University of
Cambridge

Mathematics Tripos

Part II

Logic and Set Theory

Lent, 2018

Lectures by
I. B. Leader

Notes by
Qiangru Kuang

Contents

1 Propositional Logic

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

Definition (Language)**.** The *language* or *set of propositions* $L = L(P)$ is defined inductively by

1. for every $p \in P$, $p \in L$,
2. $\bot \in L$ (reads "false"),
3. if $p, q \in L$ then $(p \implies q) \in L$.

Example. $(p_1 \implies \bot)$, $((p_1 \implies p_2) \implies (p_1 \implies p_3))$, $((p_1 \implies \bot) \implies \bot)$ are elements of L.

Note.

1. Each proposition is a finite string of symbols from the alphabet $(,), \implies, p_1, p_2, ...$
2. "Inductively defined" means more precisely that we set
$$L_1 = P \cup \{\bot\}$$
$$L_{n+1} = L_n \cup \{(p \implies q) : p, q \in L_n\}$$
and then set $L = L_1 \cup L_2 \cup ....$ L_n can be seen as "things born by time n".
3. Each proposition is built up *uniquely* from (1), (2) and (3). For example, $((p_1 \implies p_2) \implies (p_1 \implies p_3))$ came from $(p_1 \implies p_2)$ and $(p_1 \implies p_3)$.

Note that we often omit outer brackets or use different brackets for clarity.

We can now define for example, $\neg p$ (reads "not p") as an abbreviation for $p \implies \bot$, $p \vee q$ (reads "p or q") for $(\neg p) \implies q$, $p \wedge q$ (reads "p and q") for $\neg(p \implies (\neg q))$.

1.1 Semantic Entailment

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

1. $v(\bot) = 0$,
2. $v(p \implies q) = \begin{cases} 0 & \text{if } v(p) = 1, v(q) = 0 \\ 1 & \text{otherwise} \end{cases}$ for all $p, q \in L$.

Remark. On $\{0, 1\}$, we could define a constant $\bot$ by $\bot = 0$ and an operation $\implies$ by
$$(a \implies b) = \begin{cases} 0 & \text{if } a = 1, b = 0 \\ 1 & \text{otherwise} \end{cases}$$
Then a valuation is a function $L \to \{0, 1\}$ that preserves the structure ($\bot$ and $\implies$), i.e. it is a homomorphism.

Proposition 1.1.

1. *If v and v' are valuations with $v(p) = v'(p)$ for all $p \in P$, then $v = v'$.*
2. *For any $w : P \to \{0, 1\}$, there exists a valuation v with $v(p) = w(p)$ for all $p \in P$.*

In other words, a valuation is determined by its values on P and any values will do.

Proof.

1. We have for all $p \in L_1$, $v(p) = v'(p)$. But if $v(p) = v'(p)$ and $v(q) = v'(q)$ then $v(p \implies q) = v'(p \implies q)$ so $v = v'$ on L_2. Continue inductively, we have $v = v'$ on L_n for all n.
2. Set $v(p) = w(p)$ for all $p \in P$ and $v(\bot) = 0$. This defines v on L_1. Having defined v on L_2, use $v(p \implies q) = \begin{cases} 0 & \text{if } v(p) = 1, v(q) = 0 \\ 1 & \text{otherwise} \end{cases}$ to define v on L_{n+1}.

□

Example. In a valuation given by

$$\begin{aligned} v(p_1) &= 1 \\ v(p_2) &= 1 \\ v(p_n) &= 0 \text{ for all } n \geq 3 \end{aligned}$$

we have $v(\underbrace{(p_1 \implies p_2)}_{1} \implies \underbrace{p_3}_{0}) = 0$.

Definition (Tautology)**.** p is a *tautology*, written $\vDash p$ if $v(p) = 1$ for all valuations v.

Example.

1. $p \implies (q \implies p)$. "A true statement is implied by anything". To show this we could write donw a truth table

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

2. $(\neg\neg p) \implies p$, i.e. $((p \implies \bot) \implies \bot) \implies p$. "Law of excluded middle".

