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. $\Rightarrow p_1 p_3 (\in \Sigma \setminus L$.

Note that the introduction rules for the language are injective and have disjoint ranges, so there is exactly one way in which any element of the language can be constructed using rules (i) to (iii).

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

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 $t_1, \ldots, t_n$ 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 $t_1, \ldots, t_n$ 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 $t_1, t_2, \ldots$. 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. □

§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 ordered by $<$, or simply say X is a linearly 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 totally 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 $\mathbb{R}$, or $\{x : x > 0, x^2 < 2\}$ in $\mathbb{Q}$.

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. ↯ 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 $f : X \to Y$ be an order-isomorphism between X and I.
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 $a = f(z) > f(x)$ ↯, so $z \leq x$. If $z < x$ then $a = f(x) \in f(t) : t < x$ ↯ 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 $x \in X$

$$(\forall\, y < x,\, y \in S) \Rightarrow x \in S$$

Then $S = X$.

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, 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 $y < x$. We must have that $f(x) = a$, where a is the least element of $Y \setminus \{f(y) : y < x\}$. Indeed, if not, we have $f(x') = a$ for some $x' > x$ by bijectivity, contradicting the order-preserving property. Note that the set $Y \setminus \{f(x) : y < x\}$ is nonempty as it contains $f(x)$. So $f(x) = a = g(x)$, as required. □

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. For uniqueness, we apply induction on x. If f, f' agree below x, then they must agree at x since $f(x) = G\Big(f|_{I_x}\Big) = G\Big(f'|_{I_x}\Big) = f'(x)$.

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 $x \in \mathrm{dom}(h) \cap \mathrm{dom}(h')$ then $h(x) = h'(x)$ ($\mathrm{dom}(h)$ is the domain of h, i.e. I above). Fix $x \in \mathrm{dom}(h) \cap \mathrm{dom}(h')$ and assume $h(y) = h'(y)$ for every $y < x$ (note $y < x$ implies $y \in \mathrm{dom}(h) \cap \mathrm{dom}(h')$). 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. $x \in \mathrm{dom}(h)$. 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 $f(x) = h(x) = G(h|_{I_x}) = G(f|_{I_x})$. □

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.

Existence: If f is some such isomorphism, we must have that $f(x)$ is the least element of X not of the form $f(y)$ for $y < x$. We define f in this way by recursion, and this is an isomorphism as required. Note that this is always well-defined as $f(y) \leq y$, so there is always some element of X (namely, x) not of the form $f(y)$ for $y < x$. □

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 $f : X \to Y$ by

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

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 $Y \leq X$ ↯.

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 $f : X \to Y$ and $g : Y \to X$ 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) 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 $x_0 \notin X$.
$X \cup \{x_0\}$ 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 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 $X < Y$.
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

$\left\{\text{O. T.}(Y) : Y \in \widetilde{X}\right\}$ is a set of ordinals $< \alpha$, and $Y \mapsto \text{O. T.}(Y)$ 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.., so it has an order type, say α. Then X is o.i. to I_α, which is a proper i.s. of X. ↯ □

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 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 $\sup\{\omega + n : n < \omega\}$.

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

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

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

Continue in the same way.

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

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

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

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

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

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

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

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

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

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

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.

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 $A = \{(Y, <) \in \mathcal{P}(\mathbb{N}) \times \mathcal{P}(\mathbb{N} \times \mathbb{N}) : Y \text{ is a w.o. by } <\}$. 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 $\omega_1 \in B$, then $\omega_1^+ \in B$ ↯ as ω countable $\Rightarrow \omega^+$ countable. □

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 order.
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 it is bounded by $\sup\{\alpha_1, \alpha_2, \dots\}$, which is a countable ordinal.

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 $\mathbb{N}$. □

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

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

§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 $\alpha = (\sup I_\alpha)^+$. 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. $\omega = \sup\{n < \omega\}$ 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 ($\{\beta \leq \alpha : P(\beta)$ 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.

