II Logic & Set Theory

Ishan Nath, Lent 2023

Based on Lectures by Prof. Imre Leader

June 26, 2023

Contents

1 Propositional Logic

Let P be a set of *primitive propositions*: unless otherwise stated, $P = \{p_1, p_2, p_3, \ldots\}$.

The *language* L or $L(P)$ is defined inductively as:

1. If $p \in P$, then $p \in L$.
2. $\bot \in L$.
3. If $p, q \in L$, then $(p \Rightarrow q) \in L$.

For example, $((p_1 \Rightarrow p_2) \Rightarrow (p_1 \Rightarrow p_3))$, $(p_4 \Rightarrow \bot)$, $(\bot \Rightarrow \bot)$ are in L.

Remark.

1. Each proposition (member of L) is a finite string of symbols from the language: $(,), \Rightarrow, \bot, p_1, p_2, \ldots$. For clarity, we often omit the outer brackets.
2. 'L is defined inductively' means as follows: Put $L_1 = P \cup (\bot)$. Having defined L_n, we define
$$L_{n+1} = L_n \cup \{(p \Rightarrow q) \mid p, q \in L_n\}.$$
Then, we set $L = L_1 \cup L_2 \cup \cdots$.
3. Every $p \in L$ is built up uniquely from properties 1 and 2, using 3.

We would like to introduce our common logical symbols, such as AND, NOT and OR. We can do this by introducing:

- $\neg p$ ("NOT p"), as an abbreviation for $(p \Rightarrow \bot)$,
- $p \vee q$ ("p OR q") as an abbreviation for $(\neg p) \Rightarrow q$,
- $p \wedge q$ ("p AND q") as an abbreviation for $\neg(p \Rightarrow (\neg q))$.

1.1 Semantic Implication

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

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

(ii) Implies behaves correctly, i.e.

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

Remark. On $\{0,1\}$, we can define a constant $\bot = 0$, and an operation $\Rightarrow$ by

$$(a \Rightarrow b) = \begin{cases} 0 & a = 1, b = 0, \\ 1 & \text{otherwise.} \end{cases}$$

Then a valuation is precisely a mapping $L \to \{0,1\}$ that preserves this structure, i.e. a 'homomorphism'.

Similarly to linear maps on a basis, we now have the following proposition about valuations on the primitives.

Proposition 1.1.

(i) *If v, v' are valuations which agree on the primitives, so $v(p) = v'(p)$ for all $p \in P$, then $v = v'$.*

(ii) *For any function $w : P \to \{0,1\}$, we can extend it to a valuation, i.e. there exists a valuation v with $v(p) = w(p)$ for all $p \in P$.*

The motto is: a valuation is often defined by its values on the primitives, and any values will do.

Proof:

(i) We prove this by induction. We have $v(p) = v'(p)$ for all $p \in L_1$, as we force $v(\bot) = v'(\bot) = 0$. But, if $v(p) = v'(p)$ and $v(q) = v'(q)$, then $v(p \Rightarrow q) = v'(p \Rightarrow q)$. Hence $v(p) = v'(p)$ for all $p \in L_2$.

Then inductively, we obtain $v(p) = v'(p)$ for all $p \in L_n$, and then this gives $v(p) = v'(p)$ for all $p \in L$.

(ii) Set $v(p) = w(p)$ for all $p \in P$ and $v(\bot) = 0$ to obtain a valuation on L_1. Now define

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

to obtain a valuation on L_2. Continuing inductively gives our result.

Hence we can say things like '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

$$v((p_1 \Rightarrow p_3) \Rightarrow p_2) = 0.$$

A *tautology* is a $t \in L$ such that $v(t) = 1$ for all valuations v. If t is a tautology, we write $\models t$.

Example 1.1.

1. $p \Rightarrow (q \Rightarrow p)$. This means "a true statement is implied by anything". To prove this, we can check all 4 truth values of $v(p)$ and $v(q)$:

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

2. $(\neg\neg p) \Rightarrow p$, or $((p \Rightarrow \bot) \Rightarrow \bot) \Rightarrow p$. This is the "law of the excluded middle". Again, we can check every truth value, of which there are 2:

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

3. $(p \Rightarrow (q \Rightarrow r)) \Rightarrow ((p \Rightarrow q) \Rightarrow (p \Rightarrow r))$. This shows how implications chain.

   Instead of checking all truth values, we will suppose this is not a tautology. Then we have v with $v(p \Rightarrow (q \Rightarrow r)) = 1$ and $v((p \Rightarrow q) \Rightarrow (p \Rightarrow r)) = 0$.

   The latter implies $v(p \Rightarrow q) = 1$ and $v(p \Rightarrow r) = 0$, so $v(p) = 1$ and $v(r) = 0$. The former statement then gives $v(q) = 1$. However, these valuations imply $v(p \Rightarrow (q \Rightarrow r)) = 0$, a contradiction.

For $S \subset L$ and $t \in L$, we say that S *entails* or *semantically implies* t, written $S \models t$, if every valuation v such that $v(s) = 1$ for all $s \in S$ has $v(t) = 1$.

Example 1.2.

We have $\{p \Rightarrow q, q \Rightarrow r\} \models p \Rightarrow r$. Indeed, suppose v has $v(p \Rightarrow q) = 1$ and $v(p \Rightarrow r) = 1$, but $v(p \Rightarrow r) = 0$. Then $v(p) = 1$ and $v(r) = 0$.

This implies $v(q) = 1$ from the first statement, but this contradicts $v(q \Rightarrow r) = 1$.

We say v is a *model* of $S \subset L$, so S is *true* in v, if $v(s) = 1$ for all $s \in S$. Thus $S \models t$ says that every model of S is also a model of t.

Remark. $\models t$ says that $\emptyset \models t$, as expected.

1.2 Syntactic Implication

For a notion of a proof, we need axioms, and deduction rules. The axioms we take are the tautologies from before:

1. $p \Rightarrow (q \Rightarrow p)$,
2. $[p \Rightarrow (q \Rightarrow r)] \Rightarrow [(p \Rightarrow q) \Rightarrow (p \Rightarrow r)]$,
3. $(\neg\neg p) \Rightarrow p$.

Remark.

1. Sometimes we call these 'axiom schemes', as each is really a set of axioms.
2. Each axiom is a tautology.

For the deduction rules, we have only *modus ponens*: from p and $p \Rightarrow q$, we can deduce q.

Now for $S \subset L$ and $t \in S$, we say S *proves* or *syntactically implies* t, written $S \vdash t$, if there exists a sequence $t_1, \ldots, t_n$ in L, with $t_n = t$, such that every t_n is either:

(i) an axiom, or

(ii) a member of S, or

(iii) obtained by modus ponens.

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

Example 1.3.

We want to show $\{p \Rightarrow q, q \Rightarrow r\} \vdash p \Rightarrow r$. We go for the sentence $(p \Rightarrow q) \Rightarrow (p \Rightarrow r)$, and then say we known $(p \Rightarrow q)$. Hence our proof is as follows:

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).
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).
6. $p \Rightarrow q$ (hypothesis).
7. $p \Rightarrow r$ (modus ponens).

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

Example 1.4.

We will show $\vdash (p \Rightarrow p)$, from the axioms. Our method is, we try to get to the right hand side of axiom 2, where we replace q with $(p \Rightarrow p)$ and r with p. Indeed,

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).
4. $p \Rightarrow (p \Rightarrow p)$ (axiom 1).
5. $p \Rightarrow p$ (modus ponens).

Often, showing $S \vdash p$ is made easier by the following:

Proposition 1.2 (Deduction theorem)**.** *Let $S \subset L$ and $p, q \in L$. Then $S \vdash (p \Rightarrow q)$ if and only if $S \cup \{p\} \vdash q$.*

In a sense, provability corresponds to the connective '$\Rightarrow$' in L.

Proof: Suppose we have a proof of $p \Rightarrow q$ from S. Then write down:

1. p (hypothesis).
2. q (modus ponens).

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

Now we show the more interesting part, the converse. Suppose we have a proof of q from $S \cup \{p\}$. Enumerate the proof $t_1, \ldots, t_n$. Then we will show that $S \vdash (p \Rightarrow t_i)$ for all i. There are three cases for what we can obtain in our proof:

(i) If t_1 is an axiom, then we can write down

1. t_i (axiom).
2. $t_i \Rightarrow (p \Rightarrow t_i)$ (axiom 1).
3. $p \Rightarrow t_i$ (modus ponens).

(ii) If $t_i \in S$, then we can write down the same thing as above, but replacing "t_i (axiom)" with "t_i (hypothesis)".

(iii) If $t_i = p$, then we already know $S \vdash (p \Rightarrow p)$, as $\vdash (p \Rightarrow p)$.

(iv) If t_i is obtained by modus ponens, then we have an earlier t_j and $t_k = (t_j \Rightarrow t_i)$ for some $j, k < i$. By induction, we can assume $S \vdash (p \Rightarrow t_j)$ and $S \vdash (p \Rightarrow (t_j \Rightarrow t_i))$. But then we can write down

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).

This shows $S \vdash (p \Rightarrow t_i)$.

Hence all cases are proven, so $S \vdash (p \Rightarrow q)$.

As an example, to show $\{p \Rightarrow q, q \Rightarrow r\} \vdash p \Rightarrow r$, it is sufficient to show $\{p \Rightarrow q, q \Rightarrow r, p\} \vdash r$, which is trivial by modus ponens twice.

From all the examples we have seen, it is natural to think that $\models$ and $\vdash$ are the same. So our aim is to show that

$$S \models t \iff S \vdash t.$$

This is known as the *completeness theorem*.

This is made up of $S \vdash t \implies S \models t$, which is known as *soundness*. Essentially, this validates that our axioms and deduction rule are not silly.

The other way is $S \models t \implies S \vdash t$, which is *adequacy*. This says our axioms are strong enough to be able to deduce every semantic consequence of S.

Proposition 1.3 (Soundness)**.** *Let $S \subset L$, $t \in L$. Then $S \vdash t \implies S \models t$.*

Proof: Since $S \vdash t$, we have a proof $t_1, \ldots, t_n$ of t from S. To show every model of S is a model of t, we need to show that if v is a valuation with $v(s) = 1$ for all $s \in S$, then also $v(t) = 1$.

But $v(p) = 1$ for each axiom p (as they are tautologies) and for each $p \in S$ (given), and we know $v(p)$ and $v(p \Rightarrow q) = 1$ implies $v(q) = 1$.

So $v(t_i) = 1$ for all i, by induction.

Before proving adequacy, we look at a special case when $t = \bot$. More specifically, we show if $S \models \bot$ then $S \vdash \bot$. We say S is *consistent* if S does not prove $\bot$.

Note $S \models \bot$ means there are no models of S, so this statement means that if S

has no model, then S is inconsistent. In other words, if S is consistent, then S has a model.

In fact, this implies adequacy in general. Indeed, if $S \models t$, then $S \cup \{\neg t\}$ has no models. Hence, by our special case $S \cup \{\neg t\} \vdash \bot$. By deduction theorem, $S \vdash (\neg t \Rightarrow \bot)$, or rewriting $S \vdash (\neg\neg t)$.

By by axiom 3, $S \vdash (\neg\neg t) \Rightarrow t$, so $S \vdash t$.

Hence, our task is as follows: given S consistent, find a model of S. We can try defining $v(t) = 1$ if $t \in S$, and $v(t) = 0$ otherwise. However, this fails as S might not be *deductively closed*, meaning $S \vdash p \implies p \in S$.

We could first replace S with its *deductive closure*, which is the set $\{t \in L \mid S \vdash t\}$. This is consistent as S is.

However this still fails: if S does not mention p_3, then S doesn't prove p_3 or $\neg p_3$. So we would be setting both $v(p_3) = 0$ and $v(\neg p_3) = 0$, which does not work.

Thus, we can extend S to 'swallow up' one of p or $\neg p$, for each p.

Theorem 1.1 (Model Existence Lemma)**.** *Let $S \subset L$ be consistent. Then S had a model.*

Proof: We claim that for any consistent $S \subset L$ and $p \in L$, then either $S \cup \{p\}$ or $S \cup \{\neg p\}$ is consistent.

If not, then both $S \cup \{p\} \vdash \bot$ and $S \cup \{\neg p\} \vdash \bot$. So from deduction theorem, $S \vdash (p \Rightarrow \bot)$, or $S \vdash (\neg p)$.

Hence since $S \cup \{\neg p\} \Rightarrow \bot$, we have $S \Rightarrow \bot$, implying S is inconsistent, a contradiction.

Now note L is countable as $L_1, L_2, \ldots$ are countable. Hence we can list L as $t_1, t_2, \ldots$.

Let $S_0 = S$, $S_1 = S_0 \cup \{t_1\}$ or $S_0 \cup \{\neg t_1\}$, with S_1 consistent, $S_2 = S_1 \cup \{t_2\}$ or $S_1 \cup \{\neg t_2\}$ with S_2 consistent, and define S_n similarly.

Then we can set $\overline{S} = S_0 \cup S_1 \cup S_2 \cup \cdots$. By construction, for each $t \in L$, either $t \in \overline{S}$ or $(\neg t) \in \overline{S}$.

Then $\overline{S}$ is consistent: if $\overline{S} \vdash \bot$, then as proofs are finite, we would have $S_n \vdash \bot$ for some n, however this is impossible by our induction.

Moreover, $\overline{S}$ is deductively closed as if $\overline{S} \vdash p$, we must have $p \in \overline{S}$, as otherwise $(\neg p) \in \overline{S}$, whence $\overline{S} \vdash p$, $\overline{S} \vdash (p \Rightarrow \bot)$ gives $\overline{S} \vdash \bot$.

Now define a valuation $v : L \to \{0, 1\}$ by $v(t) = 1$ if $t \in \overline{S}$, and $v(t) = 0$ otherwise. We show this is a model of S. Clearly $v(s) = 1$ for all $s \in S$.

