type}(X) : \ X \in A\}$$

consists of all countable ordinals. Set $\omega_1 = \sup B$. If ω_1 is countable, then so is ω_1^+, and hence $\omega_1^+ \in B$. It follows that $\omega_1^+ \leqslant \omega_1$, which is a contradiction. □

Note. The ordinal ω_1 constructed in the proof above is the least uncountable ordinal. Every proper initial segment of ω_1 is countable. If $\alpha_1, \alpha_2, \dots$ are countable ordinals, then so is $\sup\{\alpha_1, \alpha_2, \dots\}$ being the order-type of $\bigcup_{n\in\mathbb{N}} I_{\alpha_n}$. (Here we are using the fact that a countable union of countable sets is countable.)

Theorem 13. (Hartogs' Lemma) For every set X, there is an ordinal γ which does not inject into X.

Proof. This is a generalisation of Theorem 12. We form the set B of order-types of well-orderings of subsets of X. We let $\gamma = (\sup B)^+$. If there is an injection from γ to X, then this induces a well-ordering on a subset of X with order-type γ. Thus, $\gamma \in B$, and so $\gamma \leqslant \sup B < \gamma$ — contradiction. □

Types of ordinals. Let α be an ordinal. We have two cases according to whether α (or, more precisely, any well-ordered set of order-type α, *e.g., I_α*) has a greatest element or not.

If I_α has greatest element β, then $I_\alpha = I_\beta \cup \{\beta\}$, and hence $\alpha = \beta^+$. In this case, we say α is a *successor ordinal.* Note that in this case $\beta = \sup I_\alpha < \alpha$.

If I_α has no greatest element, then $\alpha = \sup I_\alpha$ and we say α is a *limit ordinal*

Examples. Rather trivially, 0 is a limit ordinal, whereas $1 = 0^+$ is a successor. Any ordinal n with $0 < n < \omega$ is a successor since any non-empty, finite well-ordered set has a greatest element. Since $\omega = \sup\{n : \ n < \omega\}$ is a limit and ω^+ is a successor.

Ordinal Arithmetic. For ordinals α, β, we define $\alpha + \beta$ by recursion on β, with α fixed, as follows.

$$\begin{aligned}
&\alpha + 0 = \alpha \\
&\alpha + \beta^+ = (\alpha + \beta)^+ \\
&\alpha + \lambda = \sup\{\alpha + \beta : \ \beta < \lambda\} \qquad \text{for a non-zero limit ordinal } \lambda \ .
\end{aligned}$$

Remark. Technically, since the ordinals do not form a set, we need to fix an ordinal γ and define $\alpha+\beta$ for $\beta<\gamma$ by recursion in the well-ordered set I_γ. The definition is then independent of γ by uniqueness of recursion. This justifies the recursive definition above and others given below.

In a similar way, induction on ordinals works even though ordinals do not form a set. Indeed, assume p is a property of ordinals. Then

$$(\forall\,\alpha)((\forall\,\beta)((\beta<\alpha)\Rightarrow p(\beta))\Rightarrow p(\alpha)) \implies (\forall\,\alpha)p(\alpha)$$

since otherwise we have $\neg p(\gamma)$ for some γ and the non-empty set $\{\beta\leqslant\gamma:\,\neg p(\beta)\}$ has a least element α. By minimality, we then have $(\forall\,\beta)((\beta<\alpha)\Rightarrow p(\beta))$ which implies $p(\alpha)$ by above — a contradiction.

Examples. $\alpha+1=\alpha+0^+=(\alpha+0)^+=\alpha^+$ for every ordinal α.

$m+0=m$ and $m+(n+1)=m+n^+=(m+n)^+=(m+n)+1$ for any $m,n<\omega$. This is the usual inductive definition of integer addition.

$\omega+1=\omega^+$ by above, and so $\omega+2=\omega+1^+=(\omega+1)^+=\omega^{++}$.

$\omega+\omega=\sup\{\omega+n:\,n<\omega\}=\sup\{\omega,\omega+1,\omega+2,\dots\}$.

$1+\omega=\sup\{1+n:\,n<\omega\}=\sup\{1,2,3,\dots\}=\omega\neq\omega+1$. Thus ordinal addition is not commutative.

Proposition 14. $\beta\leqslant\gamma$ implies $\alpha+\beta\leqslant\alpha+\gamma$.

Proof. We proceed by induction on γ with α fixed. We consider three cases.

$\gamma=0$: if $\beta\leqslant\gamma$, then $\beta=0$, so $\alpha+\beta=\alpha+\gamma=\alpha$.

$\gamma=\delta^+$ is a successor: if $\beta\leqslant\gamma$, then either $\beta=\gamma$ and the result is clear, or $\beta\leqslant\delta$, and so by induction,

$$\alpha+\beta\leqslant\alpha+\delta<(\alpha+\delta)^+=\alpha+\beta\ .$$

γ is a non-zero limit ordinal: if $\beta\leqslant\gamma$, then again we can assume that $\beta<\gamma$, and so $\beta<\delta$ for some $\delta<\gamma$ by the definition of a limit ordinal. It follows by induction that

$$\alpha+\beta\leqslant\alpha+\delta\leqslant\alpha+\gamma\ .$$

□

Remark. It follows directly from the result above that $\beta<\gamma$ implies $\alpha+\beta<\alpha+\gamma$. Indeed, we have $\beta^+\leqslant\gamma$, and hence

$$\alpha+\beta<(\alpha+\beta)^+=\alpha+\beta^+\leqslant\alpha+\gamma\ .$$

Note, however, that $\beta<\gamma$ does not imply $\beta+\alpha<\gamma+\alpha$ in general. For example, $1<2$ but $1+\omega=2+\omega=\omega$. On the other hand, $\beta\leqslant\gamma$ does imply $\beta+\alpha\leqslant\gamma+\alpha$ (by induction on α).

Lemma 15. Let S be a non-empty set of ordinals. Then

$$\alpha + \sup S = \sup\{\alpha + \beta : \beta \in S\}$$

for any ordinal α,

Proof. Put $T = \{\alpha + \beta : \beta \in S\}$.

For $\beta \in S$, we have $\beta \leqslant \sup S$, and hence $\alpha + \beta \leqslant \alpha + \sup S$ by Proposition 14. It follows that $\sup T \leqslant \alpha + \sup S$.

To show the reverse inequality, first assume that S has a greatest element γ. Then $\gamma = \sup S$, and $\alpha + \gamma$ is also the greatest element of T by Proposition 14. It follows that $\sup T = \alpha + \gamma = \alpha + \sup S$.

Now assume that S has no greatest element. Then $\lambda = \sup S$ is a non-zero limit ordinal. Indeed, $\lambda \notin S$ as S has no greatest element, and so $S \subset I_\lambda$, which implies that $\lambda = \sup S \leqslant \sup I_\lambda$, so I_λ has no greatest element. Next, given $\gamma < \lambda$, there exists $\beta \in S$ with $\gamma < \beta$. Hence $\alpha + \gamma \leqslant \alpha + \beta \leqslant \sup T$. It follows that

$$\alpha + \sup S = \alpha + \lambda = \sup\{\alpha + \gamma : \gamma < \lambda\} \leqslant \sup T \ .$$

□

Proposition 16. $\alpha + (\beta + \gamma) = (\alpha + \beta) + \gamma$.

Proof. We proceed by induction on γ with α, β fixed. As usual, there are three cases.

$\gamma = 0$: $\alpha + (\beta + 0) = \alpha + \beta = (\alpha + \beta) + 0$.

$\gamma = \delta^+$ is a successor: Then by induction, we get

$$\alpha + (\beta + \gamma) = \alpha + (\beta + \delta)^+ = \big[\alpha + (\beta + \delta)\big]^+ = \big[(\alpha + \beta) + \delta\big]^+ = (\alpha + \beta) + \gamma \ .$$

γ is a non-zero limit: Then using Lemma 15 and induction, we get

$$\begin{aligned} \alpha + (\beta + \gamma) &= \alpha + \sup\{\beta + \delta : \delta < \gamma\} \\ &= \sup\{\alpha + (\beta + \delta) : \delta < \gamma\} \\ &= \sup\{(\alpha + \beta) + \delta) : \delta < \gamma\} = (\alpha + \beta) + \gamma \ . \end{aligned}$$

□

Remark. The definition of ordinal addition we gave above is called *inductive definition.* We now give an alternative.

Synthetic definition of ordinal addition. For well-ordered sets X, Y, we write $X \sqcup Y$ for $X \times \{0\} \cup Y \times \{1\}$ (the disjoint union of X and Y) well-ordered by the relation

$$(x, i) < (y, j) \iff \begin{cases} \text{either} & i = j = 0 \text{ and } x < y \text{ in } X \\ \text{or} & i = j = 1 \text{ and } x < y \text{ in } Y \\ \text{or} & i = 0, j = 1 \ . \end{cases}$$

Informally, X comes before Y. For ordinals α, β, we define $\alpha + \beta$ to be the order-type of $\alpha \sqcup \beta$ or, more precisely, the order-type of $X \sqcup Y$, where X is a well-ordered set of order-type α and Y is a well-ordered set of order-type β.

Remark. In a moment we show that the synthetic and inductive definitions coincide. For now, observe that $\alpha+1 = \alpha^+$ for any ordinal α holds by definition. Also, associativity is easy in the synthetic definition, since $X \sqcup (Y \sqcup Z)$ is order-isomorphic to $(X \sqcup Y) \sqcup Z$ for any well-ordered sets X, Y, Z. The inequality $\alpha + \beta \leqslant \alpha + \gamma$ for $\beta \leqslant \gamma$ also follows easily since if Y is an initial segment of Z, then $X \sqcup Y$ is an initial segment of $X \sqcup Z$.

Proposition 17. The synthetic and inductive definitions of ordinal addition coincide.

Proof. Let us temporarily write $\alpha \dotplus \beta$ for the synthetic definition while keeping $\alpha + \beta$ for the inductive definition of addition. We prove $\alpha + \beta = \alpha \dotplus \beta$ by induction on β with α fixed.

$\beta = 0$: $\alpha + 0 = \alpha = \text{order-type}(\alpha \sqcup 0) = \alpha \dotplus 0$.

$\beta = \delta^+$ is a successor: by induction, we have $\alpha + \beta = (\alpha + \delta)^+ = (\alpha \dotplus \delta)^+$. The latter is the order-type of $(\alpha \sqcup \delta) \sqcup 1$ which is order-isomorphic to $\alpha \sqcup (\delta \sqcup 1)$. Thus, $\alpha + \beta = \alpha \dotplus \delta^+ = \alpha \dotplus \beta$.

β is a non-zero limit: By induction we have

$$\alpha + \beta = \sup\{\alpha + \delta : \delta < \beta\} = \sup\{\alpha \dotplus \delta : \delta < \beta\} \ .$$

For $\delta < \beta$, $\alpha \dotplus \delta$ is the order-type of $\alpha \sqcup \delta$, and the supremum of the nested set $\{\alpha \sqcup \delta : \delta < \beta\}$ of well-ordered sets is $\bigcup_{\delta<\beta}(\alpha \sqcup \delta) = \alpha \sqcup \beta$ which has order-type $\alpha \dotplus \beta$. This completes the proof that $\alpha + \beta = \alpha \dotplus \beta$ in this case. □

Ordinal multiplication. As for addition, we give an inductive and a synthetic definition. We start with the former: we define $\alpha \cdot \beta$ by recursion on β:

$$\begin{aligned} &\alpha \cdot 0 = 0 \\ &\alpha \cdot \beta^+ = \alpha \cdot \beta + \alpha \\ &\alpha \cdot \lambda = \sup\{\alpha \cdot \delta : \delta < \lambda\} \qquad \text{for a non-zero limit ordinal } \lambda \ . \end{aligned}$$

For the synthetic definition, we first define the product $X \times Y$ of well-ordered sets to be their Cartesian product well-ordered as follows:

$$(x, y) < (u, v) \iff \begin{cases} \text{either} & y < v \text{ in } Y \\ \text{or} & y = v \text{ and } x < u \text{ in } X \ . \end{cases}$$

We then define $\alpha \cdot \beta$ to be the order-type of $\alpha \times \beta$ or, more precisely, the order-type of $X \times Y$, where X is a well-ordered set of order-type α and Y is a well-ordered set of order-type β. It is then straightforward to verify (by induction on β) that the two definitions coincide.

Examples. For $m, n < \omega$, $m \cdot 0 = 0$ and $m \cdot (n + 1) = m \cdot n^+ = m \cdot n + m$, which is the usual inductive definition of integer multiplication.

$\omega \cdot 2 = \omega \cdot 1^+ = \omega \cdot 1 + \omega = \omega \cdot 0^+ + \omega = (\omega \cdot 0 + \omega) + \omega = \omega + \omega.$

$2 \cdot \omega = \sup\{2 \cdot n : n < \omega\} = \omega.$ So ordinal multiplication is not commutative.

Properties. The following can be verified either by induction on γ or by using the synthetic definition.

$\alpha(\beta\gamma) = (\alpha\beta)\gamma$

$\beta \leqslant \gamma \Rightarrow \alpha\beta \leqslant \alpha\gamma.$

$\beta < \gamma \Rightarrow \alpha\beta < \alpha\gamma$ provided $\alpha > 0$.

$\beta \leqslant \gamma \Rightarrow \beta\alpha \leqslant \gamma\alpha.$

Note, however, that the last inequality cannot be strengthened. *E.g.,* $1 < 2$ but $1 \cdot \omega = 2 \cdot \omega = \omega$.

Exercise. Which, if any, of the following distributive laws hold: $\alpha(\beta + \gamma) = \alpha\beta + \alpha\gamma$ and $(\alpha + \beta)\gamma = \alpha\gamma + \beta\gamma$?

Ordinal exponentiation. We define α^β by recursion on β:

$$\begin{aligned} \alpha^0 &= 1 \\ \alpha^{\beta^+} &= \alpha^\beta \cdot \alpha \\ \alpha^\lambda &= \sup\{\alpha^\delta : \delta < \lambda\} \qquad \text{for a non-zero limit ordinal } \lambda\ . \end{aligned}$$

A synthetic definition also exists.

Examples. For $m, n < \omega$ we recover the usual meaning of m^n.

$\omega^2 = \omega^{1^+} = \omega^1 \cdot \omega = \omega^{0^+} \cdot \omega = (1 \cdot \omega) \cdot \omega = \omega \cdot \omega$

$2^\omega = \sup\{2^n : n < \omega\} = \omega$ is countable!

An application to Functional Analysis (non-examinable)

Remark. This section is for those who attended the Part II Linear Analysis course in the Michaelmas Term.

Fact. Every separable Banach space embeds isometrically into the separable Banach space $C[0,1]$ of continuous functions on $[0,1]$ with the uniform norm. Thus, the class $\mathcal{SB}$ of all separable Banach spaces has a *universal* element: there is a member Z of $\mathcal{SB}$ that contains isomorphic (or, in this case, even isometric) copies of every other member of the class.

Question. Does the class $\mathcal{SR}$ of separable reflexive spaces have a universal member?

Solution by W. Szlenk. Szlenk introduced an ordinal index, known today as the Szlenk index, which associates and ordinal $\mathrm{Sz}(X)$ to every Banach space X. This has the following key properties.

(i) $\mathrm{Sz}(X) \leqslant \omega_1$ and furthermore, $\mathrm{Sz}(X) < \omega_1$ if and only if the dual space X^* of X is separable;

(ii) if the Banach space X isomorphically embeds into the Banach space Y, then $\mathrm{Sz}(X) \leqslant \mathrm{Sz}(Y)$;

(iii) for every countable ordinal α, there exists a separable, reflexive Banach space X_α such that $\mathrm{Sz}(X_\alpha) > \alpha$.

It follows immediately that the answer to the question posed above is 'no'. Indeed, if Z is a universal member of the class $\mathcal{SR}$, then each X_α embeds isomorphically into Z. Then by property (ii) above, $\mathrm{Sz}(Z) > \alpha$ for all countable ordinal α. Thus, $\mathrm{Sz}(Z) = \omega$, which implies that Z^* is not separable by property (i). Since Z is reflexive, $Z^{**} = Z$ is separable, and hence Z^* is separable (as the dual of a space is always at least as 'big' as the space itself) — contradiction.

END OF NON-EXAMINABLE SECTION

3 Posets and Zorn's Lemma

Definition. A *partial order* on a set X is a relation $\leqslant$ on X that

reflexive: $x \leqslant x$ for all $x \in X$;

antisymmetric: if $x \leqslant y$ and $y \leqslant x$, then $x = y$ for all $x, y \in X$;

transitive: if $x \leqslant y$ and $y \leqslant z$, then $x \leqslant z$ for all $x, y, z \in X$.

We then define $x < y$ to mean $x \leqslant y$ and $x \neq y$ for $x, y \in X$. Then $<$ is a relation on X that is irreflexive and transitive.

Definition. A *partially ordered set* or *poset* is a set with a partial order on it.

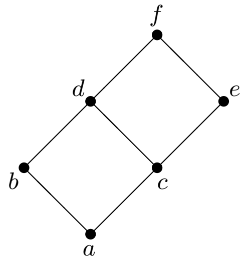
Examples. 1. Any linearly ordered set is a poset.

2. On $\mathbb{N}$ letting $a \leqslant b$ iff $a|b$ is a partial order.

3. For any set S, the power set $X = \mathbb{P}S$ is a poset with $a \leqslant b$ iff $a \subset b$.

4. Any subset of a poset is a poset by restricting the partial order to the subset. *E.g.,* if G is a group, then the subset $\{H \subset G : H \text{ a subgroup of } G\}$ of $\mathbb{P}G$ is a poset.

5. This is an example of a poset given by a *Hasse diagram*:



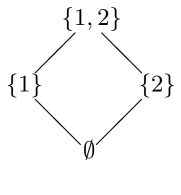
$X = \{a, b, c, d, e, f\}$

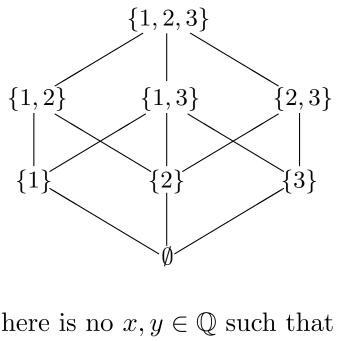
$a \leqslant b,\ a \leqslant c,\ b \leqslant d,\ c \leqslant d,\ c \leqslant e,\ d \leqslant f,\ e \leqslant f$

and all consequences by transitivity

In general, a *Hasse diagram* of a poset X is a drawing of the points of X with x joined to y with an upward line if y *covers* x meaning: $x < y$ and there is no z with $x < z < y$.

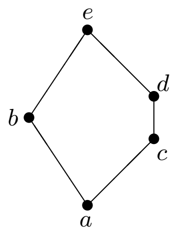
For example, here are drawings of $\mathbb{P}\{1, 2\}$ and $\mathbb{P}\{1, 2, 3\}$.





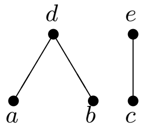
$\mathbb{Q}$ has no Hasse diagram as there is no $x, y \in \mathbb{Q}$ such that y covers x.

6. The following Hasse diagram shows that unlike in the previous examples, in general there is no sensible notion of 'height' or 'rank' in a Hasse diagram.



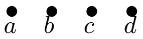
If the ‘height’ of a, c, d, e is $0, 1, 2, 3$, respectively, then what should be the ‘height’ of b?

7.



There needs to be no relation between different parts.

8.



There needs to be no relation at all other than what is necessary by reflexivity.

Definition. A subset S of a poset X is a *chain* if it is linearly ordered by the partial order of X.

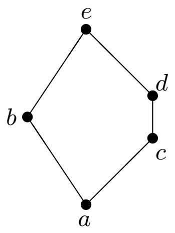
Examples. 1. Every subset of a linearly ordered set is a chain.

2. Every subset of a chain in a poset is a chain.

3. In $\mathbb{N}$ with $a \leqslant b$ iff $a|b$, the set $\{1, 2, 4, 8, \dots\}$ of powers of 2 is a chain.

4. The set $\{\emptyset, \{1\}, \{1, 2\}, \{1, 2, 3\}\}$ is a chain in $\mathbb{P}\{1, 2, 3\}$.

5. In the poset



$\{a, c, d, e\}$ is a chain.

