Part II — Logic and Set Theory

Based on lectures by Dr Zsak and notes by thirdsgames.co.uk

Lent 2023

Contents

§1 Propositional Logic

We build a language consisting of statements/propositions;
We will assign truth values to statements;
We build a deduction system so that we can prove statements that are true (and only those). These are also features of more complicated languages.

§1.1 Languages

Let P be a set of **primitive propositions**. Unless otherwise stated, we let $P = \{p_1, p_2, \dots\}$ (i.e. countable). The **language** $L = L(P)$ is a set of **propositions** (or **compound propositions**) and is defined inductively by

1. if $p \in P$, then $p \in L$;
2. $\bot \in L$, where the symbol $\bot$ is read 'false'/ 'bottom';
3. if $p, q \in L$, then $(p \Rightarrow q) \in L$.

Example 1.1
$((p_1 \Rightarrow p_2) \Rightarrow (p_1 \Rightarrow p_3)) \in L$. $(p_4 \Rightarrow \bot) \in L$.

If $p \in L$ then $((p \Rightarrow \bot) \Rightarrow \bot) \in L$.

Remark 1. Note that the phrase 'L is defined inductively' means more precisely the following. Let $L_1 = P \cup \{\bot\}$, and define $L_{n+1} = L_n \cup \{(p \Rightarrow q) : p, q \in L_n\}$. We set $L = \bigcup_{n=1}^{\infty} L_n$.

Note that the elements of L are just finite strings of symbols from the alphabet $P \cup \{(,), \Rightarrow, \bot\}$. Brackets are only given for clarity; we omit those that are unnecessary, and may use other types of brackets such as square brackets.
We can prove that L is the smallest (w.r.t. inclusion) subset of the set Σ of all finite strings in $P \cup \{(,), \Rightarrow, \bot\}$ s.t. the properties of a language hold.
Note that $L \subsetneq \Sigma$. E.g. $\Rightarrow p_1 p_3 (\in \Sigma \setminus L$.

Every $p \in L$ is uniquely determined by the properties of a language above, i.e. either $p \in P$ or $p = \bot$ or $\exists$ unique $q, r \in L$ s.t. $p = (q \Rightarrow r)$.

We can now introduce the abbreviations $\neg, \wedge, \vee, \top$, which are not, and, or and true/top respectively, defined by

Notation.

$$\neg p = (p \Rightarrow \bot); \quad p \vee q = \neg p \Rightarrow q; \quad p \wedge q = \neg(p \Rightarrow \neg q), \top = (\bot \Rightarrow \bot)$$

§1.2 Semantic implication

Definition 1.1 (Valuation)

A **valuation** is a function $v\colon L \to \{0,1\}$ s.t.

1. $v(\bot) = 0$;
2. If $p, q \in L$ then $v(p \Rightarrow q) = \begin{cases} 0 & v(p) = 1 \text{ and } v(q) = 0 \\ 1 & \text{else} \end{cases}$

Example 1.2

If $v(p_1) = 1$, $v(p_2) = 0$. Then

$$v\left(\underbrace{(\bot \Rightarrow p_1)}_{1} \Rightarrow \underbrace{(p_1 \Rightarrow p_2)}_{0}\right) = 0$$

Remark 2. On $\{0,1\}$, we can define the constant $\bot = 0$ and the operation $\Rightarrow$ in the obvious way. Then, a valuation is precisely a mapping $L \to \{0,1\}$ preserving all structure, so it can be considered a homomorphism.

Proposition 1.1

Let $v, v'\colon L \to \{0,1\}$ be valuations that agree on the primitives p_i. Then $v = v'$. Further, any function $w\colon P \to \{0,1\}$ extends to a valuation $v : L \to \{0,1\}$ s.t. $v|_P = w$.

Remark 3. This is analogous to the definition of a linear map by its action on the basis vectors.

Proof. Clearly, v, v' agree on L_1 as $v(\bot) = v'(\bot) = 0$, the set of elements of the language of length 1. If v, v' agree at $p, q \in L_n$, then they agree at $p \Rightarrow q$. So by induction, v, v' agree on L_{n+1} for all n, and hence on L.

Let $v(p) = w(p)$ for all $p \in P$, and $v(\bot) = 0$ to obtain v on the set L_1. Assuming v is defined on $p, q \in L_n$ we can define it at $p \Rightarrow q$ in the obvious way. This defines v on L_{n+1}, hence v is defined on $\cup L_n = L$. By construction, v is a valuation on L and $v|_P = w$. ☐

Example 1.3

Let v be the valuation with $v(p_1) = v(p_3) = 1$, and $v(p_n) = 0$ for all $n \neq 1, 3$. Then, $v((p_1 \Rightarrow p_3) \Rightarrow p_2) = 0$.

Definition 1.2 (Tautology)

A **tautology** is $t \in L$ s.t. $v(t) = 1$ $\forall$ valuations v. We write $\models t$.

Example 1.4

$p \Rightarrow (q \Rightarrow p)$ (a true statement is implied by any true statement).

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

Since the right-hand column is always 1, $\models p \Rightarrow (q \Rightarrow p)$.

Example 1.5 (Law of Excluded Middle)

$\neg\neg p \Rightarrow p$, which expands to $((p \Rightarrow \bot) \Rightarrow \bot) \Rightarrow p$.

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

Hence $\models \neg\neg p \Rightarrow p$.

Example 1.6

$\neg p \vee p$, which expands to $((p \Rightarrow \bot) \vee p)$.

$v(p)$	$v(\neg p)$	$v(\neg p \vee p)$
0	1	1
1	0	1

Hence $\models \neg p \vee p$.

Example 1.7

$(p \Rightarrow (q \Rightarrow r)) \Rightarrow ((p \Rightarrow q) \Rightarrow (p \Rightarrow r))$. Suppose this is not a tautology. Then we have a valuation v s.t. $v(p \Rightarrow (q \Rightarrow r)) = 1$ and $v((p \Rightarrow q) \Rightarrow (p \Rightarrow r)) = 0$. Hence, $v(p \Rightarrow q) = 1, v(p \Rightarrow r) = 0$, so $v(p) = 1, v(r) = 0$, giving $v(q) = 1$, but then $v(p \Rightarrow (q \Rightarrow r)) = 0$ contradicting the assumption.

Definition 1.3 (Semantic Implication)
Let $S \subseteq L$ and $t \in L$. We say S **entails** or **semantically implies** t, written $S \models t$, if for every valuation v on L, $v(s) = 1\ \forall s \in S \Rightarrow v(t) = 1$.

Example 1.8
$\{p, p \Rightarrow q\} \models q$.

Example 1.9
Let $S = \{p \Rightarrow q, q \Rightarrow r\}$, and let $t = p \Rightarrow r$. Suppose $S \not\models t$, so there is a valuation v s.t. $v(p \Rightarrow q) = 1, v(q \Rightarrow r) = 1, v(p \Rightarrow r) = 0$. Then $v(p) = 1, v(r) = 0$, so $v(q) = 1$ and $v(q) = 0$ ↯.

Definition 1.4 (Model)
Given $t \in L$, say a valuation **v is a model for t** (or **t is true in v**) if $v(t) = 1$.

Definition 1.5 (Model)
We say that **v is a model of S** in L if $v(s) = 1$ for all $s \in S$.

Thus, $S \models t$ is the statement that every model of S is also a model of t/ t is true in every model of S.

Remark 4. The notation $\models t$ is equivalent to $\varnothing \models t$.

§1.3 Syntactic implication

For a notion of proof, we require a system of axioms and deduction rules. As axioms, we take (for any $p, q, r \in L$),

1. $p \Rightarrow (q \Rightarrow p)$;
2. $(p \Rightarrow (q \Rightarrow r)) \Rightarrow ((p \Rightarrow q) \Rightarrow (p \Rightarrow r))$;
3. $((p \Rightarrow \perp) \Rightarrow \perp) \Rightarrow p$.

Remark 5. Sometimes, these three axioms are considered axiom **schemes**, since they are really a different axiom for each $p, q, r \in L$.

These are all tautologies.

For deduction rules, we will have only the rule **modus ponens (MP)**, that from p and $p \Rightarrow q$ one can deduce q.

Definition 1.6 (Proof)

Let $S \subseteq L$, $t \in L$. A **proof of t from S** is a finite sequence $t_1, \ldots, t_n$ of propositions in L s.t. $t_n = t$ and every t_i is either

1. an axiom;
2. an element of S (t_i is a premise or hypothesis); or
3. follows by MP, where $t_j = p$ and $t_k = p \Rightarrow q$ where $j, k < i$.

We say that S is the set of **premises** or **hypotheses**, and t is the **conclusion**.

We say S **proves** or **syntactically implies** t, written $S \vdash t$, if there exists a proof of t from S.

Example 1.10

We will show $\{p \Rightarrow q, q \Rightarrow r\} \vdash (p \Rightarrow r)$.

1. $q \Rightarrow r$ (hypothesis)
2. $(q \Rightarrow r) \Rightarrow (p \Rightarrow (q \Rightarrow r))$ (axiom 1)
3. $p \Rightarrow (q \Rightarrow r)$ (modus ponens on lines 1, 2)
4. $(p \Rightarrow (q \Rightarrow r)) \Rightarrow ((p \Rightarrow q) \Rightarrow (p \Rightarrow r))$ (axiom 2)
5. $(p \Rightarrow q) \Rightarrow (p \Rightarrow r)$ (modus ponens on lines 3, 4)
6. $p \Rightarrow q$ (hypothesis)
7. $p \Rightarrow r$ (modus ponens on lines 5, 6)

Definition 1.7 (Theorem)

If $\varnothing \vdash t$, we say t is a **theorem**, written $\vdash t$.

Example 1.11

$\vdash (p \Rightarrow p)$.

1. $(p \Rightarrow ((p \Rightarrow p) \Rightarrow p)) \Rightarrow ((p \Rightarrow (p \Rightarrow p)) \Rightarrow (p \Rightarrow p))$ (axiom 2)
2. $p \Rightarrow ((p \Rightarrow p) \Rightarrow p)$ (axiom 1)
3. $(p \Rightarrow (p \Rightarrow p)) \Rightarrow (p \Rightarrow p)$ (modus ponens on lines 1, 2)

4. $p \Rightarrow (p \Rightarrow p)$ (axiom 1)
5. $p \Rightarrow p$ (modus ponens on lines 3, 4)

§1.4 Deduction theorem

Theorem 1.1 (Deduction Theorem)
Let $S \subseteq L$, and $p, q \in L$. Then $S \vdash (p \Rightarrow q)$ iff $S \cup \{p\} \vdash q$.

Remark 6. This show '$\Rightarrow$' really does behave like implication in formal proofs.

Proof. ($\Rightarrow$): Given a proof of $p \Rightarrow q$ from S, add the line p to the hypothesis and deduce q from modus ponens, to obtain a proof of q from $S \cup \{p\}$.

($\Leftarrow$): Suppose we have a proof of q from $S \cup \{p\}$. Let $t_1, \dots, t_n$ be the lines of the proof. We will prove that $S \vdash (p \Rightarrow t_i)$ for all i by induction.

- If t_i is an axiom, we write t_i (axiom); $t_i \Rightarrow (p \Rightarrow t_i)$ (axiom 1); $p \Rightarrow t_i$ (modus ponens).
- If $t_i \in S$, we write t_i (hypothesis); $t_i \Rightarrow (p \Rightarrow t_i)$ (axiom 1); $p \Rightarrow t_i$ (modus ponens).
- If $t_i = p$, we write the proof of $\vdash p \Rightarrow p$ given above.
- Suppose t_i is obtained by modus ponens from t_j and $t_k = t_j \Rightarrow t_i$ where $j, k < i$. We may assume by induction that $S \vdash p \Rightarrow t_j$ and $S \vdash p \Rightarrow (t_j \Rightarrow t_i)$. We write
    1. $(p \Rightarrow (t_j \Rightarrow t_i)) \Rightarrow ((p \Rightarrow t_j) \Rightarrow (p \Rightarrow t_i))$ (axiom 2)
    2. $(p \Rightarrow t_j) \Rightarrow (p \Rightarrow t_i)$ (modus ponens)
    3. $p \Rightarrow t_i$ (modus ponens)

  giving $S \vdash p \Rightarrow t_i$.

□

Example 1.12
Consider $\{p \Rightarrow q, q \Rightarrow r\} \vdash p \Rightarrow r$. By the Deduction Theorem, it suffices to prove $\{p \Rightarrow q, q \Rightarrow r, p\} \vdash r$, which is obtained easily from modus ponens.

§1.5 Soundness

We aim to show $S \models t$ iff $S \vdash t$. The direction $S \vdash t$ implies $S \models t$ is called **soundness**, which is a way of verifying that our axioms and deduction rule make sense. The direction $S \models t$ implies $S \vdash t$ is called **adequacy**, which states that our axioms are powerful enough to deduce everything that is (semantically) true.

Proposition 1.2 (Soundness Theorem)
Let $S \subseteq L$ and $t \in L$. Then $S \vdash t$ implies $S \models t$.

Proof. We have a proof $t_1, \dots, t_n$ of t from S. We aim to show that any model of S is also a model of t, so if v is a valuation that maps every element of S to 1, then $v(t) = 1$.

We show this by induction on the length of the proof. $v(p) = 1$ for each axiom p (as axioms are tautologies) and for each $p \in S$. Further, $v(t_i) = 1, v(t_i \Rightarrow t_j) = 1$, then $v(t_j) = 1$. Therefore, $v(t_i) = 1$ for all i. □

§1.6 Adequacy

Consider the case of adequacy where $t = \bot$. If our axioms are adequate, $S \models \bot$ implies $S \vdash \bot$. We say S is **consistent** if $S \nvdash \bot$ and **inconsistent** if $S \vdash \bot$. Therefore, in an adequate system, if S has no models then S is inconsistent; equivalently, if S is consistent then it has a model.

In fact, the statement that consistent axiom sets have a model implies adequacy in general. Indeed, if $S \models t$, then $S \cup \{\neg t\}$ has no models, and so it is inconsistent by assumption. Then $S \cup \{\neg t\} \vdash \bot$, so $S \vdash \neg t \Rightarrow \bot$ by the deduction theorem, giving $S \vdash t$ by axiom 3.

We aim to construct a model of S given that S is consistent. Intuitively, we want to write

$$v(t) = \begin{cases} 1 & t \in S \\ 0 & t \notin S \end{cases}$$

but this does not work on the set $S = \{p_1, p_1 \Rightarrow p_2\}$ as it would evaluate p_2 to false.

We say a set $S \subseteq L$ is **deductively closed** if $p \in S$ whenever $S \vdash p$. Any set S has a **deductive closure**, which is the (deductively closed) set of statements $\{t \in L : S \vdash t\}$ that S proves. If S is consistent, then the deductive closure is also consistent. Computing the deductive closure before the valuation solves the problem for $S = \{p_1, p_1 \Rightarrow p_2\}$. However, if a primitive proposition p is not in S, but $\neg p$ is also not in S, this technique still does not work, as it would assign false to both p and $\neg p$.

Theorem 1.2 (Model Existence Lemma)

Every consistent set $S \subseteq L$ has a model.

Remark 7. We use the fact that P is a countable set in order to show that L is countable. The result does in fact hold if P is uncountable, but requires Zorn's Lemma and will be proved in Chapter 3. Some sources call this theorem the 'completeness theorem'.

Proof. First, we claim that for any consistent $S \subseteq L$ and proposition $p \in L$, either $S \cup \{p\}$ is consistent or $S \cup \{\neg p\}$ is consistent. If this were not the case, then $S \cup \{p\} \vdash \bot$, and also $S \cup \{\neg p\} \vdash \bot$. By the deduction theorem, $S \vdash p \Rightarrow \bot$ and $S \vdash (\neg p) \Rightarrow \bot$. But then $S \vdash \neg p$ and $S \vdash \neg\neg p$, so $S \vdash \bot$ contradicting consistency of S.

Now, L is a countable set as each L_n is countable, so we can enumerate L as $t_1, t_2, \dots$. Let $S_0 = S$, and define $S_1 = S_0 \cup \{t_1\}$ or $S_1 = S_0 \cup \{\neg t_1\}$, chosen s.t. S_1 is consistent. Continuing inductively, define $\overline{S} = \bigcup_i S_i$.
Then, $\forall t \in L$, either $t \in \overline{S}$ or $\neg t \in \overline{S}$.
Note that $\overline{S}$ is consistent since proofs are finite; indeed, if $\overline{S} \vdash \bot$, then this proof uses hypotheses only in S_n for some n, but then $S_n \vdash \bot$ contradicting consistency of S_n.
Note also that $\overline{S}$ is deductively closed, so if $\overline{S} \vdash p$, we must have $p \in \overline{S}$; otherwise, $\neg p \in \overline{S}$ so $\overline{S} \vdash \neg p$, giving $\overline{S} \vdash \bot$ by MP, contradicting consistency of $\overline{S}$.

Now, define the function

$$v(t) = \begin{cases} 1 & t \in \overline{S} \\ 0 & t \notin \overline{S} \end{cases}$$

We show that v is a valuation, then the proof is complete as $v(s) = 1$ for all $s \in S$. Since $\overline{S}$ is consistent, $\bot \notin \overline{S}$, so $v(\bot) = 0$.

Suppose $v(p) = 1, v(q) = 0$. Then $p \in \overline{S}$ and $q \notin \overline{S}$, and we want to show $(p \Rightarrow q) \notin \overline{S}$. If this were not the case, we would have $(p \Rightarrow q) \in \overline{S}$ and $p \in \overline{S}$, so $q \in \overline{S}$ as $\overline{S}$ is deductively closed.

Now suppose $v(q) = 1$, so $q \in \overline{S}$, and we need to show $(p \Rightarrow q) \in \overline{S}$. Then $\overline{S} \vdash q$, and by axiom 1, $\overline{S} \vdash q \Rightarrow (p \Rightarrow q)$. Therefore, as $\overline{S}$ is deductively closed, $(p \Rightarrow q) \in \overline{S}$.

Finally, suppose $v(p) = 0$, so $p \notin \overline{S}$, and we want to show $(p \Rightarrow q) \in \overline{S}$. We know that $\neg p \in \overline{S}$, so it suffices to show that $(p \Rightarrow \bot) \vdash (p \Rightarrow q)$. By the deduction theorem, this is equivalent to proving $\{p, p \Rightarrow \bot\} \vdash q$, or equivalently, $\bot \vdash q$. But by axiom 1, $\bot \Rightarrow (\neg q \Rightarrow \bot)$ where $(\neg q \Rightarrow \bot) = \neg\neg q$, so the proof is complete by axiom 3. □

Corollary 1.1 (Adequacy)
Let $S \subseteq L$ and let $t \in L$, s.t. $S \models t$. Then $S \vdash t$.

Proof. $S \cup \{\neg t\} \models \bot$, so Model Existence Lemma, $S \cup \{\neg t\} \vdash \bot$. Then by Deduction Theorem $S \vdash \neg\neg t$. $\neg\neg t \Rightarrow t$ by Axiom 3 and so by MP $S \vdash t$. □

§1.7 Completeness

Theorem 1.3 (Completeness Theorem for Propositional Logic)
Let $S \subseteq L$ and $t \in L$. Then $S \models t$ iff $S \vdash t$.

Proof. Follows from soundness and adequacy. □

Theorem 1.4 (Compactness Theorem)
Let $S \subseteq L$ and $t \in L$ with $S \models t$. Then there exists a finite subset $S' \subseteq S$ s.t. $S' \models t$.

Proof. Trivial after applying the completeness theorem, since proofs depend on only finitely many hypotheses in S. □

Corollary 1.2 (Compactness Theorem, Equivalent Form)
Let $S \subseteq L$. Then if every finite subset $S' \subseteq S$ has a model, then S has a model.

Proof. Let $t = \bot$ in the compactness theorem. Then, if $S \models \bot$, some finite $S' \subseteq S$ has $S' \models \bot$. But this is not true by assumption, so there is a model for S. □

Remark 8. This corollary is equivalent to the more general compactness theorem, since the assertion that $S \models t$ is equivalent to the statement that $S \cup \{\neg t\}$ has no model, and $S' \models t$ is equivalent to the statement that $S' \cup \{\neg t\}$ has no model.