- Since $\bot \notin \overline{S}$, as $\overline{S}$ is consistent, we have $v(\bot) = 0$.
- We now have to show $v(p \Rightarrow q)$ behaves correctly. We have three cases:

  (i) If $v(p) = 1$ and $v(q) = 0$, then $p \in \overline{S}$ and $q \notin \overline{S}$. Then we need to show $v(p \Rightarrow q) = 0$, i.e. $(p \Rightarrow q) \notin \overline{S}$.

  But if $(p \Rightarrow q) \in \overline{S}$, then from $p \in \overline{S}$, and deductive closure, we would get $q \in \overline{S}$.

  (ii) If $v(q) = 1$, then $q \in \overline{S}$, and we need to show $(p \Rightarrow q) \in \overline{S}$. But since $\overline{S} \vdash (q \Rightarrow (p \Rightarrow q))$ from axiom 1, $\overline{S} \vdash (p \Rightarrow q)$, i.e. $(p \Rightarrow q) \in \overline{S}$ as $\overline{S}$ is deductively closed.

  (iii) If $v(p) = 0$, then $p \notin \overline{S}$, and we again need to show $(p \Rightarrow q) \in \overline{S}$. We know that $(p \Rightarrow \bot) \in \overline{S}$, so it is enough to show that $p \Rightarrow \bot \vdash p \Rightarrow q$.

  Hence, it is enough to show that $\{p, p \Rightarrow \bot\} \vdash q$ by deduction theorem, or $\bot \vdash q$. But $\bot \vdash \neg\neg q$ from axiom 1, and $\neg\neg q \vdash q$ from axiom 3, hence we are done.

Remark.

1. We used the fact that the primitive propositions $P = \{p_1, p_2, \ldots\}$ were countable. In fact, this holds if P is not countable, as we will see later.
2. Sometimes this theorem is called the completeness theorem.

By the remarks before theorem 4, we get adequacy.

Corollary 1.1 (Adequacy)**.** *Let $S \subset L$, $t \in L$ with $S \models t$. Then $S \vdash t$.*

Combing this with soundness, we get the completeness theorem (for propositional logic):

Theorem 1.2 (Completeness theorem)**.** *Let $S \subset L$, $t \in L$. Then*

$$S \vdash t \iff S \models t.$$

We can apply the completeness theorem to transform trivial statements about syntactic implication, to non-trivial statements about semantic implication, and vice versa.

Corollary 1.2 (Compactness theorem)**.** *Let $S \subset L$, $t \in L$ with $S \models t$. Then there exists some finite $S' \subset S$ with $S' \models t$.*

Proof: This is trivial for semantic implication, as proofs are finite things. Hence this follows from completeness theorem.

If we take $t = \bot$ then this says that if S has no models, then there is some finite $S' \subset S$ such that S' has no models. That is, if every finite $S' \subset S$ has a model, then S has a model.

This is a very useful form of compactness, and in fact is equivalent to compactness in general, as $S \models t$ is equivalent to $S \cup \{\neg t\}$ has no models, and $S' \models t$ is equivalent to $S' \cup \{\neg t\}$ has no models.

Corollary 1.3 (Compactness theorem, equivalent form)**.** *Let $S \subset L$, then if every finite subset of S has a model, then so does S.*

Another application of the completeness theorem is the following:

Corollary 1.4 (Decidability theorem)**.** *Let $S \subset L$ be a finite set, and let $t \in L$. Then there is an algorithm to decide in a finite time whether or not $S \vdash t$.*

Proof: This is trivial for syntactic implication, as we can check all $2^{|S|}$ truth values. Then this follows from completeness theorem.

Remark. This is very surprising!

2 Well-Ordering & Ordinals

A *total order* or *linear order* is a pair $(X, <)$ where X is a set and $<$ is a relation on X that is:

(i) *Irreflexive*: for all $x \in X$, we don't have $x < x$.

(ii) *Transitive*: for all $x, y, z \in X$, if $x < y$ and $y < z$, then $x < z$.

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

Remark.

1. In property (iii), at most one of these can hold. So we could have equally said exactly one of these holds.
2. We can write $x > y$ if $y < x$, and $x \leq y$ if $x < y$ or $x = y$.
3. In terms of $\leq$, a total order is:

   (i)' *Reflexive*: for all $x \in x$, $x \leq x$.

   (ii)' *Transitive*: for all $x, y, z \in X$, if $x \leq y$ and $y \leq z$, then $x \leq z$.

   (iii)' *Antisymmetric*: for all $x, y \in X$, if $x \leq y$ and $y \leq x$, then $x = y$.

   (iv)' *Trichotomous*: for all $x, y \in X$, $x \leq y$ or $y \leq x$.

Example 2.1.

Examples of total orders:

1. $(\mathbb{N}, \leq)$.
2. $(\mathbb{Q}, \leq)$.
3. $(\mathbb{R}, \leq)$.
4. Not $(\mathbb{N}^+, |)$, as 2 and 3 are not related.
5. Not $(\mathcal{P}(S), \subset)$, as it fails trichotomy (for $|S| > 1$).

A total order $(X, <)$ is a *well-ordering* if every (non-empty) subset has a least element, so for all $S \subset X$ with $S \neq \emptyset$, there exists $x \in S$ such that $x \leq y$ for all $y \in S$.

Example 2.2.

Examples of well-orders:

1. $\mathbb{N}$.
2. Not $\mathbb{Z}$.
3. Not $\mathbb{Q}$.
4. Not $\mathbb{R}$.
5. Not $[0,1] \subset \mathbb{R}$ (as $(0,1)$ doesn't have a least element).
6. $\{\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \ldots\}$.
7. $\{\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \ldots\} \cup \{1\}$.
8. $\{\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \ldots\} \cup \{2\}$.
9. $\{\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \ldots\} \cup \{1 + \frac{1}{2}, 1 + \frac{2}{3}, 1 + \frac{3}{4}, \ldots\}$.

Remark. $(X, <)$ is a well-ordering if and only if there is no infinite strictly decreasing sequence.

Indeed, if $x_1 > x_2 > x_3 > \cdots$, then $\{x_1, x_2, \ldots\}$ has no least element.

Conversely, if $S \subset X$ has no least element, then for each $x \in S$, there is an $x' \in S$ with $x' < x$. Then $x > x' > x'' > \cdots$.

It turns out this equivalent definition is oddly unhelpful.

Say total order X, Y are *isomorphic* if there exists a bijection $f : X \to Y$ such that

$$x < y \iff f(x) < f(y).$$

For example, well-orders 1 and 6, and 7 and 8 are isomorphic. But 1 and 7 are not isomorphic, as 7 has a greatest element, but 1 does not.

Proposition 2.1 (Proof by Induction)**.** *Let X be well-ordered, and let $S \subset X$ such that whenever $y \in S$ for all $y < x$, then $x \in S$. Then $S = X$.*

Equivalently, if $P(x)$ is a property such that $P(y)$ for all $y < x$ implies $P(x)$, then $P(x)$ holds for all x.

Proof: Suppose $S \neq X$, and let x be least in $X \setminus S$. Then $y \in S$ for all $y < x$, but $x \notin S$, a contradiction.

Now we look at an example of induction.

Proposition 2.2. *Let X, Y be isomorphic well-orderings. Then there exists a unique isomorphism from X to Y.*

Note this is false for general total order, e.g. from $\mathbb{Z}$ to $\mathbb{Z}$, we could have $x \mapsto x - t$ for any t, or from $\mathbb{R}$ to $\mathbb{R}$ we could have all of these, as well as $x \mapsto x^3$.

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

We are given that $f(y) = g(y)$ for all $y < x$, and we want to show $f(x) = g(x)$.

Then we must have $f(x) = a$, the least element of $Y \setminus \{f(y) \mid y < x\}$, which is non-empty as it contains $f(x)$. Indeed, if not then $f(x') = a$ for some $x' > x$, as f is bijective, contradicting f being order-preserving.

Similarly, we have $g(x) = a$.

A subset I of a total order X is an *initial segment* if $x \in I$, $y < x$ implies $y \in I$.

For example, $I_x = \{y \in X \mid y < x\}$ for any $x \in X$ is an initial segment, but not every initial segment is of this form, e.g. in $\mathbb{R}$ we can have $\{x \mid x \leq 3\}$, and in $\mathbb{Q}$ we can have $\{x \mid x \leq 0 \text{ or } x^2 < 2\}$.

However, in a well-ordering, every proper initial segment is of the form I_x for some $x \in X$. Indeed, let x be the least element of $X \setminus I$. Then $y < x$ implies $y \in I$, by our choice of x. Conversely, if $y \in I$, then we must have $y < x$, as if $y \geq x$ then $x \in I$, a contradiction. So $I = I_x$.

In fact, any subset of a well-ordering X is isomorphism to an initial segment of X. This is false in general for total orders: for example $(1, 2, 3)$ in $\mathbb{Z}$, or $\mathbb{Q}$ in $\mathbb{R}$.

To prove this, we need some theorem about recursion, as that is how we will prove this proposition. Recall for $f : A \to B$ and $C \subset A$, the *restriction* of f to C is

$$f|_C = \{(x, f(x)) \mid x \in C\}.$$

Theorem 2.1 (Definition by Recursion)**.** *Let X be a well-ordering, and Y any set. Let $G : \mathcal{P}(X \times Y) \to Y$ be a rule. Then there exists a function $f : X \to Y$ such that*

$$f(x) = G(f|_{I_x}),$$

for all x. Moreover, f is unique.

In defining $f(x)$, we make use of f on $I_x = \{y \mid y < x\}$.

Proof: First, we show existence. We say h is an attempt if $h : I \to Y$ for some initial segment I of X, and for all $x \in I$, we have $h(x) = G(h|_{I_x})$.

Note that if h, h' are attempts both defined at x, then $h(x) = h'(x)$, by induction on x. Indeed, if $h(y) = h'(y)$ for all $y < x$, then $h(x) = h'(x)$.

Also, for all x there exist attempts defined at x, also by induction. Indeed, suppose that for all $y < x$, there exists an attempt defined at y. So for all $y < x$ there exists a unique attempt h_y with domain $\{z \mid z \leq y\}$. Let

$$h = \bigcup_{y<x} h_x$$

be an attempt with domain I_x. Thus $h \cup \{(x, G(h))\}$ is an attempt defined at x.

We can now defined $f : X \to Y$ by setting $f(x) = y$ if there exists and attempt h, defined at x, with $h(x) = y$.

Uniqueness follows from the above claim: if f, f' are suitable then $f(x) = f'(x)$ for all x by induction.

Proposition 2.3 (Subset Collapse)**.** *Let X be a well-ordering, and $Y \subset X$. Then Y is isomorphic to an initial segment I of X. Moreover I is unique.*

Proof: To have $f : Y \to X$ being an isomorphism with an initial segment of X, we need precisely that, for all $x \in Y$,

$$f(x) = \min X \setminus \{f(y) \mid y < x\}.$$

We are done (for existence and uniqueness), by recursion.

Note that $X \setminus \{f(y) \mid y < x\}$ is non-empty, as $f(y) \leq y$ for all y by induction, so $x \notin \{f(y) \mid y < x\}$.

In particular, X itself cannot be isomorphic to a proper initial segment, by uniqueness.

2.1 Relating Well-Orderings

For well-orderings X, Y, we write $X \leq Y$ if X is isomorphic to an initial segment of Y. For example, if $X = \mathbb{N}$, $Y = \{\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \ldots\}$, then $X \leq Y$.

Proposition 2.4. *Let X, Y be well-orderings. Then $X \leq Y$ or $Y \leq X$.*

Proof: Suppose $Y \not\leq X$. We will show that $X \leq Y$.

To obtain an isomorphism $f : X \to Y$ to an initial segment of Y, we need precisely that for all $x \in X$,

$$f(x) = \min Y \setminus \{f(y) \mid y < x\}.$$

Note that if $\{f(y) \mid y < x\} = Y$, then Y is isomorphic to I_x, so we are done by recursion.

Proposition 2.5. *Let X, Y be well-orderings with $X \leq Y$ and $Y \leq X$. Then X and Y are isomorphic.*

Proof: We have isomorphisms $f : X \to$ an initial segment of Y, and $g : Y \to$ an initial segment of X.

Then $g \circ f : X \to X$ is an isomorphism from X to an initial segment of X. But by uniqueness, $g \circ f = \mathrm{id}_X$. Similarly $f \circ g = \mathrm{id}_Y$, so f and g are inverses, and f is a bijection.

2.2 Making Well Orderings

For two well-orderings X, Y, we say $X < Y$ is $X \leq Y$ and X not isomorphic to Y. Equivalently, $X < Y$ if and only if X is isomorphic to a proper initial segment of Y.

Given a well-ordering X, we can pick $x \notin X$, and well-order $X \cup \{x\}$ by setting $y < x$ for all $y \in X$. This is a well-ordering which contains X, and is called the *successor* of X, written X^+.

Now given well-orderings X_i for $i \in I$, we seek X with $X_i \subset X$ for all i.

For well-orderings $(X, <_X)$ and $(Y, <_Y)$, we say $(Y, <_Y)$ *extends* $(X, <_X)$ if $X \subset Y$, $<_Y|\,|_X = <_X$, and X is an initial segment of $(Y, <_Y)$. We say well-orderings X_i for $i \in I$ are *nested* if for all i, j, X_i extends X_j or X_j extends X_i.

Proposition 2.6. *Let $(X_i \mid i \in I)$ be a nested set of well-orderings. Then there exists a well-ordering X such that $X_i \subset X$ for all i.*

Proof: Let $X = \bigcup_{i \in I} X_i$, with orderings $<_X = \bigcup_{i \in I} <_i$, meaning $x < y$ in X if there exists i such that $x, y \in X_i$ and $x <_i y$. This is a total order.

Now given $S \subset X$ with $S \neq \emptyset$, we have $S \cap X_i \neq \emptyset$, for some i. Let x be the least element of $S \cap X_i$ with ordering $<_i$. Thus x is the least element of S, as X_i is an initial segment of X, by nestedness.

So X is a well-ordering, and $X_i \subset X$ for all i.

Remark. The above proposition holds if we do not know if the X_i are nested.