6. The set $\{(-\infty, x) \cap \mathbb{Q} : x \in \mathbb{R}\}$ is an uncountable chain in $\mathbb{P}\mathbb{Q}$.

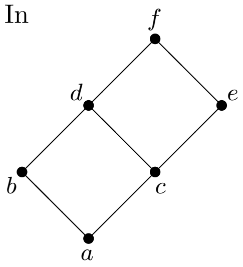
Definition. A subset S of a poset X is an *antichain* if no two members of S are related: for all $x, y \in S$, if $x \leqslant y$, then $x = y$.

Examples. 1. In a linearly ordered set, there is no antichain of size > 1.

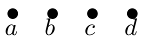
2. In $\mathbb{N}$ with $a \leqslant b$ iff $a|b$, the set of primes is antichain.

3. in $\mathbb{P}\{1, 2, \dots, n\}$, the family $\mathcal{F}_k = \{A \subset \{1, \dots, n\} : |A| = k\}$ is an antichain for each $k = 0, 1, \dots, n$.

4. In



$\{b, c\}$ and $\{d, e\}$ are antichains.

5. In 

the whole set $\{a, b, c, d\}$ is an antichain.

Definition. Let X be a poset, $S \subset X$ and $x \in X$.

x is an *upper bound* for S if $y \leqslant x$ for all $y \in S$.

x is a *least upper bound* or *supremum* for S if x is an upper bound for S and $x \leqslant y$ for every upper bound y for S.

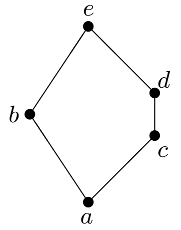
If a supremum for S exists, it is unique and is denoted $\sup S$ or $\bigvee S$.

Examples. 1. If $S \subset \mathbb{P}X$, then $\sup S = \bigcup\{A : A \in S\}$.

2. In $\mathbb{R}$, $\sup(0, 1) = 1$ and $\sup[0, 1] = 1$.

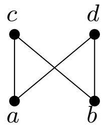
3. In $\mathbb{Q}$, the set $\{x : x^2 < 2\}$ has an upper bound, *e.g.,* 2, but it has no supremum.

4. In the poset



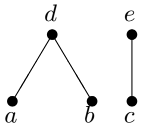
$\sup\{a, b, c\} = e$.

5. In the poset



$\{a, b\}$ has upper bounds c, d but
$\{a, b\}$ has no supremum.

6. In the poset



$\{b, c\}$ has no upper bounds.

Definition. A poset X is *complete* if every subset of X has a supremum.

Examples. 1. The real interval $[0, 1]$ in the usual order is complete.

2. $\mathbb{R}$ is not complete; *e.g.,* $\mathbb{R}$ has no upper bound.

3. $[0, 2] \cap \mathbb{Q}$ is not complete; *e.g.,* $\{x : x^2 < 2\}$ has no supremum.

4. For any set S, the power set $X = \mathbb{P}S$ is complete.

Note. If X is complete, then X has greatest element $\sup X$ and least element $\sup \emptyset$. In particular, X is non-empty.

Definition. A function $f : X \to Y$ between posets X and Y is *order-preserving* if $f(x) \leqslant f(y)$ whenever $x \leqslant y$.

Examples. 1. $f : \mathbb{N} \to \mathbb{N}$, $f(n) = n + 1$.

2. $f : \mathbb{P}(S) \to \mathbb{P}(S)$, $f(A) = A \cup B$, where $B \subset S$ is fixed.

Note. If $f\colon X \to Y$ is order-preserving and injective, then $x < y$ implies $f(x) < f(y)$. The converse holds if X is linearly ordered.

Theorem 1. (Knaster–Tarski fixed point theorem) Let X be a complete poset and $f\colon X \to X$ order-preserving. Then f has a fixed point.

Proof. Set $S = \{x \in X : x \leqslant f(x)\}$ and let $z = \sup S$. For $x \in S$, we have $x \leqslant z$, and hence $x \leqslant f(x) \leqslant f(z)$. Thus, $f(z)$ is an upper bound for S, and so $z \leqslant f(z)$. It follows that $f(z) \leqslant f(f(z))$, and so $f(z) \in S$ and $f(z) \leqslant z$. □

Corollary 2. (Schröder–Bernstein theorem) Suppose that A, B are sets such that there are injections $f\colon A \to B$ and $G\colon B \to A$. Then there is a bijection $h\colon A \to B$.

Proof. We seek partitions $A = P \cup Q$ of A and $B = R \cup S$ of B such that $f(P) = R$ and $g(S) = Q$. If we then define $h\colon A \to B$ by

$$h(x) = \begin{cases} f(x) & \text{if } x \in P \\ g^{-1}(x) & \text{if } x \in Q \end{cases}$$

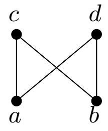
then h is a bijection.

To see that such partitions exists, let $X = \mathbb{P}(A)$ and define $H\colon X \to X$ by $H(P) = A \setminus g(B \setminus f(P))$. Then H is order-preserving and X is a complete poset. Hence by Knaster–Tarski, H has a fixed point P. Setting $Q = A \setminus P$, $R = f(P)$ and $S = B \setminus R$, we get the desired partitions. □

Zorn's Lemma

Definition. Let X be a poset. Say $x \in X$ is a *maximal* element of X if $x \leqslant y$ implies $x = y$ for every $y \in X$, *i.e.*, there is no y in X such that $x < y$.

Examples. 1. For a set S, the poset $X = \mathbb{P}(S)$ has maximal element X. In fact, X is a maximum element.

2. In general, a greatest element is maximal, but the converse is false in general. *E.g.,* in the poset 

c and d are maximal, but there is no maximum.

3. In $\mathbb{N}$ with $a \leqslant b$ iff $a|b$, there are no maximal elements.

Theorem 3. (Zorn's Lemma) Let X be a (non-empty) poset in which every chain has an upper bound. Then X has a maximal element.

Remark. The empty set is a chain in X, so it has an upper bound by assumption. Thus $X \neq \emptyset$. However, we sometimes verify the conditions of Zorn's lemma by checking that $X \neq \emptyset$ and that every non-empty chain has an upper bound.

Proof. Assume that X has no maximal elements. Then for every $x \in X$ we can fix an element $x' \in X$ with $x < x'$. Let us also fix, for every chain C, an upper bound $u(C)$ for C in X. Let $\gamma = \gamma(X)$ (from Hartogs' lemma) and define $f\colon \gamma \to X$ by recursion as follows:

$$\begin{aligned} f(0) &= u(\emptyset) \\ f(\alpha+1) &= f(\alpha)' \\ f(\lambda) &= u(\{f(\alpha) : \alpha < \lambda\}) \qquad \lambda \neq 0 \text{ limit.} \end{aligned}$$

An easy induction (on β with α fixed) shows that $f(\alpha) < f(\beta)$ for all $\alpha < \beta$. It follows that f is injective contradicting the choice of γ. □

Remark. Technically, the definition of $f(\lambda)$ above is only valid when $\{f(\alpha) : \alpha < \lambda\}$ is a chain, otherwise we should define $f(\lambda)$ differently, *e.g.,* we could set $f(\lambda) = u(\emptyset)$ in that case. Then an easy induction shows that the 'otherwise' clause never arises.

Theorem 4. Every vector space V has a basis.

Proof. We seek a maximal (with respect to inclusion) linearly independent set B. Any such set is a basis, *i.e.*, span V otherwise for any $x \in V \setminus \text{span}B$ the set $B \cup \{x\}$ is linearly independent contradicting the maximality of B.

Let $X = \{A \subset V : A \text{ is linearly independent}\}$ partially ordered by inclusion. Let $C = \{A_i : i \in I\}$ be a chain in X. We show that $A = \bigcup_{i \in I} A_i$ is linearly independent. Then A is an upper bound for C. So assume that $\sum_{j=1}^n \lambda_j x_j = 0$ is a linear relation on A. For each $j = 1, \dots, n$ we have $x_j \in A_{i_j}$ for some $i_j \in I$. Since C is a chain, there is a k, $1 \leqslant k \leqslant n$, such that $A_{i_j} \subset A_{i_k}$ for each $j = 1, \dots, n$. Since A_{i_k} is linearly independent, it follows that $\lambda_j = 0$ for all j, and hence A is linearly independent.

We showed that every chain in X has an upper bound. By Zorn's Lemma, X has a maximal element. □

Remarks. 1. A similar argument shows that every linearly independent subset B_0 of V is contained in a basis B of V.

2. $\mathbb{R}$ is a vector space over the field $\mathbb{Q}$. A basis of $\mathbb{R}$ over $\mathbb{Q}$ is called a *Hamel basis.*

3. The real vector space $\mathbb{R}^{\mathbb{N}}$ of all real sequences has no countable basis. By Theorem 4, a basis of $\mathbb{R}^{\mathbb{N}}$ does exist.

4. There are many examples for the use of Zorn's Lemma in different areas of mathematics. *E.g.,* the existence of maximal ideals in rings with 1, the existence of continuous linear functionals in normed spaces, the compactness of a product of compact topological spaces (Tychonov's theorem).

We next use Zorn's Lemma to complete the proof of the Model Existence Lemma from Chapter 1 without the assumption that the set of primitive propositions is countable.

Theorem 5. Let P be any set of primitive propositions. Let $S \subset L = L(P)$ be consistent. Then there exists $\overline{S} \subset L$ such that $S \subset \overline{S}$ and for all $t \in L$, either $t \in \overline{S}$ or $\neg t \in \overline{S}$.

Proof. We seek a maximal consistent set $\overline{S} \supset S$. We then complete the proof as follows. Given $t \in L$, one of $\overline{S} \cup \{t\}$ and $\overline{S} \cup \{\neg t\}$ is consistent since otherwise $\overline{S} \cup \{t\} \vdash \perp$ and $\overline{S} \cup \{\neg t\} \vdash \perp$, which implies $\overline{S} \vdash \neg t$ and $\overline{S} \vdash \neg\neg t$ by the Deduction Theorem, and hence $\overline{S} \vdash \perp$ by modus ponens. It follows by maximality of $\overline{S}$ that either $t \in \overline{S}$ or $\neg t \in \overline{S}$.

Let $X = \{T \subset L : S \subset T,\ T \text{ is consistent}\}$ partially ordered by inclusion. Note that $X \neq \emptyset$ since $S \in X$. Next, let $C = \{T_i : i \in I\}$ be a non-empty chain in X. We show that $T = \bigcup_{i \in I} T_i$ is an upper bound of C. We have $S \subset T$ (as $C \neq \emptyset$) and $T_i \subset T$ for all $i \in I$. So we just need to show that $T \in X$, *i.e.*, T is consistent. If $T \vdash \perp$, then since proofs are finite, there exist $i_1, \dots, i_n$ in I such that $\bigcup_{j=1}^n T_{i_j} \vdash \perp$. Since C is a chain, for some $k = 1, \dots, n$ we have $\bigcup_{j=1}^n T_{i_j} = T_{i_k} \vdash \perp$ — contradiction.

By Zorn's Lemma, X has a maximal element as required. □

Theorem 6. (Well-ordering principle (WO)) Every set A can be well-ordered.

Proof. Let $X = \{(B, R) \in \mathbb{P}A \times \mathbb{P}(A \times A) : R \text{ well-orders } B\}$. We partially order X by $(B_1, R_1) \leqslant (B_2, R_2)$ if and only if B_2 extends B_1 (*i.e.*, $B_1 \subset B_2$, $R_1 = R_2 \cap (B_1 \times B_1)$ and B_1 is an initial segment of B_2).

Note that $X \neq \emptyset$ as $(\emptyset, \emptyset) \in X$. Assume that $\{(B_i, R_i) : i \in I\}$ is a non-empty chain in X, *i.e.*, a nested set of well-ordered sets. Then $\left(\bigcup_{i \in I} B_i, \bigcup_{i \in I} R_i\right)$ is an upper bound (*cf.* Proposition 8 in Chapter 2).

By Zorn's Lemma, there is a maximal element (B, R) in X. If $B \neq A$, then for any $x \in A \setminus B$, the extension $(B, R)^+ = \left(B \cup \{x\}, R \cup \{(b, x) : b \in B\}\right)$ contradicts the maximality of (B, R). Thus, $A = B$ and R is a well-ordering of A. □

Example. $\mathbb{R}$ can be well-ordered which is surprising.

Remark. In applications of Zorn's Lemma, like the previous example, the maximal object whose existence it asserts cannot be described explicitely.

The Axiom of Choice (AC)

Note. In the proof of ZL we used two functions: $X \to X$, $x \mapsto x' \in \{y \in X : x < y\}$, and $u \colon \{C \subset X : C \text{ is a chain}\} \to X$ with $u(C) \in \{x \in X : x \text{ is an upper bound for } C\}$. These are examples of choice functions. Another example from IA Numbers and sets: to show that $\bigcup_{n \in \mathbb{N}} A_n$ is countable if each A_n is countable, we chose, for each $n \in \mathbb{N}$, and injection $f_n \colon A_n \to \mathbb{N}$.

Axiom of Choice (AC). This is the assertion that for every set $X = \{A_i : i \in I\}$ of non-empty sets, there is a function $f \colon I \to \bigcup_{i \in I} A_i$ with $f(i) \in A_i$ for all $i \in I$, called a *choice function* for X.

Note. This rule differs in character from other rules for building sets (*e.g.,* union, power set) in that the object whose existence it asserts is not unique. It is therefore it is in general of interest whether a result whose proof uses AC can be proved without AC. We show in a moment that ZL and WO both need AC.

Remark. We don't need AC when I is finite. In this case, the existence of a choice function can be proved by induction on $|I|$.

Theorem 7. AC $\iff$ ZL $\iff$ WO

Proof. We have already proved the implications AC$\Rightarrow$ZL (Theorem 3) and ZL$\Rightarrow$WO (Theorem 6). It remains to show WO$\Rightarrow$AC. Let $X = \{A_i : i \in I\}$ be a set of non-empty sets. Let $Y = \bigcup_{i \in I} A_i$ and fix a well-ordering of Y. Define $f\colon I \to Y$ by letting $f(i)$ be the least element of A_i. Then f is a choice function of X. □

Exercise. Prove the remaining three implications directly.

Remark. We finish this chapter with some additional material.

START OF NON-EXAMINABLE SECTION

Definition. A poset X is *chain-complete* if $X \neq \emptyset$ and every non-empty chain in X has a supremum.

Examples. Every complete poset and every non-empty finite poset is chain-complete. In general, if X is a poset, then

$$Y = \{C \subset X : C \text{ is a chain}\}$$

partially ordered by inclusion is chain-complete.

Definition. A function $f\colon X \to X$ on a poset X is *inflationary* if $x \leqslant f(x)$ for all $x \in X$.

Theorem 8. (Bourbaki–Witt fixed-point theorem) If X is a chain-complete poset and $f\colon X \to X$ is inflationary, then f has a fixed point.

Proof 1 (with AC). By ZL, X has a maximal element x. Then $x \leqslant f(x)$, and hence $x = f(x)$. □

Proof 2 (without AC). Let $\gamma = \gamma(X)$ (from Hartogs' Lemma). Fix $x_0 \in X$ ($X \neq \emptyset$) ande define $g\colon \gamma \to X$ recursively as follows:

$$\begin{aligned} g(0) &= x_0 \\ g(\alpha^+) &= f(g(\alpha)) \\ g(\lambda) &= \sup\{g(\alpha) : \alpha < \lambda\} \qquad \lambda \neq 0 \text{ limit} \end{aligned}$$

An easy induction shows that g is increasing. (Note that in particular, for a non-zero limit λ, the set $\{g(\alpha) : \alpha < \lambda\}$ is a chain, and thus the definition of $g(\lambda)$ makes sense).

If f has no fixed point, then g is strictly increasing, and hence injective which contradicts the choice of γ. □

Theorem 9. AC+Bourbaki–Witt $\implies$ ZL

Remark. For this reason, the Bourbaki–Witt fixed point theorem is sometimes called the 'choice-free' part of ZL.

Proof. Let X be a poset in which every chain has an upper bound.

Case 1: X is chain-complete. Fix a choice function $g\colon \mathbb{P}X \setminus \{\emptyset\} \to X$. Assume that X has no maximal element. We can then define $f\colon X \to X$ by $f(x) = g(\{y \in X : x < y\})$ for every $x \in X$. Then $x < f(x)$ for all $x \in X$ contradicting Bourbaki–Witt.

Case 2: general case. Define

$$\mathcal{C} = \{C \subset X : C \text{ is a chain}\}$$

partially ordered by inclusion. Then $\mathcal{C}$ is chain-complete, and hence it contains a maximal element C by Case 1. Let x be an upper bound for C in X. If $x < y$ for some $y \in X$, then $C \cup \{y\}$ is a chain contradicting the maximality of C. Thus, x is a maximal element of X. □

Remark. As a by-product, we proved the Hausdorff Maximality Principle (which is also equivalent to AC): every poset contains a maximal chain.

END OF NON-EXAMINABLE SECTION

4 First-order Predicate Logic

Introduction. In Propositional Logic we had a set P of primitive propositions and we combined them using $\perp$ and $\Rightarrow$ (and later $\wedge, \vee, \neg, \top, \Leftrightarrow$) to build the language $L = L(P)$ of all (compound) propositions. The primitive propositions, however, had no meanings attached to them.

Our aim now is to describe a wide variety of mathematical theories. We replace primitive propositions with statements like

$$m(x, m(y,z)) = m(m(x,y),z) \ , \qquad m(x, i(x)) = e$$

in the language of groups, or $x \leqslant y$ in the language of posets. These statements are built using variables (like $x, y, z, \dots$ above), operation symbols (like the binary symbol m, the unary symbol i and the nullary symbol, *i.e.*, constant, e) and predicate symbols (like the binary predicate $\leqslant$).

We then combine these statements into formulae. For example, in the language of groups we have

$$(\forall\, x)(\forall\, y)(\forall\, z)(m(x, m(y,z)) = m(m(x,y),z))$$

to describe associativity, or

$$(\forall\, x)(m(x,x) = e) \Rightarrow (\forall\, x)(\forall\, y)(m(x,y) = m(y,x))$$

to describe the statement you proved in IA Groups that if every non-identity element has order 2, then the group is abelian. An example from the language of posets is the sentence

$$(\forall\, x)(\forall\, y)(\forall\, z)(((x \leqslant y) \wedge (y \leqslant z)) \Rightarrow (x \leqslant z))$$

to describe transitivity.

In Propositional Logic we had valuations. A valuation v can be thought of as a choice of constant $v(p)$, for every proposition p, from the set $\{0,1\}$. In turn, a choice of constant from $\{0,1\}$ is a function from a singleton set to $\{0,1\}$. In first-order logic we will have a structure which will be a set A together with a choice of function $p_A \colon A^n \to \{0,1\}$ for every formula p, where n is the number of variables in p. For a set S of formulae, we will define the notion of model for S and we will define semantic and syntactic consequences of S in a way very similar to what we did in Chapter 1. It is now time to give the formal definitions.

The Language

The language is specified by two disjoint sets: the set Ω of *operation symbols* and the set Π of *predicate symbols* together with an *arity function* $\alpha \colon \Omega \cup \Pi \to \mathbb{N}_0 = \mathbb{N} \cup \{0\}$. The language $L = L(\Omega, \Pi)$ then consists of the following.

Variables. This is a countable set disjoint from Ω and Π. We usually denote variables as $x_1, x_2, \dots$ or $x, y, z, \dots$.

Terms. Also called Ω-terms, these are defined inductively as follows.

(i) Every variable is a term.

(ii) If $\omega \in \Omega$, $n = \alpha(\omega)$ and $t_1, t_2, \ldots, t_n$ are terms, then $\omega t_1 t_2 \ldots t_n$ is also a term.

For example, the language of groups consists of $\Omega = \{m, i, e\}$ with arities $\alpha(m) = 2$, $\alpha(i) = 1$, $\alpha(e) = 0$, and $\Pi = \emptyset$. The following then are examples of terms in the language of groups.

$$mxmyz \ , \quad mmxyz \ , \quad mxe \ , \quad mxix \ , \quad mee$$

Note that brackets are not needed as their positions are uniquely determined by the order of operation symbols and variables. The following strings of operation symbols and variables are not terms: $mxymz$, emx, $mxyz$.

Atomic formulae. There are of two kinds.

(i) If s, t are terms, then $(s = t)$ is an atomic formula.

(ii) If $\varphi \in \Pi$, $\alpha(\varphi) = n$ and $t_1, t_2, \ldots, t_n$ are terms, then $\varphi t_1 t_2 \ldots t_n$ is an atomic formula.