Note. The use of the word compactness is more than a fanciful analogy. See Sheet 1.

Theorem 1.5 (Decidability Theorem)
Let $S \subseteq L$, S finite and $t \in L$. Then, there is an algorithm to decide (in finite time) if $S \vdash t$.

Proof. Trivial after replacing $\vdash$ with $\models$, and checking all valuations by drawing the relevant truth tables. $\square$

§2 Well-Orderings

§2.1 Definition

Definition 2.1 (Linear Order)
A **linear order** or **total order** is a pair $(X, <)$ where X is a set, and $<$ is a relation on X s.t.

- (irreflexivity) $\forall x \in X, \neg(x < x)$;
- (transitivity) $\forall x, y, z \in X, (x < y \wedge y < z) \Rightarrow (x < z)$;
- (trichotomy) $\forall x, y \in X$, either $x < y$, $y < x$, or $x = y$.

We say X is linearly/totally ordered by $<$, or simply say X is a linearly/totally ordered set.

Note. In trichotomy, exactly one holds, e.g. if $x < y$ and $y < x$, then $x < x$ by transitivity contradicting irreflexivity.

If X is linearly ordered by $<$, we use the obvious notation $x > y$ to denote $y < x$. In terms of the $\leq$ relation, we can equivalently write the axioms of a linear order as

- (reflexivity) $\forall x \in X, x \leq x$;
- (transitivity) $\forall x, y, z \in X, (x \leq y \wedge y \leq z) \Rightarrow (x \leq z)$;
- (antisymmetry) $\forall x, y \in X$, if $(x \leq y \wedge y \leq x) \Rightarrow (x = y)$.
- (trichotomy, or totality) $\forall x, y \in X$, either $x \leq y$ or $y \leq x$.

Example 2.1

1. $(\mathbb{N}, \leq)$ is a linear order.
2. $(\mathbb{Q}, \leq)$ is a linear order.
3. $(\mathbb{R}, \leq)$ is a linear order.
4. $(\mathbb{N}^+, |)$ is not a linear order, where $|$ is the divides relation, since 2 and 3 are not related.
5. $(\mathcal{P}(S), \subseteq)$ is not a linear order if $|S| > 1$, since it fails trichotomy.

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

Definition 2.2 (Well-Ordering)

A linear order $(X, <)$ is a **well-ordering** if every nonempty subset $S \subseteq X$ has a least element.

$$\forall S \subseteq X,\ S \neq \varnothing \Rightarrow \exists x \in S,\ \forall y \in S,\ x \leq y$$

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

Note. This least element is unique by antisymmetry.

Example 2.2

1. $(\mathbb{N}, <)$ is a well-ordering.
2. $(\mathbb{Z}, <)$ is not a well-ordering, since $\mathbb{Z}$ has no least element.
3. $(\mathbb{Q}, <)$ is not a well-ordering.
4. $(\mathbb{R}, <)$ is not a well-ordering.
5. $[0, 1] \subset \mathbb{R}$ with the usual order is not a well-ordering, since $(0, 1]$ has no least element.
6. $\left\{\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \dots\right\} \subset \mathbb{R}$ with the usual order is a well-ordering.
7. $\left\{\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \dots\right\} \cup \{1\}$ with the usual order is also a well-ordering.
8. $\left\{\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \dots\right\} \cup \{2\}$ with the usual order is another example.
9. $\left\{\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \dots\right\} \cup \left\{1 + \frac{1}{2}, 1 + \frac{2}{3}, 1 + \frac{3}{4}, \dots\right\}$ is another example.

Note. Every subset of a well-ordered set is well-ordered.

Remark 9. Let $(X, <)$ be a linear order. $(X, <)$ is a well-ordering iff there is no infinite decreasing sequence $x_1 > x_2 > \dots$. Indeed, if $(X, <)$ is a well-ordering, then the set $\{x_1, x_2, \dots\}$ has no minimal element, contradicting the assumption. Conversely, if $S \subseteq X$ has no minimal element, then we can construct an infinite decreasing sequence by arbitrarily choosing points $x_1 > x_2 > \dots$ in S, which exists as S has no minimal element.

Definition 2.3 (Order-Isomorphism)

Linear ordered sets X, Y are **order-isomorphic** if there $\exists$ bijection $f : X \to Y$ which is **order-preserving**: $\forall x < y$ in X, $f(x) < f(y)$. Such an f is an **order-isomorphism** and f^{-1} is also an order-isomorphism.

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

Examples (1) and (6) are isomorphic, and (7) and (8) are isomorphic. Examples (1) and (7) are not isomorphic, since example (7) has a greatest element and (1) does not. Example (9) is not isomorphic to (6) or (7).

Example 2.3

1. $\mathbb{N}, \mathbb{Q}$ are not order-isomorphic.
2. $\mathbb{Q}, \mathbb{Q} \setminus \{0\}$ are.

Definition 2.4 (Initial Segment)

A subset I of a linearly ordered set X is an **initial segment** (i.s.) if $x \in I$ implies $y \in I$ for all $y < x$.

Example 2.4

$\{1, 2, 3, 4\}$ is an i.s. of $\mathbb{N}$. $\{1, 2, 3, 5\}$ is not.

Remark 10. In any linear ordering X and element $x \in X$, the set $\{y : y < x\}$ is an initial segment by transitivity.

Not every initial segment is of this form, for instance $\{x : x \leq 3\}$ in $\mathbb{R}$.

Remark 11. In a well-ordering, every proper initial segment $I \neq X$ is of this form. Indeed, letting $I_x = \{y : y < x\}$ where x is the least element of $X \setminus I$ we see $I_x = I$.
If $y \in I_x$ then $y < x$ so $y \in I$ by choice of x, i.e. $I_x \subseteq I$. If $y \in I$ and $y \geq x$, then $x \in I$ as I is an i.s. $\lightning$ so $y < x$, i.e. $y \in I_x$ and $I \subseteq I_x$.

Lemma 2.1

Let X, Y be well-ordered sets, I an i.s. of Y and $f : X \to I$ be an order-isomorphism. Then $\forall x \in X$, $f(x)$ is the least element of $Y \setminus \{f(t) : t < x\}$.

Proof. The set $A = Y \setminus \{f(t) : t < x\}$ is non-empty, e.g. $f(x) \in A$. Let a be the least element of A. Then $a \leq f(x)$ and $f(x) \in I$ and so $a \in I$. Thus $a = f(z)$ for some $z \in X$. Note that $z > x$ implies that $a = f(z) > f(x)$ $\lightning$, so $z \leq x$. If $z < x$ then $a = f(z) \in \{f(t) : t < x\}$ $\lightning$ as $a \in A$. So $z = x$ and $a = f(z) = f(x)$. ∎

Proposition 2.1 (Proof by Induction)
Let X be a well-ordered set, and let $S \subseteq X$ be s.t. for every $x \in X$

$$(\forall y < x,\ y \in S) \Rightarrow x \in S^{a}$$

Then $S = X$.

[a] If $y \in S$ for all $y < x$, then $x \in S$

Remark 12. Equivalently, if $p(x)$ is a property s.t. if $p(y)$ is true for all $y < x$ then $p(x)$, then $p(x)$ holds for all x.

Formally, if S is given by a property p, $S = \{x \in X : p(x)\}$.
$(\forall x \in X)((\forall y < x, p(y)) \Rightarrow p(x)) \Rightarrow (\forall x \in X, p(x))$ (base case is included).

Proof. Suppose $S \neq X$. Then $X \setminus S$ is nonempty, and therefore has a least element x. But all elements $y < x$ lie in S as o/w they are the least element, and so by the property of S, we must have $x \in S$, contradicting the assumption. □

Proposition 2.2
Let X, Y be order-isomorphic well-orderings. Then there is exactly one order-isomorphism between X and Y.

Note that this does not hold for general linear orderings, such as $\mathbb{Q}$ to itself or $[0,1]$ to itself by $x \mapsto x$ or $x \mapsto x^2$.

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

Suppose $f(y) = g(y)$ for all $y < x$. By lemma 2.1, $f(x)$ is the least element of $A = Y \setminus \{f(y) : y < x\}$, and $g(x)$ the least element of $B = Y \setminus \{g(y) : y < x\}$. By the induction hypothesis, $A = B$ and hence $f(x) = g(x)$. □

Remark 13. Induction proves things. We need a tool to construct things.

§2.2 Initial segments

Note. A function from a set X to a set Y is a subset of f of $X \times Y$ s.t.

1. $\forall x \in X\ \exists y \in Y\ (x, y) \in f$;
2. $\forall x \in X\ \forall y, z \in Y\ ((x, y) \in f \wedge (x, z) \in f) \Rightarrow (y = z)$.

Of course we write $y = f(x)$ instead of $(x, y) \in f$. Note that $f \in \mathcal{P}(X \times Y)$.

For $Z \subseteq X$, the restriction of f to Z is $f|_Z = \{(x, y) \in f; x \in Z\}$. $f|_Z$ is a fcn $Z \to Y$, so $f|_Z \subseteq Z \times Y \subseteq X \times Y$ so $f_Z \in \mathcal{P}(Z \times Y)$.

Theorem 2.1 (Definition by Recursion)

Let X be a w.o. set and Y be any set. Then for any fcn $G \colon \mathcal{P}(X \times Y) \to Y$ there's a unique fcn $f : X \to Y$ s.t. $f(x) = G(f|_{I_x})$ for every $x \in X$.

Remark 14. What this means in defining $f(x)$, we may use the value of $f(y)$ for all $y < x$.

Proof. We say that h is an **attempt** to mean that $h \colon I \to Y$ where I is some i.s. of X, s.t. $\forall x \in I,\ h(x) = G\big(h|_{I_x}\big)$ (note $I_x \subseteq I$).

Let h, h' be attempts. We show that $\forall x \in X$ if $x \in \operatorname{dom}(h)$[a] $\cap \operatorname{dom}(h')$ then $h(x) = h'(x)$. Fix $x \in \operatorname{dom}(h) \cap \operatorname{dom}(h')$ and assume $h(y) = h'(y)$ for every $y < x$ (note $y < x$ implies $y \in \operatorname{dom}(h) \cap \operatorname{dom}(h')$). Then $h|_{I_x} = h'|_{I_x}$ so $h(x) = G(h|_{I_x}) = G(h'|_{I_x}) = h'(x)$. Done by induction.

Now we need to show that $\forall x \in X\ \exists$ attempt h s.t. $x \in \operatorname{dom}(h)$. We prove this by induction. Fix $x \in X$ and assume that for $y < x$ there's an attempt defined at y, and let h_y be the unique attempt with domain $\{z \in X : z \leq y\} = I_y \cup \{y\}$. Then $h = \bigcup_{y<x} h_y$ is a well defined fcn on I_x and it is an attempt since for $y < x$, $h(y) = h_y(y) = G(h_y|_{I_y}) = G(h|_{I_y})$.

The attempt $h' = h \cup \{(x, G(h))\}$ is an attempt with domain $I_x \cup \{x\}$. Therefore, there is an attempt defined at each x, so we can define $f \colon X \to Y$ by $f(x) = h(x)$ where h is some attempt defined at x. This is well defined by above and $f(x) = h(x) = G(h|_{I_x}) = G(f|_{I_x})$. □

[a] $\operatorname{dom}(h)$ is the domain of h, i.e. I above.

Proposition 2.3 (Subset Collapse)

Let Y be a w.o. set where $X \subseteq Y$. Then X is order-isomorphic to a unique initial segment of Y.

This is not true for general linear orderings, such as $\{1, 2, 3\} \subset \mathbb{Z}$, or $\mathbb{Q}$ in $\mathbb{R}$.

Proof. WLOG $X \neq \varnothing$.

Uniquness: Assume $f : X \to I$ is an o.i. where I is an i.s. of Y. By lemma 2.1, $f(x) = \min(Y \setminus \{f(y) : y < x, y \in X\})$. So by induction, f and hence I are uniquely determined.

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

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

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

Given $y < x$ in X, since

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

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 15. A w.o. set X cannot be isomorphic to a proper i.s. by uniqueness as it is isomorphic to itself.

§2.3 Relating well-orderings

Definition 2.5 (Less than or equal)

For well-ordered sets X, Y, we will write $X \leq Y$ if X is o.i. to an i.s. of Y.

$X \leq Y$ iff X is o.i. to some subset of Y.

Example 2.5

$\mathbb{N} \leq \left\{\frac{1}{2}, \frac{2}{3}, \dots\right\}$.

Proposition 2.4

Let X, Y be well-ordered sets. Then either $X \leq Y$ or $Y \leq X$.

Proof. Assume $Y \not\leq X$. Then in particular, $Y \neq \varnothing$. Fix $y_0 \in Y$ and define by recursion $f\colon X \to Y$ by

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

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 occur. It follows as in the proof of Subset Collapse that f is an o.i. from I_x to Y, so $Y \leq X$ $\lightning$.

Hence, the 'otherwise' clause never arises, and so it follows as in the proof of Subset Collapse that f is an o.i. from X to an i.s. of Y. □

Proposition 2.5

Let X, Y be well-ordered sets s.t. $X \leq Y$ and $Y \leq X$. Then X is o.i. to Y.

Proof. Let $f\colon X \to Y$ and $g\colon Y \to X$ be o.i.s to i.s. of Y and X respectively. Then $g \circ f$ is an o.i. from X to some i.s. of X. So by uniqueness in Subset Collapse, $g \circ f = \mathrm{id}|_X$. Similarly, $f \circ g = \mathrm{id}|_Y$, so f and g are inverses. □

Remark 16. This shows that $\leq$ is a linear-order (reflexive, antisymmetric, transitive and trichotomous) on the collection of well-ordered sets provided we identified w.o. sets that are o.i. to each other.

§2.4 Constructing larger well-orderings

Definition 2.6 (Less than)

For w.o. sets X, Y, we write $X < Y$ if $X \leq Y$ and X not o.i. to Y.

So $X < Y \iff X$ o.i. to a proper i.s. of Y.

Question

Do the w.o. sets form a set? If so, is it a w.o. set?

Answer

First we construct new w.o. sets from old. "There is always another": Let X be w.o. and let $x_0 \notin X$.
$X^+ = X \cup \{x_0\}$ is w.o. by setting $x < x_0$ for all $x \in X$. This is unique up to o.i. and $X < X^+$.

Upper Bounds: Given set $\{X_i : i \in I\}$ of w.o. sets. We seek a w.o. set X s.t. $X_i \leq X \ \forall i \in I$.

Definition 2.7 (Extends)

For well-orderings $(X, <_X), (Y, <_Y)$, we say that $(Y, <_Y)$ **extends** $(X, <_X)$ if $X \subseteq Y$, $<_Y|_X = <_X$, and X is an i.s. of Y.
Then $\{X_i : i \in I\}$ is **nested** if $\forall i, j \in I$ either X_i extends X_j or X_j extends X_i.

Proposition 2.6

Let $\{X_i : i \in I\}$ be a nested set of w.o. sets. Then, $\exists$ w.o. set X s.t. $X_i \leq X \ \forall i \in I$.

Proof. Let $X = \bigcup_{i \in I} X_i$ with $x < y$ iff $\exists i \in I$ s.t. $x, y \in X_i$ and $x <_i y$ where $<_i$ the well-ordering of X_i. Since the X_i's are nested, this is a well-defined linear order s.t. each X_i is an i.s. of X.

We show that this is a well-ordering. Let $S \subseteq X$ be a nonempty set. Since $S = \bigcup_{i \in I}(S \cap X_i)$, $\exists i \in I$ s.t. $S \cap X_i \neq \varnothing$. Let x be a least element of $S \cap X_i$ (since X_i is w.o.). Then x is a least element of S since X_i is an i.s. and if $y < x$, $y \in X_i$. □

Remark 17. The proposition holds without the nestedness assumption (see Section 5).

§2.5 Ordinals

Definition 2.8 (Ordinal)

An **ordinal** is a w.o. set, where we regard two ordinals as equal if they are o.i.

Remark 18. We cannot construct ordinals as equivalence classes of well-orderings, due to Russell's paradox. Later, we will see a different construction that deals with this problem in Section 5.

Definition 2.9 (Order Type)

The **order type** of a w.o. set X is the unique ordinal α o.i. to X. Let X be a well-ordering corresponding to an ordinal α.

Notation. Write "α is the O.T. of X".

Example 2.6

For $k \in \mathbb{N}_0$, we let k be the O.T. of a w.o. set of size k (this is unique).
Let ω be the O.T. of $\mathbb{N}$ (also of $\mathbb{N}_0$).

Example 2.7

In the reals, the set $\{-2, 3, -\pi, 5\}$ has order type 4. The set $\left\{\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \dots\right\}$ has order type ω.

Note. For ordinals α, β write $\alpha \leq \beta$ if $X \leq Y$ where X is a w.o. set with O.T. α and Y has O.T. β. This does not depend on the choice of representative X or Y.

We define $\alpha < \beta$ for $X < Y$.

Notation. Let α^+ be the O.T. of X^+.

Remark 19. Note that $\leq$ is a linear order; if $\alpha \leq \beta, \beta \leq \alpha$ then $\alpha = \beta$.

Theorem 2.2

Let α be an ordinal. Then the set of ordinals less than α form a w.o. set of O.T. α.

Proof. Let X be a w.o. set with O.T. α.

Then, w.o. sets less than X are the proper i.s. of X, up to o.i.. Let $\widetilde{X} = \{Y \subset X : Y \text{ a proper i.s. of } X\}$. Then $<$ (for w.o. sets) is a linear order on $\widetilde{X}$.

Note the fcn $X \to \widetilde{X}$ defined by $x \mapsto I_x$ is an o.i. So $\widetilde{X}$ is a w.o. set of O.T. α. So $\left\{\mathrm{O.T.}(Y) : Y \in \widetilde{X}\right\}$ is a set of ordinals $< \alpha$, and $Y \mapsto \mathrm{O.T.}(Y)$ is an o.i. from $\widetilde{X}$ to this set. ☐

Notation. We define $I_\alpha = \{\beta : \beta < \alpha\}$, which is a nice example of a w.o. set of O.T. α. This is often a convenient representative to choose for an ordinal.

Proposition 2.7

Every nonempty set S of ordinals has a least element.

Proof. Let $\alpha \in S$. Suppose α is not the least element of S. Then $S \cap I_\alpha$ is nonempty. But I_α is w.o., so $S \cap I_\alpha$ has a minimal element β. Then β is a least element of S, as if $\gamma \in S$ s.t. $\gamma < \alpha$, then $\gamma \in I_\alpha \cap S$ and so $\beta \leq \gamma$. ☐

Theorem 2.3 (Burali-Forti paradox)
The ordinals do not form a set.

Proof. Suppose X is the set of all ordinals. Then X is a w.o. by proposition 2.7, so it has an order type, say α. Then X is o.i. to I_α, which is a proper i.s. of X. $\lightning$ $\square$

Remark 20. Let $S = \{\alpha_i : i \in I\}$ be a set of ordinals. Then by proposition 2.6, the nested set $\{I_{\alpha_i} : i \in I\}$ has an upper bound. So $\exists$ ordinal α s.t. $\alpha_i \leq \alpha$ $\forall i \in I$. By theorem 2.2, I_α is w.o., so we can take the least such α:
Take the least element of $\{\beta \in I_\alpha \cup \{\alpha\} : \forall i \in I, \alpha_i \leq \beta\}$.
We denote by "$\sup S$" the **least upper bound on S**.

Note if $\alpha = \sup S$, then $I_\alpha = \cup_{i \in I} I_{\alpha_i}$.

Example 2.8
$\sup \{2, 4, 6, \dots\} = \omega$.

§2.6 Some ordinals

$$0, 1, 2, 3, \dots, \omega$$

Write $\alpha + 1$ for the successor α^+ of α.

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

where $\omega + \omega = \omega \cdot 2$ is defined by $\sup \{\omega + n : n < \omega\}$.

$$\omega \cdot 2 + 1, \omega \cdot 2 + 2, \dots, \omega \cdot 3, \omega \cdot 4, \omega \cdot 5, \dots, \omega \cdot \omega = \omega^2$$

where we define $\omega \cdot \omega = \sup \{\omega \cdot n : n < \omega\}$.