3. $(p \implies (q \implies r)) \implies ((p \implies q) \implies (p \implies r))$. This is an example where writing down a truth table is not so desirable. Instead, this is not a tautology only if we have v with

$$v(p \implies (q \implies r)) = 1$$
$$v((p \implies q) \implies (p \implies r)) = 0$$

so $v(p \implies q) = 1, v(p \implies r) = 0$ whence $v(p) = 1, v(r) = 0$, so also $v(q) = 1$. But then $v(q \implies r) = 0$ so $v(p \implies (q \implies r)) = 0$. Absurd.

Definition (Semantic entailment)**.** For $S \subseteq L, t \in L$, we say S *entails* or *semantically implies* t, written $S \vDash t$, if $v(s) = 1$ for all $s \in S$ then $v(t) = 1$ for each valuation v.

This says whenever all of S is true, t is true as well.

Example. $\{p \implies q, q \implies r\} \vDash (p \implies r)$. Indeed, suppose not. So have v with $v(p \implies q) = v(q \implies r) = 1, v(p \implies r) = 0$. Then $v(p) = 1, v(r) = 0$, whence $v(q) = 0$ (from $v(q \implies r) = 1$), so $v(p \implies q) = 0$. Absurd.

Definition (Model)**.** If $v(t) = 1$, we say t is *true in* v or that v is a *model* of t.

For $S \subseteq L$, v is a *model* of S if $v(s) = 1$ for all $s \in S$.

Using this terminology, $S \vDash t$ says that every model of S is a model of t.

Note. $\vDash t$ is equivalent to $\emptyset \vDash t$.

1.2 Syntactic Implication

For a notion of "proof", we'll need axioms and deduction rules. As axioms, we'll take

1. $p \implies (q \implies p)$ for all $p, q \in L$.
2. $(p \implies (q \implies r)) \implies ((p \implies q) \implies (p \implies r))$ for all $p, q, r \in L$.
3. $(\neg\neg p) \implies p$ for all $p \in L$.

Note. We have already checked that these are all tautologies. Sometimes we say 3 axiom schemes to mean 3 infinite sets of axioms.

As deduction rules, we'll take just *modus ponens*: from p and $(p \implies q)$ we can deduce q.

Definition (Proof)**.** For $S \subseteq L$ and $t \in L$, a *proof* of t from S consists of a finite sequence $t_1, \ldots, t_n$ of propositions, with $t_n = t$ such that for every i, the proposition t_i is an axiom, or a member of S, or there exists $j, k < i$ with $t_j = (t_k \implies t_i)$.

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

Definition (Syntactical implication)**.** If there is a proof of t from S, say S *proves* or *syntactically implies* t, written $S \vdash t$.

Definition (Theorem)**.** t is a *theorem* if $\emptyset \vdash t$, written $\vdash t$.

Example. $\{p \implies q, q \implies r\} \vdash p \implies r$

1. $(q \implies r) \implies (p \implies (q \implies r))$, A1
2. $q \implies r$, hypothesis
3. $p \implies (q \implies r)$, MP
4. $(p \implies (q \implies r)) \implies ((p \implies q) \implies (p \implies r))$, A2
5. $(p \implies q) \implies (p \implies r)$, MP
6. $p \implies q$, hypothesis
7. $p \implies r$, MP

Example. $\vdash p \implies p$

1. $p \implies ((p \implies p) \implies p)$, A1
2. $(p \implies ((p \implies p) \implies p)) \implies ((p \implies (p \implies p)) \implies (p \implies p))$, A2
3. $(p \implies (p \implies p)) \implies (p \implies p)$, MP
4. $p \implies (p \implies p)$, A1
5. $p \implies p$, MP

The following theorem allows us to prove things much more easily:

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

Proof.