For example, the language of posets consists of $\Omega = \emptyset$ and $\Pi = \{\leqslant\}$ with arity $\alpha(\leqslant) = 2$. The following are then atomic formulae in the language of posets.

$$(x_1 = x_2) \ , \qquad (x_1 \leqslant x_2)$$

The latter example should really be written as $\leqslant x_1 x_2$. However, here and below we shall often revert to the more conventional way of writing algebraic expressions which then may require the insertion of brackets to avoid ambiguity.

Formulae. These are defined inductively as follows.

(i) Atomic formulae are formulae.

(ii) $\bot$ is a formula.

(iii) If p, q are formulae, then $(p \Rightarrow q)$ is a formula. As before, we introduce $\wedge, \vee, \neg, \top, \Leftrightarrow$ in order to abbreviate certain formulae.

(iv) If p is a formula and x is a free variable in p, then $(\forall\, x)p$ is a formula. We also introduce $(\exists\, x)p$ as an abbreviation for $\neg(\forall\, x)\neg p$.

To explain clause (iv), note that a formula is a string of symbols from $\Omega \cup \Pi$, variables and from the set $\{\bot, \Rightarrow, (,), =\}$ (and other sybmols introduced as abbreviations). The occurence of a variable x in a term is always *free*. The occurence of a variable x in a formula p is always free except in the case p is of the form $(\forall\, x)q$ in which case every occurence of x that was free in q becomes a *bound* occurence in p. We denote by $\mathrm{FV}(p)$ the set of free variables of p, *i.e.*, the set of variables that have at least one free occurence in p.

Examples. The following are formulae in the language of groups. In each case we indicate for every occurence of the variable x whether it is a free or bound occurence.

$$\begin{array}{c}(mxix = e) \Rightarrow (mixx = e)\\ \uparrow\ \uparrow \qquad\qquad\quad \uparrow\uparrow\\ \text{free} \qquad\qquad\quad \text{free}\end{array}$$

$$\begin{array}{c}(\forall\, x)(mxx = e) \Rightarrow (\forall\, x)(\forall\, y)(mxy = myx)\\ \uparrow\uparrow \qquad\qquad\qquad\qquad \uparrow \qquad \uparrow\\ \text{bound} \qquad\qquad\qquad\quad \text{bound} \quad \text{bound}\end{array}$$

$$\begin{array}{c}(mxx = y) \Rightarrow \neg(\exists\, x)(mmxxx = y)\\ \uparrow\uparrow \qquad\qquad\qquad \uparrow\uparrow\uparrow\\ \text{free} \qquad\qquad\qquad \text{bound}\end{array}$$

Note that in the last example, the variable x has both free and bound occurences. Although such formulae are technically allowed, it is usual mathematical practice to avoid them. In the second example, there are no free variables: both x and y only have bound occurences. Such formulae have a special name.

Definition. A formula with no free variables is called a *sentence*.

Structures

Definition. A *structure* in a language $L = L(\Omega, \Pi)$, or L-*structure*, is a nonempty set A together with functions

$$\omega_A \colon A^n \to A \qquad (n \text{ the arity of } \omega)$$

for each $\omega \in \Omega$, and subsets $\varphi_A \subset A^n$ (n the arity of φ) or equivalently, identifying a subset with its indicator function, functions

$$\varphi_A \colon A^n \to \{0, 1\} \qquad (n \text{ the arity of } \varphi)$$

for each $\varphi \in \Pi$.

Note. An operation symbol ω of arity 0 is called a *constant*. Its interpretation in a structure A is a function ω_A from A^0, a singleton set, to A, *i.e.*, ω_A is simply an element of A.

Note. In normal mathematical practice, we allow the empty set to be a structure. Its exclusion here is a simplifying assumption (see later for an explanation).

Examples. 1. If L is the language of groups, then an L-structure is a set A together with functions

$$m_A \colon A \times A \to A\ , \qquad i_A \colon A \to A$$

and an element $e_A \in A$. Note that A is not a group yet!

2. If L is the language of posets, then an L-structure is a set A together with a subset $\leqslant_A$ of $A \times A$, *i.e.*, a relation on A. Again, note that A is not yet a poset.

Motivation. Given an L-structure A and a formula p in the language L, we want to define what it means that 'p is satisfied in A'. For example, if p is the formula $(mxix = e)$ in the language of groups, then we let p_A be the subset $\{a \in A : m_A(a, i_A(a)) = e_A\}$ of A, and then say that p is satisfied in A if $p_A = A$. Equivalently, identifying p_A with its indicator function, we say p is satisfied in A if $p_A \colon A \to \{0,1\}$ is the constant function with value 1. If q now is the sentence $(\forall\, x)p$, then its interpretation in A should be a function $A^0 \to \{0,1\}$, *i.e.*, q_A is simply an element of $\{0,1\}$. We set $q_A = 1$ if $p_A(a) = 1$ for every $a \in A$, *i.e.*, if p holds in A, otherwise we set $q_A = 0$. We now give the formal description of how to interpret formulae in a structure. This is rather dry, and it is best to let these motivating examples be the guide.

Definitions. Let L be a first-order language and A be an L-structure. Let t be a term in L and p be a formula in L both with free variables contained in the set $\{x_1, \dots, x_n\}$. We define interpretations $t_A \colon A^n \to A$ of t in A and $p_A \colon A^n \to \{0,1\}$ (equivalently, $p_A \subset A^n$) of p in A by induction on the language L as follows.

If t is x_i for some $i = 1, \dots, n$, then

$$t_A \colon A^n \to A\ , \qquad t_A(a_1, \dots, a_n) = a_i$$

is projection on the i^{th} coordinate.

If t is $\omega t_1 \dots t_m$ for some $\omega \in \Omega$ with arity m and terms $t_1, \dots, t_m$, then

$$t_A \colon A^n \to A\ , \quad t_A(a_1, \dots, a_n) = \omega_A\big((t_1)_A(a_1, \dots, a_n), \dots, (t_m)_A(a_1, \dots, a_n)\big)\ .$$

If p is $(u = v)$ for terms u, v, then

$$p_A = \{(a_1, \dots, a_n) \in A^n : u_A(a_1, \dots, a_n) = v_A(a_1, \dots, a_n)\}\ .$$

$\perp_A$ is the constant function with value 0.

If p is $(q \Rightarrow r)$, then

$$p_A = \{(a_1, \dots, a_n) \in A^n : q_A(a_1, \dots, a_n) = 0 \text{ or } r_A(a_1, \dots, a_n) = 1\}\ .$$

If p is $(\forall\, x_{n+1})q$ with $FV(q) \subset \{x_1, \dots, x_n, x_{n+1}\}$, then

$$p_A = \{(a_1, \dots, a_n) \in A^n : (a_1, \dots, a_n, a_{n+1}) \in q_A \text{ for all } a_{n+1} \in A\}\ .$$

Theories and Models

Definition. Let $L = L(\Omega, \Pi)$ be a first-order language and A be an L-structure. Given a formula p in the language L, we say p is *satisfied* in A (or p *holds* in A, or p is *true* in A or A is a *model* of p) if $p_A = A^n$ or, equivalently, $p_A \colon A^n \to \{0,1\}$ is the constant function with value 1, where n is the number of free variables in p.

Note. When p is a sentence, then p holds in A if $p_A = 1$.

Definition. A *theory* in L is a set of sentences in L. A *model* for a theory T is an L-structure in which each sentence in T is satisfied.

Examples. 1. Theory of groups The language of groups has been defined above: it consists of $\Omega = \{m, i, e\}$ and $\Pi = \emptyset$ with the arities of m, i, e being $2, 1, 0$, respectively. The theory of groups is the following set of sentences.

$$\begin{aligned} T = \{&(\forall\, x)(\forall\, y)(\forall\, z)(mxmyz = mmxyz)\ , \\ &(\forall\, x)(mxe = x \wedge mex = x)\ , \\ &(\forall\, x)(mxix = e \wedge mixx = e)\} \end{aligned}$$

Then every model of T is a group and every group is a model of T. So groups can be axiomatised as a first-order theory.

2. Theory of posets The language is as defined above: $\Omega = \emptyset$, $\Pi = \{\leqslant\}$ and $\leqslant$ has arity 2. The theory of posets is then the following set of sentences.

$$\begin{aligned} T = \{&(\forall\, x)(x \leqslant x)\ , \\ &(\forall\, x)(\forall\, y)((x \leqslant y \wedge y \leqslant x) \Rightarrow x = y)\ , \\ &(\forall\, x)(\forall\, y)(\forall\, z)((x \leqslant y \wedge y \leqslant z) \Rightarrow x \leqslant z)\} \end{aligned}$$

This axiomatises posets: every (non-empty) poset is a model of T and every model of T is a poset.

3. Theory of rings with 1 The language consists of $\Omega = \{+, \times, -, 0, 1\}$ with arities $2, 2, 1, 0, 0$, respectively, and $\Pi = \emptyset$. The theory is as follows.

$$\begin{aligned} T = \{&(\forall\, x)(\forall\, y)(\forall\, z)((x + (y + z)) = ((x + y) + z))\ , \\ &(\forall\, x)(\forall\, y)(x + y = y + x)\ , \\ &(\forall\, x)(x + 0 = x)\ , \\ &(\forall\, x)(x + (-x) = 0)\ , \\ &(\forall\, x)(\forall\, y)(\forall\, z)(x(yz) = (xy)z)\ , \\ &(\forall\, x)(\forall\, y)(\forall\, z)((x(y + z) = xy + xz) \wedge ((x + y)z = xz + yz))\ , \\ &(\forall\, x)((x1 = x) \wedge (1x = x))\} \end{aligned}$$

The models of T are precisely the rings with 1. Note that here we reverted to writing $x + y$ instead of $+xy$, xy instead of $x \times y$, etc. This and the theory of groups are examples of *algebraic theories*: the sentences only involve equations.

4. Theory of fields The language is the same as for rings with 1. The theory is the theory for rings with 1 together with the following three sentences.

$$\begin{gathered} (\forall\, x)(\forall\, y)(xy = yx) \\ \neg(0 = 1) \\ (\forall\, x)(\neg(x = 0) \Rightarrow (\exists\, y)(xy = 1)) \end{gathered}$$

This axiomatises fields. Note that this is not an algebraic theory, and indeed fields cannot be axiomatised by an algebraic theory. This is because every field has at least two elements, and it is easy to see that the singleton set is a model for every algebraic theory.

5. Graph theory The language consists of $\Omega = \emptyset$, $\Pi = \{a\}$ with a having arity 2 (a is the adjacency predicate). The theory is

$$\begin{aligned}\{(\forall\, x)\neg(axx)\ ,\\ (\forall\, x)(\forall\, y)(axy \Rightarrow ayx)\}\end{aligned}$$

So graphs can also be axiomatised as a first-order theory.

6. Propositional theories This example shows that predicate logic contains propositional logic as a subset. Let P be a set of primitive propositions. Set $\Omega = \emptyset$ and $\Pi = P$ with each primitive proposition having arity 0. In the language $L = L(\Omega, \Pi)$ every primitive proposition is an atomic formula, and thus every proposition in $L(P)$ (as defined in Chapter 1) is a formula. An L-structure is a nonempty set A together with a function $v\colon P \to \{0,1\}$ that maps $p \in P$ to $p_A \in \{0,1\}$ (as usual we identify a function $A^0 \to \{0,1\}$ with the value it takes in $\{0,1\}$). Given a (compound) proposition $t \in L(P)$, its interpretation t_A in A is an element of $\{0,1\}$ (*i.e.*, a function $A^0 \to \{0,1\}$), and the map $t \mapsto t_A$ is precisely the extension $v\colon L(P) \to \{0,1\}$ of v to $L(P)$ as defined in Chapter 1. Given $S \subset L(P)$, a model of S is a structure A with a function $v\colon P \to \{0,1\}$ such that for every $s \in S$ we have $s_A = v(s) = 1$. Thus, the meaning of model coincides with the definition of model in Chapter 1. Note that the underlying set A here is irrelevant.

Semantic entailment

Definition. Given a first-order language $L = L(\Omega, \Pi)$, a set S of sentences in L and a sentence t in L, we say S (*semantically*) *entails* t, written $S \models t$, if t holds in every model of S.

Examples. 1. Let T be the theory of groups (in the language of groups). Then

$$T \models \big((\forall\, x)(xx = e) \Rightarrow (\forall\, x)(\forall\, y)(xy = yx)\big)$$

2. Let T be the theory of fields (in the language of rings with 1). Then

$$T \models (\forall\, x)\big(\neg(x = 0) \Rightarrow (\forall\, y)(\forall\, z)((xy = 1 \wedge xz = 1) \Rightarrow (y = z))\big)$$

We will also need the define $S \models t$ in the case when $S \cup \{t\}$ contains formulae with free variables. The following example motivates the definition.

Example. Let T be the theory of fields (in the language of rings with 1). Let p be the formula $\neg(x = 0)$, let t be the formula $(\exists\, y)(xy = 1)$ and let $S = T \cup \{p\}$. It ought to be the case that $S \models t$ because, given a field F, if we assign a value $a \in F$ to the variable x, then according to the field axioms, if $p_A(a)$ is true (*i.e.*, $p_A(a) = 1$), then $t_F(a, b)$ is true for some $b \in F$.

Definition. Let $L = L(\Omega, \Pi)$ be a first-order language, S a set of formulae in L and t a formula in L. Introduce a new constant to L for each free variable occuring in $S \cup \{t\}$. For a formula $u \in S \cup \{t\}$, let u' be the sentence obtained from u by replacing each free occurence of a variable with the corresponding constant and set $S' = \{s' : s \in S\}$. We then say S (*semantically*) *entails* t, written $S \models t$, if $S' \models t'$.

Definition. A formula t in a first-order language L is a *tautology* if $\emptyset \models t$, *i.e.*, t holds in every L-structure.

Note. In the definition of semantic entailment we substituted constants into formulae. Later we will need a more general notion of substitution.

Definition. Let p be a formula in a first-order language L. If x is a free variable in p and t is a term in L whose variables do not occur bound in p, then $p[t/x]$ is the formula obtained from p by replacing each free occurence of x with t.

Examples. Let p be the formula $(\forall\, y)(mmxyy = mmyxy)$ in the language of groups. If t is the term mzz, then

$$p[t/x] \quad \text{is} \quad (\forall\, y)(mmmzzyy = mmymzzy)$$

If t is the term mxx, then

$$p[t/x] \quad \text{is} \quad (\forall\, y)(mmmxxyy = mmymxxy)$$

Finally, if t is the term myy, then $p[t/x]$ is not defined as the variable y in t occurs bound in p.

Syntactic entailment

Axioms.

(A1) $p \Rightarrow (q \Rightarrow p)$ (p, q any formulae)

(A2) $(p \Rightarrow (q \Rightarrow r)) \Rightarrow ((p \Rightarrow q) \Rightarrow (p \Rightarrow r))$ (p, q, r any formulae)

(A3) $\neg\neg p \Rightarrow p$ (p any formula)

(A4) $(\forall\, x)(x = x)$

(A5) $(\forall\, x)(\forall\, y)((x = y) \Rightarrow (p \Rightarrow p[y/x]))$

(x, y distinct variables, p a formula, y does not occur bound in p)

(A6) $((\forall\, x)p) \Rightarrow p[t/x]$

(p a formula, x a free variable of p, t a term no variable of which occurs bound in p)

(A7) $(\forall\, x)(p \Rightarrow q) \Rightarrow (p \Rightarrow (\forall\, x)q)$

(p, q formulae, x a free variable of q that does not occur free in p)

Note. Every axiom is a tautology.

Rules of deduction.

Modus ponens (MP): from p and $p \Rightarrow q$, we can deduce q.

Generalisation (Gen): from p, we can deduce $(\forall\, x)p$ provided x does not occur free in any premiss used in the proof of p.

Definition. Let S be a set of formulae and p be a formula. A *proof of p from S* is a finite sequence $t_1, \ldots, t_n$ of formulae such that $t_n = p$ and for every i,

(i) either t_i is an axiom, or

(ii) t_i is a premiss (member of S), or

(iii) t_i follows by modus ponens ($\exists\ j, k < i$ such that t_k is $(t_j \Rightarrow t_i)$), or

(iv) t_i follows by Generalisation ($\exists\, j < i$ such that $t_i = (\forall\, x)t_j$ where x is a free variable in t_j that does not occur free in any premiss t_k, $k < j$).

We say S *proves* p, and write $S \vdash p$, if there is a proof of p from S. If S is a theory, p is a sentence and $S \vdash p$, then we say p is a *theorem* of S.

Remark. We can now explain why a structure in a language L had to be non-empty. Suppose we allowed the empty set as a structure. Then $(\forall\, x)\neg(x = x)$ is satisfied in $\emptyset$, whereas $\bot$ is not. Thus, $\{(\forall\, x)\neg(x = x)\} \not\models \bot$. On the other hand, $\{(\forall\, x)\neg(x = x)\} \vdash \bot$. Indeed, we have

$$\begin{array}{cl}
(\forall\, x)\neg(x = x) & \text{(premiss)} \\
(\forall\, x)\neg(x = x) \Rightarrow \neg(x = x) & \text{(A6)} \\
\neg(x = x) & \text{(MP)} \\
(\forall\, x)(x = x) & \text{(A4)} \\
(\forall\, x)(x = x) \Rightarrow (x = x) & \text{(A6)} \\
(x = x) & \text{(MP)} \\
\bot & \text{(MP)}
\end{array}$$

Example. $\{(x = y)\} \vdash (y = x)$

$$\begin{array}{cl}
(\forall\, x)(\forall\, y)((x = y) \Rightarrow ((x = z) \Rightarrow (y = z))) & \text{(A5)} \\
(x = y) \Rightarrow ((x = z) \Rightarrow (y = z)) & \text{(A6+MP twice)} \\
(x = y) & \text{(premiss)} \\
(x = z) \Rightarrow (y = z) & \text{(MP)} \\
(\forall\, z)((x = z) \Rightarrow (y = z)) & \text{(Gen)} \\
(\forall\, z)((x = z) \Rightarrow (y = z)) \Rightarrow ((x = x) \Rightarrow (y = x)) & \text{(A6)} \\
(x = x) \Rightarrow (y = x) & \text{(MP)} \\
(\forall\, x)(x = x) & \text{(A4)} \\
(\forall\, x)(x = x) \Rightarrow (x = x) & \text{(A6)} \\
(x = x) & \text{(MP)} \\
(y = x) & \text{(MP)}
\end{array}$$

Proposition 1. (Deduction Theorem) Let S be a set of formulae and p, q be formulae. Then $S \vdash (p \Rightarrow q)$ if and only if $S \cup \{p\} \vdash q$.

Proof. Assume that $S \vdash (p \Rightarrow q)$. Write down a proof of $(p \Rightarrow q)$ from S and append the lines

$$\begin{array}{ll} p & \text{(premiss)} \\ q & \text{(MP)} \end{array}$$

to obtain a proof of q from $S \cup \{p\}$.

Now assume that $S \cup \{p\} \vdash q$ and let $t_1, \ldots, t_n$ be a proof of q from $S \cup \{p\}$. We show that $S \vdash (p \Rightarrow t_i)$ for all i by induction.

Induction hypothesis before the i^{th} step: for each $j < i$, we have $S \vdash (p \Rightarrow t_j)$ such that if a variable x did not occur free in any premiss used in the proof of t_j from $S \cup \{p\}$, then x does not occur free in any premiss used in the proof of $(p \Rightarrow t_j)$ from S.

We now show $S \vdash (p \Rightarrow t_i)$ by considering a number of cases.

Case 1: If t_i is an axiom or $t_i \in S$, then

$$\begin{array}{ll} t_i \Rightarrow (p \Rightarrow t_i) & \text{(A1)} \\ t_i & \text{(axiom or premiss)} \\ p \Rightarrow t_i & \text{(MP)} \end{array}$$

is a proof of $p \Rightarrow t_i$ from S.

Case 2: If $t_i = p$, then $S \vdash (p \Rightarrow t_i)$ since $\vdash (p \Rightarrow p)$.