2.3 Ordinals

From the above, we can determine that the collections of all well-orderings are a total order, up to isomorphism. Let us make this more explicit.

An *ordinal* is a well-ordered set, with two well-ordered sets regarded as the same if they are isomorphic.

Just as a rational is an expression $\frac{m}{n}$, with two regarded as the same if $mn' = m'n$, but we cannot formalize this with equivalence classes.

For a well-ordering X corresponding to an ordinal α, we say X has *order-type* α.

Example 2.3.

1. For any $k \in \mathbb{N}$, we write k for the order-type of the unique (up to isomorphism) well-ordering on a set of size k.
2. Write ω for the order-type of $\mathbb{N}$.
3. In $\mathbb{R}$, $\{-2, 3, \pi, 5\}$ has order-type 4.
4. $\{\frac{1}{2}, \frac{2}{3}, \ldots\}$ has order type ω.

We write $\alpha \leq \beta$ if $X \leq Y$, where X has order-type α and Y has order-type β. This does not depend on the choice of X and Y.

Similarly we define $\alpha < \beta$ and α^+. For all ordinals α, β, $\alpha \leq \beta$ or $\beta \leq \alpha$. Moreover, if $\alpha \leq \beta, \beta \leq \alpha$, then $\alpha = \beta$.

Proposition 2.7. *For any ordinal α, the ordinals less than α form a well-ordered set of order-type α.*

Proof: Let X have order-type α. Then the well-ordered sets less than X are precisely (up to isomorphism) the proper initial segments of X, which are I_x for $x \in X$.

But these are in order-bijection with X itself, via $I_x \to x$.

For any α, we have $I_\alpha = \{\beta \mid \beta < \alpha\}$ is a well-ordered set with order-type α.

Proposition 2.8. *Every non-empty set S of ordinals has a least element.*

Proof: Choose $\alpha \in S$. If α is minimal in S, then we have a least element. If not, $S \cap I_\alpha$ is non-empty, so it has a minimal element as I_α is well-ordered.

Now all the conditions are correct for the ordinals to form a well-ordered set: they are transitive, reflexive, symmetric and every non-empty set has a least element. However,

Theorem 2.2 (Burali-Fortis Paradox)**.** *The ordinals do not form a set.*

Proof: Suppose we have a set X, the set of all ordinals. Then X is a well-ordered set, so it has an order type, say α.

Thus X is order isomorphic to I_α, so X is order isomorphic to a proper initial segment of X, a contradiction.

Note that given a set $S = (\alpha_i \mid i \in I)$ of ordinals, there exists an upper bound α for S, by applying proposition 2.6 to the nested family I_{α_n}.

Hence, applying the above proposition, it has a least upper bound, written $\sup S$.

For example, $\sup\{2, 4, 6, \dots\} = \omega$.

Example 2.4.

We list some examples of ordinals.

0	$\omega + 1$	$\omega \cdot 2 + 2$	$\omega \cdot 5$
1	$\omega + 2$	$\vdots$	$\vdots$
2	$\omega + 3$	$\omega \cdot 3$	$\omega \cdot \omega = \omega^2$
3	$\vdots$	$\vdots$	$\omega^2 + 1$
$\vdots$	$\omega + \omega = \omega \cdot 2$	$\omega \cdot 4$	$\omega^2 + 2$
ω	$\omega \cdot 2 + 1$	$\vdots$	$\vdots$

$\vdots$	$\omega^\omega \cdot 3$	$\omega^2 + \omega$
$\epsilon_0 + \omega$	$\vdots$	$\vdots$
$\vdots$	$\omega^\omega \cdot \omega = \omega^{\omega+1}$	$\omega^2 + \omega \cdot 2$
$\epsilon_0 + \omega^\omega$	$\vdots$	$\omega^2 + \omega \cdot 3$
$\epsilon_0 + \epsilon_0 = \epsilon_0 \cdot 2$	$\omega^{\omega+2}$	$\vdots$
$\vdots$	$\vdots$	$\omega^2 + \omega^2 = \omega^2 \cdot 2$
$\epsilon_0 \cdot 3$	$\omega^{\omega+3}$	$\vdots$
$\vdots$	$\omega^{\omega \cdot 2}$	$\omega^2 \cdot 3$
$\epsilon_0 \cdot \omega$	$\vdots$	$\vdots$
$\vdots$	$\omega^{\omega \cdot 3}$	$\omega^2 \cdot 4$
ϵ_0^2	$\vdots$	$\vdots$
$\vdots$	ω^{ω^2}	ω^3
ϵ_0^3	$\vdots$	$\vdots$
$\vdots$	ω^{ω^3}	$\omega^2 + \omega^2 \cdot 7 + \omega \cdot 4 + 13$
ϵ_0^ω	$\vdots$	$\vdots$
$\vdots$	ω^{ω^ω}	ω^4
$\epsilon_0^{\epsilon_0}$	$\vdots$	$\vdots$
$\vdots$	$\omega^{\omega^{\omega^2}}$	ω^5
$\epsilon_0^{\epsilon_0^{\epsilon_0}}$	$\vdots$	$\vdots$
$\vdots$	$\omega^{\omega^{\omega^\omega}}$	ω^ω
$\epsilon_1 = \epsilon_0^{\epsilon_0^{\epsilon_0^{\cdot^{\cdot^{\cdot}}}}}$	$\vdots$	$\vdots$
$\vdots$	$\omega^{\omega^{\omega^{\cdots}}} = \epsilon_0$	$\omega^\omega \cdot 2$
etc.	$\epsilon_0 + 1$	$\vdots$

All of the above ordinals are countable, as they are unions of countable sets. Hence, we want to find whether there are uncountable ordinals, or an uncountable

well-ordered set.

It is easy to well order $\mathbb{N}$ and $\mathbb{Q}$. However, we cannot show we can well order $\mathbb{R}$, as we haven't found an uncountable ordinal. Amazingly, we can prove the following:

Theorem 2.3. *There is an uncountable ordinal.*

Proof: We want to take the set B, of the countable ordinals, and take their supremum.

However, to do this, we need to show B is a set, which is completely unobvious. Indeed, the whole of the ordinals in not a set, so why cannot B be the entirety of the ordinals?

Let $A = \{R \in \mathcal{P}(\omega \times \omega) \mid R \text{ is a well ordering of some subset of } \omega\}$. Then let $B = \{\text{order-type}(R) \mid R \in A\}$.

So $\alpha \in B$ if and only if α is a countable ordinal. Let $\omega_1 = \sup B$. We must have ω_1 uncountable. If not, then it would be the greatest element of B, contradicting $\omega_1 < \omega_1^+$.

Remark. Or, having our set B, we could say that B cannot be all of the ordinals by Burali-Fortis, so there exists an uncountable ordinal.

In fact, ω_1 is the least uncountable ordinal by definition of B.

The ordinal ω_1 has some remarkable properties, for example:

1. ω_1 is uncountable, but $\{\beta \mid \beta < \alpha\}$ is countable for all $\alpha < \omega_1$.
2. Any sequence $\alpha_1, \alpha_2, \alpha_3, \dots$ in w_1 is bounded, namely by $\sup(\alpha_1, \alpha_2, \alpha_3, \dots)$.

The same argument shows:

Theorem 2.4 (Hartogs' Lemma)**.** *For every set X, there exists an ordinal α that does not inject into X.*

The least such ordinal is written $\gamma(X)$.

2.4 Successors and Limits

Say α is a *successor* if there exists β such that $\alpha = \beta^+$. Otherwise, α is a *limit*.

In other words, an ordinal α has a greatest element if and only if α is a successor. So α is a limit if and only if it has no greatest element, if and only if, for all $\beta < \alpha$, there exists $\gamma < \alpha$ with $\gamma > \beta$.

Example 2.5.

5 is a successor, as $5 = 4^+$. Similarly $\omega + 2$ is a successor.

ω and 0 are limits.

2.5 Ordinal Arithmetic

Define $\alpha + \beta$ by induction on β (with α fixed) by:

1. $\alpha + 0 = \alpha$.
2. $\alpha + (\beta^+) = (\alpha + \beta)^+$.
3. $\alpha + \lambda = \sup\{\alpha + \gamma \mid \gamma < \lambda\}$ for λ a (non-zero) limit.

Example 2.6.

1. $\omega + 1 = \omega + 0^+ = (\omega + 0)^+ = \omega^+$.
2. $\omega + 2 = \omega + 1^+ = (\omega + 1)^+ = \omega^{++}$.
3. $1 + \omega = \sup\{1 + \gamma \mid \gamma < \omega\} = \omega$, so addition is not commutative.

Remark. Officially (as the ordinals do not form a set), this means to define $\alpha + \beta$, we define $\alpha + \gamma$ on $\{\gamma \mid \gamma \leq \beta\}$, and uniqueness.

Similarly for the proof by induction, if for some α we have $p(\alpha)$ false, then on $\{\gamma \mid \gamma \leq \alpha\}$, p is not everywhere true.

We have seen that addition is not commutative. However, we can prove it is associative.

Proposition 2.9. *$\alpha + (\beta + \gamma) = (\alpha + \beta) + \gamma$ for all ordinals α, β, γ.*

Proof: We induct on γ, with α, β fixed.

1. If $\gamma = 0$, then $\alpha + (\beta + 0) = \alpha + \beta = (\alpha + \beta) + 0$.
2. If γ^+ is a successor, then

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

3. If $\gamma = \lambda$ is a non-zero limit, then

$$(\alpha + \beta) + \lambda = \sup\{(\alpha + \beta) + \gamma \mid \gamma < \lambda\} = \sup\{\alpha + (\beta + \gamma) \mid \gamma < \lambda\}.$$

Now we would be done if $\beta + \lambda$ was a limit. But this is true: $\beta + \lambda = \sup\{\beta + \gamma \mid \gamma < \lambda\}$, so for all $\gamma < \lambda$, there exists $\gamma' < \lambda$ with $\gamma < \gamma'$, whence $\beta + \gamma < \beta + \gamma'$.

Thus there is no greatest element of $\{\beta + \gamma \mid \gamma < \lambda\}$, so $\beta + \lambda = \sup\{\beta + \gamma \mid \gamma < \lambda\}$ is a limit.

Hence $\alpha + (\beta + \gamma) = \sup\{\alpha + \delta \mid \delta < \beta + \lambda\}$, and we are done if we can show $\sup\{\alpha + (\beta + \gamma) \mid \gamma < \lambda\} = \sup\{\alpha + \delta \mid \delta < \beta + \lambda\}$.

First, $\gamma < \lambda \implies \beta + \gamma < \beta + \lambda$, so the left hand set is a subset of the right hand set. Conversely, for all $\delta < \beta + \lambda$, $\delta \leq \beta + \gamma$ for some $\gamma < \lambda$ (by definition of $\beta + \lambda$). So $\alpha + \delta \leq \alpha + (\beta + \gamma)$. Therefore each member of the right hand set is less than some member of the left hand set, completing the proof.

Remark. We used the facts that:

- $\beta \leq \gamma \implies \alpha + \beta \leq \alpha + \gamma$, which is trivial by induction on γ, and
- $\beta < \gamma \implies \alpha + \beta < \alpha + \gamma$, since

$$\begin{aligned} \beta < \gamma &\implies \beta^+ \leq \gamma \implies \alpha + \beta^+ \leq \alpha + \gamma \\ &\implies \alpha + \beta < (\alpha + \beta)^+ = \alpha + \beta^+ \leq \alpha + \gamma. \end{aligned}$$

However, $1 < 2$, but $1 + \omega = 2 + \omega = \omega$.

The above is the inductive definition of $+$. There is also a synthetic definition of $+$, by taking $\alpha + \beta$ the order type of the well-ordering of the disjoint union of α and β, taking everything in α less than everything of β.

Using this viewpoint, we can easily see why $\alpha + (\beta + \gamma) = (\alpha + \beta) + \gamma$: these both are the order type of α, followed by β, followed by γ.

Proposition 2.10. *The two definitions of $+$ coincide.*

Proof: Write $+$ for the inductive proof, and $+'$ for the synthetic proof. We will show $\alpha + \beta = \alpha +' \beta$ for all α and β, by induction on β.

1. If $\beta = 0$, then $\alpha + 0 = \alpha = \alpha +' 0$.
2. If β^+ is a successor, then

$$\alpha + (\beta^+) = (\alpha + \beta)^+ = (\alpha +' \beta)^+ = \alpha +' \beta^+,$$

by the associativity of $+'$.

3. If $\beta = \lambda$, a non-zero limit, then

$$\alpha + \lambda = \sup\{\alpha + \gamma \mid \gamma < \lambda\} = \sup\{\alpha +' \gamma \mid \gamma < \lambda\} = \alpha +' \lambda,$$

which is true as the supremum is a union, and all the sets are nested.

Now we can define ordinal multiplication. Define $\alpha\beta$, with α fixed and by recursion on β, by

1. $\alpha 0 = 0$.
2. $\alpha(\beta^+) = (\alpha\beta) + \alpha$.
3. $\alpha\lambda = \sup\{\alpha\gamma \mid \gamma < \lambda\}$ for λ a non-zero limit.

Example 2.7.

1. $\omega 2 = \omega 1 + \omega = (\omega 0 + \omega) + \omega = \omega + \omega$.
2. $\omega 3 = \omega + \omega + \omega$.
3. $\omega\omega = \sup\{0, \omega, \omega + \omega, \omega + \omega + \omega, \ldots\}$.
4. $2\omega = \sup\{0, 2, 4, 6, 8, \ldots\} = \omega$, hence multiplication is not commutative.

We can show that $\alpha(\beta\gamma) = (\alpha\beta)\gamma$.

Moreover, we have a synthetic definition (which we can show coincides) by letting $\alpha\beta$ be the order type of $\alpha \times \beta$, ordered by $(x, y) < (z, w)$ if $y < w$ or $y = w$ and $x < z$.

We can also define exponentiation and towers similarly, for example we can define α^β by:

1. $\alpha^0 = 1$.
2. $\alpha^{\beta^+} = \alpha^\beta \alpha$.
3. $\alpha^\lambda = \sup\{\alpha^\gamma \mid \gamma < \lambda\}$ for a non-zero limit.