- $\Longleftarrow$: Given a proof of $p \implies q$ from S, append the lines
  1. p, hypothesis
  2. q, MP

  to obtain a proof of q from $S \cup \{p\}$.
- $\implies$: Let $t_1, \ldots, t_n = q$ be a proof of q from $S \cup \{p\}$. We'll show that $S \vdash p \implies t_i$ for all i. Split into cases
  - t_i is an axiom: write down
    1. $t_i \implies (p \implies t_i)$, A1
    2. t_i, axiom
    3. $p \implies t_i$, MP

- $t_i \in S$: identical as above.
- $t_i = p$: write down the proof $p \implies p$.
- t_i is obtained by MP: there exist $j, k < i$ such that $t_k = (t_j \implies t_i)$. By induction $S \vdash p \implies t_j$ and $S \vdash p \implies t_k$. Now write down
  1. $(p \implies (t_j \implies t_i)) \implies ((p \implies t_j) \implies (t \implies t_i))$, A1
  2. $p \implies (t_j \implies t_i)$, known
  3. $(p \implies t_j) \implies (p \implies t_i)$, MP
  4. $p \implies t_j$, known
  5. $p \implies t_i$, MP

and we can conclude $S \vdash p \implies t_i$ for all i.

□

Example. In order to show $\{p \implies q, q \implies r\} \vdash p \implies r$, it suffices to show $\{p \implies q, q \implies r, p\} \vdash r$ by deduction theorem, which is easy by using MP twice.

Now we have two turnstiles $\vDash$ and $\vdash$, how are they related? The aim of the rest of the chapter is to prove

Theorem 1.3 (Completeness theorem)**.** *$S \vDash t$ if and only if $S \vdash t$.*

We break this down into two directions:

- $\implies$: adequacy
- $\impliedby$: soundness

The easy part is

Proposition 1.4 (Soundness)**.** *If $S \vdash t$ then $S \vDash t$.*

Proof. Given v that models S and a proof $t_1, \dots, t_n = t$ of $S \vdash t$, we will show that $v(t_i) = 1$ for all i.

If t_i is an axiom then $v(t_i) = 1$ since it is tautology. If t_i is a hypothesis then $v(t_i) = 1$ by assumption. Finally, if t_i is obtained by MP, say from $t_j \implies t_i$, since $v(t_j) = 1$ and $v(t_j \implies t_i) = 1$ by induction, $v(t_i) = 1$. □

Note that soundness holds whenever our axioms are tautologies.

To prove adequacy, which is a bit harder, we need a few lemmas.

Definition (Consistency)**.** S is *inconsistent* if $S \vdash \bot$. Otherwise S is *consistent*.

Theorem 1.5 (Model existence lemma)**.** *Let $S \subseteq L$ be consistent, then S has a model.*

The first idea is to define a valuation v by $v(p) = 1$ if and only if $p \in S$. As 1 is preserved under $\vDash$ and thus $\vdash$, a more sensible aim is $v(p) = 1$ if and only if $S \vdash p$.

But maybe neither $S \vdash p$ nor $S \vdash \neg p$. So we want to "grow" S to contain one of p or $\neg p$ for each $p \in L$ (while remaining consistent).

Proof. Claim that for any consistent $S \subseteq L$, $S \cup \{p\}$ or $S \cup \{\neg p\}$ is consistent: if not, then $S \cup \{p\} \vdash \bot$ and $S \cup \{\neg p\} \vdash \bot$. But then $S \vdash (p \implies \bot)$ by deduction theorem, i.e. $S \vdash \neg p$. Then $S \vdash \bot$. Absurd.

Now as L is countable, we can list L as $t_1, t_2, ...$ Put $S_0 = S$. Set $S_1 = S_0 \cup \{t_1\}$ or $S_0 \cup \{\neg t_1\}$ such that S_1 is consistent. Then let $S_2 = S_1 \cup \{t_2\}$ or $S_1 \cup \{\neg t_2\}$ such that S_2 is consistent and continue inductively. Let $\overline{S} = S_0 \cup S_1 \cup ....$ Then $\overline{S} \supseteq S$ and $\overline{S}$ is consistent (as each S_n is consistent and proofs are finite). For all $p \in L$ either we have $p \in \overline{S}$ or $\neg p \in \overline{S}$. Also $\overline{S}$ is *deductively closed*, meaning that if $\overline{S} \vdash p$ then $p \in \overline{S}$. Indeed if $p \notin \overline{S}$ then $\neg p \in \overline{S}$, so $\overline{S} \vdash p, \overline{S} \vdash (\neg p)$, whence $\overline{S} \vdash \bot$. Absurd.