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

$1 + \omega = \sup\{1 + \gamma : \gamma < \omega\} = \sup 1, 2, 3, \ldots = \omega \neq \omega + 1$.
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.

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

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

$\gamma \neq 0$ 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.

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

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

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

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

□

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 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.

§2.10 Definitions

Definition 2.11

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.

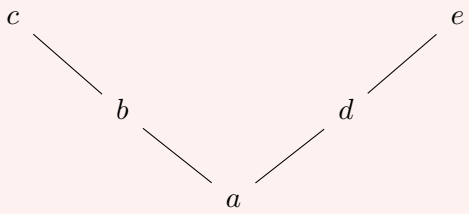
- (reflexivity) for all $x \in X$, $x \leq x$;
- (transitivity) for all $x, y, z \in X$, $x \leq y$ and $y \leq z$ implies $x \leq z$;
- (antisymmetry) for all $x, y \in X$, $x \leq y$ and $y \leq x$ implies $x = y$.

We write $x < y$ for $x \leq y$ and $x \neq y$. Alternatively, a post 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 2.12

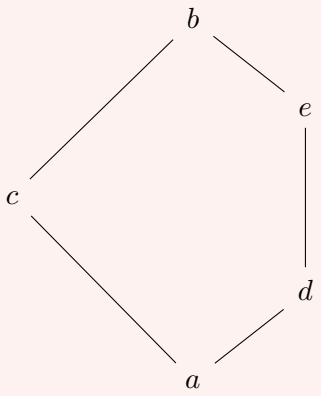
1. Any total order is a poset.
2. $\mathbb{N}^+$ with the divides relation is a poset.
3. $(\mathcal{P}(S), \subseteq)$ is a poset.
4. $(X, \subseteq)$ is a poset where $X \subseteq \mathcal{P}(S)$, such as the set of vector subspaces of a vector space.
5. The following diagram is also a poset, where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



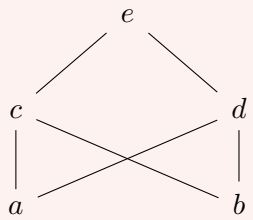
This is called a **Hasse diagram**. An upwards line from x to y is drawn if y **covers** x, so $y > x$ and no z has $y > z > x$. The natural numbers can be represented as a Hasse diagram.

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 2.12

A subset S of a poset X is a **chain** if it is totally ordered.

Example 2.13

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

Definition 2.13

A subset S of a poset X is an **antichain** if no two distinct elements are related.

Example 2.14

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

Definition 2.14

For $S \subseteq X$, an **upper bound** for S is an $x \in X$ s.t. $x \geq y$ for all $y \in S$. A **least upper bound** is an upper bound $x \in X$ for S s.t. for all upper bounds $y \in X$ for S, $x \leq y$.

Example 2.15

If $S = \left\{x \mid x < \sqrt{2}\right\} \subset \mathbb{R}$, 7 is an upper bound, and $\sqrt{2}$ is a least upper bound. We write $\sqrt{2} = \sup S = \bigvee S$ for the least upper bound or **join** of S.

In $\mathbb{Q}$, the set $\{x \mid x^2 < 2\}$ has 7 as an upper bound but has no least upper bound.

In example (v), $\{a, b\}$ has upper bounds b and c, so the least upper bound is b. $\{b, d\}$ has no upper bound. In example (vii), $\{a, b\}$ has upper bounds c, d, e, so does not have a least upper bound.

Definition 2.15

A poset X is **complete** if every $S \subseteq X$ has a least upper bound.

Example 2.16

$\mathbb{R}$ 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 2.17

$X = \mathcal{P}(S)$ is always complete as a poset under inclusion, with $\sup\{A_i \mid i \in I\} = \bigcup_{i \in I} A_i$.