Case 3: If there exist $j, k < i$ such that $t_k = (t_j \Rightarrow t_i)$, then by induction hypothesis, there are proofs of $p \Rightarrow t_j$ and $p \Rightarrow (t_j \Rightarrow t_i)$ from S. Adding the lines

$$\begin{array}{ll} \big(p \Rightarrow (t_j \Rightarrow t_i)\big) \Rightarrow \big((p \Rightarrow t_j) \Rightarrow (p \Rightarrow t_i)\big) & \text{(A2)} \\ (p \Rightarrow t_j) \Rightarrow (p \Rightarrow t_i) & \text{(MP)} \\ p \Rightarrow t_i & \text{(MP)} \end{array}$$

we obtain a proof of $p \Rightarrow t_i$ from S.

Case 4: Finally, if for some $j < i$ we have $t_i = (\forall\, x)t_j$, then x occurs free in t_j and x does not occur free in any premiss used in the proof of t_j from $S \cup \{p\}$. We now have two further cases.

Case 4(a): If x occurs free in p, then p was not used in the proof of t_j from $S \cup \{p\}$, and so we have a proof of t_j from S and x does not occur free in any premiss used in this proof. We write down this proof and append the lines

$$\begin{array}{ll} (\forall\, x)t_j & \text{(Gen)} \\ (\forall\, x)t_j \Rightarrow (p \Rightarrow (\forall\, x)t_j) & \text{(A1)} \\ p \Rightarrow (\forall\, x)t_j & \text{(MP)} \end{array}$$

to obtain a proof of $p \Rightarrow t_i$ from S.

Case 4(b): If x does not occur free in p, then we have the following proof from S.

$$\begin{array}{ll} \quad p \Rightarrow t_j & \text{(induction hypothesis)} \\ (\forall\, x)(p \Rightarrow t_j) & \text{(Gen)} \\ (\forall\, x)(p \Rightarrow t_j) \Rightarrow (p \Rightarrow (\forall\, x)t_j) & \text{(Gen)} \\ \quad p \Rightarrow (\forall\, x)t_j & \text{(MP)} \end{array}$$

In all cases it is easy to check that the induction hypothesis is valid up to and including the i^{th} step. □

Aim. We now embark on the proof of the Completeness Theorem that states that, for first-order logic, $\vdash$ and $\models$ coincide.

Proposition 2. (Soundness Theorem) Let S be a set of formulae (in a first-order language $L = L(\Omega, \Pi)$) and p be a formula. If $S \vdash p$, then $S \models p$.

*Proof (**non-examinable**).* Let $t_1, \ldots, t_n$ be a proof of p from S. A straightforward induction shows that $S \models t_i$ for all i. Note that in the case $t_i = (\forall\, x)t_j$ for some $j < i$, there exists $S_1 \subset S$ such that $t_1, \ldots, t_j$ is a proof of t_j from S_1, and x does not occur free in S_1. By induction hypothesis, $S_1 \models t_j$ and since x does not occur free in S_1, it follows that $S_1 \models t_i$, which in turn implies $S \models t_i$. □

Theorem 3. (Model existence lemma) If S is a consistent theory (in a first-order language $L = L(\Omega, \Pi)$), then S has a model.

Idea of proof. We will build a model from the language L itself. We initially choose our structure to be the set A of all *closed terms* of L, *i.e.*, terms not involving variables. Examples of closed terms in the language of commutative rings with 1:

$$1+1\ , \quad (1+0)+1\ , \quad 1\cdot 1\ , \quad (1+0)\cdot 0+1\ , \quad \text{etc.}$$

We turn A into a structure by interpreting the operation symbols in the obvious way. In the example above, we would have $(1+0) +_A (1+1) = (1+0)+(1+1)$ or $0_A = 0$, $1_A = 1$.

However, if S is the theory of fields, for example, then A is not a model of S. *E.g.,* we have $S \vdash (1+0=1)$ but $1+0=1$ is not satisfied in A since the closed terms $1+0$ and 1 are different terms. There is an easy remedy. Introduce the equivalence relation $s \sim t$ if and only if $S \vdash (s=t)$ and replace A by its quotient $A/\sim$. Operation symbols are now interpreted on representatives of equivalence classes. *E.g.,* $[1] +_A [1] = [1+1]$ in our example. Two issues remain.

For an example of the first issue, consider the theory S of fields with characteristic 2 or 3, which consists of the theory of fields together with the sentence $(1+1=0)\vee(1+1+1=0)$. Then $S \not\vdash (1+1=0)$ since S has models that are fields of characteristic 3. Similarly, $S \not\vdash (1+1+1=0)$. It follows that in our structure A, we have $[1] +_A [1] = [1+1] \neq [0]$ and $[1] +_A [1] +_A [1] = [1+1+1] \neq [0]$, and thus A is not a model. As for propositional logic, we will first extend the theory S to a consistent theory that is complete. In general, we say that a theory S in a first-order language L is *complete* if for every sentence p, either $S \vdash p$ or $S \vdash \neg p$.

For an example of the second issue, consider the theory S of fields in which 2 has a square root. This consists of the theory of fields together with the sentence $(\exists\, x)(xx = 1+1)$. Then the structure A defined above consisting of $\sim$-equivalence classes of closed terms is not a model, since there is no closed term t such that $[tt] = [1+1]$. In other words, we lack a *witness* to the sentence $(\exists\, x)(xx = 1+1)$, *i.e.*, we lack a closed term t such that $S \vdash p[t/x]$ where p is the formula $(xx = 1+1)$. The solution is to add a new constant c to our language and the new sentence $(cc = 1+1)$ to our theory.

The problem is that the two processes of adding witnesses and of completion pull in different directions. When we add witnesses to a complete theory, the new theory may no longer be complete. When we complete a theory which has witnesses, the new theory may lack witnesses.

*Proof of Theorem 3 (**non-examinable**).* We begin with the observation that if S is a consistent theory, then for any sentence p, one of $S \cup \{p\}$ and $S \cup \{\neg p\}$ is consistent. Indeed, otherwise, by the Deduction Theorem, we have $S \vdash \neg p$ and $S \vdash \neg\neg p$, and hence an application of modus ponens yields $S \vdash \perp$ — a contradiction. An argument using Zorn's Lemma now shows that S is contained in a consistent theory $\overline{S}$ such that for every sentence p, either $p \in \overline{S}$ or $\neg p \in \overline{S}$. In particular, $\overline{S}$ is complete.

Next assume that S is a consistent theory and $S \vdash (\exists\, x)p$ where p is a formula with one free variable x. We add a new constant c to the language L and show that $S \cup \{p[c/x]\}$ is consistent. Indeed, otherwise, by the Deduction Theorem, we get $S \vdash \neg p[c/x]$. Since c does not occur in S, if we replace every occurence of c with x in the proof of $\neg p[c/x]$ from S, we obtain a proof of $\neg p$ from S. It follows that $S \vdash (\forall\, x)\neg p$ by Generalisation, and since $S \vdash (\exists\, x)p$ by assumption, we deduce $S \vdash \perp$ by modus ponens — a contradiction. Applying this to all theorems of S of the form $(\exists\, x)p$, we obtain a new language $\overline{L} = L(\Omega \cup C, \Pi)$ where C is a set of new constants and a consistent theory $\overline{S} \supset S$ such that for every sentence of the form $(\exists\, x)p$ in the language L that is deducible from S, there is a closed term t in $\overline{L}$ such that $\overline{S} \vdash p[t/x]$.

Let us now start with a consistent theory S in a language $L = L(\Omega, \Pi)$. Put $S_0 = S, L_0 = L$ and inductively define, using the two processes described above, theories $S_0 \subset S_1 \subset T_1 \subset S_2 \subset T_2 \subset \dots$ and languages $L_n = L(\Omega \cup C_1 \cup \dots \cup C_n, \Pi)$, where $C_1, C_2, \dots$ are pairwise disjoint sets of constants each disjoint from Ω (and from Π), for each $n \in \mathbb{N}$, S_n is a consistent, complete theory in L_{n-1} and T_n is a consistent theory in L_n which has witnesses for S_n. We put $L^* = \bigcup_n L_n$ and $S^* = \bigcup_n S_n$. It is straightforward to verify that S^* is a consistent theory in the language L^* that is complete and has witnesses. Since every model of S^* is also a model of S, to complete the proof, we may assume that S is complete and has witnesses.

We let A be the set of equivalence classes of closed terms in L under the equivalence relation $s \sim t$ if and only if $S \vdash (s = t)$. We turn A into an L-structure as follows. For $\omega \in \Omega$ with arity n, we define $\omega_A \colon A^n \to A$ by setting

$$\omega_A([t_1], \dots, [t_n]) = [\omega t_1 \dots t_n] ,$$

and for $\varphi \in \Pi$ with arity n, we let

$$\varphi_A([t_1], \ldots, [t_n]) = 1 \qquad \text{if and only if} \qquad S \vdash (\varphi t_1 \ldots t_n) \ .$$

An easy induction shows that $s_A = [s]$ for any closed term s. Another induction on the language then shows that for any sentence p, we have $S \vdash p$ if and only if $p_A = 1$ (*i.e.*, p holds in A). In particular, A is a model of S. □

Corollary 4. (Adequacy Theorem) Let S be a set of formulae (in a first-order language $L = L(\Omega, \Pi)$) and p be a formula. If $S \models p$, then $S \vdash p$.

*Proof (**non-examinable**).* We first reduce to the case when S is a theory and p is a sentence. Indeed, by definition, $S \models p$ means that $S' \models p'$, where S' and p' are obtained from S and p by replacing free occurences of variables with constants added to the language. Then if we know that there is a proof of p' from S', then we can put the free variables back in to replace the new constants, and thus obtain a proof of p from S.

Now since $S \models p$, the theory $S \cup \{\neg p\}$ has no models, and so $S \cup \{\neg p\} \vdash \perp$ by Theorem 3. By the Deduction Theorem, we have $S \vdash \neg\neg p$, and hence $S \vdash p$ using the axiom $\neg\neg p \Rightarrow p$ and modus ponens. □

Theorem 5. (Gödel's Completeness Theorem for first-order logic) Let S be a set of formulae (in a first-order language $L = L(\Omega, \Pi)$) and p be a formula. Then $S \models p$ if and only if $S \vdash p$. □

Corollary 6. (Compactness Theorem) Let S be a theory in a first-order language L. If every finite subset of S has a model, then S had a model.

Proof. If S has no model, then $S \vdash \perp$ by Theorem 3. As proofs are finite, there is a finite subset S' of S such that $S' \vdash \perp$. By the Soundness Theorem, $S' \models \perp$, *i.e.*, S' has no model — contradiction. □

Applications of completeness/compactness

Question. Can we axiomatise the theory of finite groups? In other words, does there exist a first-order theory T in a suitable language such that every finite group is a model of T and every model of T is a finite group?

Consider, for each $n \in \mathbb{N}$, the following sentence:

$$t_n : \qquad (\exists\, x_1)(\exists\, x_2) \ldots (\exists\, x_n)(\forall\, x)(x = x_1 \vee x = x_2 \vee \ldots \vee x = x_n)$$

which says that there are at most n elements. So we would like to take T to be the theory of groups (in the language of groups) and add the 'sentence' $t_1 \vee t_2 \vee \ldots$. Obviously, this is not possible as sentences are finite strings of symbols.

Corollary 7. The theory of finite groups is not axiomatisable as a first-order theory.

Proof. Assume that T is a first-order theory whose models are the finite groups. Let

$$S = T \cup \{\neg t_1, \neg t_2, \neg t_3, \dots\}$$

where the t_n are the sentences defined above. Note that $\neg t_n$ says that there are more than n elements in any model. Thus, every finite subset of S has a model: *e.g.,* the cyclic group of order N for sufficiently large N. By the Compactness Theorem, S has a model — contradiction. □

A similar argument shows the following.

Corollary 8. Let S be a theory in a first-order language. If S has arbitrarily large finite models, then S has an infinite model.

Proof. Add the sentences $\neg t_1, \neg t_2, \dots$ to S. Observe that every finite subset of the new theory has a model, and thus it has a model by the Compactness Theorem. A model of the new theory is an infinite model of S. □

Corollary 9. (The Upward Löwenheim–Skolem Theorem) If a first-order theory S has an infinite model, then it has an uncountable model.

Proof. Add an uncountable set $\{c_i : i \in I\}$ of new constants to the language and set

$$S' = S \cup \{\neg c_i = c_j : i, j, \in I,\ i \neq j\}$$

By assumption, every finite subset of S' has a model, and hence S' has a model by compactness. Note that a model of S' is a model A of S together with an injection $I \to A$. □

Note. For any set X we can take $I = \gamma(X)$ (from Hartogs' Lemma). This shows that S has models that do not inject into X.

Remark. We can easily write down uncountable groups or vector spaces, but already for fields, the Upward Löwenheim–Skolem Theorem is not obvious.

Corollary 10. (The Downward Löwenheim–Skolem Theorem) Let S be a first-order theory in a countable language (*i.e.*, $\Omega \cup \Pi$ is countable). If S has a model, then it has a countable model.

Proof. By the Soundness Theorem, S is consistent since it has a model. Then the model constructed in the proof of the Model Existence Lemma is countable. □

Peano Arithmetic

We finish this chapter with another worked example. Our aim is to axiomatise the set of natural numbers. The key defining property of $\mathbb{N}$ is induction which we try to emulate with an axiom-scheme.

The language consists of $\Omega = \{0, s, +, \times\}$ with arities $0, 1, 2, 2$, respectively, and $\Pi = \emptyset$.

Peano Arithmetic (PA) (also known as *formal number theory*) is the theory in the language above with axioms, *i.e.*, sentences, as follows.

$(\forall\, x)(\neg sx = 0)$

$(\forall\, x)(\forall\, y)(sx = sy \Rightarrow x = y)$

$(\forall\, x)(x + 0 = 0)$

$(\forall\, x)(\forall\, y)(x + sy = s(x + y))$

$(\forall\, x)(x \times 0 = 0)$

$(\forall\, x)(\forall\, y)(x \times sy = (x \times y) + x)$

$(\forall\, y_1) \dots (\forall\, y_n)\Big(\big(p[0/x] \wedge (\forall\, x)(p \Rightarrow p[sx/x])\big) \Rightarrow (\forall\, x)p\Big)$
for any formula p with $\mathrm{FV}(p) = \{x, y_1, \dots, y_n\}$.

Remark. The last axiom is the axiom-scheme for induction. The variables $y_1, \dots, y_n$ are parameters. To see why they are needed, consider the following formula p: $(x + y) + z = x + (y + z)$ with free variables x, y, z. We can prove $(\forall\, x)(\forall\, y)(\forall\, z)p$ by iduction on z with x, y treated as parameters. Formally, we verify that

$$p[0/z] \wedge (\forall\, z)(p \Rightarrow p[sz/z])$$

holds in any model of PA, and hence it is provable by completeness. We then use the induction-scheme to deduce that $(\forall\, z)p$ holds, and hence so does $(\forall\, x)(\forall\, y)(\forall\, z)p$ by Generalisation.

Note. An obvious model of PA is the set $\mathbb{N}_0$ of non-zero integers. ($\mathbb{N}$ is also a model but it is more natural to include 0 in this context.) Then by the Upward Löwenheim–Skolem Theorem, PA has uncountable models. Doesn't this contradict the fact that induction (and the other axioms of PA) uniquely determine $\mathbb{N}_0$? The answer is 'no' because *true* induction says:

$$(\forall\, A \subset \mathbb{N}_0)\Big(\big(0 \in A \wedge (\forall\, x)(x \in A \Rightarrow sx \in A)\big) \Rightarrow A = \mathbb{N}_0\Big)$$

whereas in first-order theory we cannot quantify over subsets of a structure. Since the language of PA is countable, the induction axiom-scheme can only capture countable many subsets of $\mathbb{N}_0$.

Definition. A subset A is $\mathbb{N}_0$ is *definable* in PA if there is a formula p with free variable x such that $p_{\mathbb{N}_0}$, the interpretation of p in $\mathbb{N}_0$, is A.

Examples. The following sets are definable using the given formula:

set of squares: $(\exists\, y)(y \times y = x)$

set of primes: $\neg(x = 1) \wedge (\forall\, y)\big((y|x) \Rightarrow (y = 1 \vee y = x)\big)$ where $y|x$ is shorthand for $(\exists\, z)(z \times y = x)$ and 1 is $s0$.

powers of 2: $(\forall\, y)\big((\text{`}y \text{ is a prime'} \wedge y|x) \Rightarrow y = 2\big)$ where 2 is $ss0$.

Remark. Gödel's Incompleteness Theorem says (amongst other things) that PA is not a complete theory. Thus, there is a formula p in PA that holds in $\mathbb{N}_0$ but PA$\nvdash p$.

5 Set Theory

We will axiomatise set theory as a first-order theory. So this is just another piece of mathematical theory like group theory, topology, etc. So we could think of this chapter as just another worked example of first-order logic. Since any model of set theory should contain all of mathematics, it will obviously be a very complicated example of a first-order theory.

Zermelo–Fraenkel Set Theory (ZF). This is the version of set theory we will be studying. The language has no operation symbols ($\Omega = \emptyset$) and a single predicate $\in$ of arity 2. There will be 9 axioms: 2 to get started, 4 for building new sets and 3 further ones.

A model of ZF will be denoted V. This is a non-empty set together with an interpretation $\in_V \subset V \times V$ of the predicate $\in$ satisiying the axioms. Elements of V will be called 'sets'. When (a, b) is in $\in_V$, we say that 'a belongs to b' or that 'a is a member of b'. We refer to V as the 'set-theoretic universe' or the 'universe of sets'. So 'set' and 'belongs to' now have technical meanings inside the universe, but they retain their usual meaning in the world of mathematics of which V is a part. We will be interested in the question of what the universe of all sets looks like. We begin by listing the axioms of ZF.

1. Axiom of extensionality (Ext).
'If two sets have the same members, then they are equal.'

$$\text{(Ext)} \qquad (\forall\, x)(\forall\, y)\big((\forall\, z)(z \in x \Leftrightarrow z \in y) \Rightarrow (x = y)\big)$$

2. Axiom of separation (Sep).
'Can form a subset of a set.' This is in fact an axiom-scheme:

$$\text{(Sep)} \qquad (\forall\, t_1) \dots (\forall\, t_n)\Big((\forall\, x)(\exists\, y)(\forall\, z)\big(z \in y \Leftrightarrow ((z \in x) \wedge p)\big)\Big)$$

for any formula p with $\mathrm{FV}(p) = \{z, t_1, \dots, t_n\}$, where $t_1, \dots, t_n$ should be thought of as parameters. The set y whose existence is asserted in the sentence above is unique by (Ext) and is denoted by $\{z \in x : p\}$. (Formally, we are adding an $(n+1)$-ary operation symbol to the language of ZF.)

Note. Parameters are needed. For example, in a model of ZF we can form, given sets t and x, the subset $\{z \in x : t \in z\}$ of x.

3. Empty-set axiom (Emp).
'There is a set with no element.'

$$\text{(Emp)} \qquad (\exists\, x)(\forall\, y)\neg(y \in x)$$

The set x whose existence this axiom asserts is unique by (Ext). We call this set the *empty set* denoted by $\emptyset$. (Formally, we add the constant $\emptyset$ to the language of ZF with the sentence $(\forall\, y)\neg(y \in \emptyset)$.)

Strictly speaking, this axiom is not needed as it follows from (Sep). Indeed, in a structure V, we can pick any set x and form the set $\{y \in x : \neg(y = y)\}$ by (Sep). However, if in first-order logic we allow the empty set as a structure, then (Emp) is needed (or some axiom asserting the existence of some set).

4. Pair-set axiom (Pair).
'For any sets x, y, we can form $\{x, y\}$.'

$$\text{(Pair)} \qquad (\forall\, x)(\forall\, y)(\exists\, z)(\forall\, t)\Big(t \in z \Leftrightarrow \big((t = x) \vee (t = y)\big)\Big)$$

The unique set z whose existence is asserted here is denoted by $\{x, y\}$. We shall write $\{x\}$ for $\{x, x\}$. (Formally, we add a binary operation $\{,\}$ and a unary operation $\{\}$ to the language of ZF.)

It follows from (Ext) that $\{x, y\} = \{y, x\}$ for all x, y. Thus, (Pair) gives us unordered pairs. Ordered pairs can be constructed using the following device (due, independently, to K. Kuratowski and N. Wiener): for sets x, y define $(x, y) = \{\{x\}, \{x, y\}\}$. This satisfies:

$$(\forall\, x)(\forall\, y)(\forall\, z)(\forall\, t)\big((x, y) = (z, t) \Leftrightarrow ((x = z) \wedge (y = t))\big)$$

as one would expect. We now introduce a number of abbreviations.