Example 2.8.

1. $\omega^2 = \omega^1 = \omega = (\omega^0\omega)\omega = \omega\omega$.
2. $2^\omega = \sup\{1, 2, 4, 8, \ldots\} = \omega$.

3 Posets and Zorn's Lemma

A *partially ordered set* or *poset* is a pair $(X, \leq)$, where X is a set and $\leq$ is a relation on X that is:

(i) Reflexive: $x \leq x$ for all $x \in X$.

(ii) Transitive: $x \leq y$ and $y \leq z$ implies $x \leq z$ for all $x, y, z \in X$.

(iii) Antisymmetric: $x \leq y$ and $y \leq x$ implies $x = x$ for all $x, y \in X$.

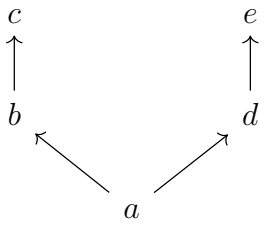
We write $x < y$ is $x \leq y$ and $x \neq y$. In terms of $<$, a poset is:

(i) Irreflexive: Not $x < x$ for all $x \in X$.

(ii) Transitive: $x < y$ and $y < z$ implies $x < z$ for all $x, y, z \in X$.

Example 3.1.

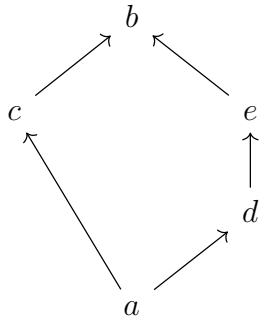
Examples of posets:

1. Any total order.
2. $\mathbb{N}^+$ with 'divides'.
3. $\mathcal{P}(S)$ with $\subset$ (for any set S).
4. In fact, any subset X of $\mathcal{P}(S)$ with $\subset$ is a poset, for example the subspaces of a vector space V, with $\subset$.
5. 

   Let $a \leq b$, $b \leq c$, $a \leq d$, $d \leq e$, and everything else that follows from transitivity. For example, $a \leq c$, but b and d are unrelated.

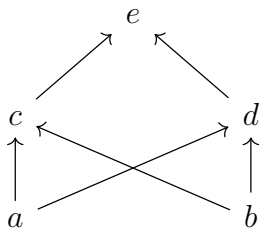
   In general, the *Hasse diagram* if a poset is a drawing of its points with an upwards line from x to y if y *covers* x (meaning $x < y$ and there is no z such that $x < z < y$). Hasse diagrams can be useful, for example $\mathbb{N}$ with $\leq$, or useless, for example $\mathbb{Q}$ with $\leq$.

6.



This example shows that there is no notion of a 'height' or a 'rank' in a poset.

7.



8. We can also have the trivial poset:

$$a \qquad b \qquad c \qquad d \qquad e$$

A subset S of a poset X is a *chain* if it is totally ordered. For example, in example 2, $\{1, 2, 4, 8, 16\}$ is a chain, and in example 5, $\{a, b, c\}$ or $\{a, c\}$ is a chain.

A subset S is an *antichain* if no two elements are related. For example, in example 2 $\{p \mid p \text{ prime}\}$ is an antichain, in example 5, $\{c, e\}$ is an antichain, and in example 8, the whole poset is an antichain.

For $S \subset X$, an *upper bound* for S is an $x \in X$ such that $x \geq y$ for all $y \in S$. x is a *least upper bound* for S if x is an upper bound for S, and for every upper bound of y, we have $x \leq y$.

Example 3.2.

We look at examples of upper bounds.

1. In $\mathbb{R}$, if $S = \{x \mid x < \sqrt{2}\}$, then 7 is an upper bound, and $\sqrt{2}$ is the least upper bound. We write $\sqrt{2} = \sup S$ or $\vee S$.

2. In $\mathbb{Q}$, $\{x \mid x^2 < 2\}$ has 7 as an upper bound, but has no least upper bound.
3. In example 5, $\{a, b\}$ has upper bounds b and c, so it has a least upper bound b. But $\{b, d\}$ has not upper bound.
4. In example 7, (a, b) has upper bounds c, d, e, but it does not have a least upper bound

We say X is *complete* if every $S \subset X$ has a least upper bound.

Example 3.3.

We look at examples of complete and incomplete spaces.

1. $\mathbb{R}$ is not complete, as $\mathbb{Z}$ has no upper bound.
2. $[0, 1]$ is complete.
3. $(0, 1)$ is not complete, as $(0, 1)$ itself has no upper bound.
4. $X = \mathcal{P}(S)$ is always complete: we have

$$\sup\{A_i \mid i \in I\} = \bigcup_{i \in I} A_i.$$

Note every complete poset X has a greatest element x, namely $\sup X$. Moreover, it also has a least element y, namely $\sup \emptyset$.

For $f : X \to Y$, where X, Y are posets, we say f is *order-preserving* if $x \leq y$ implies $f(x) \leq f(y)$.

Example 3.4.

We look at order-preserving functions:

1. Take $f : \mathbb{N} \to \mathbb{N}$ by $n \mapsto n + 1$.
2. Take $f : [0, 1] \to [0, 1]$ by $x \mapsto \frac{1+x}{2}$.
3. Take $f : \mathcal{P}(S) \to \mathcal{P}(S)$ by $A \mapsto A \cup \{i\}$ for some fixed $i \in S$.

Note that not every order-preserving $f : X \to X$ has a *fixed point*. For example, $x \mapsto x + 1$ has no fixed point. But we can prove the following:

Theorem 3.1 (Knaster-Tarski fixed point theorem)**.** *Let X be a complete poset. Then any order-preserving $f : X \to X$ has a fixed point.*

Proof: Let $E = \{x \in X \mid x \leq f(x)\}$, and let $s = \sup E$. We will show that $f(s) = s$, by showing $f(s) \geq s$, and then $f(s) \leq s$.

To show $s \leq f(s)$, it is enough to show that $f(s)$ is an upper bound for E, as then $s \leq f(s)$ since s is the least upper bound. We can show this through the following:

$$x \in E \implies x \leq s \implies f(x) \leq f(s) \implies x \leq f(x) \leq f(s).$$

Now to show $f(s) \leq s$, it is enough to show $f(s) \in E$. Note we have $s \leq f(s)$, so $f(s) \leq f(f(s))$, implying $f(s) \in E$ and so $f(s) \leq s$.

We can apply Knaster-Tarski as follows:

Corollary 3.1 (Schröder-Bernstein)**.** *Let $f : A \to B$ and $g : B \to A$ be injections. Then there exists a bijection $h : A \to B$.*

Proof: We seek disjoint partitions $A = P \cup Q$ and $B = R \cup S$, such that $f(P) = R$ and $g(S) = Q$. Then we can set $h = f$ on P, and $h = g^{-1}$ on R.

Thus, we seek exactly a fixed point $\theta : \mathcal{P}(A) \to \mathcal{P}(A)$, by

$$P \mapsto A \setminus g(B \setminus (f(P))).$$

But $\mathcal{P}(A)$ is complete, and θ is order-preserving, hence we are done by Knaster-Tarski.

3.1 Zorn's Lemma

For a poset X, $x \in X$ is *maximal* if no $y \in X$ has $y > x$. For example, $[0, 1]$ has 1 maximal, and example 3.1.5 has both c and e maximal.

We have seen many posets without a maximal element, for example $(\mathbb{R}, \leq)$ and $(\mathbb{N}^+, |)$. But in each case, we can find a chain with no upper bound, for example $(\mathbb{N}, |)$ has the powers of 2.

Theorem 3.2 (Zorn's lemma)**.** *Let X be a non-empty poset in which every chain has an upper bound. Then there exists a maximal element of X.*

Proof: Suppose not, so for each $x \in X$ we have x' with $x' > x$, and each chain C has an upper bound $u(C)$. Fix $x \in X$ 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 non-zero limit.

Note that the $x_\beta, \beta < \lambda$ do form a chain by induction.

Then we have injected $\gamma(X)$ into X, a contradiction.

Remark. While the proof looks easy, it requires well-orderedness, recursion, and Hartogs' lemma.

A typical application of Zorn's: does every vector space have a basis? Recall that a basis is a linearly independent spanning set.

Example 3.5.

1. $\mathbb{R}^3$ has a basis e_1, e_2, e_3.
2. The space of all real polynomials has basis $1, x, x^2, \ldots$
3. Consider the space S of all real sequences. Setting $e_1 = (1, 0, 0, \ldots)$, the set $\{e_1, e_2, e_3, \ldots\}$ is not a basis, as it does not contain $(1, 1, 1, \ldots)$. In fact, there is no countable basis, and moreover there are no explicit basis.
4. $\mathbb{R}$ as a vector space over $\mathbb{Q}$ has no explicit basis. A basis here is called a *Hamel basis*.

Theorem 3.3. *Every vector space V has a basis.*

Proof: Let $X = \{A \subset V \mid A \text{ is linearly independent}\}$, ordered by inclusion. We seek a maximal element of X; then we would be done. If A is maximal, then A must span - if not, then for any $x \in V - \langle A \rangle$ we have $A \cup \{x\}$ is linearly independent, contradiction maximality of A.

Now $X \neq \emptyset$, as $\emptyset \in X$. Given a chain $\{A_i \mid i \in I\}$ in X, let $A = \bigcup A_i$. So certainly $A \supset A_i$, hence we just need to show that $A \in X$, i.e. that A is linearly independent.

Suppose A is not linearly independent. Then we have $x_1, \ldots, x_n \in A$ which are linearly dependent. We have $x_1 \in A_{i_1}, \ldots, x_n \in A_{i_n}$ for some $x_1, \ldots, x_n$.

Then we get some A_{i_k} contains all of $A_{i_1}, \ldots, A_{i_n}$, as the A_i for a chain.

So $x_1, \ldots, x_n \in A_{i_k}$, contradicting the fact A_{i_k} is linearly independent. Hence by Zorn, there is a maximal element in X.

Remark.

1. The only actual linear algebra in the proof was showing the maximal element was indeed linearly independent.
2. 'X non-empty' is not strictly needed, as $\emptyset$ is a chain, so it has an upper bound, but it is safer to check it directly, as often we only want to look at non-empty chains.

Another application is proving the completeness theorem for propositional logic, with no restriction on P.

Theorem 3.4. *Let $S \subset L = L(P)$ for any set P, that is consistent. Then S has a model.*

Proof: We'll extend S to $\bar{S}$ such that for all $t \in L$, either $t \in \bar{S}$ or $(\neg t) \in \bar{S}$. Let $X = \{T \supset S \mid T \text{ consistent}\}$, ordered by inclusion. We seek a maximal element of X.

Then we are done: let $\bar{S}$ be maximal. If $t \notin \bar{S}$, then $\bar{S} \cup \{t\} \vdash \bot$, by maximality. Hence then $\bar{S} \vdash (\neg t)$, and so $(\neg t) \in \bar{S}$ by maximality of $\bar{S}$.

Now $X \neq \emptyset$ as $S \in X$. Given non-empty chain $\{T_i \mid i \in I\}$, put $T = \bigcup T_i$. We have $T \supset T_i$, so we just need to show $T \in X$.

Since our chain is non-empty, $S \subset T$. Also, T is consistent. Suppose $T \vdash \bot$. Then some $\{t_1, \ldots, t_n\} \vdash \bot$, as proofs are finite. Now $t_1 \in T_{i_1}, \ldots, t_n \in T_{i_n}$ for some $i_1, \ldots, i_n \in I$. Whence $t_1, \ldots, t_n \in T_{i_k}$ for some k, as the T_i form a chain.

This contradicts the fact T_{i_k} is consistent. Hence by Zorn's, X has a maximal element.

Another application:

Theorem 3.5 (Well-ordering theorem)**.** *Every set S can be well-ordered.*

Remark. This is very surprising, e.g. take $S = \mathbb{R}$, until one has met Hartogs' lemma.

Proof: Let $X = \{(A, R) \mid A \subset S, R \text{ is a well-ordering of } A\}$, ordered by $(A, R) \leq (A', R')$ if the latter extends the former.

We have $X \neq \emptyset$, as $(\emptyset, \emptyset) \in X$, and given a chain $\{(A_i, R_i) \mid i \in I\}$, we have an upper bound $(\bigcup A_i, \bigcup R_i)$, because our family is a nested family.

Hence by Zorn there exists a maximal element, say (A, R). We must have $A = S$: if not, let $x \in S \setminus A$ and 'take the successor': well-order $A \cup \{x\}$ by making $x > y$ for all $y \in A$, contradicting the maximality of (A, R).

3.2 Zorn's Lemma and the Axiom of Choice

In our proof of Zorn's lemma, we made infinitely many arbitrary choices, when we selected the x'. We did this also when showing a countable union of countable sets is countable: we have sets $A_1, A_2, \ldots$, each with a listing, and we fixed, all at once, a listing for each one.

In terms of 'rules for building sets', we are appealing to the *axiom of choice*, which states that, given a family of non-empty sets, one can choose an element from each. More precisely, for any family $\{A_i \mid i \in I\}$ of non-empty sets, there is a *choice function*, meaning a function $f : I \to \bigcup A_i$, such that $f(i) \in A_i$ for all i.

This is different in character from the other 'rules for building sets' (for example forming $A \cup B$ or $\mathcal{P}(A)$), in that the object whose existence is asserted is not uniquely specified by its properties.

Many proofs in mathematics, even without the axiom of choice, are non-constructive, e.g. the proof that there exists a transcendental number, or proof that in $\mathbb{Q}[X_1, \ldots, X_n]$, every ideal is finitely generated (from Hilbert's basis theorem). It is often nice to know whether a proof needed the axiom of choice.

Hence it is natural to ask whether our proof of Zorn's lemma needed to use axiom of choice. The answer is yes, because we can deduce the axiom of choice from Zorn's lemma.