Define a valuation

$$v : L \to \{0, 1\}$$
$$p \mapsto \begin{cases} 1 & p \in \overline{S} \\ 0 & \text{otherwise} \end{cases}$$

Indeed, $v(\bot) = 0$ as $\bot \notin \overline{S}$. For $v(p \implies q)$:

- if $v(p) = 1, v(q) = 0$, we have $p \in \overline{S}, q \notin \overline{S}$, and we want $v(p \implies q) = 0$, i.e. $(p \implies q) \notin \overline{S}$. But if $(p \implies q) \in \overline{S}$ then $\overline{S} \vdash q$, $q \in \overline{S}$. Absurd.
- if $v(q) = 1$, we have $q \in \overline{S}$, and we want $v(p \implies q) = 1$, i.e. $(p \implies q) \in \overline{S}$. But $\vdash q \implies (p \implies q)$ so $\overline{S} \vdash (p \implies q)$.
- if $v(p) = 0$, we have $p \notin \overline{S}$. Then $(\neg p) \in \overline{S}$. We want $(p \implies q) \in \overline{S}$. Thus we need $(p \implies \bot) \vdash (p \implies q)$, which by deduction theorem is equivalent to $\{p \implies \bot, p\} \vdash q$. Thus suffices to show that $\bot \vdash q$. But we have $\vdash (\neg\neg q) \implies q$, and $\vdash (\bot \implies (\neg\neg q))$. Thus $\vdash (\bot \implies q)$, i.e. $\bot \vdash q$. Done.

□

Remark.

1. Sometimes this is called completeness theorem as it contains the majority of the work.
2. What would happen if P is uncountable? In fact, the result still holds. See chapter 3.

By remark before the above theorem, we now have

Corollary 1.6 (Adequacy)**.** *Let $S \subseteq L, t \in L$. Then if $S \vDash t$ then $S \vdash t$.*

Theorem 1.7 (Completeness theorem)**.** *Let $S \subseteq L, t \in L$. Then $S \vdash t$ if and only if $S \vDash t$.*

Proof. By soundness and adequacy. □

Some consequences:

Corollary 1.8 (Compactness theorem)**.** *Let $S \subseteq L, t \in L$ with $S \vDash t$. Then there exists a finite $S' \subseteq S$ with $S' \vDash t$.*

Proof. Trivial if we replace $\vDash$ with $\vdash$ as proofs are finite. □

Specialising to $t = \bot$, this theorem says that if S has no model then some finite $S' \subseteq S$ has no model. Equivalently,

Corollary 1.9 (Compactness theorem, equivalent form)**.** *Let $S \subseteq L$. If every finite finite subset of S has a model then S has a model.*

Proof. This is equivalent to the previous corollary because $S \vDash t$ if and only if $S \cup \{\neg t\}$ has no model and $S' \vDash t$ if and only if $S' \cup \{\neg t\}$ has no model. □

Corollary 1.10 (Decidability theorem)**.** *There is an algorithm to determine (in finite time) whether or not, for a given $S \subseteq L, t \in L$, we have $S \vdash t$.*

Remark. Highly non-obvious.

Proof. Trivial to decide if $S \vDash t$, just by drawing a truth table. □

2 Well-orderings and Ordinals

2.1 Definitions

Definition (Total order)**.** A *total order* or *linear order* on a set X is a relation $<$ on X that is

1. irreflexive: for all x, not $x < x$,
2. transitive: for all x, y, z, $x < y, y < z$ implies $x < z$,
3. trichotomous: for all x, y, $x < y, x = y$ or $y < x$.

Note. Any two of 3 cannot hold: if $x < y, y < x$ then $x < x$, absurd.

Notation. We write $x \leq y$ if $x < y$ or $x = y$. Write $y > x$ if $x < y$ etc.
In terms of $\leq$, a total order is

1. reflexive: for all x, $x \leq x$,
2. transitive: for all x, y, z, $x \leq y, y \leq z$ implies $x \leq z$,
3. antisymmetric: for all x, y, $x \leq y, y \leq x$ implies $x = y$,
4. trichotomous: for all x, y, $x \leq y$ or $y \leq x$.

Example.