'x is an ordered pair' means

$$(\exists\, y)(\exists\, z)(x = (y, z))$$

'f is a function' means

$$(\forall\, x)\big((x \in f) \Rightarrow (x \text{ is an ordered pair})\big) \\ \wedge (\forall\, x)(\forall\, y)(\forall\, z)\Big(\big((x, y) \in f \wedge (x, z) \in f\big) \Rightarrow (y = z)\Big)$$

$x = \operatorname{dom} f$ ('x is the domain of f') means

$$(f \text{ is a function}) \wedge (\forall\, y)\big((y \in x) \Leftrightarrow (\exists\, z)((y, z) \in f)\big)$$

$f\colon x \to y$ ('f is a function from x to y') means

$$(x = \operatorname{dom} f) \wedge (\forall\, z)\big((\exists\, t)((t, z) \in f) \Rightarrow (z \in y)\big)$$

5. Union axiom (Un).
'Can form the union of a set.'

$$\text{(Un)} \qquad (\forall\, x)(\exists\, y)(\forall\, z)\Big((z \in y) \Leftrightarrow (\exists\, t)\big((t \in x) \wedge (z \in t)\big)\Big)$$

Note that the set y is the union of the members of x. We denote this unique (by Extensionality) set y by $\bigcup x$. (Formally, we add the unary symbol $\bigcup$ to the language of ZF.)

As an example, note that if $x = \{a, b\}$, then $t \in \bigcup x$ if and only if $t \in a$ or $t \in b$. We shall denote $\bigcup\{a, b\}$ by $a \cup b$. (Formally, we introduce a binary operation $\cup$ to the language of ZF.)

Note. We do not need a separate axiom for intersections. Indeed, the following sentence follows from the axioms so far:

$$(\forall\, x)\Big(\neg(x = \emptyset) \Rightarrow (\exists\, y)(\forall\, z)\big((z \in y) \Leftrightarrow (\forall\, t)(t \in x \Rightarrow z \in t)\big)\Big)$$

Indeed, given a non-empty set x, we can form the set

$$y = \Big\{ z \in \bigcup x : (\forall\, t)(t \in x \Rightarrow z \in t) \Big\}$$

using (Un) and (Sep). Note that technically we work in a model here to construct the set y; then the sentence above follows by the Completeness Theorem. We will denote the unique (by Extensionality) set y constructed above by $\bigcap x$. We will write $a \cap b$ for $\bigcap\{a, b\}$.

6. Power-set axiom (Pow).

'Can form the power set of a set.'

(Pow) $$(\forall\, x)(\exists\, y)(\forall\, z)(z \in y \Leftrightarrow z \subset x)$$

where $z \subset x$ is an abbreviation for $(\forall\, t)(t \in z \Rightarrow t \in x)$. The unique set y is denoted by $\mathbb{P}x$.

We can now form the Cartesian product of sets x and y. First note that for $u \in x$ and $v \in y$, the ordered pair $(u, v) = \{\{u\}, \{u, v\}\}$ belongs to $\mathbb{PP}(x \cup y)$. We can then form the set

$$x \times y = \Big\{ z \in \mathbb{PP}(x \cup y) : (\exists\, t)(\exists\, u)\big((z = (t, u)) \wedge (t \in x) \wedge (u \in y)\big) \Big\}$$

using (Un), (Pow) and (Sep).

In turn, we can form the set y^x of all functions from x to y:

$$y^x = \{ f \in \mathbb{P}(x \times y) : f \colon x \to y \}$$

7. Axiom of infinity (Inf). With the first six axioms we can already do quite a bit of mathematics. Also, in any model V there will be infinitely many elements. For example, it is easy to show that the sets $\emptyset, \mathbb{P}\emptyset, \mathbb{PP}\emptyset, \dots$ are pairwise distinct.

For another example, let us first introduce for a set x, the *successor* of x to be the set $x^+ = x \cup \{x\}$. Then the sets $\emptyset, \emptyset^+, \emptyset^{++}, \emptyset^{+++}, \dots$ are pairwise distinct. We shall denote these sets by $0, 1, 2, 3, \dots$, respectively. Thus,

$$0 = \emptyset \ , \quad 1 = \{0\} \ , \quad 2 = \{0, 1\} \ , \quad 3 = \{0, 1, 2\} \ , \dots$$

These examples show that from the outside, V is an infinite set. However, V is not a set, *i.e.*, the sentence $(\exists\, x)(\forall\, y)(y \in x)$ does not hold in V. This is known as Russell's paradox. (Indeed, if the sentence holds in V, then we can form the set $y = \{z \in x : \neg(z \in z)\}$ by (Sep), and get a contradiction by considering whether $y \in y$.) So we need an axiom that says that, for example, the elements $0, 1, 2, 3, \dots$ form a set. We begin with a definition.

Say that 'x is a successor set' if

$$(0 \in x) \wedge (\forall\, y)(y \in x \Rightarrow y^+ \in x)$$

Any successor set will contain $0, 1, 2, 3, \dots$, and hence any successor set is infinite. The axiom of infinity states that successor sets exist.

(Inf) $$(\exists\, x)(x \text{ is a successor set})$$

It follows from this axiom (and the ones listed so far) that there is a smallest successor set:

$$(\exists\, x)\big(\text{`}x \text{ is a successor set'} \wedge (\forall\, y)(\text{`}y \text{ is a successor set'} \Rightarrow x \subset y)\big)$$

To prove this, fix a successor set y and form the set

$$z = \{t \in \mathbb{P}y : t \text{ is a successor set}\}$$

by (Pow) and (Sep). It is easy to check that $x = \bigcap z$ is the smallest successor set which will be denoted by ω.

Note that every successor set contained in ω is ω:

$$(\forall\, x \subset \omega)(0 \in x \wedge (\forall\, y)(y \in x \Rightarrow y^+ \in x)) \Rightarrow x = \omega)$$

where $(\forall\, x \subset y)p$ is shorthand for $(\forall\, x)(x \subset y \Rightarrow p)$ for any formula p. So we have true induction in V as we quantify over all subsets of ω. (Note, however, that from the outside ω may have subsets that do not form a set inside V.) We refer to this as ω-induction.

An example of a statement that can be proved by ω-induction is that every non-zero element of ω is a successor:

$$(\forall\, x \in \omega)(\neg(x = 0) \Rightarrow (\exists\, y \in \omega)(x = y^+))$$

where $(\forall\, x \in y)p$ is shorthand for $(\forall\, x)(x \in y \Rightarrow p)$ for any formula p.

It is straightforward to check that $(\forall\, x \in \omega)\neg(x^+ = 0)$. One can also prove that $(\forall\, x \in \omega)(\forall\, y \in \omega)(x^+ = y^+ \Rightarrow x = y)$. Thus, ω satisfies the usual rules for the natural numbers (*cf.* axioms of Peano Arithmetic).

We now define abbreviations 'x is finite' to mean $(\exists\, y \in \omega)(x$ bijects with $y)$ and 'x is countable' to mean that $(\exists\, f)(f \colon x \to \omega$ is an injection).

8. Axiom of replacement (Rep).

With the axioms so far, we can form the set ω containing $0, 1, 2, \ldots$. How about the sets $\emptyset, \mathbb{P}\emptyset, \mathbb{P}\mathbb{P}\emptyset, \ldots$? Do they form a set? We will see later that the map (inside V) $0 \mapsto \emptyset, 1 \mapsto \mathbb{P}\emptyset, 2 \mapsto \mathbb{P}\mathbb{P}\emptyset, \ldots$ can be expressed by a formula. Such a map is called a function-class. So we need an axom saying that the image of a set under a function-class is a set. It will then follow that $\emptyset, \mathbb{P}\emptyset, \mathbb{P}\mathbb{P}\emptyset, \ldots$ form a set, namely the image of ω under the function-class $0 \mapsto \emptyset, 1 \mapsto \mathbb{P}\emptyset, 2 \mapsto \mathbb{P}\mathbb{P}\emptyset, \ldots$.

Digression on classes.

Definition. A *class* is a collection C of elements of V such that there is a formula p with one free variable satisfying $C = p_V$, *i.e.*, $x \in C$ if and only if $p(x)$ holds in V.

Examples. 1. V is a class: we can take p to be the formula $(x = x)$.

2. The collection of sets of size one is a class: we can take p to be the formula $(\exists\, y)(x = \{y\})$.

Definition. A class C, given by a formula p with free variable y, is a set if $(\exists\, x)(\forall\, y)(y \in x \Leftrightarrow p)$. Otherwise we say C is a *proper class*.

Example. V is a proper class.

Definition. A *function-class* is a subset F of $V \times V$ such that $F = p_V$ for some formula p with two free variables x, y satisfying

$$(\forall\, x)(\forall\, y)(\forall\, z)\big((p \wedge p[z/y]) \Rightarrow (y = z)\big)$$

Thus, $(x, y) \in F$ if and only if $p(x, y)$ holds in V.

Example. The map $x \mapsto \{x\}$ is a function class: we can take p to be $(y = \{x\})$.

End of digression.

The Axiom of Replacement is an axiom-scheme stating that the image of a set under a function-class is a set. As usual, we use parameters.

$$\text{(Rep)} \quad (\forall\, t_1) \dots (\forall\, t_n)\Big[(\forall\, x)(\forall\, y)(\forall\, z)\big((p \wedge p[z/y]) \Rightarrow (y = z)\big) \\ \Rightarrow (\forall\, x)(\exists\, y)(\forall\, z)\big(z \in y \Leftrightarrow (\exists\, u)(u \in x \wedge p[u/x, z/y])\big)\Big]$$

for any formula p with $\mathrm{FV}(p) = \{x, y, t_1, \dots, t_n\}$.

9. Axiom of foundation (Fnd).

We want a picture of the universe in which sets appear at a certain 'time', and a set cannot appear before all its elements do. So we want to avoid pathological behaviour like $x \in x$, *i.e.*, we want to avoid $\{x\}$ having no $\in$-minimal member. Similarly, we don't want $x \in y$ and $y \in x$, *i.e.*, we want to avoid $\{x, y\}$ having no $\in$-minimal member. The Axiom of Foundation (or Axiom of Regularity) states that every non-empty set has an $\in$-minimal member:

$$\text{(Fnd)} \qquad (\forall\, x)(\neg(x = \emptyset) \Rightarrow (\exists\, y)(y \in x \wedge (\forall\, z)(z \in x \Rightarrow \neg(z \in y))))$$

Remark. The nine axioms and axiom-schemes above form ZF set theory. Note that the Axiom of Choice is not included. We shall write ZFC for ZF+AC, *i.e.*, ZF set theory with the Axiom of Choice.

$$\text{(AC)} \quad (\forall\, x)((\forall\, y \in x)\neg(y = \emptyset) \Rightarrow (\exists\, f)(f \colon x \to \bigcup x \wedge (\forall\, y \in x)(f(y) \in y)))$$

Remark. For the rest of this chapter we work within ZF. Our ultimate aim is to describe the set-theoretic universe V. We first prove versions of induction and recursion similar to but more general than those introduced for well-ordered sets in Chapter 2. This will eventually lead to a proper definition of ordinals thereby filling the gap from Chapter 2. We then describe a picture of the universe in which sets appear in 'time' measured by ordinals where no set appears before all its members do.

Definition. A set x is *transitive* if every member of a member of x is a member of x. Thus, 'x is transitive' is shorthand for

$$(\forall\, y)((\exists\, z)(z \in x \wedge y \in z) \Rightarrow y \in x)$$

or equivalently $\bigcup x \subset x$.

Remark. This is not the same as saying that $\in$ is a transitive relation on x.

Examples. An easy ω-induction shows that $(\forall\, x \in \omega)(x \subset \omega)$. It follows that ω is a transitive set. Another ω-induction shows that every member of ω is a transitive set.

Lemma 1. Every set x is contained in a transitive set:

$$(\forall\, x)(\exists\, y)(\text{`}y \text{ is transitive'} \wedge x \subset y)$$

Remark. Since the intersection of a non-empty set of transitive sets is transitive, Lemma 1 implies that there is a smallest transitive set containing x called the *transitive closure of* x denoted $\mathrm{TC}(x)$.

Idea of proof. If y is a transitive set with $x \subset y$, then $\bigcup x \subset y$, and in turn $\bigcup\bigcup x \subset y$, etc. So we would like to form the set $\bigcup\{x, \bigcup x, \bigcup\bigcup x, \dots\}$ which is indeed transitive and contains x. However, for this to work, we need $\{x, \bigcup x, \bigcup\bigcup x, \dots\}$ to be a set. This will follow by the Axiom of Replacement. So we need to show that the map sending $0 \mapsto x, 1 \mapsto \bigcup x, 2 \mapsto \bigcup\bigcup x, \dots$ is a function-class.

Proof of Lemma 1. We introduce the abbreviation 'f is an attempt' to mean

$$\begin{aligned}&(f \text{ is a function}) \wedge (\operatorname{dom} f \in \omega) \wedge (x = f(0)) \\ &\quad \wedge (\forall\, m)(\forall\, n)\Big(\big(m \in \operatorname{dom} f \wedge n \in \operatorname{dom} f \wedge n = m^+\big) \Rightarrow f(n) = \bigcup f(m)\Big)\end{aligned}$$

A straightforward ω-induction shows that any two attempts agree on the intersection of their domains:

$$(*) \quad (\forall\, f)(\forall\, g)(\forall\, n)\Big(\Big[(f \text{ is an attempt}) \wedge (g \text{ is an attempt}) \\ \wedge (n \in \operatorname{dom} f) \wedge (n \in \operatorname{dom} g)\Big] \Rightarrow (f(n) = g(n))\Big)$$

Another ω-induction shows that every member of ω is in the domain of some attempt:

$$(**) \qquad (\forall\, n)\Big((n \in \omega) \Rightarrow (\exists\, f)\big[(f \text{ is an attempt}) \wedge (n \in \operatorname{dom} f)\big]\Big)$$

To see this, form the set

$$w = \big\{n \in \omega : (\exists\, f)((f \text{ is an attempt}) \wedge (n \in \operatorname{dom} f))\big\}$$

by (Sep). We show that w is a successor set from which it will follow that $w = \omega$, as required. Since $f = \{(0, x)\}$ is an attempt, $0 \in w$. If $n \in w$, then fix an attempt f with $n \in \operatorname{dom} f$. Since every member of ω is transitive, we have $n \subset \operatorname{dom} f$, and hence $n^+ \subset \operatorname{dom} f$. By restricting f to n^+, we can assume that $\operatorname{dom} f = n^+$ in which case

$$g = f \cup \big\{\big(n^+, \textstyle\bigcup f(n)\big)\big\}$$

is an attempt with $n^+ \in \operatorname{dom} g$. Thus, $n^+ \in w$.

We now let p be the following formula with free variables y, z:

$$(\exists\, f)\big((f \text{ is an attempt}) \wedge ((y, z) \in f)\big)$$

It follows from $(*)$ that

$$(\forall\, y)(\forall\, z)(\forall\, u)((p \wedge p[u/z]) \Rightarrow u = z)$$

and thus p defines a function-class. By Replacement the image of ω under this function-class is a set: we can form the set w such that $z \in w$ if and only if $(\exists\, y)p$ holds. Informally, w is the set $\{x, \bigcup x, \bigcup\bigcup x, \dots\}$. We now form the set $t = \bigcup w$ and show that t is transitive with $x \subset t$ which will complete the proof.

Since $\{(0, x)\}$ is an attempt, it follows that $x \in w$ and $x \subset t$. Now assume that $a \in t$. Then $a \in z$ for some $z \in w$, and in turn $z = f(n)$ for some attempt f with $n \in \operatorname{dom} f$. By $(**)$, there is an attempt g with $n^+ \in \operatorname{dom} g$. Then $n \in \operatorname{dom} g$ since members of ω are transitive, and $g(n) = f(n) = z$ by $(*)$. It follows that $g(n^+) = \bigcup g(n) = \bigcup z$ by the definition of an attempt, and so $\bigcup z \in w$. Thus for any $b \in a$, we have $b \in \bigcup z \in w$, and hence $b \in t$ as required. □

Remark. The set t constructed in the proof above is in fact the transitive closure $\mathrm{TC}(x)$ of x.

Theorem 2. (Principle of $\in$-induction) For each formula p with free variables $\mathrm{FV}(p) = \{x, t_1, \dots, t_n\}$, the following sentence holds in ZF.

$$(\forall\, t_1)\dots(\forall\, t_n)\Big((\forall\, x)\big[(\forall\, y \in x)p[y/x] \Rightarrow p\big] \Rightarrow (\forall\, x)p\Big)$$

Proof. Fix values $t_1, \dots, t_n$ of the parameters and assume that $p(x)$ holds whenever $p(y)$ holds for all members y of x, *i.e.*, $(\forall\, x)\big[(\forall\, y \in x)p[y/x] \Rightarrow p\big]$ holds. Assume for a contradiction that $\neg(\forall\, x)p(x)$ holds and fix any set x such that $p(x)$ fails.

(At this point we would like to take a minimal counterexample, *i.e.*, an $\in$-minimal member of $\{y : \neg p(y)\}$. However, $\{y : \neg p(y)\}$ may not be a set. This is where transitive closure comes in.)

By Lemma 1 we can form the set $t = \mathrm{TC}(\{x\})$, and by (Sep) we can form the set $u = \{y \in t : \neg p(y)\}$. Then $x \in u$, and hence u has an $\in$-minimal member z. If $y \in z$, then $y \in t$ since t is transitive, and thus $y \notin u$ since z is $\in$-minimal in u. It follows that $p(y)$ holds for all $y \in z$. By assumption on p, we deduce $p(z)$ contradicting the choice of z. □

Remark. In the presence of the first eight axioms of ZF, the Principle of $\in$-induction is equivalent to the Axiom of Foundation. One direction is Theorem 2. For the converse, assume the Principle of $\in$-induction. Say that a set x is *regular* if

$$(\forall\, y)(x \in y \Rightarrow (y \text{ has an } \in\text{-minimal member}))$$

(this definition is the clever bit). Then (Fnd) is equivalent to the assertion that $(\forall\, x)(x \text{ is regular})$ which we prove by $\in$-induction. Fix a set x and assume that every $y \in x$ is regular (the induction hypothesis). Let z be a set with $x \in z$.

We need to show that z has an $\in$-minimal member. This is obviously true if x itself is an $\in$-minimal member of z. If not, then we have $y \in z$ for some $y \in x$, in which case z has an $\in$-minimal member since y is regular by the induction hypothesis. This shows that x is regular, as required.

We now turn to $\in$-recursion. Informally, this is the statement that a function f can be defined so that for every x, the value $f(x)$ is given in terms of the values $f(y)$, $y \in x$.

Theorem 3. ($\in$-recursion theorem) Let G be a function-class (given by a formula p in two variables) defined everywhere (*i.e.*, $(\forall\, x)(\exists\, y)p(x, y)$), then there is a function-class F (given by a formula q in two variables) such that $(\forall\, x)(F(x) = G(F{\restriction}_x))$. Moreover, F is unique.

Note. The restriction $F{\restriction}_x$ of F to x is $\{(s, F(s)) : s \in x\}$ which is a set by Replacement: It is the image of the set x under the function-class $s \mapsto (s, F(s))$ which is given by the formula $(\exists\, z)(q(s, z) \wedge t = (s, z))$.

Proof. We first prove uniqueness. Assume that F_1 and F_2 both satisfy the conclusions of the theorem. We show $(\forall\, x)(F_1(x) = F_2(x))$ by $\in$-induction. Fix a set x and assume that $(\forall\, y \in x)(F_1(y) = F_2(y))$. Then $F_1{\restriction}_x = F_2{\restriction}_x$, and hence $F_1(x) = G(F_1{\restriction}_x) = G(F_2{\restriction}_x) = F_2(x)$.