Indeed, the axiom of choice follows from the well-ordering theorem: given our family $\{A_i \mid i \in I\}$, then we can well-order $\bigcup A_i$, and hence we can define $f(i)$ to be the least element of A_i.

Remark.

1. The axiom of choice is trivial if $|I| = 1$. It is also easy to prove for all I finite, by induction on $|I|$. But in general, it turns out that AC cannot be deduced from the other set-building rules.

2. Zorn's lemma is hard from first principles because it needed ordinals, recursion and Hartogs' lemma, not because it is equivalent to the axiom of choice.

3. No theorem in chapter two used axiom of choice. Indeed, axiom of choice was used only twice in the remark in chapter two: the fact that in a non-well-ordered set there exists an infinite decreasing sequence, and the fact that ω_1 is not a countable supremum.

4 Predicate Logic

Recall that a *group* is a set A equipped with functions $m : A^2 \to A$ with arity 2, $i : A^1 \to A$ with arity 1, and a constant $e \in A$, which can be viewed as a function $A^0 \to A$, i.e. with arity 0, such that some axioms hold.

A *poset* is a set A equipped with a relation $\leq \subset A^2$, with arity 2, such that some axioms hold.

In more rigorous terms, let Ω be a set of *functional symbols* at Π be a set of *relation symbols*, and $\alpha : \Omega \cup \Pi \to \mathbb{N}$ be the *arity*.

The *language* $L = L(\Omega, \Pi, \alpha)$ is the set of formulae, defined as follows:

The variables are $x_1, x_2, x_3, \ldots$, and they are defined inductively as:

1. Each variable is a term.
2. For $f \in \Omega$, $\alpha(f) = n$ and terms $t_1, \ldots, t_n$, $f(t_1, \ldots, t_n)$ (or equivalently $ft_1 \ldots t_n$) is a term.

Example 4.1.

In the language of groups, $\Omega = (m, i, e)$, $\Pi = \emptyset$, with arity $\alpha(m, i, e) = (2, 1, 0)$.

Some terms include $m(x_1, x_2)$, $m(x_1, i(x_1))$, e, $m(e, e)$, $m(e, x_1)$.

We also define *atomic formulae* as follows:

1. $\perp$ is an atomic formula.
2. For terms s and t, $(s = t)$ is an atomic formula.
3. For $\phi \in \Pi$ with $\alpha(\phi) = n$, and terms $t_1, \ldots, t_n$, $\phi(t_1, \ldots, t_n)$ is an atomic formula.

Example 4.2.

In the language of groups, $e = m(e, e)$ and $m(x, y) = m(y, x)$ are atomic formulae.

In the language of posets, $\Omega = \emptyset$, $\Pi = \{\leq\}$. Some atomic formulae are $x = y$, and $x \leq y$.

We can also define *formulae* inductively by:

1. Each atomic formula is a formula.

2. If p, q are formulae, then $(p \Rightarrow q)$ is a formula.
3. If p is a formula, x is a variable, then $(\forall x)p$ is a formula.

Example 4.3.

In the language of groups, $(\forall x)(m(x,x) = e)$ and $m(x,x) = e \Rightarrow (\exists y)(m(y,y) = x)$ are formulae.

Remark.

1. A formula is a finite string of symbols.
2. Can define $(\neg p)$, $p \wedge q$, $p \vee q$ as before, and also $(\exists x)p$ as $\neg(\forall x)(\neg p)$.

A term is *closed* if it contains no variables.

An occurrence of a variable x in a formula p is *bound* if it is inside the brackets of a $(\forall x)$ quantifier. Otherwise, it is *free*.

Example 4.4.

$e, m(e, i(e))$ are closed, but neither $m(x, e)$, or $m(x, i(x))$, are.

In the formula $(\forall x)(m(x,x) = e)$, all occurrences of x are bound. But in $m(x,x) = e \Rightarrow (\exists y)(m(y,y) = x)$, all occurrences of x are free, while all occurrences of y are bound.

In the sentence

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

the first two occurrences of x are free, while the rest of the variables are bound. Note that having the same variable both free and bound is unhelpful.

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

Example 4.5.

$(\forall x)(m(x,x) = e)$ and $(\forall x)(m(x,x) = e \Rightarrow (\exists y)(m(y,y) = x))$ are sentences.

In the language of posets, $(\forall x)(\exists y)(x \leq y \wedge \neg x = y)$ is a sentence.

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

Example 4.6.

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

4.1 Semantic Implication

Let $L=(\Omega,\Pi,\alpha)$ be a language. An *L-structure* is a non-empty set A equipped with, for each $f\in\Omega$, a function $f_A:A^n\to A$ (with $n=\alpha(f)$), and for each $\phi\in\Pi$, a subset $\phi_A\subset A^n$ (with $n=\alpha(\phi)$).

Example 4.7.

In the language of groups, an L-structure is an A with $m_A:A^2\to A$, $i_A:A\to A$, $e_A\in A$.

In the language of posets, and L-structure is an A, with $\leq_A\subset A^2$.

We want to define, for an L-structure A, and a sentence p, what it means for p to be true in A.

Example 4.8.

For $(\forall x)m(x,x)=e$ to hold in A, we expect for each $a\in A$, that $m_A(a,a)=e_A$.

So to informally define whether a sentence is true, we insert '$\in A$' after each '$\forall x$', and add subscripts '$_A$', and read it aloud. However we obviously cannot define it like this.

Define the *interpretation* $p_A\in\{0,1\}$ of a sentence p in an L-structure A as follows:

The interpretation $t_A\in A$ of a closed term in an L-structure A is defined inductively, as

$$(ft_1,\ldots,t_n)_A=f_A((t_1)_A,\ldots,(t_n)_A),$$

for $f\in\Omega$, $\alpha(f)=n$ and $t_1,\ldots,t_n$ closed terms. Moreover c_A is already defined for each constant symbol $c\in\Omega$.

We will define our valuation as follows. For atomic sentences:

(i) Let $\bot_A=0$.

(ii) Let $(s=t)_A=1$ if $s_A=t_A$, and 0 if else.

(iii) Let $\phi(t_1,\ldots,t_n)=1$ if $((t_1)_A,\ldots,(t_n)_A)\in\phi_A$, and 0 if not.

Now for compound sentences, we can inductively define:

(i) $(p \Rightarrow q)_A = 0$ if $p_A = 1$, $q_A = 0$, and 1 otherwise.

(ii) $((\forall x)p)_A = 1$ if $p[\bar{a}/x]_A = 1$ for all $a \in A$, and 0 otherwise, where we add a constant symbol $\bar{a}$ to L (for a fixed $a \in A$), to form a language L', and make it into an L'-structure by setting $\bar{a}_A = a$.

Remark. For a formula p with n free variables, we can define $p_A \subset A^n$, for example if p is $m(x, x) = e$, then $p_A = \{a \in A \mid m_A(a, a) = e\}$.

If $p_A = 1$, we say that p *holds* in A, or p is *true* in A, or A is a *model* of p.

For a *theory* T (which is a set of sentences), A is a *model* of T if $p_A = 1$ for all $p \in T$.

For a theory T and a sentence p, we write $T \models p$ if every model of T is a model of p. For example, the three group axioms $\models m(e, e) = e$.

Example 4.9.

We look at some theories:

1. For groups, we let L be the language of groups, defined above, and we have
$$\begin{aligned} T = \{&(\forall x)(\forall y)(\forall z)(m(x, m(y, z))) = m(m(x, y), z), \\ &(\forall x)(m(x, e) = x \wedge m(e, x) = x), \\ &(\forall x)(m(x, i(x)) = e \wedge m(i(x), x) = e)\}. \end{aligned}$$
Then an L-structure is a model of T if and only if it is a group. We say T *axiomatizes* the theory of groups (the class of groups). Often, the elements of T are called the *axioms* of T.

2. For posets, let L be the language of posets, and T the usual poset axioms. Then T axiomatizes the class of posets.

3. For fields, let L be the language of field, with $\Omega = \{0, 1, +, \times, -\}$, and let T be the usual field axioms, including
$$(\forall x)(\neg(x = 0) \implies (\exists y)(xy = 1)).$$

4. For graphs, let L be the language with $\Omega = \emptyset$, and $\Pi = \{a\}$ with $\alpha(a) = 2$. Then we can define
$$T = \{(\forall x)(\neg a(x, x)), (\forall x)(\forall y)(a(x, y) \Rightarrow a(y, x))\}.$$
Then T axiomatizes the theory of graphs.

4.2 Syntactic Entailment

To prove things, we will need our (logical) axioms and deduction rules. With the addition of more structure, we now have 7 axioms, 3 usual ones, 2 defining how $=$ works, and 2 defining how $\forall$ works. The axioms are:

1. $p \Rightarrow (q \Rightarrow p)$, for each p, q formulae.
2. $[p \Rightarrow (q \Rightarrow r)] \Rightarrow [(p \Rightarrow q) \Rightarrow (p \Rightarrow r)]$, for each p, q formulae.
3. $(\neg\neg p) \Rightarrow p$, for p any formula.
4. $(\forall x)(x = x)$, for x any variable.
5. $(\forall x)(\forall y)(x = y \Rightarrow (p \Rightarrow p[y/x]))$, for any variables x, y, and a formula p with y not occurring bound.
6. $[(\forall x)p] \Rightarrow p[t/x]$ for any variable x, formula p, and term t with no free variable of t occurring bound in p.
7. $(\forall x)(p \Rightarrow q) \Rightarrow (p \Rightarrow (\forall x)q)$, for any variable x, formulae p, q with x not occurring free in p.

It can be checked that each of these is a tautology, meaning they are true in every structure.

We can now define the deduction rules:

1. Modus ponens: from p and $p \Rightarrow q$, we can deduce q.
2. Generalization: from p, we can deduce $(\forall x)p$, provided x does not occur free in any premise or hypothesis used to prove p.

For $S \subset L$ and $t \in L$, we say S *proves* p, written $S \vdash p$, if there exists a proof of p from S, meaning a finite sequence of formula, ending with p, such that each formula is either:

- a logical axiom, or
- a member of S, or
- obtained from earlier lines by one of the two deduction rules.

Remark. Suppose we allowed the empty structure A, then $\perp$ is false in A, and $(\forall x)\perp$ is true in A. So $((\forall x)\perp) \Rightarrow \perp$ is false in A.

But this is an instance of axiom 6.

We look at some examples of proofs.

Example 4.10.

Let us show that $\{x = y, x = z\} \vdash y = z$. Our strategy is to go for axiom 5, to get $y = z$ from $x = z$. Beginning:

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).
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).
6. $x = y$ (hypothesis).
7. $x = z \Rightarrow y = z$ (modus ponens).
8. $x = z$ (hypothesis).
9. $y = z$ (modus ponens).

We now immediately state and prove the deduction theorem.

Proposition 4.1 (Deduction theorem)**.** *Let $S \subset L$ and $p, q \in L$. Then $S \vdash (p \Rightarrow q)$ if and only if $S \cup \{p\} \vdash q$.*

Proof: As before, given a proof of $p \Rightarrow q$ from S, we can then write down p (hypothesis), and then q (modus ponens), to obtain a proof of q from $S \cup \{p\}$.

For the converse, we prove the same result: that if our proof if q from $S \cup \{p\}$ is $t_1, t_2, \ldots, t_n = q$, then we show S proves $p \Rightarrow t_1, p \Rightarrow t_2, \ldots, p \Rightarrow t_n$ by induction.

However our only new case is generalization, the other cases are the same as in the deduction theorem for propositional logic. Hence if in the proof of q from $S \cup \{p\}$, suppose we have r, followed by $(\forall x)r$.

Now we have a proof of $p \Rightarrow r$ from S by induction, and in the proof of r from $S \cup \{p\}$, no hypothesis had x free. So the same is true for the proof of $p \Rightarrow r$ in S. Hence by generalization, we can write down $S \vdash (\forall x)(p \Rightarrow r)$. There are now two cases:

- If x is free in p, then we get $S \vdash p \Rightarrow (\forall x)r$ by axiom 7, and modus

ponens.

- If x occurs free in p, then the proof of r from $S \cup \{p\}$ cannot have used hypothesis p, so in fact $S \vdash r$, and hence $S \vdash (\forall x)r$ by generalization.

  Therefore $S \vdash p \Rightarrow (\forall x)r$ by axiom 1, and modus ponens.

Our aim is now the same in propositional logic: we want to prove

$$S \models p \iff S \vdash p.$$

This is a strong statement. It says, for example, if p is true in all groups, then we can deduce p from the group axioms.

We again prove soundness and adequacy.

Proposition 4.2 (Soundness)**.** *Let S be a set of sentences and p a sentence in language L. Then,*

$$S \vdash p \implies S \models p.$$

Proof: Consider our proof $t_1, \ldots, t_n = p$ of p from S, and we want to know if A is a model of S, then A is a model of t_i for all i. This is an easy induction.

For adequacy, we run the same trick as we did for propositional logic. We want:

$$\begin{aligned} S \models p &\implies S \vdash p, \\ \iff S \cup \{\neg p\} \models \bot &\implies S \cup \{\neg p\} \vdash \bot, \\ \iff S \cup \{\neg p\} \text{ consistent} &\implies S \cup \{\neg p\} \text{ has a model,} \end{aligned}$$

by taking the contrapositive.

Theorem 4.1 (Model Existence lemma)**.** *Let S be a set of sentences in language L. Then if S is consistent, then S has a model.*

Some ideas are as follows:

1. We can build our structure out of the language itself, using the closed terms of L.

   For example, if L is the language of fields, and S is the field axioms, then we can take the closed terms with $+$ and $\times$, in the obvious way:

   $$`(1+1)' + `(1+1)' = `(1+1) + (1+1)'.$$

2. However, some terms in our above structure may be equal in every theory, for example the closed terms $1+0$ and 1 are distinct, and $S \vdash 1+0=1$.

   Hence another idea is to quotient out by equivalence relation on closed terms given by $s \sim t$. If this set is A, then we can define $[s] +_A [t] = [s+t]$.

3. Suppose S are the field of characteristic 2 or 3. Then S is the field axioms with the sentence $1+1=0 \wedge 1+1+1=0$.