Note that every complete poset X has a greatest element $\sup X$. A complete poset also has a least element $\sup \varnothing$. In the case $X = \mathcal{P}(S)$, $\sup X = S$ and $\sup \varnothing = \varnothing$.

Definition 2.16

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

Example 2.18

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

Not all order-preserving functions have a fixed point x s.t. $f(x) = x$, for example $f(x) = x + 1$ on $\mathbb{N}$.

Theorem 2.6 (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 $E = \{x \in X \mid x \leq f(x)\}$, and let $s = \sup E$. We show that s is a fixed point for f.

First, we show $s \leq f(s)$, so $s \in E$. It suffices to show $f(s)$ is an upper bound for E, then the result holds as s is the least such upper bound. If $x \in E$, we know $x \leq s$, so $f(x) \leq f(s)$ as f is order-preserving, as required.

Now, we show $f(s) \leq s$. It suffices to show $f(s) \in E$, as s is an upper bound for E. Since $s \leq f(s)$, we have $f(s) \leq f(f(s))$, but this is precisely the fact that $f(s) \in E$. □

Corollary 2.1 (Schröder–Bernstein theorem)

Let $f\colon A \to B$ and $g\colon B \to A$ be injections. Then there is a bijection $A \to B$.

Proof. We seek partitions $A = P \sqcup Q$, $B = R \sqcup S$ s.t. $f(P) = R$ and $g(S) = Q$; then we define h to equal to f on P and g^{-1} on Q. Thus, we need a set P that is a fixed point of $\theta\colon \mathcal{P}(A) \to \mathcal{P}(A)$ given by $P \mapsto A \setminus g(B \setminus f(P))$. But θ is order-preserving and $\mathcal{P}(A)$ is a complete poset. So P exists by the Knaster–Tarski fixed point theorem. □

§2.11 Zorn's lemma

Definition 2.17

Let X be a poset. We say that $x \in X$ is **maximal** if there is no $y \in X$ with $y > x$.

Example 2.19

In $[0, 1]$, 1 is maximal. In example (v), there are two maximal elements c and e.

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 2.7 (Zorn's lemma)

Let X be a poset in which every chain has an upper bound. Then X has a maximal element.

The empty chain must have an upper bound in X, so X must be nonempty to apply Zorn's lemma. Zorn's lemma can be equivalently be stated as the following.

Theorem 2.8

Let X be a nonempty poset in which every nonempty chain has an upper bound. Then X has a maximal element.

One can view Zorn's lemma as a fixed point theorem on a function $f : X \to X$ 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 $x \in X$ be any element, and define x_α for each $\alpha < \gamma(X)$ by recursion.

- $x_0 = x$;
- $x_{\alpha+1} = x'_\alpha$;
- $x_\lambda = u\{x_\beta \mid \beta < \lambda\}$ for λ a nonzero limit.

Note that $\{x_\beta \mid \beta < \lambda\}$ forms a chain, so it has an upper bound as required. Then, we have an injection from $\gamma(X)$ into X, contradicting the definition of $\gamma(X)$. □

Remark 26. 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 2.9

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 27. 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 2.10

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 $t \in S$ 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 $\{t_1, \ldots, t_n\} \vdash \bot$ as proofs are finite. Now, $t_1 \in T_{i_1}, \ldots, t_n \in T_{i_n}$ so all t_j are elements of T_{i_k} for some k. But T_{i_k} is consistent, so $\{t_1, \ldots, t_n\} \nvdash \bot$, giving a contradiction. □

§2.12 Well-ordering principle

Theorem 2.11

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 $R'|_A = A$ 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. 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. □

§2.13 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 $A_1, A_2, \ldots$ 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 function** $f \colon I \to \bigcup_{i \in I} A_i$ s.t. $f(i) \in A_i$ for all i.

Unlike the other axioms of set theory, the function 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 function $\star \mapsto 1$ 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 $\mathbb{Q}[X_1, \ldots, X_n]$ 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 function 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.