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

Continue in the same way.

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

where $\omega^3 = \sup \{\omega^2 \cdot n : n < \omega\}$.

$$\omega^3 + \omega^2 \cdot 7 + \omega \cdot 4 + 13, \dots, \omega^4, \omega^5, \dots, \omega^\omega$$

where $\omega^\omega = \sup \{\omega^n : n < \omega\}$.

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

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

where $\varepsilon_0 = \sup\{\omega, \omega^\omega, \omega^{\omega^\omega}, \dots\}$.

$$\varepsilon_0 + 1, \varepsilon_0 + \omega, \varepsilon_0 + \varepsilon_0 = \varepsilon_0 \cdot 2, \dots, \varepsilon_0^2, \varepsilon_0^3, \dots, \varepsilon_0^{\varepsilon_0}$$

where $\varepsilon_0^{\varepsilon_0} = \sup\{\varepsilon_0^\omega, \varepsilon_0^{\omega^\omega}, \dots\}$.

$$\varepsilon_0^{\varepsilon_0^{\varepsilon_0^{\cdot^{\cdot^{\cdot}}}}} = \varepsilon_1$$

All of these ordinals are countable, as each operation only takes a countable union of countable sets.

§2.7 Uncountable ordinals

Question

Can $\exists$ an uncountable ordinal/ w.o. set? Can we well order $\mathbb{R}$?

Answer

The reals cannot be explicitly well-ordered.

Theorem 2.4

There exists an uncountable ordinal.

Idea: Assume α an uncountable ordinal. Then there is a least such α: $\{\beta \in I_\alpha \cup \{\alpha\} : \beta \text{ uncountable}\} \neq \varnothing$, so has a least element, say γ. So I_γ is exactly the set of all countable ordinals, i.e. the set of order-types of well-orderings of subsets of $\mathbb{N}$.

If X is a countable w.o. set, then $\exists$ injection $f : X \to \mathbb{N}$. Then $Y = f(X)$ is w.o. by $f(x) < f(y) \iff x < y$ in X. Then Y is an o.i. to X.

Proof. Let $A = \{(Y, <) \in \mathcal{P}(\mathbb{N}) \times \mathcal{P}(\mathbb{N} \times \mathbb{N}) : Y \text{ is a w.o. by } <\}$ be the set of well-orderings of subsets of $\mathbb{N}$. Let $B = \{\text{O.T.}(Y, <) : (Y, <) \in A\}$. By above, B is exactly the set of all countable ordinals.

Let $\omega_1 = \sup B$. If $\omega_1 \in B$, i.e. ω_1 countable then so is ω_1^+ and then $\omega_1^+ \in B$. Thus $\omega_1 \leq \omega$ ↯ □

Remark 21. Without introducing A, it would be difficult to show that B was in fact a set.

Remark 22. Another ending to the proof above is as follows. B cannot be the set of all ordinals, since the ordinals do not form a set by the Burali-Forti paradox, so there exists an uncountable ordinal. In particular, there exists a least uncountable ordinal.

The ordinal ω_1 has a number of remarkable properties.

1. It is the least uncountable ordinal.
2. ω_1 is uncountable, but $\{\beta : \beta < \alpha\}$ is countable for all $\alpha < \omega_1$, i.e. every proper i.s. of ω_1 is countable.
3. There exists no sequence $\alpha_1, \alpha_2, \dots$ in I_{ω_1} with supremum ω_1, as $\sup \alpha_i$ is the O.T of $\bigcup_{i \in \mathbb{N}} I_{\alpha_i}$ which is countable.

Theorem 2.5 (Hartog's Lemma)
For every set X, $\exists$ an ordinal α that does not inject into X.

Proof. Repeat proof of theorem 2.4 with X instead of $\mathbb{N}$. Letting $\gamma = (\sup B)^+$, if there is an injection from γ to X, this induces a well-ordering on a subset of X with order type γ. Thus $\gamma \in B$, and so $\gamma \leq B < \gamma$ ↯. □

Remark 23. We write $\gamma(X)$ for the least ordinal that does not inject into X. For example $\gamma(\omega) = \omega_1$.

$0, 1, \dots, \omega, \dots, \varepsilon_0 = \omega^{\omega^{\omega^{\cdots}}}, \dots, \varepsilon_1, \dots, \varepsilon_{\varepsilon_\varepsilon}, \dots, \omega_1, \dots, \omega_1 \cdot 2, \dots, \omega_2 = \gamma(\omega_1), \dots$

§2.8 Successors and limits

Let α be an ordinal, consider whether α has a greatest element (i.e. if X has O.T. α, does X have a greatest element).

Definition 2.10 (Successor)
If $\exists$ greatest element of I_α, say β, then $I_\alpha = I_\beta \cup \{\beta\}$. So $\alpha = \beta^+$ and $\beta = \sup I_\alpha < \alpha$. We call such α a **successor**.

Else, $I_\alpha = \sup I_\alpha$. i.e. $\alpha = \sup\{\beta : \beta < \alpha\}$. Say α is a **limit**.

Example 2.9
$1 = 0^+$ is a successor. 5 is a successor. $\omega + 2 = (\omega^+)^+$ is a successor. $\omega = \sup\{n : n < \omega\}$ is a limit as it has no greatest element. ω_1 is a limit. 0 is a limit.

§2.9 Ordinal arithmetic

Let α, β be ordinals. We define $\alpha + \beta$ by induction on β with α fixed, by

- $\alpha + 0 = \alpha$;
- $\alpha + \beta^+ = (\alpha + \beta)^+$;
- $\alpha + \lambda = \sup\{\alpha + \gamma : \gamma < \lambda\}$ for $\lambda \neq 0$ a limit ordinal.

Remark 24. As the ordinals do not form a set, we must technically define addition $\alpha + \gamma$ by induction on the set $\{\gamma : \gamma \leq \beta\}$. The choice of β does not change the definition of $\alpha + \gamma$ as defined for $\gamma \leq \beta$. This gives a well-defined "+" by uniqueness in the recursion thm.

Similarly, we can prove things by induction: Let $P(\alpha)$ be a statement for each ordinal α, then

$$(\forall\alpha)((\forall\beta)[(\beta < \alpha) \Rightarrow P(\beta)] \Rightarrow P(\alpha)) \Rightarrow (\forall\alpha)P(\alpha).$$

If not, then $\exists\alpha$ s.t. $P(\alpha)$ is false. Then $\exists$ least such α ($\{\beta \leq \alpha : P(\beta) \text{ false}\} \neq \varnothing$). By proposition 2.7, α is the least element. So $P(\beta)$ is true $\forall\beta < \alpha$. By assumption $P(\alpha)$ is true.

Example 2.10

For any α, $\alpha + 1 = \alpha + 0^+ = (\alpha + 0)^+ = \alpha^+$.
If $m < \omega$, then we have $m + 0 = m$ and for $n < \omega$, $m + (n+1) = m + n^+ = (m+n)^+ = (m+n) + 1$.

So on ω, ordered addition is the normal addition.

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

$1 + \omega = \sup\{1 + \gamma : \gamma < \omega\} = \sup\{1, 2, 3, \dots\} = \omega \neq \omega + 1$.
Therefore, "+" is noncommutative.

Proposition 2.8

$\forall\alpha, \beta, \gamma$ ordinals, $\beta \leq \gamma \Rightarrow \alpha + \beta \leq \alpha + \gamma$.

Proof. We prove this by induction on γ, with α, β fixed.

$\underline{\gamma = 0}$: If $\beta \leq \gamma$, then $\beta = 0$, so the result is true.

$\underline{\gamma = \delta^+}$: If $\beta \leq \gamma$, then <u>either</u> $\beta = \gamma$ and we are done. <u>Or</u> $\beta \leq \delta$ and so $\alpha + \beta \leq \alpha + \delta$ as $\delta < \gamma$ and induction hypothesis. Further $\alpha + \delta < (\alpha + \delta)^+ = \alpha + \delta^+ = \alpha + \gamma$.

$\underline{\gamma \neq 0 \text{ limit}}$: If $\beta \leq \gamma$, then wlog $\beta < \gamma$, so $\alpha + \beta \leq \sup\{\alpha + \delta : \delta < \gamma\} = \alpha + \gamma$. □

Remark 25. From proposition 2.8, we get $\beta < \gamma \Rightarrow \alpha + \beta < \alpha + \gamma$.
Indeed, $\alpha + \beta < (\alpha + \beta)^+ = \alpha + \beta^+ \leq \alpha + \gamma$ since $\beta^+ \leq \gamma$ (from proposition 2.8).

Note that $1 < 2$ but $1 + \omega = 2 + \omega = \omega$.

Lemma 2.2

Let α be an ordinal and S a non-empty set of ordinals. Then $\alpha + \sup S = \sup\{\alpha + \beta : \beta \in S\}$.

Proof. If $\beta \in S$, then $\alpha + \beta \leq \alpha + \sup S$ (proposition 2.8). Hence $\sup\{\alpha + \beta : \beta \in S\} \leq \alpha + \sup S$.

For the reverse inequality, consider two cases. If S has greatest element, β say, then $\alpha + \sup S = \alpha + \beta$. $\forall \gamma \in S, \gamma \leq \beta$, so by proposition 2.8, $\alpha + \gamma \leq \alpha + \beta$. It follows that $\sup\{\alpha + \gamma : \gamma \in S\} = \alpha + \beta$.
If S has no greatest element, then $\lambda = \sup S$ is a $\neq 0$ limit ordinal (If $\lambda = \gamma^+$, then $\gamma < \lambda$ so $\exists \delta \in S$ s.t. $\gamma < \delta$ then $\lambda = \gamma^+ \leq \delta$ so $\lambda = \delta \in S$ ↯). So $\alpha + \sup S = \sup\{\alpha + \beta : \beta < \lambda\}$ by defn.
If $\beta < \lambda$, then $\exists \delta \in S$ s.t. $\beta < \delta$. By proposition 2.8, $\alpha + \beta \leq \alpha + \delta$. It follows that $\sup\{\alpha + \beta : \beta < \lambda\} \leq \sup\{\alpha + \delta : \delta \in S\}$. □

Proposition 2.9

$\forall \alpha, \beta, \gamma$, $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$.

Proof. By induction on γ.

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

$\underline{\gamma = \delta^+}$: $(\alpha + \beta) + \delta^+ = ((\alpha + \beta) + \delta)^+ = (\alpha + (\beta + \delta))^+ = \alpha + (\beta + \delta)^+ = \alpha + (\beta + \delta^+) = \alpha + (\beta + \gamma)$.

$\underline{\gamma \neq 0 \text{ limit}}$:

$$
\begin{aligned}
(\alpha + \beta) + \gamma &= \sup\{(\alpha + \beta) + \delta : \delta < \gamma\} \\
&= \sup\{\alpha + (\beta + \delta) : \delta < \gamma\} \\
&= \alpha + \sup\{\beta + \delta : \delta < \gamma\} \text{ by lemma 2.2} \\
&= \alpha + (\beta + \gamma).
\end{aligned}
$$

□

The above is the **inductive** definition of addition; there is also a **synthetic** definition of addition. We can define $\alpha + \beta$ to be the order type of $\alpha \sqcup \beta$, where every element of α is taken to be less than every element of β.

For instance, $\omega + 1$ is the order type of ω with a point afterwards, and $1 + \omega$ is the order type of a point followed by ω, which is clearly isomorphic to ω. Associativity is clear, as $(\alpha + \beta) + \gamma$ and $\alpha + (\beta + \gamma)$ are the order type of $\alpha \sqcup \beta \sqcup \gamma$.

Proposition 2.10

The inductive and synthetic definitions of addition coincide.

Proof. We write $+'$ for synthetic addition, and aim to show $\alpha + \beta = \alpha +' \beta$. We perform induction on β.

For $\beta = 0$, $\alpha + 0 = \alpha$ and $\alpha +' 0 = \alpha$. For successors, $\alpha + \beta^+ = (\alpha + \beta)^+ = (\alpha +' \beta)^+$, which is the order type of $\alpha \sqcup \beta \sqcup \{\star\}$, which is equal to $\alpha +' \beta^+$.

Let λ be a nonzero limit. We have $\alpha + \lambda = \sup\{\alpha + \gamma : \gamma < \lambda\}$. But $\alpha + \gamma = \alpha +' \gamma$ for $\gamma < \lambda$, so $\alpha + \lambda = \sup\{\alpha +' \gamma : \gamma < \lambda\}$. As the set $\{\alpha +' \gamma : \gamma < \lambda\}$ is nested, it's sup is equal to its union, which is $\alpha +' \lambda$. □

Synthetic definitions can be easier to work with if such definitions exist. However, there are many definitions that can only easily be represented inductively, and not synthetically.

We define multiplication inductively by

- $\alpha 0 = 0$;
- $\alpha \beta^+ = \alpha\beta + \alpha$;
- $\alpha\lambda = \sup\{\alpha\gamma : \gamma < \lambda\}$ for λ a nonzero limit.

Example 2.11

$\omega 2 = \omega 1 + \omega = \omega 0 + \omega + \omega = \omega + \omega$. Similarly, $\omega 3 = \omega + \omega + \omega$. $\omega\omega = \sup\{0, \omega 1, \omega 2, \dots\} = \{0, \omega, \omega + \omega, \dots\}$. Note that $2\omega = \sup\{0, 2, 4, \dots\} = \omega$. Multiplication is noncommutative. One can show in a similar way that multiplication is associative.

We can produce a synthetic definition of multiplication, which can be shown to coincide with the inductive definition. We define $\alpha\beta$ to be the order type of the Cartesian product $\alpha \times \beta$ where we say $(\gamma, \delta) < (\gamma', \delta')$ if $\delta < \delta'$ or $\delta = \delta'$ and $\gamma < \gamma'$. For instance, $\omega 2$ is the order type of two infinite sequences, and 2ω is the order type of a sequence of pairs.

Similar definitions can be created for exponentiation, towers, and so on. For instance, α^β can be defined by

- $\alpha^0 = 1$;
- $\alpha^{(\beta^+)} = \alpha^\beta \alpha$;
- $\alpha^\lambda = \sup \{\alpha^\gamma : \gamma < \lambda\}$ for λ a nonzero limit.

For example, $\omega^2 = \omega^1\omega = \omega^0\omega\omega = \omega\omega$. Further, $2^\omega = \sup \{2^0, 2^1, \dots\} = \omega$, which is countable.

§3 Posets

§3.1 Definitions

Definition 3.1 (Poset)
A **partially ordered set** or **poset** is a pair $(X, \leq)$ where X is a set, and $\leq$ is a relation on X s.t.

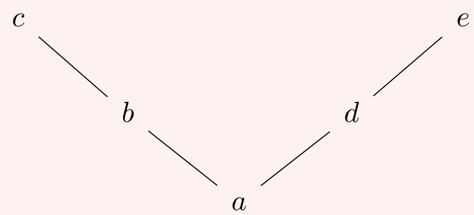
- (reflexivity) for all $x \in X$, $x \leq x$;
- (transitivity) for all $x, y, z \in X$, $x \leq y$ and $y \leq z$ implies $x \leq z$;
- (antisymmetry) for all $x, y \in X$, $x \leq y$ and $y \leq x$ implies $x = y$.

We write $x < y$ for $x \leq y$ and $x \neq y$. Alternatively, a poset is a pair $(X, <)$ where X is a set, and $<$ is a relation on X s.t.

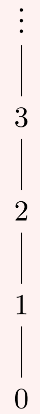
- (irreflexivity) for all $x \in X$, $x \not< x$;
- (transitivity) for all $x, y, z \in X$, $x < y$ and $y < z$ implies $x < z$.

Example 3.1

1. Any total order is a poset.
2. $\mathbb{N}^+$ with the divides relation is a poset.
3. $(\mathcal{P}(S), \subseteq)$ is a poset.
4. $(X, \subseteq)$ is a poset where $X \subseteq \mathcal{P}(S)$, such as the set of vector subspaces of a vector space.
5. The following diagram is also a poset, where the lines from a upwards to b denote relations $a \leq b$.

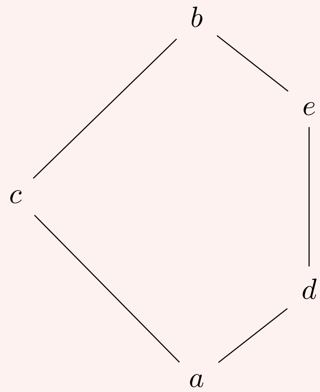


This is called a **Hasse diagram**. An upwards line from x to y is drawn if y **covers** x, so $y > x$ and no z has $y > z > x$. The natural numbers can be represented as a Hasse diagram.

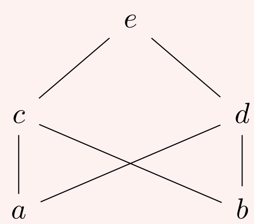


The rationals cannot, since no element covers another.

6. There is no notion of 'height' in a poset, illustrated by the following diagram.



7.



Definition 3.2 (Chain)
A subset S of a poset X is a **chain** if it is linearly ordered by the partial order.

Example 3.2

Every linearly ordered set is a chain in itself.

Example 3.3

Any subset of a chain in a poset is a chain.

Example 3.4

The powers of 2 in $(\mathbb{N}^+, |)$ is a chain.

Example 3.5

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

Definition 3.3 (Antichain)

A subset S of a poset X is an **antichain** if no two distinct elements are related: $\forall x, y \in S, x \leq y \Rightarrow x = y$.

Example 3.6

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

Example 3.7

The set of primes in $(\mathbb{N}^+, |)$ is an antichain.

Definition 3.4 (Supremum)

For $S \subseteq X$ where X a poset, an **upper bound** for S is an $x \in X$ s.t. $y \leq x \ \forall y \in S$. A **least upper bound** or **supremum** is an upper bound $x \in X$ for S s.t. for all upper bounds $y \in X$ for S, $x \leq y$.

Note. S can be $\varnothing$.

Notation. If the supremum exists, denote it by $\sup S$ or $\bigvee S$ ("join" of S).

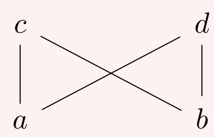
Example 3.8
In $\mathbb{R}$, $\sup[0,1] = 1$, $\sup(0,1) = 1$.

Example 3.9
$\mathbb{Q}$ has no sup in $\mathbb{Q}$, it doesn't have any upper bound.

Example 3.10
If $S = \left\{x : x < \sqrt{2}\right\} \subset \mathbb{R}$, 7 is an upper bound, and $\sup S = \sqrt{2}$.

In $\mathbb{Q} \cap [0,2]$, the set $\{x : x^2 < 2\}$ has 2 as an upper bound but no supremum.

Example 3.11



$\{a, b\}$ has upper bounds, e.g. c and d but no sup.

Example 3.12
If $X = \mathcal{P}(A)$ where A is any set and $S \subseteq X$, then $\sup S = \cup\{B \subseteq A : B \in S\}$.

Definition 3.5 (Complete)
A poset X is **complete** if every $S \subseteq X$ has a sup.

Example 3.13
$\mathbb{R}$ is not complete, as $\mathbb{Z}$ has no upper bound.
$[0,1] \subseteq \mathbb{R}$ is complete.
$(0,1) \subseteq \mathbb{R}$ is not complete, as $(0,1)$ has no upper bound.

Example 3.14
$\mathbb{Q} \cap [0,2]$ is not complete by earlier example.

Example 3.15

$\mathcal{P}(A)$ is complete under inclusion for any A.

Remark 26. Note that every complete poset X has a greatest element $\sup X$. A complete poset also has a least element $\sup \varnothing$. In particular, $X \neq \varnothing$.

Definition 3.6 (Order-Preserving)

Let $f \colon X \to Y$ be a fcn where X, Y are posets. We say f is **order-preserving** if $x \leq y \Rightarrow f(x) \leq f(y)$.

Note. f need not be injective. But f order-preserving and injective implies $x < y \Rightarrow f(x) < f(y)$.

Example 3.16

The fcn $f \colon \mathbb{N} \to \mathbb{N}$ defined by $f(x) = x + 1$ is order-preserving.
The fcn $f \colon [0,1] \to [0,1]$ defined by $x \mapsto \frac{x+1}{2}$ is order-preserving.
The fcn $f \colon \mathcal{P}(S) \to \mathcal{P}(S)$ defined by $f(A) = A \cup B$ for some fixed $B \subseteq S$ is order-preserving.