   Then $S \nvdash 1+1=0$, and $S \nvdash 1+1+1=0$, so A is not characteristic 2 or 3. The solution is to extend S to a *maximal* consistent set first.

4. Now suppose S are the fields with $\sqrt{2}$, i.e. S is the field axioms with $(\exists x)(x \times x = 1+1)$. Then no closed term t has $[t \times t] = [1+1]$.

   The problem is that S 'lacks witnesses'. To solve this, for each $(\exists x)p \in S$, we add a new constant c to the language, and add to S the sentence $p[c/x]$. It is still easy to check this is still consistent.

5. But now our new S is not maximal consistent, as we extended it. So we must loop back to step 3. The problem is, this process might not terminate.

Proof: We have a consistent S in language $L = L(\Omega, \Pi)$. Extend S to a maximal consistent S, in L via Zorn's lemma. Then for each sentence $p \in L$, we have $p \in S_1$ or $(\neg p) \in S_1$, so S_1 is complete.

Now add witnesses to S_1: for each $(\exists x)p \in S_1$, add a new constant c, and add sentence $p[c/x]$. Then we obtain a theory T_1 in the language $L_1 = L(\Omega \cup C_1, \Pi)$ that has witnesses for S_1. Then it is easy to check that T_1 is consistent.

We can then extend T_1 to a maximal consistent S_2 in L_1, and add witnesses to form T_2 in language $L_2 = L(\Omega \cup C_1 \cup C_2, \Pi)$. Then we can continue inductively.

Let $\bar{S} = S_1 \cup S_2 \cup \cdots$ in the language $\bar{L} = L(\Omega \cup C_1 \cup C_2 \cup \cdots, \Pi)$. Then we claim that $\bar{S}$ is consistent, complete and has witnesses.

- It is consistent: if $\bar{S} \vdash \bot$, then some $S_n \vdash \bot$ for some $n \in \mathbb{N}$, a contradiction.
- It is complete: for a sentence $p \in \bar{L}$, we have $p \in L_n$ for some n. So $S_{n+1} \vdash p$ or $S_{n+1} \vdash (\neg p)$, so $\bar{S} \vdash p$ or $\bar{S} \vdash (\neg p)$.
- It has witnesses: if $(\exists p) \in \bar{S}$, then it is in S_n for some n. So $p[t/x] \in T_n$, for some closed term t.

Now we can begin taking quotients in $\bar{L}$. On the closed terms of $\bar{L}$, define $s \sim t$ if $\bar{S} \vdash (s = t)$. This is an equivalence relation. Now let A be the set of equivalence classes, made into an $\bar{L}$-structure by:

- $f_A([t_1], \ldots, [t_n]) = [ft_1 \ldots t_n]$, for each $f \in \Omega \cup C_1 \cup C_2 \cup \cdots$, $\alpha(f) = n$, and $t_1, \ldots, t_n$ closed terms.
- $\phi_A = \{([t_1], \ldots, [t_n]) \in A^n \mid \bar{S} \vdash \phi(t_1, \ldots, t_n)\}$, for each $\phi \in \Pi$, $\alpha(\phi) = n$, $t_1, \ldots, t_n$ closed terms.

Now we claim that for a sentence $p \in \bar{L}$, we have $p_a = 1 \iff \bar{S} \vdash p$. Then we are done, as certainly $p_A = 1$ for all $p \in S$, i.e. A is a model of S.

The proof of this claim is an easy induction. For atomic sentences,

- $\bot_A = 0$, and $\bar{S} \nvdash \bot$.
- For closed terms such that $\bar{S} \vdash (s = t)$, we have

$$\bar{S} \vdash (s = t) \iff [s] = [t] \iff s_A = t_A \iff s = t \text{ in } A.$$

- The case for $\phi(t_1, \ldots, t_n)$ is the same.

Now we have the induction step.

- For $p \Rightarrow q$, we have

$$\begin{aligned} \bar{S} \vdash (p \Rightarrow q) &\iff \bar{S} \vdash (\neg p) \text{ or } \bar{S} \vdash q \iff p_A = 0 \text{ or } q_A = 1 \\ &\iff (p \Rightarrow q) \text{ in } A. \end{aligned}$$

- For the statements involving $\forall$, we instead use $\exists$, and have

$$\begin{aligned} \bar{S} \vdash (\exists x)p &\iff \bar{S} \vdash p[t/x] \text{ (as } \bar{S} \text{ has witnesses)} \\ &\iff p[t/x]_A = 1 \text{ for some closed term } t \\ &\iff (\exists x)p \text{ in } A, \end{aligned}$$

  as A is the set of all equivalence classes of all closed terms.

Hence, we get adequacy.

Corollary 4.1 (Adequacy)**.** *For S a theory, and p a sentence, in our language we have*

$$S \models p \implies S \vdash p.$$

Thus, combining these two, we get:

Theorem 4.2 (Completeness theorem)**.** *For S a theory, and p a sentence in language L, we have*

$$S \vdash p \iff S \models p.$$

Remark.

1. If L is countable, then Zorn is not needed.
2. 'First-order' means that our variables range over elements, not subsets.

Theorem 4.3 (Compactness theorem)**.** *Let S be a theory in a language L. Then if every finite subset of S has a model, then S has a model.*

Proof: This is trivial if we replace 'has a model' with 'is consistent', as proofs are finite.

Note there is no decidability theorem, as we cannot check if $S \models p$.

4.3 Applications of Compactness

Corollary 4.2. *The class of finite groups is not axiomatizable (in the language of groups).*

It is remarkable we can prove this, as opposed to merely guessing it.

Proof: Suppose S axiomatizes the theory of finite groups. Consider S together with the sentences:

1. $(\exists x_1)(\exists x_2)(x_1 \neq x_2)$.
2. $(\exists x_1)(\exists x_2)(\exists x_3)(x_1, x_2, x_3 \text{ distinct})$.
3. $\cdots$

This gives a set S' of axioms, for which any finite subset of S' has a model (for example, $\mathbb{Z}_n$ for some n large enough). So S' has a model - a finite group which has $\geq n$ elements for all $n \in \mathbb{N}$, contradiction.

Similarly,

Corollary 4.3. *Let S be a theory with arbitrarily large finite models. Then S has an infinite model.*

Proof: As above, add sentences and apply compactness.

The slogan is:

finiteness is not a first-order property.

We can also go up one cardinality.

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

Proof: Add constants $\{c_i \mid i \in I\}$ to the language, where I is an uncountable set, and form theory S' by adding to S the sentences $c_i \neq c_j$ for each $i, j \in I$, with $i \neq j$. Then any finite subset of S' has a model (indeed, our infinite model of S will do. So S' has a model.

Similarly, we can get a model of S that does not inject into X, for any fixed set X: just choose $\gamma(X)$ constants, or $\mathcal{P}(X)$ constants.

Example 4.11.

There exists an infinite field, $\mathbb{Q}$, so there exists an uncountable field, e.g. $\mathbb{R}$, and also, say, a field that doesn't inject into $\mathcal{P}(\mathcal{P}(\mathbb{R}))$.

There is also a partial converse to this.

Theorem 4.5 (Downward Löwenheim-Skolem Theorem)**.** *Let S be a theory in a countable language. Then if S has a model, it also has a countable model.*

Proof: We have S consistent, and then the model constructed in the proof of adequacy is countable.

4.4 Peano Arithmetic

We try to make the global axioms to make $\mathbb{N}$ into a first-order theory.

Firstly, our language is $L : \Omega = \{0, s, +, \cdot\}$, and $\Pi = \emptyset$, where $\alpha(\Omega) = (0, 1, 2, 2)$. The axioms are:

1. $(\forall x)(s(x) \neq 0)$.
2. $(\forall x)(\forall y)(s(x) = s(y) \Rightarrow x = y)$.

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

These axioms are called *Peano Arithmetic* or PA or *formal number theory.*

Regarding axiom 3, our first guess would be the same formula without the parameters, but then we would be missing sets such as $\{x \mid x \geq y\}$, where y is a variable.

Now PA has an infinite model, namely $\mathbb{N}$, so by upward Löwenheim-Skolem, it has an uncountable model, which in particular is not isomorphic to $\mathbb{N}$. But this contradicts the fact that the usual axioms for $\mathbb{N}$ characterizes $\mathbb{N}$ uniquely.

The answer is that axiom 3 is not 'true' induction (over all subsets). Even in $\mathbb{N}$ itself, axiom 3 applies to only countably many subsets.

Say $S \subset \mathbb{N}$ is *definable* or *definable in the language of* PA, if there exists a formula p with free variable x, such that for every $m \in \mathbb{N}$,

$$m \in S \iff p[m/x]$$

holds in $\mathbb{N}$ (where $m = s(s(\cdots s(0)\cdots))$). As there are countably many formula, there are only countable many definable sets.

Example 4.12.

Some examples of definable sets:

1. The set if squares: $(\exists y)(y \cdot y = x)$.
2. The set of primes: $x \neq 0 \wedge x \neq 0 \wedge (\forall y)(y \mid x \Rightarrow y = 1 \vee y = x)$.
3. The set of powers of 2: $(\forall y)(y \text{ is prime} \wedge y \mid x \Rightarrow y = 2)$.
4. Similarly we can define the set of powers of 4 (the powers of 2 which are squares), and powers of 6.

The question is whether PA is complete. For many years, mathematicians tried to prove this. However,

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

So we have a sentence p such that $\mathsf{PA} \nvdash p$, $\mathsf{PA} \nvdash \neg p$. But one of $p, \neg p$ holds in $\mathbb{N}$. Therefore, there exists a sentence p that is true in $\mathbb{N}$, but which PA does not prove.

This does not contradict the completeness theorem, which would tell us that if p is true in *every* model of PA, then $\mathsf{PA} \vdash p$.

5 Set Theory

Our goal to find out what the universe of sets looks like. While this sounds meaningless, we can take a viewpoint of set theory as a first-order theory.

We will be looking at *Zermelo-Fraenkel* set theory, which has language $\Omega = \emptyset$, $\Pi = \{\in\}$, where $\alpha(\in) = 2$, and a 'universe of sets' is a model $(V, \in)$ of the ZF axioms.

There are 9 axioms: 2 to get started, 4 to build things, and 3 subtle axioms.

We can view the entirety of this chapter as a worked example of a first-order theory, but much scarier, since (hopefully) every model of ZF will contain 'all of mathematics', and so will be very complicated.

The axioms of ZF are as follows (they all have fancy names):

1. Axiom of extension: sets with the same members are equal.

   $$(\forall x)(\forall y)[(\forall z)(z \in x \iff z \in y) \Rightarrow x = y].$$

   Note the converse is an instance of a logical axiom.

2. Axiom of separation (also called comprehension or subset selection): we can form subsets of a set, or more precisely, for a set x and a property p, we can form $\{z \in x \mid p(z)\}$.

   $$(\forall t_1) \cdots (\forall t_n)(\forall x)(\exists y)(\forall z)(z \in y \iff z \in x \wedge p),$$

   for each formula p and free variables $x_1, \ldots, x_n$. Note that we do need parameters as we may want to form $\{z \in x \mid z \in t\}$, for some variable t.

3. Empty-set axiom: there is an empty set.

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

   We write $\emptyset$ for the (unique, by extension) set guaranteed by this axiom. This is an abbreviation, so $p(\emptyset)$ means $(\exists x)(x$ has no members $\wedge\, p(x))$. Similarly, we write $\{z \in x \mid p(x)\}$ for the set guaranteed by the axiom of separation.

4. Pair-set axiom: We can form $\{x, y\}$:

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

   We write $\{x, y\}$ for this z. We write $\{x\}$ for $\{x, x\}$.

We can now define the ordered pair $(x, y) = \{\{x\}, \{x, y\}\}$. It follows from the axioms so far that $(x, y) = (z, t) \iff x = z, y = t$.

Say x is an *ordered pair* if $(\exists y)(\exists z)(x = (y, z))$, and f is a *function* if

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

Call x the *domain* of f, written $x = \mathrm{Dom}(f)$ if

$$(f \text{ if a function}) \wedge (\forall y)(y \in x \iff (\exists z)((y, z) \in f)),$$

and then $f : x \to y$ means

$$(f \text{ is a function}) \wedge (x = \mathrm{Dom}(f)) \wedge (\forall z)(\forall t)((z, t) \in f \Rightarrow t \in y).$$

5. Union axiom: we can form unions.

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

6. Power-set axiom: we can form power-sets.

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

We write $\bigcup x$ and $\mathcal{P}(x)$ for the sets guaranteed by these axioms. We can write $x \cup y$ for $\bigcup\{x, y\}$, etc.

No new axioms are needed for intersection. We can form $\bigcap x$ for x any non-empty set, as a subset of y for any $y \in x$, so we are done by separation.

We can now form $x \times y$, as a subset of $\mathcal{P}(\mathcal{P}(x \cup y))$, because if $t \in x$, $z \in y$, then $(t, z) \in \mathcal{P}(\mathcal{P}(x \cup y))$. Moreover the st of all functions from x to y exists as a subset of $\mathcal{P}(x \times y)$.

7. Axiom of infinity: so far, any model V must be infinite. For example, writing x^+ for $x \cup \{x\}$, the successor of x, we have $\emptyset, \emptyset^+, \emptyset^{++}$ distinct.

   We often write 0 for $\emptyset$, 1 for $\emptyset^+$, 2 for $\emptyset^{++}$, and so on.

   However, V may not have an infinite set. In the world of maths, we know V infinite. But no $x \in V$ has all $y \in V$ as members, by Russell's paradox $(\forall x)\neg(\forall y)(y \in x)$.