We now turn to existence. We say that a set f is an *attempt* if

$$(f \text{ is a function}) \wedge (\operatorname{dom} f \text{ is transitive}) \wedge (\forall\, x)(x \in \operatorname{dom} f \Rightarrow f(x) = G(f{\restriction}_x))$$

Note that if $x \in \operatorname{dom} f$, then $x \subset \operatorname{dom} f$ since $\operatorname{dom} f$ is transitive, and hence $f{\restriction}_x$ makes sense. Now a straightforward $\in$-induction (as in the proof of uniqueness) shows that

$$(*) \quad (\forall\, f)(\forall\, g)(\forall\, x)\Big(\big[(f \text{ is an attempt}) \wedge (g \text{ is an attempt}) \\ \wedge (x \in \operatorname{dom} f) \wedge (x \in \operatorname{dom} g)\big] \Rightarrow f(x) = g(x)\Big)$$

Another $\in$-induction shows that

$$(\forall\, x)(\exists\, f)((f \text{ is an attempt}) \wedge (x \in \operatorname{dom} f))$$

To see this, fix a set x and assume that for all $y \in x$ there is an attempt defined at y. Note that an attempt defined at y is defined on $\mathrm{TC}(\{y\})$ since the domain of an attempt is transitive, and the restriction to $\mathrm{TC}(\{y\})$ of this attempt is still an attempt. Hence by $(*)$, for each $y \in x$ there is a unique attempt f_y defined on $\mathrm{TC}(\{y\})$. Then $\{f_y : y \in x\}$ is a set by Replacement and $f' = \bigcup\{f_y : y \in x\}$ is an attempt whose domain contains x. Finally, $f = f' \cup \{(x, G(f'{\restriction}_x))\}$ is an attempt defined at x.

We now let q be the formula $(\exists\, f)((f \text{ is an attempt}) \wedge y = f(x))$. It is now straightforward to verify that q defines a function-class F with the required properties. □

Remark. We can generalize $\in$-induction and $\in$-recursion to other relations. By a relation we mean a formula r in two variables (whose interpretation in a model V is a subset of $V \times V$). For example, if r is $(x \in y)$, then r is the $\in$-relation. The next definition identifies the two properties of the $\in$-relation that were crucial in proving induction and recursion.

Definition. A relation r is *well-founded* if every non-empty set has an r-minimal member:

$$(\forall x)\Big(\neg(x = \emptyset) \Rightarrow (\exists y)\big[y \in x \wedge (\forall z \in x)\neg r(z, y)\big]\Big)$$

A relation r is *local* if the r-predecessors of a set form a set:

$$(\forall x)(\exists y)(\forall z)(z \in y \Leftrightarrow r(z, x))$$

Remark. Given a local relation r, we can define the notion of r-closure similar to transitive closure. Then if in addition r is well-founded, we can prove r-induction and r-recursion.

In what follows we can restrict attention to relations defined on sets. If r is a relation on a set a, *i.e.*, $r \subset a \times a$, then r is automatically local since for any $x \in a$, we can form the set $\{y \in a : y\, r\, x\}$ by Separation. (Here we use the familiar notation $y\, r\, x$ instead of $(y, x) \in r$.) So in this case we just need r to be well-founded (every non-empty subset of a has an r-minimal member) in order to prove r-induction and r-recursion.

The next result shows that a well-founded relation on a set can always be modelled by $\in$ provided we assume a further property.

Definition. A relation r on a set a is *extensional* if members of a are uniquely determined by their set of r-predecessors:

$$(\forall x \in a)(\forall y \in a)\big((\forall z \in a)(z\, r\, x \Leftrightarrow z\, r\, y) \Rightarrow x = y\big)$$

Theorem 4. (Mostowski's Collapsing Theorem) Let r be a well-founded, extensional relation on a set a. Then there is a transitive set b and a bijection $f\colon a \to b$ such that

$$(\forall x \in a)(\forall y \in a)\big(x\, r\, y \Leftrightarrow f(x) \in f(y)\big)$$

Moreover, the pair (b, f) is unique.

Remark. The sentence above states that f is an order-isomorphism between a with relation r and b with relation $\in$. Thus, well-foundedness and extensionality of r are necessary conditions.

Proof. We begin with existence. By r-recursion there is a function-class f such that

$$(\forall x \in a)\big(f(x) = \{f(y) : y \in a,\ y\, r\, x\}\big)$$

Since $\{(x, f(x)) : x \in a\}$ is a set by (Rep), we can take f to be a function. We next set $b = \{f(x) : x \in a\}$ which is also a set by (Rep). We verify that (b, f) satisfies the conclusions of the theorem.

We first show that b is transitive. Given $z \in b$, $z = f(x)$ for some $x \in a$, and hence $z = \{f(y) : y \in a,\ y\, r\, x\}$. It follows that $w \in b$ whenever $w \in z$.

By definition, f is surjective and $x\,r\,y$ implies $f(x) \in f(y)$ for all $x, y \in a$. It remains to show that f is injective. This will also show that $f(x) \in f(y)$ implies $x\,r\,y$ for all $x, y \in a$. Indeed, if $f(x) \in f(y)$, then $f(x) = f(z)$ for some $z \in a$ with $z\,r\,y$. Then by injectivity $z = x$, and thus $x\,r\,y$.

For $x \in a$, say that f is *injective at a* if $(\forall\, y \in a)(f(y) = f(x) \Rightarrow y = x)$. Then f is injective if and only if $(\forall\, x \in a)(f$ is injective at $x)$ which we show by r-induction. Fix $x \in a$ and assume that f is injective at s for all $s \in a$ with $s\,r\,x$. Assume that $f(x) = f(y)$ for some $y \in a$. Then

$$\{f(s) : s \in a,\ s\,r\,x\} = \{f(t) : t \in a,\ t\,r\,y\}$$

Since f is injective at every $s \in a$ with $s\,r\,x$, it follows that

$$\{s : s \in a,\ s\,r\,x\} = \{t : t \in a,\ t\,r\,y\}$$

Since the relation r is extensional, we have $x = y$ as required.

We complete the proof by showing uniqueness. Let (b, f) and (b', f') be pairs both satisfying the conclusions of the theorem. We show

$$(\forall\, x \in a)(f(x) = f'(x))$$

by r-induction. Fix $x \in a$ and assume that $f(y) = f'(y)$ for all $y \in a$ with $y\,r\,x$. Given $w \in f(x)$, we have $w \in b$ since b is transitive, and $w = f(z)$ for some $z \in a$ since f is surjective. Then $f(z) \in f(x)$, and hence $z\,r\,x$. It follows by the induction hypothesis that $w = f(z) = f'(z) \in f'(x)$. Similarly, $w \in f'(x)$ implies $w \in f(x)$. Thus, by the Axiom of Extensionality, we have $f(x) = f'(x)$. This completes the r-induction which shows that $f = f'$, and thus $b = b'$. □

Definition. An *ordinal* is a transitive set well-ordered by $\in$ (equivalently, linearly ordered by $\in$ since $\in$ is well-founded by (Fnd)).

Note. Suppose a is a set and r is a well-ordering on a. By Mostowski, there is a transitive set b and a bijection $f\colon a \to b$ such that $x\,r\,y \Leftrightarrow f(x) \in f(y)$ for all $x, y \in a$. Thus, (a, r) is order-isomorphic to $(b, \in)$. It follows that b is an ordinal. Moreover, by the uniqueness in Theorem 4, this is the unique ordinal to which (a, r) is order-isomorphic. This unique ordinal is called the *order-type* of the well-ordered set a.

Remark. We denote by ON the class of all ordinals. ON is a proper class by the Burali-Forti paradox. Note that each order-isomorphism class of well-ordered sets contains exactly one ordinal.

Proposition 5. Let $\alpha, \beta \in \mathrm{ON}$ and a be a set of ordinals.

(i) Every member of α is an ordinal.

(ii) $\beta \in \alpha \Leftrightarrow \beta < \alpha$

(iii) $\beta \in \alpha$ or $\beta = \alpha$ or $\alpha \in \beta$

(iv) $\alpha^+ = \alpha \cup \{\alpha\}$

(v) $\bigcup a$ is an ordinal and $\bigcup a = \sup a$.

Remarks. 1. Recall that the notation $\beta < \alpha$ in part (ii) means that β is order-isomorphic to a proper initial segment of α. Parts (i) and (ii) together show that the ordinal α really *is* the set of ordinals strictly less than α.

2. Part (iii) shows that $\in$ is a linear order on the class ON.

3. Part (iv) reconciles two definitions. According to the definition in Chapter 2, α^+ is the unique (up to order-isomorphism) well-ordered set that consists of α as a proper initial segment and one extra element that is a maximum. By Mostowski, this well-ordered set is order-isomorpic to a unique ordinal (its order-type). Part (iv) shows that this ordinal is the successor of the set α as defined in this chapter. In particular, this shows that the successor of an ordinal is an ordinal.

4. Part (v) shows that any set x of well-ordered sets has an upper bound. This was owed from Chapter 2 (see the Remark following Proposition 2.8). Indeed, $a = \{\text{order-type}(y) : y \in x\}$ is a set of ordinals by (Rep) which by part (v) has an upper bound.

Proof. (i) Fix $\gamma \in \alpha$. Then $\gamma \subset \alpha$ since α is transitive, and so γ is linearly ordered by $\in$. It remains to show that γ is transitive. Let $\eta \in \delta$ and $\delta \in \gamma$. Since α is transitive, we have $\delta \in \alpha$, and in turn $\eta \in \alpha$. Since $\in$ linearly orders α, it is in particular a transitive relation on α, and thus $\eta \in \gamma$.

(ii) Recall that for a well-ordered set x, for every $y \in x$ we denote by I_y the proper initial segment $\{z \in x : z < y\}$ of x, and moreover, every proper initial segment is of this form. It follows that if $\beta \in \alpha$, then $I_\beta = \{\gamma \in \alpha : \gamma \in \beta\}$ is a proper initial segment of α. Since α is transitive, $I_\beta = \beta$, and thus $\beta < \alpha$. Conversely, if $\beta < \alpha$, then β is order-isomorphic to $I_\delta = \delta$ for some $\delta \in \alpha$. By uniqueness, $\delta = \beta$, and thus $\beta \in \alpha$.

(iii) is immediate from parts (i) and (ii) and from Theorem 2.6. Note that if $\alpha \neq \beta$, then α and β are not order-isomorphic by uniqueness in Theorem 4.

(iv) Let β be the successor of α, *i.e.*, $\beta = \alpha \cup \{\alpha\}$. It is straightforward to check that β is an ordinal. Then α is the maximum element of β (with respect to $\in$ of course), and α is a proper initial segment of α. So β is indeed α^+ in the sense of Chapter 2.

(v) First observe the following consequences of parts (ii) and (iii). Firstly, $\beta \subset \alpha \Leftrightarrow \beta \leqslant \alpha$. Secondly, one of $\beta \subset \alpha$ and $\alpha \subset \beta$ must hold. Thus, a is a nested set of transitive sets well-ordered by $\in$. It follows that $\bigcup a$ is a transitive set well-ordered by $\in$, *i.e.*, an ordinal, and it is the supremum of a. □

Examples. We finally give some examples of ordinals. Rather trivially, $0 = \emptyset$ is an ordinal. An easy ω-induction and Proposition 5 (iv) shows that every member of ω is an ordinal. Hence, so is $\bigcup \omega = \omega$ (the equality follows from transitivity of ω). This shows that $\omega = \sup \omega$, and thus the set ω introduced in this chapter is reassuringly coincides with the ω of Chapter 2.

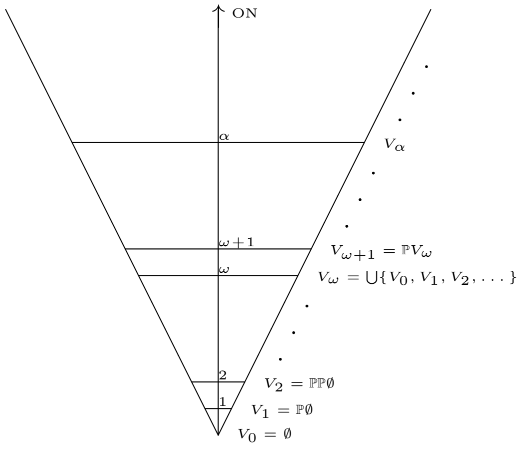
Picture of the Universe

Idea. We build the entire universe V starting from the empty set by repeatedly applying $\mathbb{P}$ and $\bigcup$. So we have $\emptyset, \mathbb{P}\emptyset, \mathbb{P}\mathbb{P}\emptyset, \dots$, then $\bigcup\{\emptyset, \mathbb{P}\emptyset, \mathbb{P}\mathbb{P}\emptyset, \dots\}$, etc.

Formally, we define V_α, $\alpha \in \mathrm{ON}$, by $\in$-recursion as follows.

$$\begin{aligned} V_0 &= \emptyset \\ V_{\alpha^+} &= \mathbb{P}V_\alpha \\ V_\lambda &= \bigcup\{V_\alpha : \alpha < \lambda\} \qquad \text{for a non-zero limit ordinal } \lambda \ . \end{aligned}$$

The class of sets V_α, $\alpha \in \mathrm{ON}$, is called the *von Neumann hierarchy.* Our aim is to show every set appears in one of the sets V_α. This leads to the following, somewhat unstable-looking, picture of the universe, which perhaps also explains why it is usually denoted by V.


ON
α
V_α
$\omega+1$
$V_{\omega+1} = \mathbb{P}V_\omega$
ω
$V_\omega = \bigcup\{V_0, V_1, V_2, \dots\}$
2
$V_2 = \mathbb{P}\mathbb{P}\emptyset$
1
$V_1 = \mathbb{P}\emptyset$
$V_0 = \emptyset$

Lemma 6. V_α is transitive for all $\alpha \in \mathrm{ON}$.

Proof. We proceed by induction on α.

$\alpha = 0$: $V_0 = \emptyset$ is transitive.

$\alpha = \beta + 1$: Assume V_β is transitive and let $x \in V_\alpha$ and $y \in x$. We need to show that $y \in V_\alpha$. Since $V_\alpha = \mathbb{P}V_\beta$, we have $x \subset V_\beta$ and $y \in V_\beta$. Since V_β is transitive, it follows that $y \subset V_\beta$, and hence $y \in \mathbb{P}V_\beta = V_\alpha$, as required.

α is a non-zero limit: Assume V_β is transitive for all $\beta < \alpha$. Then V_α is a union of transitive sets, and thus transitive. □

Lemma 7. $V_\alpha \subset V_\beta$ for all $\alpha \leqslant \beta$.

Proof. We proceed by induction on β. In each case, we can assume $\alpha < \beta$. The case $\beta = 0$ is clear.

$\beta = \gamma + 1$: Let $\alpha < \beta$. Then $\alpha \leqslant \gamma$, and so $V_\alpha \subset V_\gamma$ by the induction hypothesis. It follows that $V_\alpha \in \mathbb{P}V_\gamma = V_\beta$. By Lemma 6, the set V_β is transitive, and thus $V_\alpha \subset V_\beta$, as required.

β is a non-zero limit: If $\alpha < \beta$, then $V_\alpha \subset V_\beta$ by definition of V_β. □

Theorem 8. The von Neumann hierarchy exhausts the set-theoretic universe, *i.e.*, $(\forall\, x)(\exists\, \alpha \in \mathrm{ON})(x \in V_\alpha)$ holds in ZF.

Note. If $x \in V_\alpha$, then $x \subset V_\alpha$ since V_α is transitive. Conversely, if $x \subset V_\alpha$, then $x \in \mathbb{P}V_\alpha = V_{\alpha+1}$. Thus Theorem 8 is equivalent to the assertion that $(\forall\, x)(\exists\, \alpha \in \mathrm{ON})(x \subset V_\alpha)$ holds in ZF. For a set x, the least ordinal α with $x \subset V_\alpha$ is called the *rank* of x, denoted $\mathrm{rank}(x)$.

Proof. We prove $(\forall\, x)(\exists\, \alpha \in \mathrm{ON})(x \subset V_\alpha)$ by $\in$-induction.

Fix a set x, and assume that every $y \in x$ has a rank (the induction hypothesis). Set $\alpha = \sup\{\mathrm{rank}(y)^+ : y \in x\}$ noting that $\{\mathrm{rank}(y)^+ : y \in x\}$ is a set by Replacement. For each $y \in x$, we have $y \subset V_{\mathrm{rank}(y)}$, and so $y \in V_{\mathrm{rank}(y)^+}$. By Lemma 7 we have $V_{\mathrm{rank}(y)^+} \subset V_\alpha$ for all $y \in x$, and hence $x \subset V_\alpha$, as required. □

Corollary 9. $\mathrm{rank}(x) = \sup\{\mathrm{rank}(y)^+ : y \in x\}$.

Proof. It follows from the proof of Theorem 8 that

$$\mathrm{rank}(x) \leqslant \sup\{\mathrm{rank}(y)^+ : y \in x\} .$$

For the reverse inequality, we first show that if $x \in V_\alpha$ then $\mathrm{rank}(x) < \alpha$. So let us assume that $x \in V_\alpha$. Then $\alpha > 0$. If $\alpha = \beta^+$, then $x \subset V_\beta$, and hence $\mathrm{rank}(x) \leqslant \beta < \alpha$. If α is a limit ordinal, then $x \in V_\beta$ for some $\beta < \alpha$. It follows that $x \subset V_\beta$ since V_β is transitive, and thus $\mathrm{rank}(x) \leqslant \beta < \alpha$.

Now set $\alpha = \mathrm{rank}(x)$. Then for each $y \in x$, we have $y \in V_\alpha$, and hence $\mathrm{rank}(y) < \alpha$ by the claim above. It follows that $\sup\{\mathrm{rank}(y)^+ : y \in x\} \leqslant \alpha$. □

Example. Using the formula above, an easy induction shows that $\mathrm{rank}(\alpha) = \alpha$ for every ordinal α.

6 Cardinal Arithmetic

In this chapter we are interested in the size of sets. So we will want to identify sets that have the same size. We introduce the abbreviation '$x \equiv y$' for the formula $(\exists\, f)(f$ is a bijection from x to $y)$. Note that this is an equivalence relation on V.

Next, we wish to define the size, or cardinality, of a set x to be a set $\operatorname{card}(x)$ such that the following holds.

$$(\dagger) \qquad (\forall\, x)(\forall\, y)(\operatorname{card}(x) = \operatorname{card}(y) \Leftrightarrow x \equiv y)$$

One obvious choice for $\operatorname{card}(x)$ would be the $\equiv$-equivalence class $\{y : y \equiv x\}$ of x. However, this is always a proper class (except when $x = \emptyset$ whose $\equiv$-equivalence class is the set $\{\emptyset\}$). Another possibility is to choose a particular representative of the $\equiv$-equivalence class of x to be $\operatorname{card}(x)$. This can be done if we assume the Axiom of Choice. It turns out that something along the lines of the first possibility can be made work in ZF (this is a trick due to D. S. Scott).

Definitions. Let x be a set. In ZFC we define the *cardinality* $\operatorname{card}(x)$ of x to be the least ordinal α such that $x \equiv \alpha$. Note that such ordinals exist since in ZFC x can be well-ordered.

In ZF we first define the *essential rank* $\operatorname{ess\,rank}(x)$ of x to be the least ordinal α such that there exists a set y with $\operatorname{rank}(y) = \alpha$ and $y \equiv x$. We then define

$$\operatorname{card}(x) = \{y \in V_{\operatorname{ess\,rank}(x)+1} : y \equiv x\}$$

which is a set by Separation.

Note. In the rest of this chapter we will work in ZFC, so we could adopt the first definition of cardinality. However, the exact definition does not matter that much. What is important is property $(\dagger)$. Also, much of what we do below is valid in ZF.

Definition. Say that a set m is a *cardinal* if $m = \operatorname{card}(x)$ for some set x. In this case we say m is the *cardinality of* x.

Before discussing the arithmetic of cardinals, we introduce initial ordinals and the alephs.

The Alephs

Definition. Say $\alpha \in \mathrm{ON}$ is an *initial ordinal* if $(\forall\, \beta \in \mathrm{ON})(\beta < \alpha \Rightarrow \neg(\beta \equiv \alpha))$.

Examples. For every set x, the Hartogs' ordinal $\gamma(x)$ is an initial ordinal. Since for $n < \omega$ we have $\gamma(n) = n^+$ (easy ω-induction), it follows that all members of ω are initial ordinals, which in turn implies that ω is an initial ordinal.

The ordinals ω^2, ω^3 or $\varepsilon_0 = \omega^{\omega^{\omega^{\cdot^{\cdot^{\cdot}}}}}$ are not initial ordinals as they all biject with ω. In fact, the next initial ordinal after ω is $\gamma(\omega) = \omega_1$. More generally, we can index the infinite initial ordinals as follows.