Definition 3.7 (Fixed Point)

Let X be any set. Then a **fixed point** for a fcn $f : X \to X$ is an element $x \in X$ s.t. $f(x) = x$.

Not all order-preserving fcns have a fixed point, e.g. $f(x) = x + 1$ on $\mathbb{N}$.

Theorem 3.1 (Knaster–Tarski Fixed Point Theorem)

Let X be a complete poset. Then every order-preserving $f \colon X \to X$ has a fixed point.

Proof. Let $E = \{x \in X : x \leq f(x)\}$, and let $s = \sup E$. We show that s is a fixed point for f.

First, we show $s \leq f(s)$, so $s \in E$. If $x \in E$, we know $x \leq s$, so $f(x) \leq f(s)$. Since $x \in E$, $x \leq f(x)$, so by transitivity $x \leq f(s)$. Thus $f(s)$ an upper bound for E so $s \leq f(s)$.

Now, we show $f(s) \leq s$. Since $s \leq f(s)$, we have $f(s) \leq f(f(s))$, i.e. $f(s) \in E$ thus $f(s) \leq s$. □

Corollary 3.1 (Schröder–Bernstein Theorem)

Let A, B be sets and assume $\exists$ injections $f\colon A \to B$ and $g\colon B \to A$. Then $\exists$ bijection $h : A \to B$.

Proof. We seek partitions $A = P \cup Q$, $B = R \cup S$ s.t. $(P \cap Q = R \cap S = \varnothing)$, $f(P) = R$ and $g(S) = Q$.

Then $h = \begin{cases} f & \text{on } P \\ g^{-1} & \text{on } Q \end{cases}$ is a bijection.

Such partitions exists $\iff \exists P \subseteq A$ s.t. $A \setminus g(B \setminus f(P)) = P$.

Let $X = \mathcal{P}(A)$ with "$\subseteq$" and define $H : X \to X$ by $H(P) = A \setminus g(B \setminus f(P))$. H is order-preserving and X is a complete poset. So P exists by the Knaster–Tarski Fixed Point Theorem. □

§3.2 Zorn's lemma

Definition 3.8 (Maximal)

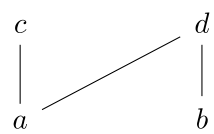
Let X be a poset. We say that $x \in X$ is **maximal** if there is no $y \in X$ with $y > x$, or $\forall y \in X, x \leq y \Rightarrow x = y$.

Example 3.17

In $\mathcal{P}(A)$, A is maximal, A is even a greatest element.

Note. In general, "greatest" $\Rightarrow$ maximal.

The converse is false, e.g.



c, d both maximal but not greatest element.

Example 3.18

In $[0, 1]$, 1 is maximal.

Note that $(\mathbb{R}, \leq)$ and $(\mathbb{N}, |)$ have no maximal elements, and they both have a chain with no upper bound, such as $\mathbb{N} \subset \mathbb{R}$, and powers of two.

Theorem 3.2 (Zorn's Lemma (ZL))

Let X be a (non-empty) poset s.t. every chain in X has an upper bound in X. Then X has a maximal element.

Remark 27. $\varnothing$ is a chain in X, so it has an UB, so $X \neq \varnothing$.
Often we check the chain condition by checking it for $\varnothing$ (i.e. that $X \neq \varnothing$) and then check for non-empty chains.

One can view Zorn's lemma as a fixed point theorem on a fcn $f\colon X \to X$ with the property that $x \leq f(x)$.

Proof. Suppose that X has no maximal element. Then for each $x \in X$, we have $x' \in X$ and $x' > x$. For each chain C, we have an upper bound $u(C)$.

Let $\gamma = \gamma(X)$ (from Hartog's lemma - the least ordinal that doesn't inject into X).

Define $f : \gamma \to X$ by recursion:

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

An easy induction (on β with α fixed) shows that $\forall \alpha < \beta$ (in γ), $f(\alpha) < f(\beta)$ (also shows that $\{f(\alpha) : \alpha < \beta\}$ is a chain $\forall \beta < \gamma$).

Hence f is an injection, ↯ defn of γ. □

Remark 28. Technically, for $\lambda \neq 0$ limit, $f(\lambda)$ should be defined as above if $\{f(\alpha) : \alpha < \lambda\}$ is a chain and $f(\lambda) = u(\varnothing)$ otherwise. Then by induction $\alpha < \beta \Rightarrow f(\alpha) < f(\beta)$, so the "otherwise" clause never happens.

Remark 29. Although this proof was short, it relied on the infrastructure of well-orderings, recursion, ordinals, and Hartogs' lemma.

We show that every vector space has a basis. Recall that a basis is a linearly independent spanning set; no nontrivial finite linear combination of basis elements is zero, and each element of the vector space is a finite linear combination of the basis elements. For instance, the space of real polynomials has basis $1, X, X^2, \dots$. The space of real sequences has a linearly independent set $(1, 0, 0, \dots), (0, 1, 0, \dots), \dots$, but this is not a basis as the sequence $(1, 1, 1, \dots)$ cannot be constructed as a finite linear combination of these vectors. In fact, there is no countable basis for this space, and no explicitly definable basis in general. $\mathbb{R}$ is a vector space over $\mathbb{Q}$. There is clearly no countable basis, and in fact no explicit basis. A basis in this case is called a **Hamel basis**.

Theorem 3.3

Every vector space V has a basis.

Proof. Let X be the set of all linearly independent subsets of V, ordered by inclusion. We seek a maximal element of X; this is clearly a basis, as any vector not in its span could be added to the set to increase the set of basis vectors. X is nonempty as $\varnothing \in X$.

We apply Zorn's lemma. Let $(A_i)_{i\in I}$ be a chain in X. We show that its union $A = \bigcup_{i\in I} A_i$ is a linearly independent set, and therefore lies in X and is an upper bound. Suppose $x_1, \dots, x_n \in A$ are linearly dependent. Then $x_1 \in A_{i_1}, \dots, x_n \in A_{i_n}$, so all x_i lie in some A_k as the A_i are a chain. But A_k is linearly independent, which is a contradiction. □

Remark 30. The only time that linear algebra was used was to show that the maximal element obtained by Zorn's lemma performs the required task; this is usual for proofs in this style.

We can now prove the completeness theorem for propositional logic with no restrictions on the size of the set of primitive propositions.

Theorem 3.4

Let $S \subseteq L = L(P)$ be consistent. Then S has a model.

Proof. We will extend S to a consistent set $\overline{S}$ s.t. for all $t \in L$, either $t \in \overline{S}$ or $\neg t \in \overline{S}$; we then complete the proof by defining a valuation v s.t. $v(t) = 1$ if $t \in \overline{S}$.

Let $X = \{T \supseteq S \mid T \text{ consistent}\}$ be the poset of consistent extensions of S, ordered by inclusion. We seek a maximal element of X. Then, if $\overline{S}$ is maximal and $t \notin \overline{S}$, then $\overline{S} \cup \{t\} \vdash \bot$ by maximality, so $\overline{S} \vdash \neg t$ by the deduction theorem, giving $\neg t \in \overline{S}$ again by maximality.

Note that $X \neq \varnothing$ as $S \in X$. Given a nonempty chain $(T_i)_{i\in I}$, let $T = \bigcup_{i\in I} T_i$. We have $T \supseteq T_i$ for all i and $T \supseteq S$ as the chain is nonempty, so it suffices to show T is consistent. Indeed, suppose $T \vdash \bot$. Then there exists a subset $\{t_1, \dots, t_n\} \in T$ with $\{t_1, \dots, t_n\} \vdash \bot$ as proofs are finite. Now, $t_1 \in T_{i_1}, \dots, t_n \in T_{i_n}$ so all t_j are elements of T_{i_k} for some k. But T_{i_k} is consistent, so $\{t_1, \dots, t_n\} \nvdash \bot$, giving a contradiction. □

§3.3 Well-ordering principle

Theorem 3.5 (Well-Ordering Principle (WP))
Every set has a well-ordering.

There exist sets with no definable well-ordering, such as $\mathbb{R}$.

Proof. Let S be a set, and let X be the set of pairs (A, R) s.t. $A \subseteq S$ and R is a well-ordering on A. We define the partial order on X by $(A, R) \leq (A', R')$ if (A', R') extends (A, R), so $R'|_A = R$ and A is an i.s. of A' for R'.

X is nonempty as the empty relation is a well-ordering of the empty set. Given a nonempty chain $(A_i, R_i)_{i\in I}$, there is an upper bound $(\bigcup_{i\in I} A_i, \bigcup_{i\in I} R_i)$, because the well-orderings are nested so by proposition 2.6. By Zorn's lemma, there exists a maximal element $(A, R) \in X$.

Suppose $x \in S \setminus A$. Then we can construct the well-ordering on $A \cup \{x\}$ by defining $a < x$ for $a \in A$, contradicting maximality of A. Hence $A = S$, so R is a well-ordering on S. □

Remark 31. Often in application of ZL, the maximal object whose existence it asserts cannot be described explicitly ("magical").

§3.4 Zorn's lemma and the axiom of choice

In the proof of Zorn's lemma, for each $x \in S$ we chose an arbitrary $x' > x$. This requires potentially infinitely many arbitrary choices. Other proofs, such as that the countable union of countable sets is countable, also required infinitely many choices; in this example, we chose arbitrary enumerations of the countable sets $A_1, A_2, \dots$ at once.

Formally, this process of making infinitely many arbitrary choices is known as the **axiom of choice** AC: if we have a family of nonempty sets, one can choose an element from each one. More precisely, for any family of nonempty sets $(A_i)_{i\in I}$, there is a **choice fcn** $f\colon I \to \bigcup_{i\in I} A_i$ s.t. $f(i) \in A_i$ for all i.

Unlike the other axioms of set theory, the fcn obtained from the axiom of choice is not uniquely defined. For instance, the axiom of union allows for the construction of $A \cup B$ given A and B, which can be fully described; but applying the axiom of choice to the family $\star \mapsto \{1, 2\}$ could give the choice fcn $\star \mapsto 1$ or $\star \mapsto 2$.

Use of the axiom of choice gives rise to nonconstructive proofs. In modern mathematics it is sometimes considered useful to note when the axiom of choice is being used. However, many proofs that do not even use the axiom of choice are nonconstructive, such as the proof of existence of transcendentals, or Hilbert's basis theorem that every ideal over $\mathbb{Q}[X_1, \dots, X_n]$ is finitely generated.

Although our proof of Zorn's lemma required the axiom of choice, it is not immediately clear that all such proofs require it. However, it can be shown that Zorn's lemma implies the axiom of choice in the presence of the other axioms of ZF set theory. Indeed, if $(A_i)_{i \in I}$ is a family of sets, we can well-order it using the well-ordering principle, and define the choice fcn by setting $f(i)$ to be the least element of A_i. Hence, Zorn's lemma, the axiom of choice, and the well-ordering principle are equivalent, given ZF.

AC can be proven trivially in ZF for the case $|I| = 1$, because a set being nonempty means precisely that there exists an element inside it. Clearly, AC holds for all finite index sets in ZF by induction on $|I|$. However, ZF does not prove the most general form of AC.

Zorn's lemma is a difficult lemma to prove from first principles because of its reliance on ordinals and Hartogs' lemma; the use of the axiom of choice does not contribute significantly to its difficulty. The construction and properties of the ordinals did not rely on the axiom of choice. The axiom of choice was only used twice in the section on well-orderings: the fact that in a set that is not well-ordered, there is an infinite decreasing sequence; and the fact that ω_1 is not a countable supremum.

Aside - Non Examinable

Definition 3.9 (Chain-Complete)

A poset X is **chain-complete** if $X \neq \varnothing$ and every non-empty chain has a sup.

Example 3.19

Every complete poset.
Finite non-empty poset.
If S in a poset, then $X = \{X \subseteq S : C \text{ is a chain}\}$ ordered by "$\subseteq$" is chain-complete, but not complete in general.

Definition 3.10 (Inflationary)

A function $f : X \to X$, X a poset, is **inflationary** if $x \leq f(x)$ $\forall x \in X$.

Theorem 3.6 (Bourbaki-Witt Fixed Point Theorem)

If X is chain-complete and $f : X \to X$ is inflationary, then f has a fixed point.

Proof (*With AC*). By ZL, X has a minimal element x. Then $x \leq f(x)$, so $x = f(x)$. □

Proof (without AC). Fix $x_0 \in X$. Let $\gamma = \gamma(X)$.

Define $g : \gamma \to X$ by recursion:

$$\begin{aligned} g(0) &= x_0 \\ g(\alpha+1) &= f(g(\alpha)) \\ g(\lambda) &= \sup\{g(\alpha) : \alpha < \lambda\} \quad \lambda \neq 0 \text{ limit} \end{aligned}$$

By induction, $\forall \alpha < \gamma$, $g(\alpha) \leq g(\alpha+1)$. Either $\exists \alpha < \gamma$, $g(\alpha+1) = g(\alpha)$. Then $g(\alpha)$ is a fixed point of f.
Otherwise g injective ↯. □

Remark 32. AC + Bourbaki-Witt ⇒ ZL. Bourbaki-Witt is the "choice-free" part of ZL.

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

Case 1: X is chain-complete.
Assume X has no maximal element. Fix a choice fcn $g : \mathcal{P}(X) \setminus \varnothing \to X$. Define $f : X \to X$, $f(x) = g(\{y \in X : x < y\})$. Then $x < f(x)$ $\forall x \in X$. ↯ of Bourbaki-Witt.

Case 2: Several case.
We first prove that $\mathcal{C} = \{C \subseteq X : C \text{ is a chain}\}$ has a maximal element. (This is the Hausdorff Maximality Principle). Follows from Case 1, since $\mathcal{C}$ is chain-complete.
Let C be a maximal chain in X, let x be an upper bound of C. If $x < y$ in X, then $C \cup \{y\}$ is a chain $\supsetneq$ C ↯. So x is maximal. □

---

§4 Predicate Logic

§4.1 Languages

In Propositional Logic we has a set P of primitive propositions and then we combined them using logical connectives $\Rightarrow$, $\perp$, ($\wedge$, $\vee$, $\neg$, $\top$) to form the language $L = L(P)$ of all (compound) propositions.

We attached no meaning to primitive propositions.
Aim: To develop languages to describe a wide range of mathematical theorems. We will replace primitive propositions with non mathematical statements.

Example 4.1

In language of groups:

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

In language of posets: $x \le y$.
This will need variables $(x, y, z, \dots)$, operation symbols $(m, i, e$ with arities $2, 1, 0$ respectively) and predicates (e.g. $\le$ with arity 2).

We will then combine these to build formulae:
In language of cosets,

$$(\forall x)(\forall y)(\forall z)((x \le y \wedge y \le z) \Rightarrow (x \le z))$$

In language of groups, $(\forall x)(m(x, i(x)) = e)$.

Valuations will be replaced by a structure, a set A, and "truth-functions" $p_A : A^n \to 0, 1$ for every formula p.

If we have set S of formulae, a model S is a structure satisfying all $p \in S$.
$S \models t$ will be same as in Section 1.
$S \vdash t$ will be same as in Section 1 but more complex.

A language in first-order logic is specified by the disjoint set Ω (set of operation symbols) and Π (set of predication) together with an arity function $\alpha \colon \Omega \cup \Pi \to \mathbb{N}_0 = \{0\} \cup \mathbb{N}$. The language $L = L(\Omega, \Pi, \alpha)$ consists of the following: **Variables** a countably infinite sets disjoint of Ω, Π. We denote variables as $x_1, x_2, x_3, \dots$ (or $x, y, z, \dots$).
Terms are defined inductively by

1. each variable is a term;
2. if $f \in \Omega$ with $\alpha(f) = n$ and terms $t_1, \dots, t_n$, then $f\ t_1 \dots\ t_n$ is a term (could write $f(t_1, \dots, t_n)$).

The **atomic formulae** are defined inductively by

1. for terms s, t, $(s = t)$ is an atomic formula;
2. if $\varphi \in \Pi$ with $\alpha(\varphi) = n$ and terms $t_1, \ldots, t_n$, then $\varphi(t_1, \ldots, t_n)$ is an atomic formula.

The **formulae** are defined inductively by

1. each atomic formula is a formula;
2. $\perp$ is a formula;
3. if p and q are formulae then $(p \Rightarrow q)$ is a formula;
4. if p is a formula and the variable x has a **free occurrence in p**, then $(\forall x)p$ is a formula.

The **language** $L = L(\Omega, \Pi, \alpha)$ is the set of formulae.

Definition 4.1 (Constant)
Every operation symbol of arity 0 is a term, and called a **constant**.

Example 4.2
In the language of groups, $\Omega = \{m, i, e\}$ and $\Pi = \varnothing$ with $\alpha(m) = 2, \alpha(i) = 1, \alpha(e) = 0$. $m(x_1, x_2), m(x_1, i(x_2)), e, m(e, e), mxmyz, mmxyz, mxix$ are examples of terms of the language. $e = m(\ell, e), m(x, y) = m(y, x)$ are atomic formulae.

Example 4.3
In the language of posets, $\Omega = \varnothing$ and $\Pi = \{\leq\}$ with $\alpha(\leq) = 2$. $x = y, x \leq y$ are atomic formulae. Technically, $x \leq y$ is written $\leq (x, y)$.

Example 4.4
In the language of groups, $(\forall x)(m(x, x) = e)$ is a formula. Another formula is $m(x, x) = e \Rightarrow (\exists y)(m(y, y) = x)$.

Remark 33. A formula is a certain finite string of symbols from the set of variables, Ω, Π, $\{(,), \Rightarrow, \perp, =, \forall\}$; it has no intrinsic semantics. We define $\neg p, p \wedge q, p \vee q$ in the usual way. We define $(\exists x)p$ to mean $\neg(\forall x)(\neg p)$.

A term is **closed** if it contains no variables. For example, $e, m(e, i(e))$ are closed in the language of groups, but $m(x, i(x))$ is not closed.

Definition 4.2 (Free/ Bound Occurence)

An occurrence of a variable x in a formula p is always **free** except if $p = (\forall x)q$ in which the $\forall x$ quantifier **binds** every free occurrence of x and then such occurrences of x are called **bound occurences**[a].

[a]The formal defn is by induction on L

Example 4.5

In the formula $(\forall x)(m(x,x) = e)$, each occurrence of x is bound.
In $m(x,x) = e \Rightarrow (\exists y)(m(y,y) = x)$, the occurrences of x are free and the occurrences of y are bound.
In the formula $m(x,x) = e \Rightarrow (\forall x)(\forall y)(m(x,y) = m(y,x))$, the occurrences of x on the left hand side are free, and the occurrences of x on the right hand side are bound.

1. $(\exists x)(m(x,x) = y) \Rightarrow (\forall z)\neg(mmzzz = y)$. The x (not the x in $\exists x$) are bound, all ys are free, the z (not the z in $\forall z$) are bound.
2. $(\forall x)(\forall y)(\forall z)(mmxyz = mxmyz)$, this is the associativity law. There are no free variables.
3. $(\exists x)(mxx = y) \Rightarrow (\forall y)(\forall x)(myz = mzy)$. The y on the LHS is free, the ys on the RHS are bound. This is technically a correct formula, but in mathematical practice we avoid this.
4. In the language of posets: $(\forall x)(\forall y)(((x \leq y) \wedge (y \leq x)) \Rightarrow (x = y))$ has no free variables.

Definition 4.3 (Sentence)

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

Definition 4.4 (Free)

A variable x in a formula is **free** if it has a free occurrence in p. Let $\mathrm{FV}(p)$ denote the set of free variables in p.

Example 4.6

For instance, $(\forall x)(m(x,x) = e)$ is a sentence, and $(\forall x)(m(x,x) \Rightarrow (\exists y)(m(y,y) = x))$ is a sentence.
In the language of posets, $(\forall x)(\exists y)(x \geq y \wedge \neg(x = y))$ is a sentence.