   We say x is a *successor set* if $(\emptyset \in x) \wedge (\forall y)(y \in x \Rightarrow y^+ \in x)$. The axiom of infinity says that there exists an infinite set/successor set:

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

Note that any intersection of successor sets is again a successor set. So there exists a least successor set, namely the intersection of all successor sets. Call this ω, then this will be our copy, in V, of the usual natural numbers. Thus,

$$(\forall x)(x \in \omega \iff (\forall y)(y \text{ a successor set} \Rightarrow x \in y)).$$

For example, $3 = \emptyset^{+++} \in \omega$. In particular, if $x \subset \omega$ is a successor set, then $x = \omega$ by the definition of ω:

$$(\forall x)((x \subset \omega \wedge \emptyset \in x \wedge (\forall y)(y \in x \Rightarrow y^+ \in x)) \Rightarrow x = \omega).$$

It is easy to check that $(\forall x)(x \in \omega \Rightarrow x^+ = \emptyset)$, and also satisfies $(\forall x)(\forall y)((x \in \omega \wedge y \in \omega \wedge x^+ = y^+) \Rightarrow x = y)$. These satisfy (in V) the usual axioms for $\mathbb{N}$.

We can now define 'x is finite' for $(\exists y)(y \in \omega \wedge x$ bijects with $y)$, and 'x is countable' for (x is finite) $\vee$ (x bijects with ω).

8. Axiom of foundation: Essentially, we want sets to be built out of simpler sets. Hence, we want to:

   - disallow $x \in x$,
   - disallow $x \in y$, $y \in x$,
   - disallow $x_0, x_1, \ldots$ with $x_1 \in x_0$, $x_2 \in x_1$, $x_3 \in x_2 \ldots$

   We can nicely summarize want we want to do by forcing every (non-empty) set to have an $\in$-minimal element:

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

9. Axiom of replacement: Often we take sets A_i for each $i \in I$, and then take $(A_i \mid i \in I)$. However, we do not know whether that is a set, or whether $i \mapsto A_i$ is a function. Why should there be a set $\{(i, A_i) \mid i \in I\}$?

   To get this, we want the image of a set, under something that looks like a function, to be a set.

Classes: Let $(V, \in)$ be an L-structure. A *class* is a collection C of elements of V such that, for some formula p, and free variables x (and possibly more), we have that

$$x \text{ belongs to } C \iff p(x) \text{ holds in } V.$$

Example 5.1.

1. The whole of V is a class. Take p to be '$x = x$'.
2. All infinite $x \in V$ is a class. Take p to be 'x is not finite'.
3. The collection of all x such that $t \in x$ is class, by taking p to be '$t \in x$'.

Note that every set $y \in V$ is a class: take p to be '$x \in y$'. Say C is a *proper class* if it is not a set in V, i.e.

$$\neg(\exists y)(\forall x)(x \in y \iff p(x)).$$

For example, V itself is a proper class. Similarly, a *function-class* F is a collection of ordered pairs from V such that for some formula p and free variables x, y (and maybe more), we have that

$$(x, y) \text{ belongs to } F \iff p(x, y),$$

and if (x, y), (x, z) belong to F, then $y = z$.

For example, $x \mapsto \{x\}$ is a function class, by taking $p(x, y)$ to be '$y = \{x\}$'. Note this is not a function: every function has a domain (obtained as a suitable subset of $\bigcup\bigcup f$), and this f would have domain V.

> Getting back to the axiom of replacement, we can rephrase what we want to say, using the theory of classes, as the image of a set under a function-class is a set.
>
> $$(\forall t_1)\dots(\forall t_n)[(\forall x)(\forall y)(\forall z)(p \wedge p[z/y] \Rightarrow y = z) \\ \Rightarrow (\forall x)(\exists y)(\forall z)(z \in y \iff (\exists t)(t \in x \wedge p[t/x, z/y]))],$$
>
> for each p, and free variables $t_1, \dots, t_n, x, y$.

Hence for any set x, we can form $\{\{t\} \mid t \in x\}$, using the function-class $t \mapsto \{t\}$. However this is a bad example, as we can form this set directly using the axiom of power-set and axiom of separation.

The above are the axioms of ZF. We write ZFC for ZF and AC, where AC is the *axiom of choice*: every family of non-empty sets has a choice function:

$$(\forall f)((f \text{ is a function}) \wedge (\forall x)(x \in \operatorname{Dom} f \Rightarrow f(x) \neq \emptyset) \Rightarrow (\exists g)((g \text{ is a function}) \\ \wedge (\operatorname{Dom} g = \operatorname{Dom} f) \wedge (\forall x)(x \in \operatorname{Dom} f \Rightarrow g(x) \in f(x)))).$$

Say x is *transitive* if each member of a member of x, is a member is x, i.e.

$$(\forall y)[(\exists z)(y \in z \wedge z \in x) \Rightarrow y \in x].$$

For example, $\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}$ are transitive, and in general, each $x \in \omega$ is transitive.

Lemma 5.1. *Every set x is contained in a transitive set.*

Remark.

1. Officially, this lemma says "let $(V, \in)$ be a model of ZF. Then..."
2. Once we know this lemma, we will know that any x is contained in a least transitive set, the *transitive closure* of x, written $TC(x)$, because any intersection of transitive sets is transitive.

Proof: We want to form $x \cup (\bigcup x) \cup (\bigcup\bigcup x) \cup \cdots$. This will be a set, by the union axiom applied to $\{x, \bigcup x, \bigcup\bigcup x, \ldots\}$. This itself is a set by replacement: it is the image of ω under the function-class $0 \mapsto x$, $1 \mapsto \bigcup x$, $2 \mapsto \bigcup\bigcup x$,...

But why is this a function class? We want $p(z, w)$ to be: $(z = 0 \wedge w = x) \vee ((\exists t)(\exists w)(z = t^+ \wedge w = \bigcup x \wedge p(t, w)))$. However, this is nonsense as it is not a formula.

Define f is an *attempt* to mean

$$(f \text{ is a function}) \wedge (\text{Dom } f \in \omega) \wedge (\text{Dom } f \neq \emptyset) \wedge (f(0) = x) \\ \wedge (\forall n \in \omega)(n \in \text{Dom } f \wedge n \neq 0 \Rightarrow f(x) = \bigcup f(n-1)).$$

Then $(\forall n \in \omega)(\exists f)(f \text{ is an attempt} \wedge n \in \text{Dom } f)$, by ω-induction. And moreover,

$$(\forall n \in \omega)(\forall f)(\forall g)(f \text{ an attempt} \wedge g \text{ an attempt} \\ \wedge n \in \text{Dom } f \wedge n \in \text{Dom } g \Rightarrow f(n) = g(n)),$$

also by ω-induction. So our function-class $p = p(z, w)$ is:

$$(\exists f)(f \text{ an attempt} \wedge z \in \text{Dom } f \wedge f(z) = w).$$

We want foundation to be capturing the idea of 'sets are built out of simpler sets'. So we want: if $p(y) \forall y \in x$, implies $p(x)$, then $p(x) \forall x$.

Theorem 5.1 (Principle of $\in$-induction)**.** *For each formula p, free variables $t_1, \ldots, t_n, x$:*

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

Proof: Given, $t_1, \ldots, t_n$, and given $(\forall x)((\forall y)(y \in x \Rightarrow p(y)) \Rightarrow p(x))$, we want $(\forall x)p(x)$.

Suppose some x has $\neg p(x)$. We want to look at $\{t \mid \neg p(t)\}$, and take an $\in$-minimal element. However this may not be a set: e.g. if $p(x)$ is $x \neq x$.

To fix this, let $u = \{t \in TC(\{x\}) \mid \neg p(t)\}$. Then $u \neq \emptyset$ as $x \in u$. Let t be a minimal element of u. Then $\neg p(t)$ as $t \in U$, but $p(z)$ for all $z \in t$ by minimality, noting that each $z \in t$ does belong to $TC(\{x\})$.

In fact, $\in$-induction is equivalent to foundation, in the presence of the other ZF axioms.

To deduce foundation, say x is *regular* if

$$(\forall y)(x \in y \Rightarrow y \text{ has a minimal element}).$$

So foundation says: every set is regular. We can prove this by $\in$-induction. Given $(\forall y \in x)(y \text{ is regular})$, we also want x regular.

Consider a set z with $x \in z$. If x is minimal in z, then we are done. Otherwise if x is not minimal in z, there exists a $y \in x$ such that $y \in z$. So z has a minimal element as y is regular.

Similarly, we can define $\in$-recursion:

Theorem 5.2 ($\in$-recursion theorem)**.** *Let G be a function-class, everywhere defined. Then there is a function class F, everywhere defined, such that*

$$(\forall x)(F(x) = G(F|x)).$$

Moreover, F is unique.

Proof: For existence, say f is an *attempt* if

$(f \text{ is a function}) \wedge (\text{Dom } f \text{ transitive}) \wedge (\forall x)(x \in \text{Dom } f \Rightarrow f(x) = G(f|x))$.

Then,

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

by $\in$-induction. Also,

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

also by $\in$-induction. Indeed, if for each $y \in x$, there exists an attempt defined at y, then for each $y \in x$ there is a unique attempt defined on $TC(\{y\})$, say f_y. Let

$$f = \bigcup \{f_y \mid y \in x\},$$

an attempt with domain $TC(\{x\})$. Then set

$$f' = f \cup \{(x, G(f|x))\},$$

is an attempt defined at x. So take $q(x, y)$ to be

$$q(x, y) = (\exists f)(f \text{ an attempt} \wedge x \in \operatorname{Dom} f \wedge f(x) = y).$$

For uniqueness, if F, F' are suitable then $(\forall x)(F(x) = F'(x))$, by $\in$-induction.

Remark. These proofs of $\in$-induction and $\in$-recursion are similar to ordinal induction and recursion.

For $\in$-induction and $\in$-recursion, we needed the following properties of $p(x, y) = x \in y$:

1. p is *well-founded*: every non-empty set has a p-minimal element.
2. p is *local*: for each y, $[x \mid p(x, y)]$ forms a set.

So actually, we have p-induction and p-recursion for any p that is well-founded and local. In particular, if r is a relation on a set a, then trivially r is local, so we just need r to be well-founded.

Thus our theorems from chapter 2 are special cases of this, as a well-ordering is a well-founded total order.

Then a natural question to ask is, what relations can be modelled by $\in$? In particular, if we have a set $\{a, b, c\}$ given by $a\, r\, b$, $b\, r\, c$, then letting $a' = \emptyset$, $b' = \{\emptyset\}$ and $c' = \{\{\emptyset\}\}$, then the map $f : \{a, b, c\} \to \{a', b', c'\}$, giving by $x \mapsto x'$, is a bijection with a transitive set such that

$$x\, r\, y \iff f(x) \in f(y).$$

Say a relation r on a set a is *extensional* if

$$(\forall x \in a)(\forall y \in a)[(\forall z \in a)(z \in x \iff z \in y) \Rightarrow x = y].$$

Then the analogue of subset collapse is:

Theorem 5.3 (Mostowski's Collapsing Theorem)**.** *Let r be a relation on a set a that is well-founded and extensional. Then there exists a transitive set b and a bijection $f : a \to b$ such that*

$$(\forall x \in a)(\forall y \in a)(x \, r \, y \Rightarrow f(x) \in f(y)).$$

Moreover, b and f are unique.

Proof: Define function f by r-recursion, as

$$f(x) = \{f(y) \mid y \, r \, x\},$$

for each $x \in a$. Note f is a function, not just a function-class, by replacement, as it is an image of a.

Similarly, $b = \{f(x) \mid x \in a\}$ is a set by replacement. Then f is surjective, by our definition of b, and b is transitive, by the definition of f. Now we need f injective. We will show that

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

by r-induction. So we are given

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

and we are given $f(x) = f(x')$. This is equivalent to

$$\{f(y) \mid y \, r \, x\} = \{f(z) \mid z \, r \, x'\},$$

so as y is unique,

$$\{y \mid y \, r \, x\} = \{z \mid z \, r \, x'\},$$

thus $x = x'$ as r is extensional.

We can prove f unique by r-induction, as we must have

$$f(x) = \{f(y) \mid y \, r \, x\},$$

for all $x \in a$.

In particular, every well-ordered set is order-isomorphic to a transitive set well-ordered by $\in$.

So say an *ordinal* is a transitive set well-ordered (or totally ordered) by $\in$. Thus each well-ordering is order-isomorphic to a unique ordinal, its order-type.

If x, y are in a well-ordered set a, with $y < x$, then the order type of I_x, has an element $f(y)$, i.e. the order type of I_y.

For ordinals α, β, $\alpha < \beta \iff \alpha \in \beta$, so $\alpha = \{\beta \mid \beta < \alpha\}$. Thus $\alpha^+ = \alpha \cup \{\alpha\}$, and

$$\sup\{\alpha_i \mid i \in I\} = \bigcup\{\alpha_i \mid i \in I\},$$

however this is unhelpful.

5.1 Picture of the Universe

We look at a picture of the universe. we hope, starting with $\emptyset$, we can keep taking power-sets to build everything.

Define sets V_α for each ordinal α, by recursion:

- $V_0 = \emptyset$,
- $V_{\alpha+1} = \mathcal{P}(V_\alpha)$,
- $V_\lambda = \bigcup\{V_\alpha \mid \alpha < \lambda\}$, for λ a non-zero limit.

Now does this hit all sets?

Lemma 5.2. *Each V_α is transitive.*

Proof: We use induction on α. First, for 0, $V_0 = \emptyset$, which is trivially transitive.

If we take a successor, then V_α transitive implies $V_{\alpha+1}$ transitive, because x transitive implies $\mathcal{P}(x)$ transitive: if $z \in y \in \mathcal{P}(x)$, then $z \in x$, so $z \subset x$, meaning $z \in \mathcal{P}(x)$.

Limits also work, as the union of transitive sets is transitive.

Lemma 5.3. *If $\alpha \leq \beta$, then $V_\alpha \subset V_\beta$.*

Proof: We use induction on β, with α fixed.

For $\beta = \alpha$, we trivially have $V_\alpha \subset V_\alpha$.

For successors, if $V_\alpha \subset V_\beta$, then $V_\beta \subset \mathcal{P}(V_\beta)$, as V_β transitive, so $V_\alpha \subset \mathcal{P}(V_\beta) = V_{\beta+1}$.

Limits are trivial, as they are the union of smaller sets, in particular V_α.