Definition. Define ω_α for $\alpha \in \mathrm{ON}$ by recursion:

$$\begin{aligned} \omega_0 &= \omega \\ \omega_{\beta^+} &= \gamma(\omega_\beta) \\ \omega_\lambda &= \sup\{\omega_\beta : \beta < \lambda\} \qquad (\lambda \text{ non-zero limit}) \end{aligned}$$

Proposition 1. The ordinals ω_α are exactly the infinite initial ordinals.

Remark. Note that for ordinals $\alpha < \beta$, if β injects into α, then by the Schröder–Bernstein theorem we have $\alpha \equiv \beta$. It follows that if $\alpha < \beta$ and β is an initial ordinal, then β cannot inject into α. We shall use this simple observation several times below.

Proof. We first show that the ω_α are initial ordinals by induction on α. We only need to check the case when α is a non-zero limit. In this case, assume that $\omega_\alpha \equiv \gamma$ for some $\gamma < \omega_\alpha$. Then $\gamma < \omega_\beta$ for some $\beta < \alpha$. Since $\omega_\beta < \omega_\alpha$ (easy induction on α), it follows that ω_β injects into γ contradicting the induction hypothesis that ω_β is an initial ordinal.

Now assume that δ is an infinite initial ordinal. An easy induction shows that $\alpha \leqslant \omega_\alpha$ for all $\alpha \in \mathrm{ON}$, and hence there is a least α with $\delta < \omega_\alpha$. Since δ is infinite, $\alpha \neq 0$, and moreover α cannot be a limit otherwise $\delta < \omega_\beta$ for some $\beta < \alpha$ contradicting the minimality of α. Thus $\alpha = \beta^+$ for some β that satisfies $\omega_\beta \leqslant \delta < \omega_{\beta^+} = \gamma(\omega_\beta)$. It follows that δ injects into ω_β, and thus $\delta = \omega_\beta$ as δ is an initial ordinal. □

Notation. For $\alpha \in \mathrm{ON}$ we denote by $\aleph_\alpha$ ('aleph-α') the cardinality of ω_α. By Proposition 1 the alephs are the cardinalities of all infinite sets. (This is true in ZFC. In ZF the alephs are the cardinalities of all infinite well-ordered sets.)

The arithmetic of cardinals

We use the letter m, n, p for cardinals, and M, N, P for sets with cardinalities m, n, p, respectively.

Definition. Write $m \leqslant n$ if M injects into N, and $m < n$ if $m \leqslant n$ and $m \neq n$.

Note. These are well defined, *i.e.*, do not depend on the choice of sets M, N. It is also easy to check that $\leqslant$ is a partial order on the class of cardinals. Antisymmetry ($m \leqslant n$ and $n \leqslant m$ imply $m = n$) follows from Schröder–Bernstein. In ZFC it is even a linear order.

Definition. We define cardinal addition, multiplication and exponentiation as follows.

$$\begin{aligned} m + n &= \mathrm{card}(M \sqcup N) \\ m \cdot n &= \mathrm{card}(M \times N) \\ m^n &= \mathrm{card}(M^N) \qquad (M^N = \{f \in \mathbb{P}(N \times M) : f \colon N \to M\}) \end{aligned}$$

These operations are well-defined, *i.e.*, they do not depend on the choice of M, N.

Properties. The following are straightforward to check by writing down a bijection between appropriate sets.

$$\begin{aligned} m+n &= n+m \\ (m+n)+p &= m+(n+p) \\ m\cdot n &= n\cdot m \\ (mn)p &= m(np) \\ m(n+p) &= mn+mp \\ (mn)^p &= m^p n^p \\ m^{n+p} &= m^n m^p \\ \big(m^n\big)^p &= m^{np} \end{aligned}$$

In addition, if $m \leqslant n$, then $m+p \leqslant n+p$, $mp \leqslant np$ and $m^p \leqslant n^p$. This is again easy to verify by writing down appropriate injections.

Note. Cantor's diagonal argument shows that $m < 2^m$ for all cardinals m (there is no surjection $M \to 2^M$). In particular, $\aleph_0 < 2^{\aleph_0}$ which contrasts with $\omega = 2^\omega$ for ordinal exponentiation.

Similarly, we have $2 \cdot \aleph_0 = \aleph_0 \cdot 2$ in contrast with $2 \cdot \omega = \omega \neq \omega \cdot 2$ for ordinal multiplication.

A consequence of the next result is that addition and multiplication of alephs is easy.

Theorem 2. $\aleph_\alpha \cdot \aleph_\alpha = \aleph_\alpha$ for all $\alpha \in \mathrm{ON}$.

Proof. We proceed by induction on α. The case $\alpha = 0$ is clear since $\omega \times \omega \equiv \omega$. Now let $\alpha > 0$ and assume that $\omega_\beta \times \omega_\beta \equiv \omega_\beta$ for all $\beta < \alpha$.

Well-order $\omega_\alpha \times \omega_\alpha$ by 'going up in squares': $(x, y) < (w, z)$ if and only if

$$\begin{array}{lllll} \text{either} & \max\{x,y\} < \max\{w,z\} & & & \\ \text{or} & \max\{x,y\} = \max\{w,z\} = \delta, \text{ say, and} & \text{either} & y < z < \delta \\ & & \text{or} & y < \delta = z \\ & & \text{or} & x < w,\ y = z = \delta \end{array}$$

Given $\delta \in \omega_\alpha \times \omega_\alpha$, the proper initial segment I_δ is contained in $\beta \times \beta$ for some $\beta < \omega_\alpha$. (*E.g.,* if $\delta = (x, y)$, then $\beta = \max\{x, y\}^+$ will do.) Then $\mathrm{card}(\beta) < \mathrm{card}(\omega_\alpha)$ since ω_α is initial. So by induction hypothesis, either β is finite, or $\beta \times \beta \equiv \beta$. It follows that $\mathrm{card}(I_\delta) \leqslant \mathrm{card}(\beta \times \beta) < \mathrm{card}(\omega_\alpha)$.

The above shows that every proper initial segment of $\omega_\alpha \times \omega_\alpha$ has order-type $< \omega_\alpha$, and hence $\omega_\alpha \times \omega_\alpha$ has order-type $\leqslant \omega_\alpha$. It follows that $\omega_\alpha \times \omega_\alpha$ injects into ω_α, and so $\aleph_\alpha \cdot \aleph_\alpha \leqslant \aleph_\alpha$.

Since $\aleph_\alpha = \aleph_\alpha \cdot 1 \leqslant \aleph_\alpha \cdot \aleph_\alpha$, the result follows. □

Corollary 3. Let $\alpha \leqslant \beta$ be ordinals. Then $\aleph_\alpha + \aleph_\beta = \aleph_\alpha \cdot \aleph_\beta = \aleph_\beta$.

Proof. $\aleph_\beta \leqslant \aleph_\alpha + \aleph_\beta \leqslant \aleph_\beta \cdot 2 \leqslant \aleph_\beta \cdot \aleph_\beta = \aleph_\beta$. □

Note. In ZFC one can define more general infinite sums and products of cardinals. In the definitions below, as earlier, lower-case letters denote cardinals and upper-case letters denote sets with cardinality the corresponding lower-case letter.

Definitions. Let I be a set, and for each $i \in I$ let m_i be a cardinal. Then

$$\sum_{i\in I} m_i = \operatorname{card}\Big(\bigsqcup_{i\in I} M_i\Big) \quad \text{and} \quad \prod_{i\in I} m_i = \operatorname{card}\Big(\prod_{i\in I} M_i\Big)$$

where $\bigsqcup_{i\in I} M_i = \bigcup_{i\in I} M_i \times \{i\}$ and

$$\prod_{i\in I} M_i = \Big\{ f \colon I \to \textstyle\bigcup_{i\in I} M_i : \ f(i) \in M_i \text{ for all } i \in I \Big\} \ .$$

Note. We need AC in these definitions twice. Firstly, we need to make a choice of sets M_i with cardinality m_i. Secondly, when we show that these definitions don't depend on the choice of the M_i, we need to make a choice of bijections $f_i \colon M_i \to M_i'$ where M_i' is another set with cardinality m_i.

Example. It is possible to show results similar to Theorem 3. For example, if $m_i \leqslant \aleph_\alpha$ for all $i \in I$ and $\operatorname{card}(I) \leqslant \aleph_\alpha$, then $\sum_{i\in I} m_i \leqslant \aleph_\alpha$.

Infinite products of cardinals relate to cardinal exponentiation which is hard. We can achieve some reduction in the problem of studying cardinal exponentiation. For example, if $\alpha \leqslant \beta$, then

$$2^{\aleph_\beta} \leqslant \aleph_\alpha^{\aleph_\beta} \leqslant \big(2^{\aleph_\alpha}\big)^{\aleph_\beta} = 2^{\aleph_\alpha \cdot \aleph_\beta} = 2^{\aleph_\beta}$$

So it is of interest to study cardinals of the form $2^{\aleph_\beta}$. We know that $\aleph_\beta < 2^{\aleph_\beta}$ but very little else is known. For example, a natural question is whether $2^{\aleph_0}$ is equal to $\aleph_1$. Since $2^{\aleph_0}$ is the cardinality of $\mathbb{R}$, this became known as the *Continuum Hypothesis* (or CH for short):

$$\text{(CH)} \qquad 2^{\aleph_0} = \aleph_1$$

P. Cohen proved in the 1960s that if ZFC is consistent, then so are ZFC+CH and ZFC+¬CH. So CH is independent of ZFC.

7 *Classical descriptive set theory*

Note. The material in this chapter is non-examinable. We study 'definable' sets in Polish spaces: Borel sets and projective sets (see definitions below). We aim to cover enough material to show the existence of analytic non-Borel sets (a sort of superficial analogue of P≠NP) and that the Continuum Hypothesis holds for analytic sets.

Polish spaces

Definition. A *Polish space* is a separable, complete metrizable topological space.

Examples. The most important example for us is *Baire space* $\mathcal{N} = \mathbb{N}^{\mathbb{N}}$, the space of sequences in $\mathbb{N}$ with the product topology: open sets are unions of basic open sets which are of the form

$$\mathcal{U}_{m_1,\dots,m_k} = \{\mathbf{n} = (n_i)_{i\in\mathbb{N}} \in \mathcal{N} : \ n_i = m_i \text{ for } 1 \leqslant i \leqslant k\}$$

for any finite sequence $m_1, \dots, m_k$ in $\mathbb{N}$. For $\mathbf{m} \neq \mathbf{n}$ in $\mathcal{N}$, let $k \in \mathbb{N}$ be minimal with $m_k \neq n_k$, and set $d(\mathbf{m}, \mathbf{n}) = 1/k$. It is easy to check that d is a complete metric on $\mathcal{N}$ inducing the product topology (the open balls are exactly the basic open sets). Moreover, the set of eventually constant sequences is a countable dense subset of $\mathcal{N}$. Thus Baire space is indeed a Polish space.

Another important example is the space $\{0, 1\}^{\mathbb{N}}$ which is compact in the product topology and can be viewed as a subspace of $\mathcal{N}$ in the obvious way. Further examples are euclidean spaces $\mathbb{R}^d$ and, more generally, separable Banach spaces and closed subsets thereof.

Lemma 1. Every (non-empty) Polish space X is the continuous image of Baire space.

Proof. By separability, we can write X as a union $\bigcup_{m\in\mathbb{N}} U_m$ of non-empty open sets each of diameter at most 1. In turn, each U_m can be written as a union $\bigcup_{n\in\mathbb{N}} U_{m,n}$ of non-empty open sets each of diameter at most 1/2. Continuing this way, we find a family of non-empty open sets $U_{m_1,\dots,m_k}$ of diameter at most $1/k$ indexed by finite sequences in $\mathbb{N}$ such that $U_{m_1,\dots,m_{k-1}} = \bigcup_{m_k\in\mathbb{N}} U_{m_1,\dots,m_k}$ for all $k, m_1, \dots, m_{k-1} \in \mathbb{N}$.

Next, fix an element $x_{m_1,\dots,m_k}$ of $U_{m_1,\dots,m_k}$ for each finite sequence $m_1, \dots, m_k$ in $\mathbb{N}$. Define $\varphi \colon \mathcal{N} \to X$ by $\varphi(\mathbf{n}) = \lim_{k\to\infty} x_{n_1,\dots,n_k}$ for $\mathbf{n} = (n_i)_{i\in\mathbb{N}} \in \mathcal{N}$. It is straightforward to verify that φ is continuous and surjective. □

Lemma 2. $\mathcal{N}$ is homeomorphic to the set of irrationals in $[0, 1]$.

Proof. Use continued fractions: Define

$$\varphi(\mathbf{n}) = \cfrac{1}{n_1 + \frac{1}{n_2 + \frac{1}{n_3 + \dots}}}$$

for $\mathbf{n} = (n_i)_{i\in\mathbb{N}} \in \mathcal{N}$. □

Borel hierarchy

Definitions. Let X be an arbitrary set. A *σ-field* (or *σ-algebra*) on X is a subset $\mathcal{F}$ of the power set $\mathbb{P}X$ such that

(i) $\emptyset \in \mathcal{F}$

(ii) $A_1, A_2, \dots \in \mathcal{F}$ implies $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{F}$

(iii) $A \in \mathcal{F}$ implies $X \setminus A \in \mathcal{F}$

Note that in particular a σ-field is closed under countable intersections (as well as countable unions).

Now assume that X is a Polish space. The *Borel σ-field $\mathcal{B}$* on X is the smallest σ-field on X containing all the open sets. (Equivalently, $\mathcal{B}$ is the intersection of all σ-fields on X that contain the open sets; there exists at least one such σ-field, namely $\mathbb{P}X$.) Members of $\mathcal{B}$ are called *Borel sets*.

Borel hierarchy. For a Polish space X, we define families Σ^0_α and Π^0_α of subsets of X for ordinals $1 \leqslant \alpha < \omega_1$ by recursion as follows.

$$\begin{aligned}
\Sigma^0_1 &= \{U \subset X : U \text{ open}\}\\
\Pi^0_1 &= \{F \subset X : F \text{ closed}\}\\
\Sigma^0_{\alpha+1} &= \{\textstyle\bigcup_{n\in\mathbb{N}} A_n : A_n \in \Pi^0_\alpha \text{ for all } n \in \mathbb{N}\}\\
\Pi^0_{\alpha+1} &= \{X \setminus A : A \in \Sigma^0_{\alpha+1}\}\\
\Sigma^0_\lambda &= \{\textstyle\bigcup_{n\in\mathbb{N}} A_n : \forall n \in \mathbb{N}\ \exists \alpha < \lambda\ A_n \in \Pi^0_\alpha\}\\
\Pi^0_\lambda &= \{X \setminus A : A \in \Sigma^0_\lambda\}
\end{aligned}$$

where in the last two lines λ is a non-zero limit. The collections of these families is the *Borel hierarchy* of X.

We define $\Delta^0_\alpha = \Sigma^0_\alpha \cap \Pi^0_\alpha$ for $1 \leqslant \alpha < \omega_1$.

Example. Σ^0_2 is the family of countable unions of closed sets known as F_σ-sets. Π^0_2 is the family of countable intersections of open sets known as G_δ-sets.

Remark. Any open set in a Polish space (or indeed in any metric space) is a countable union of closed sets, *i.e.*, $\Sigma^0_1 \subset \Sigma^0_2$. An easy induction then shows that $\Sigma^0_\alpha \subset \Delta^0_\beta$ and $\Pi^0_\alpha \subset \Delta^0_\beta$ for $1 \leqslant \alpha < \beta < \omega_1$.

$$\begin{tikzcd}
 & \Sigma^0_1 \arrow[dr, phantom, "\subset" sloped] & & \Sigma^0_2 \arrow[dr, phantom, "\subset" sloped] & & \Sigma^0_3 \arrow[dr, phantom, "\subset" sloped] & \\
\Delta^0_1 \arrow[ur, phantom, "\subset" sloped] & & \Delta^0_2 \arrow[ur, phantom, "\subset" sloped] & & \Delta^0_3 \arrow[ur, phantom, "\subset" sloped] & & \cdots \\
 & \Pi^0_1 \arrow[ul, phantom, "\supset" sloped] \arrow[ur, phantom, "\subset" sloped] & & \Pi^0_2 \arrow[ul, phantom, "\supset" sloped] \arrow[ur, phantom, "\subset" sloped] & & \Pi^0_3 \arrow[ul, phantom, "\supset" sloped] \arrow[ur, phantom, "\subset" sloped] &
\end{tikzcd}$$

Lemma 3. $\displaystyle\bigcup_{1\leqslant\alpha<\omega_1} \Sigma^0_\alpha = \bigcup_{1\leqslant\alpha<\omega_1} \Pi^0_\alpha = \mathcal{B}$ in any Polish space.

Proof. The first equality follows from the inclusions above. For the second equality, first show by induction that $\Sigma^0_\alpha \subset \mathcal{B}$ for all $1 \leqslant \alpha < \omega_1$, and then show that $\bigcup_{1\leqslant\alpha<\omega_1} \Sigma^0_\alpha$ is a σ-field containing the open sets. □

Definition. A subset $A \subset \mathcal{N} \times \mathcal{N}$ is a *universal* Σ^0_α*-set* if

(i) A is Σ^0_α, and

(ii) if $B \subset \mathcal{N}$ is Σ^0_α, then $B = \{\mathbf{n} \in \mathcal{N} : (\mathbf{m}, \mathbf{n}) \in A\}$ for some $\mathbf{m} \in \mathcal{N}$.

Theorem 4. For every $1 \leqslant \alpha < \omega$, there exists a universal Σ^0_α-set.

Proof. We first show that there exists a universal open set. Enumerate the basic open sets (recall that these are indexed by finite sequences of positive integers) as $U_1, U_2, U_3, \ldots$. Let

$$A = \{(\mathbf{m}, \mathbf{n}) \in \mathcal{N} \times \mathcal{N} : \exists\, i \in \mathbb{N}\ \mathbf{n} \in U_{m_i}\}$$

It is easy to check that A is open.

An open set $B \subset \mathcal{N}$ can be written as a union of basic open sets: $B = \bigcup_{i\in\mathbb{N}} U_{m_i}$ for some $\mathbf{m} \in \mathcal{N}$. Then

$$\mathbf{n} \in B \quad \Leftrightarrow \quad \exists\, i \in \mathbb{N}\ \mathbf{n} \in U_{m_i} \quad \Leftrightarrow \quad (\mathbf{m}, \mathbf{n}) \in A$$

and thus $B = \{\mathbf{n} \in \mathcal{N} : (\mathbf{m}, \mathbf{n}) \in A\}$.

So far we have established the theorem for $\alpha = 1$. One can prove the general statement by induction on α. This is left as an exercise. □

Corollary 5. For each $1 \leqslant \alpha < \omega_1$ there is a Σ^0_α-subset of $\mathcal{N}$ that is not Π^0_α.

Remark. This leads to the following refinement of the picture of the Borel hierarchy of $\mathcal{N}$.

$$\begin{tikzcd}
 & \Sigma^0_1 \arrow[dr, hook, "/" marking] & & \Sigma^0_2 \arrow[dr, hook, "/" marking] & & \Sigma^0_3 \arrow[dr, hook, "/" marking] & \\
\Delta^0_1 \arrow[ur, hook, "/" marking] \arrow[dr, hook, "/" marking] & & \Delta^0_2 \arrow[ur, hook, "/" marking] \arrow[dr, hook, "/" marking] & & \Delta^0_3 \arrow[ur, hook, "/" marking] \arrow[dr, hook, "/" marking] & & \cdots \\
 & \Pi^0_1 \arrow[ur, hook, "/" marking] & & \Pi^0_2 \arrow[ur, hook, "/" marking] & & \Pi^0_3 \arrow[ur, hook, "/" marking] &
\end{tikzcd}$$

Proof. Fix a universal Σ^0_α-set $A \subset \mathcal{N} \times \mathcal{N}$. Then $B = \{\mathbf{n} \in \mathcal{N} : (\mathbf{n}, \mathbf{n}) \in A\}$ is Σ^0_α since $\mathbf{n} \mapsto (\mathbf{n}, \mathbf{n})$ is continuous. If B is Π^0_α, then $B = \{\mathbf{n} \in \mathcal{N} : (\mathbf{m}, \mathbf{n}) \notin A\}$ for some $\mathbf{m} \in \mathcal{N}$. Considering whether $\mathbf{m} \in B$ leads to a contradiction. □

Projective hierarchy

Definition. A subset of a Polish space is *analytic* if it is a continuous image of $\mathcal{N}$.