For a formula p, term t, and variable x, the **substitution** $p[t/x]$ is obtained from p by replacing every free occurrence of x with t. For example,

$$p = (\exists y)(m(y,y) = x); \quad p[e/x] = (\exists y)(m(y,y) = e)$$

§4.2 Semantic implication

Definition 4.5 (Structure)
Let $L = L(\Omega, \Pi, \alpha)$ be a first-order language. An **structure** in L or **L-structure** is

- a nonempty set A;
- for each $f \in \Omega$, a function $f_A \colon A^n \to A$ where $n = \alpha(f)$;
- for each $\varphi \in \Pi$, a subset $\varphi_A \subseteq A^n$ where $n = \alpha(\varphi)$[a].

[a]Equivalently $\varphi_A : A^n \to \{0,1\}$ by identifying a set with its indicator fcn.

Remark 34. We will see later why the restriction that A is nonempty is given here.

Example 4.7
In the language of groups, a structure is a nonempty set A with fcns $m_A \colon A^2 \to A, i_A \colon A \to A, e_A \in A^a$.

Such a structure may not be a group, as we have not placed any axioms on A.

[a] A^0 is the singleton set

Example 4.8
In the language of posets, a structure is a nonempty set A with a relation $(\leq_A) \subseteq A^2$. This is not yet a poset.

Next Step: to define for a formula p what it means that 'p is satisfied in A'.

Example 4.9
$p = (\forall x)(mxix = e)$ in the language of groups. p satisfied in structure A should mean that $\forall a \in A$ we have $m_A(a, i_A(a)) = e_A$.

Let A be an L-structure. A term t in L with $FV(t) \subseteq \{x_1, \ldots, x_n\}$ has **interpretation** $t_A : A^n \to A$ defined as follows:

- If $t = x_i$ for $1 \leq i \leq n$ then $t_A(a_1, \ldots, a_n) = a_i$.

- If $t = \omega t_1 \dots t_m$ with $\omega \in \Omega$, $m = \alpha(w)$, $t_1, \dots, t_m$ terms, then $t_A(a_1, \dots, a_n) = \omega_A((t_1)_A(a_1, \dots, a_n), \dots, (t_m)_A(a_1, \dots, a_n))$.

Example 4.10

In groups $t = mx_1mx_2x_3$, $t_A(a_1, a_2, a_3) = m_A(a_1, m_A(a_2, a_3))$.

Next interpret a formula p with $FV(p) \subseteq \{x_1, \dots, x_n\}$ as a subset $p_A \subset A^n$ or equivalently a fcn $p_A : A^n \to \{0, 1\}$.

- If $p = (s = t)$, $p_A(a_1, \dots, a_n) = 1 \iff s_A(a_1, \dots, a_n) = t_A(a_1, \dots, a_n)$.
- If $p = \varphi t_1, \dots, t_m$, with $\varphi \in \Pi$, $m = \alpha(\varphi)$, $t_1, \dots, t_m$ terms. Then $p_A(a_1, \dots, a_n) = 1$ iff $\varphi_A((t_1)_A(a_1, \dots, a_n), \dots, (t_m)_A(a_1, \dots, a_n)) = 1$.
- $\bot_A \equiv 0$.
- If $p = (q \Rightarrow r)$, then $p_A(a_1, \dots, a_n) = 0$ iff $q_A(a_1, \dots, a_n) = 1$ and $r_A(a_1, \dots, a_n) = 0$.
- If $p = (\forall x_{n+1})q$ where $FV(q) \subseteq \{x_1, \dots, x_{n+1}\}$,
  $p_A = \{(a_1, \dots, a_n) \in A^n : (a_1, \dots, a_{n+1}) \in q_A \ \forall a_{n+1} \in A\}$.

Example 4.11

In groups, $p = mmxyz = mxmyz$.
$p_A = \{(a, b, c) \in A^3 : m_A(m_A(a, b), c) = m_A(a, m_A(b, c))\}$.
$q = (\forall x)(\forall y)(\forall z)p$ then $q_A = 1 \iff p_A = A^3$.

Definition 4.6 (Satisfied)

If $p_A \equiv 1$ or $p_A = A^n$[a] we say the formula p is **satisfied** in a L-structure A.
We also say p **holds** in A, p is **true** in A or A is a **model** for p.

[a] n is the number of free variables in p.

Definition 4.7 (Theory)

A **theory** is a set of sentences in L, known as L's **axioms**.

Definition 4.8 (Model)

A **model** for a theory T is an L-structure A that is a model $\forall p \in T$.

Example 4.12 (Groups)

Let L be the language of groups. The language is specified by $\Omega = \{m, i, e\}$, $\Pi = \varnothing$, α is $2, 1, 0$ for m, i, e respectively.

Let

$$\begin{aligned} T = \{&(\forall x)(\forall y)(\forall z)(m(x, m(y, z)) = m(m(x, y), z)), \\ &(\forall x)(m(x, e) = x \wedge m(e, x) = x), \\ &(\forall x)(m(x, i(x)) = e \wedge m(i(x), x) = e)\} \end{aligned}$$

Then, an L-structure is a model of T iff it is a group. Note that this statement has two assertions; every L-structure that is a model of T is a group, and that every group can be turned into an L-structure that models T.

We say that T **axiomatises** the theory of groups or the class of groups.

Example 4.13 (Posets)

Let L be the language of posets. $\Omega = \varnothing$, $\Pi = \{\leq\}$, $\alpha(\leq) = 2$.

Let

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

The models are partially ordered sets, i.e. T axiomatises the class of posets.

Example 4.14 (Rings with 1)

$\Omega = \{0, 1, +, \cdot, -\}$ with $\alpha(0) = \alpha(1) = 0, \alpha(+) = \alpha(\cdot) = 2, \alpha(-) = 1$ and $\Pi = \varnothing$.

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

The models are rings with 1.

Example 4.15 (Fields)

We use the same language as in rings with 1.

The theory is the same as rings with 1, plus

$$(\forall x)(\forall y)(x \cdot y = y \cdot x)$$
$$\neg(0 = 1)$$
$$(\forall x)(\neg(x = 0) \Rightarrow (\exists y)(x \cdot y = 1))$$

The models are fields.

Example 4.16 (Graph Theory)

Let L be the language of graphs, defined by $\Omega = \varnothing$, $\Pi = \{a\}$ ($a =$ 'is adjacent to') and $\alpha(a) = 2$. Define $T = \{(\forall x)(\neg a(x,x)), (\forall x)(\forall y)(a(x,y) \Rightarrow a(y,x))\}$. Then T axiomatises the class of graphs.

Example 4.17 (Propositional Theories)

$\Omega = \varnothing$, Π s.t. $\alpha(p) = 0\ \forall p \in \Pi$.
A structure is a nonempty set A together with $p_A \subset A^0$ for all $p \in \Pi$ (equivalently $p_A \in \{0,1\}$).
A structure is a nonempty set A together with a fcn $v : \Pi \to \{0,1\}$.

Every $p \in \Pi$ is an atomic formula. Formula w/o variables are precisely elements of $L(\Pi)$ as defined in section 1, i.e. they are propositions in Π. Interpreting these in a structure A is just a fcn $v : L(\Pi) \to \{0,1\}$ obtained from $v : \Pi \to \{0,1\}$ as in section 1, i.e. a valuation.

A **propositional theory** is a set S of formulae not using variables. A model for S is a nonempty set A with a valuation $v : L(\Pi) \to \{0,1\}$ s.t. $v(s) = 1\ \forall s \in S$ (Here A is irrelevant).

§4.3 Semantic Entailment

Definition 4.9 (Semantic Entailment)

For a set S of sentences (i.e. a theory) and a sentence t (in some first-order language L) we say S **(semantically) entails** t if t is satisfied in every model of S. We write $S \models t$.

Example 4.18 (Groups)

Let S be the theory of groups. $S \models (\forall x)(x \cdot x = e) \Rightarrow (\forall x)(\forall y)(xy = yx)$.

Example 4.19 (Fields)

Let S be the theory of fields. $S \models (\forall x)(\neg(x = 0) \Rightarrow (\forall y)(\forall z)((xy = 1 \wedge xz = 1) \Rightarrow (y = z)))$.

Next, we want to define $S \models t$ for formulae:

Example 4.20

Check Zs´ak's notes for this example, I'm not sure what's going on.

Let T be the theory of fields. Take $S = T \cup \{\neg(x = 0)\}$, $t = (\exists y)(xy = 1)$.

Does $S \models t$? Yes.

Suppose F is a structure in which all members of S are true. F is a field and for $u = \neg(x = 0)$, $u_F = \{a \in F : a \neq 0_F\} = F$ ↯.

Also, we'll soon define "$S \vdash t$", then $S \vdash t \iff T \vdash \neg(x = 0) \Rightarrow (\exists y)(xy = 1)$. This will help motivate our defn.

Let S be a set of formulae and t a formula in a language L. For every variable x that occurs free in $S \cup \{t\}$, introduce a constant c_x (add it to Ω). Let L' be our new language. For a formula p, let p' be the formula obtained from p by replacing free occurrences of x in p by c_x for every x.
Let $S' = \{s' : s \in S\}$.

Definition 4.10 (Semantic Entailment)

Say S **(semantically) entails** t, written $S \models t$, if $S' \models t'$.

Definition 4.11 (Substitution)

If x occurs free in a formula p and t is a term that contains no variable that occurs bound in p, we let the **substitution** $p[t/x]$ be the formula obtained from p by replacing free occurrences of x in p by t.

Example 4.21

Let $p = (\exists y)(m(y, y) = x)$ then $p[e/x] = (\exists y)(m(y, y) = e)$.

Example 4.22
In language of groups: $p = (\forall y)(mxx = y)$.

- $t = mzz$, then $p[t/x] = (\forall y)(mmzzmzz = y)$.
- $t = mzy$ cannot be used since y is bound in p.
- $t = mxx$, $p[t/x] = (\forall y)(mmxxmxx = y)$.

§4.4 Syntactic Entailment

We need to define (logical) axioms and deduction rules in order to construct proofs.

The Axioms (the previous 3, 2 more for "=", 2 for "$\forall$") are:

1. $p \Rightarrow (q \Rightarrow p)$ for formulae p, q.
2. $(p \Rightarrow (q \Rightarrow r)) \Rightarrow ((p \Rightarrow q) \Rightarrow (p \Rightarrow r))$ for formulae p, q, r.
3. $\neg\neg p \Rightarrow p$ for each formula p.
4. $(\forall x)(x = x)$ for any variable x.
5. $(\forall x)(\forall y)((x = y) \Rightarrow (p \Rightarrow p[y/x]))$ for any distinct variables x, y where y is not bound in the formula p and $x \in \mathrm{FV}(p)$.
6. $((\forall x)p) \Rightarrow p[t/x]$ for any variable $x \in \mathrm{FV}(p)$, formula p, and term t that has no variable that occurs bound in p.
7. $(\forall x)(p \Rightarrow q) \Rightarrow (p \Rightarrow (\forall x)q)$ for any formulae p, q and variable $x \notin \mathrm{FV}(p)$, $x \in \mathrm{FV}(q)$.

Note. Every axiom is a tautology (t is a tautology if $\varnothing \models t$, i.e. t holds in every structure).

We define the following deduction rules.

1. **Modus Ponens (MP)** From p, $p \Rightarrow q$, we can deduce q.
2. **Generalisation (Gen)** From p s.t. $x \in FV(p)$, we can deduce $(\forall x)p$ provided that x does not occur free in any premise used in the proof of p.

Definition 4.12 (Proof)
Let S be a set of formulae and p a formula. A **proof** of p from S is a finite sequence $t_1, \dots, t_n$ of formulae s.t. $t_n = p$ and $\forall i$ t_i is an axiom, a premise or deduced from previous lines (i.e. $\exists j, x < i$ s.t. $t_x = (t_j \Rightarrow t_i)$ (MP) or $\exists j < i$ s.t. $t_i = (\forall x)t_j$, $x \in \mathrm{FV}(t_j)$ and $\forall k < j$ if $t_k \in S$ then $x \notin \mathrm{FV}(t_k)$ (Gen)).
We say S **proves** p and write $S \vdash p$.

Remark 35. Suppose we allow $\varnothing$ as a structure. Then $(\forall x)\neg(x = x)$ is satisfied in $\varnothing$ whereas $\bot$ is not. So $\{(\forall x)\neg(x = x)\} \not\models \bot$. However, $\{(\forall x)\neg(x = x)\} \vdash \bot$:

1. $(\forall x)\neg(x = x)$ (premise)
2. $((\forall x)\neg(x = x)) \Rightarrow \neg(x = x)$ (A6)
3. $\neg(x = x)$ (MP)
4. $(\forall x)(x = x)$ (A4)
5. $(\forall x)(x = x) \Rightarrow (x = x)$ (A6)
6. $x = x$ (MP)
7. $\bot$ (MP)

Example 4.23

We show $\{x = y\} \vdash (y = x)$.

1. $(\forall x)(\forall y)(x = y) \Rightarrow ((x = z) \Rightarrow (y = z))$ (A5)
2. $(x = y) \Rightarrow ((x = z) \Rightarrow (y = z))$ (A6 + MP twice)[a]
3. $x = y$ (premise)
4. $(x = z) \Rightarrow (y = z)$ (MP)
5. $(\forall z)((x = z) \Rightarrow (y = z))$ (Gen)
6. $(\forall z)((x = z) \Rightarrow (y = z)) \Rightarrow ((x = x) \Rightarrow (y = x))$ (A6)
7. $(x = x) \Rightarrow (y = x)$ (MP)
8. $(\forall x)(x = x)$ (A4)
9. $x = x$ (A6 + MP)
10. $y = x$ (MP)

[a] Set $p = (\forall y)(x = y) \Rightarrow ((x = z) \Rightarrow (y = z))$ and $t = x$ so A6 gives $(\forall x)(\forall y)(x = y) \Rightarrow ((x = z) \Rightarrow (y = z)) \Rightarrow (\forall y)(x = y) \Rightarrow ((x = z) \Rightarrow (y = z))$. Then by MP we get $(\forall y)(x = y) \Rightarrow ((x = z) \Rightarrow (y = z))$. Then repeat.

Example 4.24

We show $\{x = y, x = z\} \vdash y = z$ where x, y, z are different variables.

1. $(\forall x)(\forall y)(x = y \Rightarrow (x = z \Rightarrow y = z))$ (axiom 5)
2. $((\forall x)(\forall y)(x = y \Rightarrow (x = z \Rightarrow y = z))) \Rightarrow (\forall y)(x = y \Rightarrow (x = z \Rightarrow y = z))$ (axiom 6)

3. $(\forall y)(x = y \Rightarrow (x = z \Rightarrow y = z))$ (modus ponens on lines 1, 2)
4. $((\forall y)(x = y \Rightarrow (x = z \Rightarrow y = z))) \Rightarrow (x = y \Rightarrow (x = z \Rightarrow y = z))$ (axiom 6)
5. $x = y \Rightarrow (x = z \Rightarrow y = z)$ (modus ponens on lines 3, 4)
6. $x = y$ (hypothesis)
7. $x = z \Rightarrow y = z$ (modus ponens on lines 5, 6)
8. $x = z$ (hypothesis)
9. $y = z$ (modus ponens on lines 7, 8)

§4.5 Deduction theorem

Proposition 4.1 (Deduction Theorem)
Let S be a set of formulae, and p, q formulae. Then $S \vdash (p \Rightarrow q)$ iff $S \cup \{p\} \vdash q$.

Proof. ($\Rightarrow$): Write down a proof of $p \Rightarrow q$ from S, one can establish a proof of q from $S \cup \{p\} \vdash q$ by writing p and applying modus ponens to the original proof.

($\Leftarrow$): Let $t_1, \ldots, t_n = q$ be a proof of q from $S \cup \{p\}$. We prove that $S \vdash \{p \Rightarrow t_i\}$ by induction on p_i.
Induction hypothesis at step i: for $j < i$, $S \vdash (p \Rightarrow t_j)$ s.t. if the proof of t_j from $S \cup \{p\}$ did not use any premise in which a variable x occurs free, then the proof of $(p \Rightarrow t_j)$ from S does not use any premise where x occurs free.

To see $S \vdash (p \Rightarrow t_i)$ consider the following cases:

- If $t_i \in S$ or an axiom then write
  1. t_i (premise or axiom)
  2. $t_i \Rightarrow (p \Rightarrow t_i)$ (A1)
  3. $p \Rightarrow t_i$ (MP)
- If $t_i = p$ write down a proof of $p \Rightarrow p$ from $\varnothing$.
- If $\exists j, k < i$ s.t. $t_k = (t_j \Rightarrow t_i)$ then write
  1. $(p \Rightarrow (t_j \Rightarrow t_i)) \Rightarrow ((p \Rightarrow t_j) \Rightarrow (p \Rightarrow t_i))$ (A2)
  2. $p \Rightarrow t_k$ (Induction hypothesis)
  3. $(p \Rightarrow t_j) \Rightarrow (p \Rightarrow t_i)$ (MP)
  4. $p \Rightarrow t_j$ (Induction hypothesis)

5. $p \Rightarrow t_i$ (MP)

- Finally if $\exists j < i$ s.t. $x \in \mathrm{FV}(t_j)$ and $t_i = (\forall x)t_j$, then the proof of t_j from $S \cup \{p\}$ did not use any premise where x occurs free. There are two cases
  - If x occurs free in p, p did not occur in the proof of t_j from $S \cup \{p\}$ so it is a proof of t_j from S. So by (Gen), $S \vdash (\forall x)t_j$, i.e. $S \vdash t_i$. Write lines:
    1. $t_i \Rightarrow (p \Rightarrow t_i)$ (A1)
    2. $p \Rightarrow t_i$ (MP)
  - If x doesn't occur free in p, then we have a proof of $p \Rightarrow t_j$ by induction hypothesis which does not use any premise where x occurs free. So add the lines
    1. $(\forall x)(p \Rightarrow t_j)$ (Gen)
    2. $(\forall x)(p \Rightarrow t_j) \Rightarrow (p \Rightarrow (\forall x)t_j)$ (A7)
    3. $p \Rightarrow (\forall x)t_j$, i.e. $p \Rightarrow t_i$ (MP).

  In all cases, the condition on free variables in the induction hypothesis remains true.

□

Aim: Want to show $S \vdash p$ iff $S \models p$.

§4.6 Soundness

The proofs in this section are non-examinable.

Proposition 4.2 (Soundness Theorem)
Let S be a set of formulae and p a formula. If $S \vdash t$ then $S \models t$.

Proof. We have a proof $t_1, \ldots, t_n$ of p from S. We show that if A is a model of S, A is also a model of t_i for each i (interpreting free variables as quantified); this can be shown by induction. Hence, $S \models p$. □

§4.7 Adequacy

The proofs in this section are non-examinable.

We want to show that $S \models p$ implies $S \vdash p$. Equivalently, $S \cup \{\neg p\} \models \bot$ implies $S \cup \{\neg p\} \vdash \bot$. In other words, if $S \cup \{\neg p\}$ is consistent, it has a model.

Theorem 4.1 (Model Existence Lemma)

Every consistent[a] theory has a model.

[a] If S a consistent theory then $S \nvdash \bot$.

We will need a number of key ideas in order to prove this.

1. We will construct our model out of the language itself using the closed terms of L. For instance, if L is the language of fields and S is the usual field axioms, we take the closed terms and combine them with $+$ and $\cdot$ in the obvious way.
2. However, we can prove $S \vdash 1+0 = 1$, but $1+0$ and 1 are distinct as strings. We will therefore take the quotient of this set by the equivalence relation defined by $s \sim t$ if $S \vdash s = t$. If this set is A, we define $[s] +_A [t] = [s + t]$, and this is a well-defined operation.
3. Suppose S is the set of field axioms with the statement that $1+1 = 0 \vee 1+1+1 = 0$. In this theory, $S \nvdash 1 + 1 = 0$ and $S \nvdash 1 + 1 + 1 = 0$. Therefore, $[1 + 1] \neq [0]$ and $[1+1+1] \neq [0]$, so our structure A is not of characteristic 2 or 3. We can overcome this by first extending S to a maximal consistent theory.
4. Suppose S is the set of field axioms with the statement that $(\exists x)(x \cdot x = 1 + 1)$. There is no closed term t with the property that $[t \cdot t] = [1 + 1]$. The problem is that S lacks **witnesses** to existential quantifiers. For each statement of the form $(\exists x)p \in S$, we add a new constant c to the language and add to S the sentence $p[c/x]$. This still forms a consistent set.
5. The resulting set may no longer be maximal, as we have extended our language with new constants. We must then return to step (iii) then step (iv); it is not clear if this process ever terminates.