Lemma 5.4. *Every set x belongs to some V_α.*

Remark.

1. Note $x \subset V_\alpha \iff x \in V_{\alpha+1}$, so if x is a subset of some V_α, then it is in some V_α.
2. Once we know $x \subset V_\alpha$ for some α, then there exists a least such α, called the *rank* of x. For example, $\operatorname{rank} x = x$ for all $x \in \omega$. Moreover $\operatorname{rank} \omega = \omega$, and in fact $\operatorname{rank} \alpha = \alpha$ for all ordinals by induction.

Proof: We proceed by $\in$-induction. Given a set x, we may assume that for all $y \in x$, there exists α such that $y \subset V_\alpha$, i.e. $y \subset V_{\operatorname{rank}(y)}$. Thus for all $y \in x$, $y \in V_{\operatorname{rank}(y)+1}$.

So say $\alpha = \sup\{\operatorname{rank}(y) + 1 \mid y \in x\}$. Then for all $y \in x$, $y \in V_\alpha$, so $x \subset V_\alpha$.

Remark.

1. These V_α are called the *Von Neumann hierarchy*.
2. The above shows that for all x, $\operatorname{rank}(x) = \sup\{\operatorname{rank}(y) + 1 \mid y \in x\}$. This is the right way to think about rank.

6 Cardinals

We are looking at sizes of sets, working in ZFC. We introduce new notation: write $x \leftrightarrow y$ if $(\exists f)(f$ a bijection from x to $y)$.

We want to define $\operatorname{Card} x$ or $|x|$ such that $\operatorname{Card} x = \operatorname{Card} y$ if and only if $x \leftrightarrow y$. However, we cannot put $\operatorname{Card} x = \{y \mid x \leftrightarrow y\}$, as this is not a set.

Instead, for any x, there exists α an ordinal such that $x \leftrightarrow \alpha$, by the well-ordering theorem. Hence we can just define $\operatorname{Card} x$ to be the least α such that $x \leftrightarrow \alpha$.

In ZF where we cannot use well-ordering, we use the 'Scott Trick': we consider the least α such that there exists $y \leftrightarrow x$ with $\operatorname{rank} y = \alpha$. Then let $\operatorname{Card} x = \{y \subset V_\alpha \mid y \leftrightarrow x\}$.

Say m is a *cardinality* if $m = \operatorname{Card} x$, for some x.

An ordinal is *initial* if it does not biject with any smaller ordinal. Some initial ordinals are $0, 1, 2, \ldots, \omega, \omega_1$ and $\boldsymbol{\gamma}(X)$ for any set X. However ω^2 is not initial, as $\omega^2 \leftrightarrow \omega$.

Define the initial ordinals ω_α for each ordinal α, by recursion:

- $\omega_0 = \omega$,
- $\omega_{\alpha+1} = \boldsymbol{\gamma}(\omega_\alpha)$.
- $\omega_\lambda = \sup(\omega_\alpha \mid \alpha < \lambda)$, for λ a non-zero limit.

Then each ω_α is initial, and every initial ordinal β is an ω_α. Indeed, the ω_α are unbounded, as $\omega_\alpha \geq \alpha$ for all α, by induction. So there exists a least ordinal δ with $\beta < \omega_\delta$.

We must have δ a successor, else $\omega_\delta = \sup(\omega_\alpha \mid \alpha < \delta)$, contradiction to the definition of δ. Say $\delta = \alpha + 1$, so $\omega_\alpha \leq \beta < \omega_{\alpha+1}$. Then $\beta = \omega_\alpha$; otherwise, we contradict $\omega_{\alpha+1} = \boldsymbol{\gamma}(\omega_\alpha)$.

Write $\aleph_\alpha$ for $\operatorname{Card} \omega_\alpha$, for example $\operatorname{Card} \omega = \aleph_0$, $\operatorname{Card} \omega_1 = \aleph_1$.

So the $\aleph_\alpha$ are the cardinalities of all infinite sets (in ZF, the $\aleph_\alpha$ are the cardinalities of the infinite well-ordered sets).

For cardinals m, n, we write $m \leq n$ if there exists an injection from M to N, where M, N are sets with $\operatorname{Card} M = m$, $\operatorname{Card} N = n$. This does not depend on the choice of M, N. We also write $m < n$ if $m \leq n$ and $m \neq n$, for example $\operatorname{Card} \omega < \operatorname{Card} \mathcal{P}_\omega$. Moreover, if $m \leq n$ and $n \leq m$, then by Schröder-Bernstein, $m = n$, so $\leq$ is a partial order.

In fact, $\leq$ is a total order: we may well-order M, N, and then one injects into the other (in ZF, $\leq$ need not be a total order).

6.1 Cardinal Arithmetic

For cardinals m and n, define

$$m + n = \mathrm{Card}(M \sqcup N), \qquad mn = \mathrm{Card}(M \times N), \qquad m^n = \mathrm{Card}(M^N),$$

where $\mathrm{Card}\, M = m$, $\mathrm{Card}\, N = n$, and M^N is the space of functions from N to M. This does not depend on the choice of M and N.

We could also define, for example

$$\sum_{i \in I} m_i = \mathrm{Card}\left(\bigsqcup_{i \in I} M_i \right),$$

where M_i are sets with $\mathrm{Card}\, M_i = m_i$ for all i. This is well-defined thanks to AC.

Example 6.1.

1. $\mathbb{R} \leftrightarrow \mathcal{P}(\omega) \leftrightarrow \{0,1\}^\omega$, so $\mathrm{Card}\, \mathbb{R} = \mathrm{Card}(\mathcal{P}(\omega)) = 2^{\aleph_0}$.
2. How many sequences of reals are there? It is:

$$\mathrm{Card}(\mathbb{R}^\omega) = (2^{\aleph_0})^{\aleph_0} = 2^{\aleph_0 \aleph_0} = 2^{\aleph_0},$$

where we can use simple facts like:

(i) $m + n = n + m$ (as $M \sqcup N \leftrightarrow N \sqcup M$)

(ii) $mn = nm$ (as $M \times N \leftrightarrow N \times M$)

(iii) $(m^n)^p = m^{np}$ (as $(M^N)^P \leftrightarrow M^{N \times P}$)

(iv) $\aleph_0 \aleph_0 = \aleph_0$ (as $\omega \times \omega \leftrightarrow \omega$)

Now we know $\aleph_0 \aleph_0 = \aleph_0$, how about $\aleph_1 \aleph_1$? We can calculate it and all other cardinal addition and multiplication from the following theorem:

Theorem 6.1. *$m^2 = m$ for all infinite cardinals m.*

Proof: We will show $\aleph_\alpha^2 = \aleph_\alpha$ for all α, by induction. Define a well-ordering of $\omega_\alpha \times \omega_\alpha$ by 'going up in squares':

$(x, y) < (z, w)$ if either $\max(x, y) < \max(z, w)$, or $\max(x, y) = \max(z, w) = \beta$, with

- $y < \beta$, $z < \beta$, or
- $x = z = \beta$, $y < w$, or
- $y = w = \beta$, $x < z$.

Then for any $\delta \in \omega_\alpha \times \omega_\alpha$, we have $\delta \subset \beta \times \beta$, for some $\beta < \omega_\alpha$, hence by induction, we have $\beta \times \beta \leftrightarrow \beta$ (or β is finite).

So the initial segment I_δ, since it is contained in $\beta \times \beta$, has $\mathrm{Card}(I_\delta) \leq \mathrm{Card}(\beta) < \mathrm{Card}(\omega_\alpha)$. Hence our well-ordering has order-type $\leq \omega_\alpha$. So $\omega_\alpha \times \omega_\alpha \hookrightarrow \omega_\alpha$.

Trivially $\omega_\alpha \hookrightarrow \omega_\alpha \times \omega_\alpha$, so $\omega_\alpha \times \omega_\alpha \leftrightarrow \omega_a$.

Corollary 6.1. *For any ordinals $\alpha \leq \beta$, $\aleph_\alpha + \aleph_\beta = \aleph_\alpha \aleph_\beta = \aleph_\beta$.*

Proof:

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

So, for example, $X \sqcup X \leftrightarrow X$ for any infinite set X. However, cardinal exponentiation is hard.

For example, just in ZF, $2^{\aleph_0}$ need not even be an aleph (if $\mathbb{R}$ is not well-ordered). Even in ZFC itself, $2^{\aleph_0} = \aleph_1$ is independent of the ZFC axioms. This is called the *continuum hypothesis*.

ZFC does not even decide if $2^{\aleph_0} < 2^{\aleph_1}$. Even today, not all implications about cardinal exponentiation are known.

7 Incompleteness

Our aim is to show that PA is incomplete: that is, there exists a sentence p such that $\mathsf{PA} \nvdash p$, $\mathsf{PA} \nvdash \neg p$. Equivalently, there exists a sentence p, that is true in $\mathbb{N}$, such that $PA \nvdash p$.

(See Johnstone chapter 4 and 9 for a full proof). The idea is to find p saying "I am not provable", i.e. p such that p is true if and only if p is not provable. Then we are done: if p is false, then $\mathsf{PA} \vdash p$, so p holds in every model of PA, so in particular p holds in $\mathbb{N}$, and so also p is not provable.

Recall that $S \subset \mathbb{N}$ is *definable* if there exists a formula, with free variable x, such that

$$m \in S \iff p(m) \text{ true.}$$

For example, the set of primes is definable. Take $p(x)$ to be

$$(\forall y)(\forall z)(yz = x \Rightarrow (y = 1 \vee z = 1)) \wedge (x \neq 1).$$

Hence we can say 'm is prime' is definable.

Similarly, $f : \mathbb{N} \to \mathbb{N}$ is *definable* if there exists a formula p, with free variables x, y such that for all $m, n \in \mathbb{N}$,

$$n = f(m) \iff p(m, n) \text{ holds.}$$

For example $f(x) = \lfloor \frac{x}{2} \rfloor$ is definable, as we can take

$$p(x, y) = (x = 2y) \vee (x = 2y + 1).$$

One important fact is any function given by an algorithm is definable. For example, $f(x) = 2^x$ is definable (as there exists an algorithm to compute it).

7.1 Coding

In the language of PA, L has symbols, $0, s, +, \cdot, =, \bot, \Rightarrow, (,), \forall, x$ and an operation $'$ to create new variables $x', x'', \ldots$. We can code these from 1 to 12: $v(0) = 1, v(s) = 2, v(+) = 3, \ldots, v(') = 12$.

Now for a formula $p = c_1 c_2 c_3 \ldots c_n$, where the c_i are the symbols in p respectively, then we can encode it by

$$c(p) = 2^{v(c_1)} 3^{v(c_2)} 5^{v(c_3)} \cdots (n\text{'th prime})^{v(c_n)}.$$

For example if $p = (\forall x)(x = 0)$, then

$$c(p) = 2^8 3^{10} 5^{11} 7^9 11^8 13^{11} 17^5 19^1 23^9.$$

Not every number codes a formula: for example $2^7 3^8$, which encodes $\Rightarrow$ (, or 2^{13}, or $2^7 5^7$.

Write s_n for the formula coded by n, with $s_n =$ '$\perp$' if n does not code a formula. Note that 'n codes a formula' is definable, as there exists a formula to check this. Similarly, 'l, m, n code formulae, with s_n obtained from s_l and s_m by modus ponens' is definable, and similarly for generalization. Also, 'n codes a logical axiom or an axiom of PA' is definable.

Given $p_1, \ldots, p_n$ formula, we can code that sequence as

$$s(p_1 \ldots p_n) = 2^{c(p_1)} 3^{c(p_2)} \cdots (n\text{'th prime})^{c(p_n)}.$$

So 'n codes a proof' is definable, by using the above facts. This implies 'n codes a proof of s_m' is definable: say that is $\theta(m, n)$.

So $\phi(m) =$ 's_m is provable' is definable: namely,

$$\phi(m) = (\exists n)\theta(m, n)$$

7.2 The Clever Part

Now consider a statement $\chi(m)$, that 'm codes a formula s_m with one free variable, and $s_m(m)$ is unprovable'. This is clearly definable, so is given by some formula $p(x)$, i.e.

$$\chi(m) \text{ holds} \iff p(m) \text{ holds.}$$

Let N be the code for '$p(x)$'. Then $p(N)$ is:

'N codes a formula $s_N = p(x)$, with one free variable, and $s_N(N) = p(N)$ is unprovable.'

So the sentence $p(N)$ will do. Thus, we've shown:

Theorem 7.1. PA *is incomplete.*

One subtle point: why does our proof above (that $p(N)$ is true), does not formalize into a proof within PA? It turns out that we used the existence of a model of PA (namely the naturals), i.e. we used the statement Con(PA), i.e. 'PA is consistent', or equivalently

$$(\forall x)(x \text{ does not code a proof of } \perp).$$

So our proof above actually formalizes to

$$\mathsf{PA} \cup (\mathrm{Con}(\mathsf{PA})) \vdash p(N).$$

Hence,

Theorem 7.2. $\mathsf{PA} \nvdash \mathrm{Con}(\mathsf{PA})$.

We know that PA is complete. Is it possible to add some clever sentence t (that is true in $\mathbb{N}$) to PA to get a complete theory? The answer is no: we can run the proof of theorem 1 on '$\mathsf{PA} \cup \{t\}$'.

However, we can certainly extend PA to a complete theory: just take T to be all sentences that are true in $\mathbb{N}$. Then why can't we run the proof of theorem 1, replacing PA by T, to show that T is incomplete? This can only be because:

Theorem 7.3. *T is not definable.*

In other words, there is no algorithm to decide, given n, if s_n is true or not, or in other words:

Truth is not definable.

Now what about ZFC? Does $\mathsf{ZFC} \vdash \mathrm{Con}(\mathsf{PA})$? The answer is yes, as ZFC proves 'PA has a model', namely ω. However, as with the first two theorems, we get:

Theorem 7.4. ZFC *is incomplete (if* ZFC *is consistent).*

Theorem 7.5. $\mathsf{ZFC} \nvdash \mathrm{Con}(\mathsf{ZFC})$.

Index