Examples. It follows from Lemma 1 that any Polish space, and thus any closed subset of a Polish space, is analytic.

Remark. We shall often use implicitly the following observation. The spaces $\mathbb{N} \times \mathcal{N} = \mathbb{N}^{\{0\} \sqcup \mathbb{N}}$, $\mathcal{N} \times \mathcal{N} = \mathbb{N}^{\mathbb{N} \sqcup \mathbb{N}}$ and $\mathcal{N}^{\mathbb{N}} = \mathbb{N}^{\mathbb{N} \times \mathbb{N}}$ are homeomorphic to $\mathcal{N}$ in the obvious way.

Proposition 6. Let X be a Polish space and $A \subset X$. Then TFAE.

(i) A is analytic.

(ii) A is the continuous image of a Borel subset of some Polish space.

(iii) A is the projection onto X of a Borel subset of $Y \times X$ (for some Polish space Y).

(iv) A is the projection onto X of a closed subset of $Y \times X$ (for some Polish space Y).

(v) A is the projection onto X of a Borel subset of $\mathcal{N} \times X$.

(vi) A is the projection onto X of a closed subset of $\mathcal{N} \times X$.

Proof. Together with the trivial implications, showing (ii)⇒(i)⇒(vi) will complete the proof.

(i)⇒(vi): Let $f\colon \mathcal{N} \to X$ be continuous with $f(\mathcal{N}) = A$. Then A is the projection onto X of the graph $\{(\mathbf{n}, f(\mathbf{n})) : \mathbf{n} \in \mathcal{N}\}$ of f, which is closed since f is continuous and X is Hausdorff.

(ii)⇒(i): We need to show that every Borel set is analytic. Since closed sets, *i.e.*, Π^0_1-sets, are analytic, it is enough to show that countable unions and intersections of analytic sets are analytic. Indeed, an easy induction then shows that Σ^0_α-sets and Π^0_α-sets are analytic for all $1 \leqslant \alpha < \omega_1$, and the result follows.

For each $k \in \mathbb{N}$ let $A_k \subset X$ be an analytic set, and let F_k be a closed subset of $\mathcal{N} \times X$ such that A_k is the projection onto X of F_k. Then

$$\begin{aligned} x \in \bigcup_{k \in \mathbb{N}} A_k \quad &\Leftrightarrow \quad \exists\, k \in \mathbb{N}\ \exists\, \mathbf{n} \in \mathcal{N}\ (\mathbf{n}, x) \in F_k \\ &\Leftrightarrow \quad \exists\, (k, \mathbf{n}) \in \mathbb{N} \times \mathcal{N}\ (\mathbf{n}, x) \in F_k\ , \end{aligned}$$

and thus $\bigcup_{k \in \mathbb{N}} A_k$ is the projection onto X of the closed subset

$$\{(k, \mathbf{n}, x) \in \mathbb{N} \times \mathcal{N} \times X : (\mathbf{n}, x) \in F_k\}$$

of $\mathbb{N} \times \mathcal{N} \times X$. Similarly, we have

$$\begin{aligned} x \in \bigcap_{k \in \mathbb{N}} A_k \quad &\Leftrightarrow \quad \forall\, k \in \mathbb{N}\ \exists\, \mathbf{n} \in \mathcal{N}\ (\mathbf{n}, x) \in F_k \\ &\Leftrightarrow \quad \exists\, (\mathbf{n}^{(k)})_{k \in \mathbb{N}} \in \mathcal{N}^{\mathbb{N}}\ \forall\, k \in \mathbb{N}\ (\mathbf{n}^{(k)}, x) \in F_k \end{aligned}$$

and thus $\bigcap_{k \in \mathbb{N}} A_k$ is the projection onto X of the closed subset

$$\bigcap_{k \in \mathbb{N}} \{(\mathbf{n}^{(1)}, \mathbf{n}^{(2)}, \dots, x) \in \mathcal{N}^{\mathbb{N}} \times X : (\mathbf{n}^{(k)}, x) \in F_k\}$$

of $\mathcal{N}^{\mathbb{N}} \times X$. □

Definition. We define Σ^1_1 to be the family of analytic sets (in some Polish space) and Π^1_1 to be the family of complements of analytic sets called *coanalytic* sets. Then inductively, for $1 \leqslant n < \omega$, we define Σ^1_{n+1} to be the family of continuous images of Π^1_n-sets, and Π^1_{n+1} to be the family of complements of Σ^1_{n+1}-sets. We also let $\Delta^1_n = \Sigma^1_n \cap \Pi^1_n$ for $1 \leqslant n < \omega$.

Note. It follows from Proposition 6 that $\mathcal{B} \subset \Delta^1_1$. Then an easy induction establishes the following inclusions.

$$\begin{tikzcd}
 & \Sigma^1_1 \arrow[dr, phantom, sloped, "\subset"] & & \Sigma^1_2 \arrow[dr, phantom, sloped, "\subset"] & & \Sigma^1_3 \arrow[dr, phantom, sloped, "\subset"] & \\
\Delta^1_1 \arrow[ur, phantom, sloped, "\subset"] \arrow[dr, phantom, sloped, "\subset"] & & \Delta^1_2 \arrow[ur, phantom, sloped, "\subset"] \arrow[dr, phantom, sloped, "\subset"] & & \Delta^1_3 \arrow[ur, phantom, sloped, "\subset"] \arrow[dr, phantom, sloped, "\subset"] & & \cdots \\
 & \Pi^1_1 \arrow[ur, phantom, sloped, "\subset"] & & \Pi^1_2 \arrow[ur, phantom, sloped, "\subset"] & & \Pi^1_3 \arrow[ur, phantom, sloped, "\subset"] &
\end{tikzcd}$$

It follows that $\bigcup_{1\leqslant n<\omega} \Sigma^1_n = \bigcup_{1\leqslant n<\omega} \Pi^1_n$; we denote the common union by $\mathcal{P}$.

Definition. The collection of families Σ^1_n and Π^1_n, $1 \leqslant n < \omega$, is called the *projective hierarchy.* Members of $\mathcal{P}$ are the *projective sets.*

Theorem 7. There exists a universal analytic set $A \subset \mathcal{N} \times \mathcal{N}$.

Proof. Let $U \subset \mathcal{N} \times (\mathcal{N} \times \mathcal{N})$ be a universal open set: If $V \subset \mathcal{N} \times \mathcal{N}$ is open, then $V = \{(\mathbf{m}, \mathbf{n}) \in \mathcal{N} \times \mathcal{N} : (\mathbf{p}, \mathbf{m}, \mathbf{n}) \in U\}$ for some $\mathbf{p} \in \mathcal{N}$.

Let $B \subset \mathcal{N}$ be analytic. Then there is a closed set $F \subset \mathcal{N} \times \mathcal{N}$ such that

$$B = \{\mathbf{n} \in \mathcal{N} : \exists\, \mathbf{m} \in \mathcal{N}\ (\mathbf{m}, \mathbf{n}) \in F\}$$

and by the choice of U above, there exists $\mathbf{p} \in \mathcal{N}$ such that

$$B = \{\mathbf{n} \in \mathcal{N} : \exists\, \mathbf{m} \in \mathcal{N}\ (\mathbf{p}, \mathbf{m}, \mathbf{n}) \notin U\}\ .$$

So if we set

$$A = \{(\mathbf{r}, \mathbf{s}) \in \mathcal{N} \times \mathcal{N} : \exists\, \mathbf{m} \in \mathcal{N}\ (\mathbf{r}, \mathbf{m}, \mathbf{s}) \notin U\}$$

then A is analytic (projection of closed set) and

$$B = \{\mathbf{n} \in \mathcal{N} : (\mathbf{p}, \mathbf{n}) \in A\}$$

which shows that A is universal. □

Corollary 8. There exists an analytic subset of $\mathcal{N}$ that is not coanalytic, *i.e.*, it belongs to $\Sigma^1_1 \setminus \Pi^1_1$.

Proof. Let $A \subset \mathcal{N} \times \mathcal{N}$ be a universal analytic set. Let

$$B = \{\mathbf{n} \in \mathcal{N} : (\mathbf{n}, \mathbf{n}) \in A\}\ .$$

Then B is analytic since $\mathbf{n} \mapsto (\mathbf{n}, \mathbf{n})$ is continuous.

Assume B is also coanalytic. Then $B = \{\mathbf{n} \in \mathcal{N} : (\mathbf{m}, \mathbf{n}) \notin A\}$ for some $\mathbf{m} \in \mathcal{N}$. We get a contradiction by considering whether $\mathbf{m} \in B$. □

Remark. We have already observed that Borel sets are both analytic and coanalytic. So the set B constructed in the proof above is analytic non-Borel. We will now show the converse that a set that is both analytic and coanalytic is Borel.

Definition. We say two subsets Y, Z of a Polish space X can be separated by Borel sets if there exist disjoint Borel sets B, C in X such that $Y \subset B$ and $Z \subset C$; equivalently, there exists a Borel set B in X such that $Y \subset B \subset X \setminus Z$.

Theorem 9. (Lusin's separation theorem) Disjoint analytic sets in a Polish space can be separated by Borel sets.

Proof. We first observe that given sets $Y = \bigcup_{n\in\mathbb{N}} Y_n$ and $Z = \bigcup_{n\in\mathbb{N}} Z_n$ in a Polish space, if Y_m and Z_n can be separated by Borel sets for all $m, n \in \mathbb{N}$, then Y and Z can also be separated by Borel sets. Indeed, for each $m, n \in \mathbb{N}$, choose a Borel set $B_{m,n}$ such that $Y_m \subset B_{m,n} \subset X \setminus Z_n$. Then the Borel set $B = \bigcup_{m\in\mathbb{N}} \bigcap_{n\in\mathbb{N}} B_{m,n}$ satisfies $Y \subset B \subset X \setminus Z$.

Now consider two disjoint analytic sets in a Polish space X. These have the form $f(\mathcal{N})$ and $g(\mathcal{N})$ where $f\colon \mathcal{N} \to X$ and $g\colon \mathcal{N} \to X$ are continuous with $f(\mathcal{N}) \cap g(\mathcal{N}) = \emptyset$. Recall that for any finite sequence $m_1, \dots, m_k$ of positive integers, there is a corresponding basic open set

$$\mathcal{U}_{m_1,\dots,m_k} = \{\mathbf{n} \in \mathcal{N} : n_i = m_i \text{ for } 1 \leqslant i \leqslant k\}$$

in $\mathcal{N}$. Now assume that $f(\mathcal{N})$ and $g(\mathcal{N})$ cannot be separated by Borel sets. Since $f(\mathcal{N}) = \bigcup_{n\in\mathbb{N}} f(\mathcal{U}_n)$ and $g(\mathcal{N}) = \bigcup_{n\in\mathbb{N}} g(\mathcal{U}_n)$, it follows by the initial observation that $f(\mathcal{U}_{m_1})$ and $g(\mathcal{U}_{n_1})$ cannot be separated by Borel sets for some $m_1, n_1 \in \mathbb{N}$. Repeatedly applying this reasoning, we obtain $\mathbf{m}, \mathbf{n} \in \mathcal{N}$ such that $f(\mathcal{U}_{m_1,\dots,m_k})$ and $g(\mathcal{U}_{n_1,\dots,n_k})$ cannot be separated by Borel sets for any $k \in \mathbb{N}$.

Since $f(\mathcal{N})$ and $g(\mathcal{N})$ are disjoint, it follows that $f(\mathbf{m}) \neq g(\mathbf{n})$, and hence $f(\mathbf{m})$ and $g(\mathbf{n})$ can be separated by disjoint open sets V, W (as X is Hausdorff). Hence $f^{-1}(V)$ and $g^{-1}(W)$ are disjoint open neighbourhoods of $\mathbf{m}$ and $\mathbf{n}$, respectively, and thus contain $\mathcal{U}_{m_1,\dots,m_k}$ and $\mathcal{U}_{n_1,\dots,n_k}$, respectively, for a sufficiently large $k \in \mathbb{N}$. It follows that $f(\mathcal{U}_{m_1,\dots,m_k}) \subset V$ and $g(\mathcal{U}_{n_1,\dots,n_k}) \subset W$ wich contradicts the choice of $\mathbf{m}$ and $\mathbf{n}$. □

Corollary 10. A subset of a Polish space is Borel if and only if it is both analytic and coanalytic. In symbols: $\Sigma_1^1 \cap \Pi_1^1 = \mathcal{B}$.

Remark. This completes the proof of the existence of an analytic non-Borel set. We now present a concrete example.

Example. Let Seq be the set of all finite sequences of positive integers. Then $\mathbb{P}\,\mathrm{Seq} = \{0,1\}^{\mathrm{Seq}}$ is a Polish space in the product topology (homeomorphic to $\{0,1\}^{\mathbb{N}}$ since Seq is countable).

Given $s = (m_1, \dots, m_k)$ and $t = (n_1, \dots, n_l)$ in Seq, write $s \prec t$ if $0 \leqslant k \leqslant l$ and $m_i = n_i$ for $1 \leqslant i \leqslant k$.

Say $T \subset \mathrm{Seq}$ is a *tree* if $s \in T$ whenever $s \prec t$ and $t \in T$. Say $\mathbf{n} \in \mathcal{N}$ is an *infinite branch* of T if $(n_1, \dots, n_k) \in T$ for all $k \in \mathbb{N}$. Say T is *well-founded* if T has no infinite branch.

Let $\mathcal{T}$ be the set of all trees and WFT be the set of all well-founded trees. Note that $\mathcal{T}$ is a closed subset of $\mathbb{P}\,\mathrm{Seq}$, and thus $\mathcal{T}$ is also a Polish space. We show that the subset WFT of $\mathcal{T}$ is coanalytic. For any $T \in \mathcal{T}$ we have

$$T \notin \mathrm{WFT} \quad \Leftrightarrow \quad \exists\, \mathbf{n} \in \mathcal{N}\ \forall\, k \in \mathbb{N}\ (n_1, \dots, n_k) \in T\ .$$

It follows that $\mathcal{T} \setminus \mathrm{WFT}$ is the projection onto $\mathcal{T}$ of the closed set

$$\bigcap_{k \in \mathbb{N}} \{(\mathbf{n}, T) \in \mathcal{N} \times \mathcal{T} :\ (n_1, \dots, n_k) \in T\}\ ,$$

and thus analytic. It is possible to show that WFT is not analytic, and hence WFT is a coanalytic non-Borel set.

Definition. A subset of a topological space is *perfect* if it is closed and contains no isolated points. (A point x in a subset A of a topological space X is *isolated in* A if there is an open neighbourhood U of x such that $U \cap A = \{x\}$.)

Lemma 11. A non-empty perfect subset of a Polish space has cardinality $2^{\aleph_0}$.

Proof. Let A be a non-empty perfect subset of a Polish space X. Given $x \in A$ and a radius $r > 0$, since x is not isolated in A, there exist $y, z \in A$ and a radius $s > 0$ such that the closed balls $B_s(y)$ and $B_s(z)$ are disjoint and contained in $B_r(x)$. (Note that we are implicitly assuming that X comes with a complete metric defining its topology.)

Since A is not empty, we can fix a point $x_\emptyset$ in A. Using the observation above, we inductively construct points $x_{\varepsilon_1, \dots, \varepsilon_k}$ in A indexed by finite sequences $\varepsilon_1, \dots, \varepsilon_k$ in $\{0, 1\}$ (where $\emptyset$ is the sequence of length one) and radii $r_1, r_2, \dots$ such that the closed ball $B_{r_k}(x_{\varepsilon_1, \dots, \varepsilon_k})$ contains the disjoint closed balls $B_{r_{k+1}}(x_{\varepsilon_1, \dots, \varepsilon_k, 0})$ and $B_{r_{k+1}}(x_{\varepsilon_1, \dots, \varepsilon_k, 1})$, and moreover $r_k \to 0$ as $k \to \infty$.

It is easy to verify that the function $\varphi\colon \{0, 1\}^{\mathbb{N}} \to A$ given by

$$\varphi(\varepsilon_1, \varepsilon_2, \dots) = \lim_{k \to \infty} x_{\varepsilon_1, \dots, \varepsilon_k}$$

is injective. It follows that $2^{\aleph_0} \leqslant \mathrm{card}(A)$.

Since X is the continuous image of $\mathcal{N}$, it follows that

$$\mathrm{card}(A) \leqslant \mathrm{card}(X) \leqslant \mathrm{card}(\mathcal{N}) = 2^{\aleph_0}\ ,$$

and hence $\mathrm{card}(A) = 2^{\aleph_0}$. □

Theorem 12. Every analytic set either has a perfect subset or is countable. It follows that every infinite analytic set has cardinality $\aleph_0$ or $2^{\aleph_0}$.

Proof. For a tree T let

$$[T] = \{\mathbf{n} \in \mathcal{N} :\ (n_1, \dots, n_k) \in T \text{ for all } k \in \mathbb{N}\}$$

be the set of all infinite branches of T. Note that for $T = \mathrm{Seq}$ we have $[T] = \mathcal{N}$, and for $T \in \mathrm{WFT}$ we have $[T] = \emptyset$. For a tree T and $s \in \mathrm{Seq}$, let

$$T(s) = \{t \in \mathrm{Seq} :\ t \prec s \text{ or } s \prec t\}\ .$$

Now fix an analytic set A in some Polish space X. Then $A = f(\mathcal{N}) = f([\mathrm{Seq}])$ for some continuous function $f\colon \mathcal{N} \to X$. For a tree T let

$$T' = \{s \in \mathrm{Seq} : \ f([T(s)]) \text{ is uncountable}\} \ .$$

Note that T' is a tree contained in T. Next, define trees $T^{(\alpha)}$ by recursion as follows.

$$\begin{aligned} T^{(0)} &= \mathrm{Seq} & & \\ T^{(\alpha)} &= \big(T^{(\beta)}\big)' & & \text{if } \alpha = \beta^+ \\ T^{(\alpha)} &= \bigcap_{\beta<\alpha} T^{(\beta)} & & \text{if } \alpha \text{ is a non-zero limit} \end{aligned}$$

Since Seq is countable, there exists $\alpha < \omega_1$ such that $T^{(\alpha+1)} = T^{(\alpha)}$. Set $T = T^{(\alpha)}$ and consider the following two cases.

If $T = \emptyset$, then

$$A = \bigcup_{\beta<\alpha} \Big(f\big([T^\beta]\big) \setminus f\big([T^{\beta+1}]\big)\Big)$$

and

$$f\big([T^\beta]\big) \setminus f\big([T^{\beta+1}]\big) = \bigcup \Big\{ f\big([T^\beta(s)]\big) : \ f\big([T^\beta(s)]\big) \text{ is countable}\Big\}$$

which implies that A is countable.

Now consider the case when $T \neq \emptyset$. Given $s \in T$, since $s \in T'$, it follows in particular that $f([T(s)])$ has at least two distinct elements $f(\mathbf{m})$ and $f(\mathbf{n})$. Then by the continuity of f, there exists $k \in \mathbb{N}$ such that $f([T(m_1,\dots,m_k)]) \cap f([T(n_1,\dots,n_k)]) = \emptyset$. We have thus shown that for all $s \in T$ there exist $t, u \in T$ such that $s \prec t$, $s \prec u$ and $f([T(t)]) \cap f([T(u)]) = \emptyset$.

Using the above observation, we can construct elements $s_{\varepsilon_1,\dots,\varepsilon_k}$ of T for all finite sequence $\varepsilon_1,\dots,\varepsilon_k$ in $\{0,1\}$ such that for all $k \in \mathbb{N}$, we have $s_{\varepsilon_1,\dots,\varepsilon_k} \prec s_{\varepsilon_1,\dots,\varepsilon_k,0}$, $s_{\varepsilon_1,\dots,\varepsilon_k} \prec s_{\varepsilon_1,\dots,\varepsilon_k,1}$ and $f([T(s_{\varepsilon_1,\dots,\varepsilon_k,0})]) \cap f([T(s_{\varepsilon_1,\dots,\varepsilon_k,1})]) = \emptyset$.

Finally, let

$$\mathcal{M} = \{\mathbf{n} \in \mathcal{N} : \ \exists\,(\varepsilon_i) \in \{0,1\}^{\mathbb{N}} \text{ such that } s_{\varepsilon_1,\dots,\varepsilon_k} \prec \mathbf{n} \text{ for all } k \in \mathbb{N}\} \ .$$

One can show that $\mathcal{M}$ is compact and $f(\mathcal{M})$ is a perfect set. This is left as an exercise. □