Proof. Let S be a consistent set in a language $L = L(\Omega, \Pi)$. Extend S to a maximal consistent set S_1, using Zorn's lemma. Then, for each sentence $p \in L$, either $p \in S_1$ or $\neg p \in S_1$. Such a theory is called **complete**; each sentence or its negation is proven. Now, we add witnesses to S_1: for each sentence of the form $(\exists x)p \in S_1$, we add a new constant symbol c to the language, and also add the sentence $p[c/x]$. We then obtain a new theory T_1 in the language $L_1 = L(\Omega \cup C_1 \Pi)$ that has witnesses for every existential in S_1. One can check easily that T_1 is consistent.

We then extend T_1 to a maximal consistent theory S_2 in L_1, and add witnesses to produce T_2 in the language $L_2 = L(\Omega \cup C_1 \cup C_2, \Pi)$. Continue inductively, and let $\overline{S} = \bigcup_{n \in \mathbb{N}} S_n$ in the language $\overline{L} = L(\Omega \cup \bigcup_{n \in \mathbb{N}} C_n, \Pi)$.

We claim that $\overline{S}$ is consistent, complete, and has witnesses for every existential in $\overline{S}$. Clearly $\overline{S}$ is consistent: if $\overline{S} \vdash \bot$ then $S_n \vdash \bot$ for some n as proofs are finite, contradicting consistency of S_n. For completeness, if p is a sentence in $\overline{L}$, p must lie in L_n for some n as it is a finite string of symbols. But S_{n+1} is complete in L_n, so $S_{n+1} \vdash p$ or $S_{n+1} \vdash \neg p$, so certainly $\overline{S} \vdash p$ or $\overline{S} \vdash \neg p$. If $(\exists x)p \in \overline{S}$, then $(\exists x)p \in S_n$ for some n, so T_n provides a witness.

On the closed terms of $\overline{L}$, we define the relation $s \sim t$ if $\overline{S} \vdash s = t$. This is clearly an equivalence relation, so we can define A to be the set of equivalence classes of $\overline{L}$ under $\sim$. This is an $\overline{L}$-structure by defining

- $f_A([t_1], \dots, [t_n]) = [f\ t_1 \dots t_n]$ for each $f \in \Omega \cup \bigcup_{n \in \mathbb{N}} C_n, \alpha(f) = n, t_i$ closed terms;
- $\varphi_A = \left\{([t_1], \dots, [t_n]) \in A^n \mid \overline{S} \vdash \varphi(t_1, \dots, t_n)\right\}$ for each $\varphi \in \Pi, \alpha(\varphi) = n, t_i$ closed terms.

We claim that for a sentence $p \in \overline{L}$, we have $p_A = 1$ iff $\overline{S} \vdash p$. Then the proof is complete, as $S \subseteq \overline{S}$ so $p_A = 1$ for every $p \in S$, so A is a model of S.

We prove this by induction on the length of sentences. First, suppose p is atomic. $\bot_A = 0$, as $\overline{S} \nvdash \bot$. For closed terms s, t, $\overline{S} \vdash s = t$ iff $[s] = [t]$ by definition of $\sim$. This holds iff $s_A = t_A$ by definition of the operations in A. This is precisely the statement that $s = t$ holds in A. The same holds for relations.

Now consider $p \Rightarrow q$. $\overline{S} \vdash p \Rightarrow q$ iff $\overline{S} \vdash \neg p$ or $\overline{S} \vdash q$ as $\overline{S}$ is complete and consistent; if $\overline{S} \nvdash \neg p$ and $\overline{S} \nvdash q$, then $\overline{S} \vdash p$ and $\overline{S} \vdash \neg p$. By induction on the length of the formula, this holds iff $p_A = 0$ or $q_A = 1$. This is the definition of the interpretation of $p \Rightarrow q$ in A.

Finally, consider the existential $(\exists x)p$. $\overline{S} \vdash (\exists x)p$ iff there is a closed term t s.t. $\overline{S} \vdash p[t/x]$, as $\overline{S}$ has witnesses to every existential. By induction (for example on the amount of quantifiers in a formula), this holds iff $p[t/x]_A = 1$ for some closed term t. This is true exactly when $(\exists x)p$ holds in A, as A is precisely the set of equivalence classes of closed terms. □

Corollary 4.1 (Adequacy)
Let S be a set of formulae and p a formula. If $S \models p$ then $S \vdash p$.

Proof. WLOG S is a theory and p is a sentence. Since $S \models p$, we have $S \cup \{\neg p\} \models \bot$. By Model Existence Lemma, $S \cup \{\neg p\} \vdash \bot$. So $S \vdash \neg\neg p$ by Deduction Theorem, and thus $S \vdash p$ by A3. □

§4.8 Completeness

Theorem 4.2 (Gödel's Completeness Theorem for First Order Logic)

If S is a set of formulae and p is a formula, then $S \vdash p$ iff $S \models p$.

Proof. Follows from soundness and adequacy. □

Note that **first order** refers to the fact that variables quantify over elements, rather than sets of elements.

Remark 36. If L is countable, or equivalently Ω and Π are countable, Zorn's lemma is not needed in the above proof.

Theorem 4.3 (Compactness Theorem)

Let S be a first-order theory. If every finite subset $S' \subseteq S$ has a model, S has a model.

Proof. If $S \models \bot$, then $S \vdash \bot$. Proofs are finite, so $\exists S' \subseteq S$ s.t. $S' \vdash \bot$. Hence $S' \models \bot$ $\lightning$. □

There is no decidability theorem for first order logic, as $S \models p$ can only be verified by checking its valuation in every L-structure.

Applications: Can we axiomatise finite groups? Does there exists theory T whose models are the finite groups?

For $n \in \mathbb{N}$, let $t_n = (\exists x_1) \ldots (\exists x_n)(\forall x)(x = x_1 \vee x = x_2 \vee \cdots \vee x = x_n)$. We want T to be the theory of groups $\cup\{t_1 \vee t_2 \ldots\}$. But $\{t_1 \vee t_2 \ldots\}$ is not a sentence.

Corollary 4.2

The class of finite groups is not axiomatisable as a first order theory.

Proof. Assume it is, and let T be such a theory. Consider $T' = T \cup \{\neg t_1, \neg t_2, \ldots\}$ where t_n are as above. If $\neg t_i$, then the group has at least i elements. Every finite subset of T' has a model : C_N[a] for some large N. By compactness, T' has a model $\lightning$ as this implies T has at least i elements for every i, so cannot be finite. □

[a] Cyclic group of order N.

Corollary 4.3

Let T be a first order theory with arbitrarily large finite models. Then T has an infinite model.

Proof. Consider $T' = T \cup \{(\exists x_1)(\exists x_2)(x_1 \neq x_2), (\exists x_2)(\exists x_2)(\exists x_3)(x_1 \neq x_2 \wedge x_2 \neq x_3 \wedge x_1 \neq x_3), \dots\}$. By assumption every finite subset of T has a model, so T' has a model. A model of T' is just an infinite model of T. □

Finiteness is not a first order property.

Theorem 4.4 (Upward Löwenheim–Skolem Theorem)
Let S be a first order theory with an infinite model. Then S has an uncountable model.

Proof. Add constants $\{c_i : i \in I\}$ to the language, where I is uncountable. Let $S' = S \cup \{\neg(c_i = c_j) : i, j \in I, i \neq j\}$. Any finite set of sentences in S' has a model: indeed, the infinite model of S suffices. By compactness, S' has a model. A model of S' is a model B of S together with an injection $I \to B$ so B is uncountable. □

Remark 37. Similarly, we can prove the existence of models of S that do not inject into X for any fixed set X. Adding $\gamma(X)$[1] constants or $\mathcal{P}(X)$ constants both suffice.

Example 4.25
There is an uncountable field, as there is an infinite field $\mathbb{Q}$. There is also a field that does not inject into X for any fixed set X.

Theorem 4.5 (Downward Löwenheim–Skolem Theorem)
Let S be a first order theory in a countable language L, or equivalently, Ω and Π are countable. Then if S has a model, it has a countable model.

Proof. S is consistent (by soundness), so the model constructed in the proof of the Model Existence Lemma is countable. □

§4.9 Peano Arithmetic

We want to axiomatise $\mathbb{N}$ as a first order theory.
Consider the language L given by $\Omega = \{0, s, +, \cdot\}$ with $\alpha(0) = 0, \alpha(s) = 1, \alpha(+) = \alpha(\cdot) = 2$, and $\Pi = \varnothing$.
Axioms of Peano Arithmetic (PA)

1. $(\forall x)(s(x) \neq 0)$;

[1] From Hartog's Lemma.

2. $(\forall x)(\forall y)(s(x) = s(y) \Rightarrow x = y)$;
3. $(\forall x)(x + 0 = x)$;
4. $(\forall x)(\forall y)(x + s(y) = s(x + y))$;
5. $(\forall x)(x \cdot 0 = 0)$;
6. $(\forall x)(\forall y)(x \cdot s(y) = x \cdot y + x)$.
7. $(\forall y_1) \dots (\forall y_n)[p[0/x] \wedge (\forall x)(p \Rightarrow p[s(x)/x]) \Rightarrow (\forall x)p]$ for each formula p with free variables $x, y_1, \dots, y_n$;

This is the axiom scheme for induction.

These axioms are sometimes called Peano arithmetic, PA, or formal number theory. The y_i in (7) are called **parameters**. Without the parameters, we would not be able to perform induction on sets such as $\{x : x \geq y\}$ if y is a variable.

Remark 38. Let p be the formula $x + (y + z) = (x + y) + z$. Then you can prove in PA that $(\forall x)(\forall y)(\forall z)p$ by induction on z with x, y parameters. You prove $(\forall x)(\forall y)(p[0/z] \wedge (\forall z)(p \Rightarrow p[\frac{sz}{z}]))$.

Note that $\mathbb{N}_0$ is a model of PA (so is $\mathbb{N}$). So by the upward Löwenheim–Skolem theorem, it has an uncountable model.
Didn't we learn $\mathbb{N}_0$ is uniquely determined by its properties?
Yes, but true induction says $(\forall A \subseteq \mathbb{N}_0)(((0 \in A) \wedge (\forall x)(x \in A \Rightarrow sx \in A)) \Rightarrow A = \mathbb{N}_0)$.
In first order theory we cannot quantify over subsets of structures. Axiom (7) applies only to countably many formulae p, and therefore only asserts that induction holds for countably many subsets of $\mathbb{N}_0$.

Definition 4.13 (Definable)

A subset $A \subseteq \mathbb{N}_0$ is **definable** in the language of PA if there is a formula p with a free variable x s.t. $p_{\mathbb{N}_0} = A$, i.e. $\{a \in \mathbb{N}_0 : a \text{ satisfies } p\} = A$.

Only countably many formulae exist, so only countably many sets are definable.

Example 4.26

The set of squares is definable, as it can be defined by the formula $(\exists y)(y \cdot y = x)$.
The set of primes is also definable by $x \neq 0 \wedge x \neq 1 \wedge (\forall y)((y \mid x) \Rightarrow y = 1 \wedge y = x)$, where $y \mid x$ is defined to mean $(\exists z)(z \cdot y = x)$.
The set of powers of 2 can be defined by $(\forall y)((y \text{ is prime} \wedge y \mid x) \Rightarrow y = 2)$.
The set of powers of 4 and the set of powers of 6 are also definable.

Theorem 4.6 (Gödel's Incompleteness Theorem)
PA is not complete.

This theorem shows that there is a sentence p s.t. PA $\nvdash p$ and PA $\nvdash \neg p$. However, one of $p, \neg p$ must hold in $\mathbb{N}_0$, so there is a sentence p that is true in $\mathbb{N}_0$ that PA does not prove. This does not contradict the completeness theorem, which is that if p is true in **every** model in PA then PA $\vdash p$.

§5 Set theory

§5.1 Axioms of ZF

In this section, we will attempt to understand the structure of the universe of sets. In order to do this, we will treat set theory as a first-order theory like any other, and can therefore study it with our usual tools. In particular, we will study a particular theory called **Zermelo–Fraenkel set theory**, denoted ZF. The language has $\Omega = \varnothing, \Pi = \{\in\}, \alpha(\in) = 2$. A 'universe of sets' is simply a model $(V, \in_V) = (V, \in)$ for the axioms of ZF. We can view this section as a worked example of the concepts of predicate logic, but every model of ZF will contain a copy of (most of) mathematics, so they will be very complicated.

A structure is a set V together with $[\in]_V \subset V \times V$. An element of V is called a 'set'. If $a, b \in V$ and $(a, b) \in [\in]_V$, say 'a belongs to b' or 'a is an element of b'.

We now define the axioms (there are 2 + 4 + 3 axioms) of ZF set theory.

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

   $$(\forall x)(\forall y)((\forall z)(z \in x \Leftrightarrow z \in y) \Rightarrow x = y)$$

   Note that the converse follows from the definition of equality. This implies that sets have no duplicate elements, and have no ordering.

2. **Axiom of Separation (Sep)** or **Comprehension**
   'We can have subsets of sets'

   For a set x and a property p, we can form the set of $z \in x$ s.t. $p(z)$ holds.

   $$(\forall t_1) \dots (\forall t_n)(\forall x)(\exists y)(\forall z)(z \in y \Leftrightarrow z \in x \wedge p)$$

   where the t_i are the parameters, and p is a formula with $\mathrm{FV}(p) = \{z, t_1, \dots, t_n\}$. By (Ext) the set y who existence is asserted is unique, we denote it by $\{z \in x : p\}$. (Formally we introduce an $(n+1)$-arity operation symbol to the language; informally this is an abbreviation).

   Example 5.1
   Note that we need the parameters as we may wish to form the set $\{z \in x : z \in t\}$ for some variable t.

3. **Empty-set Axiom (Emp)**

   $$(\exists x)(\forall y)(\neg y \in x)$$

   This empty set is unique by extensionality and denoted by $\varnothing$. Formally, we add a constant $\varnothing$ to the language with the sentence $(\forall y)(\neg y \in \varnothing)$.

Example 5.2

For instance, $p(\varnothing)$ is the sentence $(\exists x)((\forall y)(\neg y \in x) \wedge p(x))$.

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)**
   'We can form unordered pairs'

$$(\forall x)(\forall y)(\exists z)(\forall t)((t \in z) \Leftrightarrow (t = x \vee t = y))$$

We write $\{x, y\}$ for this set z, which is unique by Ext. $(\forall x)(\forall y)(\{x, y\} = \{y, x\})$ can be proved.

Some basic set-theoretic principles can now be defined.

- We write $\{x\} = \{x, x\}$ for the singleton set containing x.
- We can now define the ordered pair $(x, y) = \{\{x\}, \{x, y\}\}$; from the axioms so far we can prove that $(\forall x)(\forall y)(\forall t)(\forall z)((x, y) = (t, z) \iff (x = t \wedge y = z))$.

We introduce abbreviations

- "x is an ordered pair" if $(\exists y)(\exists z)(x = (y, z))$.
- "f is a function" if $(\forall x)(x \in f \Rightarrow x$ is an ordered pair$) \wedge (\forall x)(\forall y)(\forall z)((x, y) \in f \wedge (x, z) \in f \Rightarrow y = z)$
- We call a set x the domain of f, written $x = \operatorname{dom} f$, if f is a function and

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

- The notation $f\colon x \to y$ means that f is a function, if $x = \operatorname{dom} f$ and

$$(\forall z)(\forall t)((z, t) \in f \Rightarrow t \in y)$$

5. **Union Axiom (Un)**
   For each family of sets x, we can form its union $\bigcup_{t \in x} t$.

$$(\forall x)(\exists y)(\forall z)(z \in y \Leftrightarrow (\exists t)(z \in t \wedge t \in x))$$

The set y guaranteed by this axiom can be written $\bigcup x$, and we can write $x \cup y$ for $\bigcup\{x, y\}$.

Remark 39. No new axiom needed for intersection as this can be formed by (Sep)[2]. So the following line follows from the axioms so far:

$$(\forall x)(\neg x \in \varnothing \Rightarrow (\exists y)(\forall z)(z \in y \iff (\forall t)(t \in x \Rightarrow z \in t)))$$

Denote set y by $\cap x$.

To prove this, given x, form $y = \{z \in \cup x : (\forall t)(t \in x \Rightarrow z \in t)\}$ by (Sep). Check that $(\forall z)(z \in y \iff (\forall t)(t \in x \Rightarrow z \in t))$.

Given x, y, denote $z \in \cap\{x, y\} \iff (z \in x \wedge z \in y)$ by $z \in x \cap y$.

6. **Power-set Axiom (Pow)**

$$(\forall x)(\exists y)(\forall z)(z \in y \Leftrightarrow z \subseteq x)$$

where $z \subseteq x$ is an abbreviation for $(\forall t)(t \in z \Rightarrow t \in x)$. Denote y by $\mathcal{P}(x)$.

We can form the Cartesian product $x \times y$ for sets x, y: an element of $x \times y$ is an ordered pair (s, t) where $s \in x$ and $t \in y$. Note that $(s,t) = \left\{ \underbrace{\{x\}}_{\in \mathcal{P}x}, \underbrace{\{x, y\}}_{\in \mathcal{P}(x \cup y)} \right\} \in \mathcal{P}(\mathcal{P}(x \cup y))$, so by (Sep) we can form $\{z \in \mathcal{P}(\mathcal{P}(x \cup y)) : (\exists s)(\exists t)(s \in x \wedge t \in y \wedge z = (s,t))\}$.

We can also form, from sets x, y, $y^x = \{f \in \mathcal{P}(x \times y) : f : x \to y\}$ which is the set of all functions from $x \to y$.

7. **Axiom of Infinity (Inf)**
Using our currently defined axioms, any model V must be infinite, e.g. $\varnothing, \mathcal{P}\varnothing, \mathcal{P}\mathcal{P}\varnothing, \dots$ are all distinct elements of V.

Definition 5.1 (Successor)
For a set x, the **successor** of x is $x^+ = x \cup \{x\}$.

Example 5.3
Then $\varnothing, \varnothing^+, \varnothing^{++}, \dots$ are distinct elements of V.

$$\varnothing^+ = \{\varnothing\}; \quad \varnothing^{++} = \{\varnothing, \{\varnothing\}\}; \quad \varnothing^{+++} = \{\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}\}; \quad \dots$$

[2] This cannot be used to create empty intersections, as the (Sep) can only create subsets of a set that already exists.

We write $0 = \varnothing, 1 = \varnothing^+, 2 = \varnothing^{++}, \ldots$ for the successors created in this way. For instance, $3 = \{0, 1, 2\}$. We have a copy of $\mathbb{N}_0$ in V. From the outside, V is infinite. From the inside, V is not a set: $\neg(\exists x)(\forall y)(y \in x)$ (Russell's paradox) (If such x exists, form $\{y \in x : \neg y \in y\} = z\ \lightning$).

Abbreviate "x is a successor set": $\varnothing \in x$ and $(\forall y)(y \in x \Rightarrow y^+ \in x)$. Note that this is a finite-length formula that characterises an infinite set. The axiom of infinity is that there exists a successor set.

$$(\exists x)(\varnothing \in x \wedge (\forall y)(y \in x \Rightarrow y^+ \in x))$$

Note that this set is not uniquely defined, but any intersection of successor sets is a successor set. So we can construct "smallest" successor set, i.e. we can prove

$$(\exists x)((x \text{ is a successor set} \wedge (\forall y)(y \text{ is a successor set} \Rightarrow x \subseteq y)))$$

[Pick any successor set z, let $x = \cap\{y \in \mathcal{P}(z) : y \text{ is a successor set}\}$. x is then a successor set and if y is any successor set then $x \subset (y \cap z)$]. We denote the smallest successor set by ω.

If $x \subseteq \omega$ is a successor set then $x = \omega$, i.e. $(\forall x)(x \subseteq \omega \wedge \varnothing \in x \wedge (\forall y)(y \in x \Rightarrow y^+ \in x) \Rightarrow x = \omega)$. This is true induction.

It is easy to check that $(\forall x)(x \in \omega \Rightarrow x^+ \neq \varnothing)$ and $(\forall x)(\forall y)((x \in \omega \wedge y \in \omega \wedge x^+ = y^+) \Rightarrow x = y)$, so ω satisfies (in V) the usual axioms for the natural numbers.

We can define abbreviations:
"x is finite" for $(\exists y)(y \in \omega \wedge (\exists f)(f : x \to y \wedge \text{'}f \text{ is bijective'}))$;
"x is countable" for $(\exists f)(f : x \to \omega \wedge \text{'}f \text{ is injective'})$.

8. **Axiom of Replacement (Rep)**
(Inf) says that there exist sets containing $0, 1, 2, 3, \ldots$ Are there sets containing $\varnothing, \mathcal{P}\varnothing, \mathcal{P}\mathcal{P}\varnothing, \ldots$ There is a function-like[3] object that sends $0 \mapsto \varnothing$, $1 \mapsto \mathcal{P}\mathcal{P}\varnothing$, ...

We need an axiom that the image of a set under a function-like object is a set.

Definition 5.2 (Class)

A **class** is a subset C of a structure V of the language of ZF s.t. $\exists$ formula p with $\mathrm{FV}(p) = \{x\}$ s.t. $p_V = C$, i.e. $x \in C \iff p(x)$ holds in V.

C is a set outside of our model; it may not correspond to a set $x \in V$ inside the model.

Example 5.4

- For instance, V is a class, taking p to be $x = x$.

[3] We have not yet defined functions.

- The set of sets of size 1 is a class, e.g. take p to be $(\exists y)(x = \{y\})$
- There is a class of infinite sets, taking p to be 'x is not finite'.
- For any $t \in V$, the collection of x with $t \in x$ is a class; here, t is a parameter to the class.
- Every set $y \in V$ is a class by setting p to be $x \in y$.

Say the class C is a **set** if $(\exists y)(\forall x)(x \in y \iff p)$ holds in V.

Definition 5.3 (Proper Class)
If C is not a set, we say C is a **proper class**

Example 5.5
V is a proper class by Russell's paradox.

When writing about classes inside ZF, we instead write about their defining formulae, as classes have no direct representation in the language.

Definition 5.4 (Function-Class)
A **function-class** is a subset G of $V \times V$ s.t. $\exists$ formula p with $\mathrm{FV}(p) = \{x, y\}$ s.t. $(\forall x)(\forall y)(\forall z)((p \wedge p[z/y]) \Rightarrow y = z)$ holds in V and $G = p_V$.
I.e. $(x, y) \in G \iff p(x, y)$ holds in V.

Example 5.6
$G = \{(x, \{x\}) : x \in V\}$ is the fcn class that maps x to $\{x\}$ and is given by $p : (y = \{x\})$.

This is intuitively a function whose domain may not be a set. This is not a function, for example, every f has a domain which is a set in V, and this function has domain V which is not a set.

We can now define the axiom of replacement: the image of a set under a function-class is a set.

$$(\forall t_1) \dots (\forall t_n)[(\forall x)(\forall y)(\forall z)(p \wedge p[z/y] \Rightarrow y = z) \Rightarrow$$
$$(\forall x)(\exists y)(\forall z)(z \in y \Leftrightarrow (\exists t)(t \in x \wedge p[t/x, z/y]))]$$

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

9. **Axiom of Foundation (Fnd)** or **regularity**
We require that sets are built out of simpler sets. For example, we want to disallow a set from being a member of itself, and similarly forbid $x \in y$ and $y \in x$. In general, we want to forbid sets x_i s.t. $x_{i+1} \in x_i$ for each $i \in \mathbb{N}$.

Note that if $x \in x$, $\{x\}$ has no $\in$-minimal element. If $x \in y, y \in x$, $\{x, y\}$ has no $\in$-minimal element. In the last example, $\{x_0, x_1, \dots\}$ has no $\in$-minimal element.

We now define the axiom of foundation: every nonempty set has an $\in$-minimal element.

$$(\forall x)(x \neq \varnothing \Rightarrow (\exists y)(y \in x \wedge (\forall z)(z \in x \Rightarrow z \notin y)))$$

Any model of ZF without this axiom has a submodel of all of ZF.

This completes the description of the axioms of ZF. We write ZFC for ZF + AC, where AC is the axiom

$$(\forall x)(\forall y)(y \in x \Rightarrow y \neq \varnothing) \Rightarrow (\exists f)(f : x \to \cup x) \wedge (\forall y)(y \in x \Rightarrow f(y) \in y)$$

For the rest of this section we work in ZF.

Aim: Describe the set-theoretic universe, i.e. any model V of ZF.

§5.2 Transitive sets

Definition 5.5 (Transitive)
x is **transitive** if each member of a member of x is a member of x.

$$(\forall y)(\underbrace{(\exists z)(y \in z \wedge z \in x)}_{y \in \cup x} \Rightarrow y \in x)$$

Equivalently, $\cup x \subseteq x$.

Warning 5.1
This is not the same as saying $\in$ is a transitive relation on x.

Example 5.7
$\varnothing$ is a transitive set. $\{\varnothing\}$ is also transitive, and $\{\varnothing, \{\varnothing\}\}$ is transitive.

Example 5.8

ω is transitive.

Proof. NTS that $x \subseteq \omega \ \forall x \in \omega$.
Form the set $z = \{y \in \omega : y \subseteq \omega\}$ by (Sep). Check z is a successor set, so $z = \omega$, i.e. ω transitive.

Similarly, $\{x \in \omega : \text{'}x \text{ is transitive'}\}$ is also a successor set $(\cup x^+ = x)$ so it is ω. So every element of ω is a transitive set. □

Lemma 5.1

Every set is contained in a transitive set, i.e. $(\forall x)(\exists y)(\text{'}y \text{ is transitive'} \wedge x \subseteq y)$.

Remark 40. This proof takes place inside an arbitrary model of ZF. Technically, the statement of the lemma is 'let $(V, \in)$ be a model of ZF, then for all sets $x \in V$, x is contained in a transitive set $y \in V$'. By completeness, this will show that there is a proof of this fact from the axioms of ZF.

Note also that once this lemma is proven, any x is contained in a least transitive set by taking intersections, called its **transitive closure**, written $TC(x)$. This holds as any intersection of transitive sets is transitive.

Idea: If $x \subseteq y$ and y transitive then $\bigcup x \subseteq y$ so $\bigcup\bigcup x \subseteq y$, $\bigcup\bigcup\bigcup x \subseteq y$, ...
We want to form $\bigcup\{x, \bigcup x, \bigcup\bigcup x, \dots\}$, this is a set by (Rep). We need a function-class $0 \mapsto x$, $1 \mapsto \bigcup x$, $2 \mapsto \bigcup\bigcup x$, ...

Proof. We want to define the function-class $p(z, w)$ to be $(z = 0 \wedge w = x) \vee ((\exists t)(\exists u) z = t^+ \wedge w = \bigcup u \wedge p(t, u))$, but this is not a first-order formula.

Define that f is an **attempt** to mean that

$$(f \text{ is a function}) \wedge (\text{dom } f \in \omega) \wedge (\text{dom } f \neq \varnothing) \wedge (f(0) = x) \wedge$$
$$(\forall m)(\forall n)\Big((m \in \text{dom } f \wedge n \in \text{dom } f \wedge n = m^+) \Rightarrow f(n) = \bigcup f(m)\Big)$$

Then,

$$(\forall n)(n \in \omega \Rightarrow (\exists f)(f \text{ is an attempt} \wedge n \in \text{dom } f)) \quad (**)$$

can be proven by ω-induction. We can similarly prove

$$(\forall f)(\forall g)(\forall n)((f, g \text{ are attempts} \wedge (n \in \text{dom } f \cap \text{dom } g)) \Rightarrow f(n) = g(n)) \quad (*)$$

by ω-induction. We now define the function-class $p = p(y, z)$ to be

$$(\exists f)(f \text{ is an attempt} \wedge y \in \text{dom } f \wedge f(y) = z).$$

By $(*)$ we have $(\forall y)(\forall z)(\forall w)((p \wedge p[w/z]) \Rightarrow w = z)$. By (Rep) we can form $w = \{z : (\exists y)(y \in \omega \wedge p(y, z))\}$ which is $\{x, \bigcup x, \bigcup\bigcup x, \dots\}$. Also by (Un) we can form $t = \bigcup w$.

Then $x \in t$, since $x \in w$ ($\{(0, x)\}$ is an attempt). Given $a \in t$, we have $a \in z$ for some $z \in w$. Then there's an attempt f and $n \in \operatorname{dom} f$ s.t. $z = f(n)$.
By $(**)$ there's an attempt g with $n^+ \in \operatorname{dom} g$. Then $n \in \operatorname{dom} g$[a] so $\bigcup z = \bigcup f(n) = \bigcup g(n)$ by $(*)$ and $\bigcup g(n) = g(n^+) \in w$. Thus for any $b \in a$, $b \in \cup z \in w$ so $b \in t$, i.e. t transitive. □

[a] $\operatorname{dom} g$ transitive as members of ω transitive and $n \in n^+$.

Transitive closures allow us to pass from the large universe of sets, which is not a set itself, into a smaller world which is a set closed under $\in$ that contains the relevant sets in question.

§5.3 $\in$-induction

We want the axiom of foundation to capture the idea that sets are built out of simpler sets.

Theorem 5.1 (Principle of $\in$-induction)
For each formula p with free variables $x, t_1, \dots, t_n$,

$$(\forall t_1) \dots (\forall t_n)[(\forall x)[(\forall y)(y \in x \Rightarrow p(y)) \Rightarrow p(x)] \Rightarrow (\forall x)p(x)]$$

where $p(x) = p$ and $p(y) = p[y/x]$.

Proof. Fix $t_1, \dots, t_n$ and assume $(\forall x)((\forall y)(y \in x \Rightarrow p(y)) \Rightarrow p(x))$ holds, we want to show $(\forall x)p(x)$. Assume not, i.e. $\exists x$ s.t. $\neg p(x)$. We want to look at the set $\{t : \neg p(t)\}$ and take an $\in$-minimal element, but this may not be a set, e.g. for $p = (x \neq x)$.

Choose a transitive set t s.t. $x \in t$, e.g. $t = TC(\{x\})$. By (Sep) form the set $u = \{y \in t : \neg p(y)\}$; this is clearly a set in the model, and $u \neq \varnothing$ as $x \in u$. Let z be an $\in$-minimal element of u (exists by (Fnd)). If $y \in z$, then $y \in t$ (as t transitive) and $y \notin u$ (by minimality), so $p(y)$. By assumption $p(z)$ holds ↯ of $z \in u$. □

The name of this theorem should be read 'epsilon-induction', even though the membership relation is denoted $\in$ and not ε.

The principle of $\in$-induction is equivalent to the axiom of foundation (Fnd) in the presence of the other axioms of ZF.

<u>Clever Idea</u>: We say that x is **regular** if $(\forall y)(x \in y \Rightarrow y$ has a $\in$-minimal element) (this defn is the clever part). The axiom of foundation is equivalent to the assertion that

every set is regular. Given $\in$-induction, we can prove every set is regular. Fix some x and suppose $(\forall y \in x)(y \text{ is regular})$; we need to show x is regular. For a set z with $x \in z$, if x is minimal in z, then z has a minimal element. If x is not minimal in z, there exists $y \in x$ s.t. $y \in z$. So z has a minimal element as y is regular. Hence x is regular.

§5.4 $\in$-recursion

We want to be able to define $f(x)$ given $f(y)$ for all $y \in x$, i.e. $f(x)$ depends on $f|_x$.

Theorem 5.2 ($\in$-recursion Theorem)
Let G be a function-class, i.e. $(x, y) \in G$ iff $p(x, y)$ for some formula p. Suppose that G is defined for all sets. Then there is a function-class F defined for all sets by a formula $q(x, y)$ s.t.

$$(\forall x)\left(F(x) = G\left(F\Big|_x\right)\right)$$

Moreover, this F is unique.

Note that $F|_x = \{(y, F(y)) : y \in x\}$ is a set by (Rep) and is the image of the set x under the fcn class $s \mapsto (s, F(s))$.

Proof. Existence: Define that f is an **attempt** if

$$f \text{ is a function} \wedge \operatorname{dom} f \text{ is transitive} \wedge (\forall x)\left(x \in \operatorname{dom} f \Rightarrow f(x) = G\left(f\Big|_x\right)\right) \quad (*)$$

Note that $f|_x$ is defined as $\operatorname{dom} f$ is transitive. Then,

$$(\forall x)(\forall f)(\forall f')(f, f' \text{ are attempts} \wedge x \in \operatorname{dom} f \cap \operatorname{dom} f' \Rightarrow f(x) = f'(x))$$

by $\in$-induction: if $f(y) = f'(y)$ for all $y \in x$, then $f(x) = f'(x)$. Also,

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

by $\in$-induction.

Fix x. Assume every $y \in x$ is in the domain of some attempt, which is then defined on $TC(\{y\})$ and unique by $(*)$, say f_y. Then $f' = \bigcup \underbrace{\{f_y : y \in x\}}_{\text{a set by (Rep)}}$ is an attempt by $(*)$.
Finally, $f = f' \cup \{(x, G(f'))\}$ is an attempt defined at x. Note that $f|_x = f'$.
We can then take $q(x, y)$ to be the formula

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

This defines the function-class F as required.

Uniqueness: Assume F, F' satisfy the theorem, then we have $(\forall x)(F(x) = F'(x))$ by $\in$-induction. If $F(y) = F'(y)$ $\forall y \in x$, then $F|_x = F'|_x$, then $F(x) = F'(x)$. □

§5.5 Well-founded relations

Note the similarity between the proofs of $\in$-induction and $\in$-recursion and the proofs of induction and recursion on ordinals. These proofs are not specific to the relation $\in$; we only used some of its properties.

Definition 5.6 (Well-Founded)

r is **well-founded** if $(\forall x)(x \neq \varnothing \Rightarrow (\exists y)(y \in x \wedge (\forall z)(z \in x \Rightarrow \neg zry)))$, i.e. every nonempty set has an r-minimal element.

Example 5.9

If r is $(x \in y)$, then r is the $\in$-relation which is well-founded by (Fnd).

Definition 5.7 (Local)

r is **local** if $(\forall x)(\exists y)(\forall z)(z \in y \wedge zrx$[a]$)$, i.e. the r-predecessors of x form a set.

[a] $zrx = r(z, x)$

Example 5.10

Clearly $\in$.

'local' is needed for r-closure. Then we can prove r-induction and r-recursion.

We can restrict r to a class or set. If r is a relation on a set a, then for any $x \in a$, $\{y \in a : yrx\}$ is a set by (Sep). So we only need well-foundedness to have r-induction and r-recursion on a.

Is this really more general than $\in$? No, provided we also assume that r is **extensional** on a.

Definition 5.8 (Extensional)

r is **extensional** on a if

$$(\forall x \in a)(\forall y \in a)((\forall z \in a)(zrx \Leftrightarrow zry) \Rightarrow x = y)$$

This is just the axiom of extensionality applied to the relation r.

Theorem 5.3 (Mostowski's Collapsing Theorem)
Let r be a well-founded and extensional relation on a set a. Then, $\exists$ a transitive set b and a bijection $f\colon a \to b$ s.t.

$$(\forall x \in a)(\forall y \in a)(xry \Leftrightarrow f(x) \in f(y))$$

Moreover, b and f are unique.

This is an analogue of subset collapse from the section on ordinals. Transitive sets are playing the role of initial segments. Note that the well-foundedness and extensionality conditions are clearly necessary for the theorem, consider $(\mathbb{Z}, <)$ or $(\{a, b, c, \}, <)$ with $a < b, a < c$ for counterexamples.

Proof. By r-recursion on a, there's a function class f s.t. $\forall x \in a$, $f(x) = \{f(y) : y \in a \wedge yrx\}$. Note that f is a function, not just a fcn class since $\{(x, f(x)) : x \in a\}$ is a set by (Rep).

Then $b = \{f(x) : x \in a\}$ is a set by (Rep).
b is transitive: Let $z \in b$ and $w \in z$. There's $x \in a$ s.t. $z = f(x)$, and so a $y \in a$ s.t. yrx and $w = f(y) \in b$.

Clearly f surjective and $\forall x, y \in a$, $xry \Rightarrow f(x) \in f(y)$.

It remains to show that f is injective, it will then follow that $\forall x, y \in a$, $f(x) \in f(y) \Rightarrow xry$. Indeed, if $f(x) \in f(y)$, then $f(x) = f(z)$ for some $z \in a$ with zry. Since f is injective, $x = z$ so xry.

We will show

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

by r-induction on x.

Fix $x \in a$ and assume f is injective at s whenever $s \in a$ and srx. Assume $f(x) = f(y)$ for some $y \in a$, i.e. $\{f(s) : s \in a \wedge srx\} = \{f(t) : t \in a \wedge try\}$. Since f is injective at every $s \in a$ with srx, it follows that $x = y$.

Uniqueness: Assume (b, f) and (b', f') both satisfy the theorem. We prove $(\forall x \in a)(f(x) = f'(x))$ by r-induction. Fix $x \in a$ and assume $f(y) = f'(y)$ whenever $y \in a \wedge yrx$. If $z \in f(x)$ then $z \in b$ (b transitive), so $z = f(y)$ for some $y \in a$ with yrx. Then $z = f(y) = f'(y)$ (induction hypothesis). Then $z = f'(y) \in f'(x)$. Similarly, if $z \in f'(x)$ then $z \in f(x)$. By (Ext), $f(x) = f'(x)$. $\square$

In particular, every well-ordered set has a unique order isomorphism to a unique transitive set well-ordered by $\in$.

Definition 5.9 (Ordinal)

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

Note that, if a a set and r a well-ordering on a, then r is well-founded and extensional (if $x, y \in a$ and $x \neq y$, then xry or yrx, but not both). By Mostowski, $\exists$ a transitive set b and bijection $f : a \to b$ s.t. $xry \iff f(x) \in f(y)$, i.e. $f : (a, r) \to (b, \in)$ is an order isomorphism. So b is an ordinal. So by Mostowski, every well-ordered set, x, is order-isomorphic to a unique ordinal, called the order type of x.

We let ON denote the class of ordinals (given by formula 'x is an ordinal'). It is a proper class by Buralti-Forti.

Proposition 5.1

Let $\alpha, \beta \in ON$ and a a set of ordinals.

1. Every member of α is an ordinal
2. $\beta \in \alpha \iff \beta$ is o.i. to an i.s. of α ($\beta < \alpha$).
3. $\alpha \in \beta$ or $\alpha = \beta$ or $\beta \in \alpha$.
4. $\alpha^+ = \alpha \cup \{\alpha\}$ ($+$ in the sense of section 2).
5. $\cup a$ is an ordinal and $\cup a = \sup a$.

Remark 41. (2) says that α really <u>is</u> the set of ordinals $< \alpha$;
(3) says that $\in$ linearly orders the class ON;
(4) resolves the clash of notation x^+ in section 2 and 5. According to the definition in section 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). (4) shows that this ordinal is the successor of the set α as defined in this section. In particular, this shows that the successor of an ordinal is an ordinal;
(5) now shows that any set of well-ordered sets has an upper bound.

Proof.

1. Let $\gamma \in \alpha$. Then $\gamma \subseteq \alpha$ (α is transitive) and hence $\in$ linearly orders γ[a]. Given $\eta \in \delta$, $\delta \in \gamma$ then $\delta \in \alpha$ and so $\eta \in \alpha$ (α is transitive). Since $\in$ is transitive[b] on α, we have $\eta \in \gamma$. So γ is a transitive set, so γ is an ordinal.
2. If $\beta \in \alpha$, then $I_\beta = \{\gamma \in \alpha : \gamma \in$ [c]$\beta\} = \beta$ as $\beta \subset \alpha$ by transitivity of α, so $\beta < \alpha$. Any proper i.s. of α is of the form I_γ for some $\gamma \in \alpha$. So $\beta < \alpha \Rightarrow \beta \in \alpha$.
3. Done by (2) and proposition 2.4, i.e. $<$ a linear ordering.
4. Let $\beta = \alpha \cup \{\alpha\}$(successor of α in section 5). If $\gamma \in \beta$, then either $\gamma = \alpha \subseteq \beta$

or $\gamma \in \alpha$, so $\gamma \subseteq \alpha \subseteq \beta \Rightarrow \beta$ is transitive, linearly ordered by $\in$ (by (3)), and α is the greatest element. So $\beta = \alpha^+$ in section 2.

5. $\cup a$ is a union of transitive sets, hence transitive. Every member of $\cup a$ is an ordinal, so $\cup a$ linearly ordered by $\in$ (by (3)).
   If $\gamma \in a$, then $\gamma \subseteq \cup a$ so either $\gamma = \cup a$ or $\gamma \in \cup a$ (by (3)), i.e. $\gamma \leq \cup a$.
   If $\gamma \leq \delta$ for all $\gamma \in a$, then $\gamma = \delta$ or $\gamma \in \delta$ for all $\gamma \in a$, i.e. $\gamma \subseteq \delta$ (by (3)).
   So $\cup a \subseteq \delta$ i.e. $\cup a \leq \delta$. [If $\gamma < \cup a$, then $\gamma \in \cup a$, i.e. $\gamma \in \delta$ for some $\delta \in a$, i.e. $\gamma < \cup a$ (by (3)), so γ is not an upper bound for a.]

□

[a] $\in$ linearly orders α and γ a subset so $\in\big|_\gamma$ linearly orders it.
[b] As $\in$ a well ordering.
[c] We are well-ordered by $\in$ not $<$ so we use $\in$ to define initial segments

Example 5.11

$0 = \varnothing \in ON$, hence $n \in ON$ for all $n \in \omega$ (by ω-induction). ω is transitive, so $\cup\omega \subseteq \omega$. If $n \in \omega$, then $n \in n^+ \subseteq \omega$, so $n \in \cup\omega$. So $\omega \subseteq \cup\omega \Rightarrow \omega = \cup\omega$ is an ordinal and $\omega = \sup \omega$.

§5.6 The universe of sets

Idea: Everything is built up from $\varnothing$ using $\mathcal{P}$ and $\cup$. We can have $V_0 = \varnothing$, $V_1 = \mathcal{P}\varnothing = \{\varnothing\}$, $V_2 = \mathcal{P}\mathcal{P}\varnothing = \{\varnothing, \{\varnothing\}\}$, ..., $V_\omega = \cup\{V_0, V_1, \dots\}$, $V_{\omega+1} = \mathcal{P}V_\omega$ etc.

It will be (Fnd) that guarantees that every set appears in a V_α.

Define sets V_α for each ordinal $\alpha \in ON$ by $\in$-recursion:

- $V_0 = \varnothing$;
- $V_{\alpha+1} = \mathcal{P}(V_\alpha)$;
- $V_\lambda = \bigcup\{V_\alpha : \alpha < \lambda\}$ for a nonzero limit ordinal λ.

This can be viewed as a well-founded recursion on ordinals, or $\in$-recursion on the universe but mapping non-ordinals to $\varnothing$. The sets V_α form the **von Neumann Hierarchy**.

Aim: Show that every set is contained in some V_α.

Lemma 5.2

V_α is transitive $\forall \alpha \in ON$.

Proof. We show this by induction on α.
Clearly $V_0 = \varnothing$ is transitive.
Suppose V_α is transitive. If $x \in V_{\alpha^+}$ then $x \subseteq V_\alpha$. If $y \in x$, then $y \in V_\alpha$. So by the induction hypothesis, $y \subseteq V_\alpha$. So every $y \in x$ has $y \in \mathcal{P}(V_\alpha) = V_{\alpha^+}$ thus V_{α^+} is transitive.
Now suppose $\lambda \neq 0$ is a limit ordinal, if $x \in V_\lambda$ then $\exists \gamma < \alpha$ s.t. $x \in V_\gamma$. By the induction hypothesis, V_γ is transitive, so $x \subseteq V_\gamma \subseteq V_\lambda$. □

Lemma 5.3

Let $\alpha \leq \beta$. Then $V_\alpha \subseteq V_\beta$.

Proof. We show this by induction on β for a fixed α.
If $\beta = 0$, $\alpha \leq \beta \Rightarrow \alpha = 0$ so $V_\alpha = V_\beta$.
If $\beta = \gamma^+$: If $\alpha = \beta$ then $V_\alpha = V_\beta$.
If $\alpha < \beta$, then $\alpha \leq \gamma$, so by IH $V_\alpha \subseteq V_\gamma$. If $x \in V_\gamma$ then $x \subseteq V_\gamma$ (V_γ transitive) so $x \in \mathcal{P}(V_\gamma) = V_\beta$. Thus $V_\gamma \subseteq V_{\gamma^+} = V_\beta$ and hence $V_\alpha \subseteq V_\beta$.
If $\beta \neq 0$ a limit and $\alpha < \beta$ then $V_\alpha \subseteq V_\beta$ by defn. Limits are trivial. □

Theorem 5.4

The von Neumann hierarchy exhausts the set-theoretic universe V[a], i.e. $(\forall x)(\exists \alpha \in ON)(x \in V_\alpha)$. 'Every set x belongs to V_α for some α'.

[a] $V = \cup_{\alpha \in ON} V_\alpha$.

If we could construct the set V defined as the union of the V_α over all ordinals α, V would be a model of ZF.

Remark 42. If $x \in V_\alpha$ then $x \subseteq V_\alpha$ by lemma 5.2.
If $x \subseteq V_\alpha$ then $x \in \mathcal{P}(V_\alpha) = V_{\alpha+1}$. If $\exists \alpha \in ON$ s.t. $x \subseteq V_\alpha$ define the **rank** of x to be the least such α.

For example, the rank of $\varnothing$ is 0, the rank of 1 is 1, the rank of ω is ω, and in general the rank of any ordinal α is α. Intuitively, the rank of a set is the time at which it was created.

Proof. We proceed by $\in$-induction on x. Fix x and assume $\forall y \in x$, $y \subseteq V_\alpha$ for some $\alpha \in ON$. So $\forall y \in x$, $y \subseteq V_{\mathrm{rank}(y)}$.
Let $\alpha = \sup \underbrace{\left\{\mathrm{rank}(y)^+ : y \in x\right\}}_{\text{a set by (Rep)}}$.
We'll show $x \subseteq V_\alpha$. If $y \in x$, then $y \subseteq V_{\mathrm{rank}(y)}$, so $y \in \mathcal{P}(V_{\mathrm{rank}(y)}) = V_{\mathrm{rank}(y)^+} \subseteq V_\alpha$ by lemma 5.3. So $x \subseteq V_\alpha$ as required. □

Corollary 5.1

For every set x, $\text{rank}(x) = \sup\{\text{rank}(y)^+ : y \in x\}$.

Proof. "$\leq$' follows form the proof above.
"$\geq$': We first show that $x \in V_\alpha \Rightarrow \text{rank}(x) < \alpha$.
$\alpha = 0$: ✓
$\alpha = \beta^+$: $x \in \mathcal{P}(V_\beta)$ so $x \subseteq V_\beta$ so $\text{rank}(x) \leq \beta < \alpha$. ✓
$\alpha \neq 0$ a limit: $x \in V_\alpha \Rightarrow \exists\gamma < \alpha$ s.t. $x \in V_\gamma$ so $\text{rank}(x) < \gamma < \alpha$. ✓.

Now let $\alpha = \text{rank}(x)$. Then $x \subseteq V_\alpha$, so for $y \in x$, $y \in V_\alpha$ and so $\text{rank}(y) < \alpha$. Hence $\sup\{\text{rank}(y)^+ : y \in x\} \leq \alpha$. □

Example 5.12

$\text{rank}(\alpha) = \alpha \quad \forall\alpha \in ON$.
By induction:

$$\begin{aligned}\text{rank}(\alpha) &= \sup\left\{\text{rank}(\beta)^+ : \beta < \alpha\right\} \\ &= \sup\left\{\beta^+ : \beta < \alpha\right\} \text{ by IH} \\ &= \alpha.\end{aligned}$$

The ordinals can be viewed as the backbone of the universe of sets; each V_α can be thought of as resting on the ordinal α.

Example 5.13

The rank of $\{\{2,3\},6\}$ is

$$\sup\{\text{rank}\{2,3\}+1, 6+1\} = \sup\{5,7\} = 7$$

§6 Cardinals

§6.1 Definitions

We will study the possible sizes of sets in ZFC. Write $x \equiv y$ if there exists a bijection from x to y, i.e. $(\exists f)(f : x \to y \wedge$ 'f is a bijection'$)$. This is an equivalence relation class, the equivalence classes are proper classes (except for $\{\varnothing\}$).
How do we pick a representative from each equivalence class?
We seek for each set x, a set $\operatorname{card}(x)$ s.t. $(\forall x)(\forall y)(\operatorname{card} x = \operatorname{card} y \iff x \equiv y)$.

In ZFC, this is easy: given a set x, x can be well-ordered, so $x \equiv OT(x)$, i.e. $x \equiv \alpha$ for some $\alpha \in ON$. We can define $\operatorname{card} x =$ least $\alpha \in ON$ s.t. $x \equiv \alpha$.

In ZF (due to D.S. Scott), we can define the **essential rank** as follows: $\operatorname{ess\ rank}(x) =$ least α s.t. $\exists y \subseteq V_\alpha$ with $y \equiv x$, i.e. least α s.t. $y \equiv x$ and $\operatorname{rank} y = \alpha$. Note that $\operatorname{ess\ rank}(x) \leq \operatorname{rank}(x)$ (take $\alpha = \operatorname{rank}(x)$ and $y = x$). Define $\operatorname{card} x = \left\{ y \subseteq V_{\operatorname{ess\ rank}(x)} : y \equiv x \right\}$.

§6.2 The hierarchy of alephs

Definition 6.1 (Cardinal)
A set m is a **cardinal** if $m = \operatorname{card} x$ for some set x, in which case we say m is the **cardinality** of x.

Definition 6.2 (Initial Ordinal)
Say that an ordinal α is an **initial ordinal** if $(\forall \beta < \alpha)(\beta \not\equiv \alpha)$.

I.e. α is initial if there is no bijection from α to any smaller ordinal.

Example 6.1
For any set x, $\gamma(x)$ (from Hartogs' lemma) is an initial ordinal.
So $0, 1, 2, 3, \ldots$ are initial ordinals (easy ω-induction).
ω is also an initial ordinal.
ω^2 is not initial as $\omega^2 \equiv \omega$ as countable.
$\varepsilon_0 = \omega^{\omega^{\omega^{\cdots}}}$ is not as $\varepsilon_0 \equiv \omega$.
ω_1 is initial.

We define ω_α for each ordinal α by recursion.

- $\omega_0 = \omega$;
- $\omega_{\alpha+1} = \gamma(\omega_\alpha)$;

- $\omega_\lambda = \sup\{\omega_\alpha : \alpha < \lambda\}$ for a nonzero limit ordinal λ.

Each of these ordinals is initial, and every initial ordinal β is of the form ω_α.

Proposition 6.1

The ordinals ω_α are exactly the infinite initial ordinals.

Proof. First we prove that every ω_α is initial, by induction:
$\alpha = 0$ ✓
$\alpha = \beta^+$ ✓ as $\forall x\ \gamma(x)$ is initial
$\alpha \neq 0$ a limit, Fix $\beta < \omega_\alpha$. Need that $\beta \not\equiv \omega_\alpha$. Since $\beta < \omega_\alpha$, we have $\beta < \omega_\gamma$ for some $\gamma < \alpha$. Then by IH, ω_γ is initial, so $\beta \not\equiv \omega_\gamma$. If $\beta \equiv \omega_\alpha$, then $\omega_\gamma \hookrightarrow \omega_\alpha \equiv \beta$ i.e. $\exists$ injection $\omega_\gamma \hookrightarrow \beta$. By Schröder-Bernstein, $\beta \equiv \omega_\gamma$ $\lightning$.

Conversely, let δ be an infinite initial ordinal. We need $\delta = \omega_\alpha$ for some α.
Easy induction show that $\alpha \leq \omega_\alpha\ \forall \alpha$ so $\delta < \omega_{\delta+1}$. Take the least α s.t. $\delta < \omega_\alpha$. Then $\alpha \neq 0$ and α is not a limit, o/w $\delta < \omega_\gamma$ for some $\gamma < \alpha$ $\lightning$.
So $\alpha = \beta^+$ for some β. So we have $\omega_\beta \leq \delta < \omega_{\beta^+} = \gamma(\omega_\beta)$. Hence $\delta \hookrightarrow \omega_\beta$[a] and $\omega_\beta \hookrightarrow \delta$. So by Schröder-Bernstein, $\omega_\beta \equiv \delta$, so $\delta = \omega_\beta$ as δ is initial. □

[a] As $\gamma(\omega_\beta)$ minimal order which doesn't inject into ω_β

Definition 6.3 (Aleph Numbers)

We write $\aleph_\alpha$ for $\mathrm{card}(\omega_\alpha)$ where $\alpha \in ON$. The **alephs** are the cardinalities of the infinite initial ordinals.

Example 6.2

$\aleph_0 = \mathrm{card}\,\omega$, $\aleph_1 = \mathrm{card}\,\omega_1$.

In ZF without AC, the $\aleph_\alpha$ are the cardinalities of the well-orderable sets.

§6.3 Cardinal Arithmetic

We use m, n, p etc. for cardinalities and M, N, P etc. for sets with cardinalities m, n, p respectively.

We write $m \leq n$ if there exists an injection from M to N where $\mathrm{card}(M) = m, \mathrm{card}(N) = n$[4].
Similarly, we write $m < n$ if $m \leq n$ and $m \neq n$.
For example, $\mathrm{card}(\omega) < \mathrm{card}(\mathcal{P}(\omega))$. By the Schröder–Bernstein theorem, if $m \leq n$ and

[4] This is well-defined, if $M \equiv M'$ and $N \equiv N'$ then $M \hookrightarrow N$ iff $M' \hookrightarrow N'$.

$n \leq m$, then $m = n$. Hence, $\leq$ is a partial order on cardinals. This is in fact a total order in ZFC, since we can well-order the two sets in question, and one injects into the other; alternatively, the $\aleph$ numbers are clearly totally ordered.

Let m, n be cardinals. Then,

1. $m + n = \operatorname{card}(M \sqcup N)$;
2. $m \cdot n = \operatorname{card}(M \times N)$;
3. $m^n = \operatorname{card}(M^N)$;

where $m = \operatorname{card}(M), n = \operatorname{card}(N)$, and M^N is the set of functions $N \to M$. This is well-defined.

Some Properties
$m + n = n + m$ $(M \sqcup N \equiv N \sqcup M)$
$m \cdot n = n \cdot m$ $(M \times N \equiv N \times M)$
$m(n + p) = mn + mp$ $(M \times (N \sqcup P) \equiv M \times N \sqcup M \times P)$
$(m^n)^p = m^{np}$, $m^n m^p = m^{n+p}$, $m^p \cdot n^p = (mn)^p$ (not true for ordinals though)

$m \leq n \Rightarrow m + p \leq n + p$, $mp \leq np$, $m^p \leq n^p$ etc.

Note. For all cardinals m, $m < 2^m$, i.e. $M \not\equiv \mathcal{P}(M)$. In fact, there is no surjection from M to $\mathcal{P}(M)$ by Cantor's diagonal argument.

Example 6.3

$\mathbb{R}, \mathcal{P}(\omega), \{0,1\}^\omega$ biject. Hence, $\operatorname{card}(\mathbb{R}) = \operatorname{card}(\mathcal{P}(\omega)) = 2^{\aleph_0}$. In particular, cardinal exponentiation and ordinal exponentiation do not coincide, as $2^\omega = \omega$.

The cardinality of the set of sequences of reals is

$$\operatorname{card}(\mathbb{R}^\omega) = (2^{\aleph_0})^{\aleph_0} = 2^{\aleph_0 \cdot \aleph_0} = 2^{\aleph_0}$$

Note that this statement requires that addition and multiplication are commutative, $\aleph_0 \cdot \aleph_0 = \aleph_0$ as $\omega \times \omega$ bijects with ω, and that $(m^n)^p = m^{np}$. The latter holds as $(M^N)^P$ is the set of functions $P \to (N \to M)$, and $M^{N \times P}$ is the set of functions $N \times P \to M$.

Theorem 6.1

$\aleph_\alpha \cdot \aleph_\alpha = \aleph_\alpha$ for any $\alpha \in ON$.

A consequence of this result is that addition and multiplication of alephs is easy.

Proof. We show this by induction on α. Define a well-ordering of $\omega_\alpha \times \omega_\alpha$ by 'going up in squares':

$$
\begin{aligned}
(x,y) < (z,w) \iff & (\max(x,y) < \max(z,w)) \vee \\
& (\max(x,y) = \max(z,w) = \beta \\
& \quad \wedge (y < \beta, z < \beta \vee x = z = \beta, y < w \vee y = w = \beta, x < z))
\end{aligned}
$$

For any $\delta \in \omega_\alpha \times \omega_\alpha$, $\delta \in \beta \times \beta$ for some $\beta < \omega_\alpha$, as ω_α is a limit ordinal (e.g. if $\delta = (x,y)$ then $\beta = \max\{x,y\}^+$ will do). By induction, we can assume $\beta \times \beta$ bijects with β (or β is finite). Hence, the initial segment I_δ is contained in $\beta \times \beta$ and hence has cardinality at most $\operatorname{card}(\beta \times \beta) < \operatorname{card}(\omega_\alpha)$.

So 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 $\leq \omega_\alpha$. It follows that $\omega_\alpha \times \omega_\alpha$ injects into ω_α and so $\aleph_\alpha \cdot \aleph_\alpha \leq \aleph_\alpha$.

Since $\aleph_\alpha = \aleph_\alpha \cdot 1 \leq \aleph_\alpha \cdot \aleph_\alpha$ and so the result follows. □

Corollary 6.1

For any ordinals $\alpha < \beta$, we have $\aleph_\alpha + \aleph_\beta = \aleph_\alpha \cdot \aleph_\beta = \aleph_\beta$.

Proof.

$$\aleph_\beta \leq \aleph_\alpha + \aleph_\beta \leq 2 \cdot \aleph_\beta \leq \aleph_\alpha \aleph_\beta \leq \aleph_\beta^2 = \aleph_\beta$$

□

Hence, for example, $X \sqcup X$ bijects with X for any infinite set X.

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.

Definition 6.4

Let I be a set, and for each $i \in I$ let m_i be a cardinal. Then

$$\sum_{i \in I} m_i = \operatorname{card}\left(\bigsqcup_{i \in I} M_i\right) \text{ and } \prod_{i \in I} m_i = \operatorname{card}\left(\prod_{i \in I} M_i\right)$$

where $\bigsqcup_{i \in I} M_i = \bigcup_{i \in I} M_i \times \{i\}$ and

$$\prod_{i \in I} M_i = \left\{ f : I \to \bigcup_{i \in I} M_i : f(i) \in M_i \text{ for all } i \in I \right\}.$$

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 $f_i : M_i \to M_i'$ where M_i' is another set with cardinality m_i.

Example 6.4

It is possible to show results similar to corollary 6.1. 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

$$2^{\aleph_\beta} \leqslant \aleph_\alpha^{\aleph_\beta} \leqslant \left(2^{\aleph_\alpha}\right)^{\aleph_\beta} = 2^{\aleph_\alpha \cdot \aleph_\beta} = 2^{\aleph_\beta}$$

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):

$$2^{\aleph_0} = \aleph_1$$

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.

Extra stuff on cardinal exponentiation:
Cardinal exponentiation is not as simple as addition and multiplication. For instance, in ZF, $2^{\aleph_0}$ need not even be an aleph number, for instance if $\mathbb{R}$ is not well-orderable. ZFC does not even decide if $2^{\aleph_0} < 2^{\aleph_1}$. Even today, not all implications about cardinal exponentiation (such as $\aleph_\alpha^{\aleph_\beta}$) are known.