1. $\mathbb{N}$ with usual order[1].
2. $\mathbb{Q}$ and $\mathbb{R}$ with usual order.
3. $\mathbb{N}^+$ with divisibility is *not* a total order as for example, 2 and 3 are not related.
4. Given a set S, the power set $\mathcal{P}(S)$ with $x \leq y$ if $x \subseteq y$ is *not* a total order for $|S| > 1$.

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

Example.

1. $\mathbb{N}$ with usual order.
2. $\mathbb{Z}$ with usual order is *not* a well-ordering. Similar for $\mathbb{Q}$ and $\mathbb{R}$.
3. $\{x \in \mathbb{Q} : x \geq 0\}$ is *not* a well-ordering. For example, $\{x \in \mathbb{Q} : x > 0\}$ does not have a least element.
4. $\{1 - 1/n : n = 2, 3, \dots\}$ is a well-ordering. This can be thought of $\mathbb{N}$ squashed into $[0, 1]$.

[1] In this course $0 \in \mathbb{N}$. Write $\mathbb{N}^+$ for $\mathbb{N} \setminus \{0\}$.

5. $\{1 - 1/n : n = 2, 3, ...\} \cup \{1\}$ is a well-ordering.

6. In fact, we can take the union of 4 with any real larger than 1 and still have a well-ordering.

7. $\{1 - 1/n : n = 2, 3, ...\} \cup \{2 - 1/n : n = 2, 3, ...\}$, i.e. two copies of 4, is still a well-ordering.

Remark. X is well-ordered if and only if there is no $x_1 > x_2 > x_3 > ...$ in X. Indeed, if there is such a sequence then $S = \{x_1, x_2, ...\}$ has no least element.

Corollary 2.1. *If $S \subseteq X$ has no least element, then for each $x \in S$ there exists $x' \in S$ with $x' < x$. Thus we have $x > x' > x'' > ....$*

Definition (Order isomorphism)**.** Total orders X and Y are *isomorphic* if there exists a bijection $f : X \to Y$ that is order-preserving, i.e. for all $x < x'$, $f(x) < f(x')$.

Example. 1 and 4 above are isomorphic. 5 and 6 are isomorphic. 6 and 7 are not isomorphic: for example, one has a greatest element and the other one doesn't.

Proposition 2.2 (Proof by induction)**.** *Let X be a well-ordering and $S \subseteq X$ be such that if $y \in S$ for all $y < x$ then $x \in S$ for each $x \in X$, then $S = X$.*

Equivalently, if $p(x)$ is a property such that for all x, if $p(y)$ for all $y < x$ then $p(x)$, then $p(x)$ for all $x \in X$.

Proof. If $S \neq X$ then let x be the least element in $X \setminus S$. Then $x \notin S$. But $y \in S$ for all $y < x$. Absurd. □

An application:

Proposition 2.3. *Let X and Y be isomorphic well-orderings. Then there is a* unique *isomorphism from X to Y.*

Remark. This is false for total orders in general. For example, from $\mathbb{Z}$ to $\mathbb{Z}$ we could take identity or $x \mapsto x - 5$.

Proof. Let f, g be isomorphisms. We will show $f(x) = g(x)$ for all $x \in X$ by induction. Thus we may assume $f(y) = g(y)$ for all $y < x$ and want $f(x) = g(x)$.

Let a be the least element of $Y \setminus \{f(y) : y < x\}$, which is non-empty. Then we must have $f(x) = a$: if $f(x) > a$ then some $x' > x$ has $f(x') = a < f(x)$, contradicting f being order-preserving. Same holds for g. Thus $f(x) = g(x)$. □

Definition (Initial segment)**.** In a total order X, an *initial segment* I is a subset of X such that $x \in I, y < x$ implies $y \in I$.

Example.

1. For any $x \in X$, set $I_x = \{y \in X : y < x\}$.

2. Not every initial segment is of this form. For example, in $\mathbb{R}$ take $\{x : x \leq 3\}$, or in $\mathbb{Q}$, take $\{x : x^2 < 2 \text{ or } x < 0\}$.

Note. In a well-ordering, every proper initial segment I is of the form I_x for some x. Indeed, let x be the least element of $X \setminus I$. Then $y < x$ implies $y \in I$ (by definition of y). Also if $y \in I$ then must have $y < x$: if $y = x$ or $y > x$ then $x \in I$ which is a contradiction.

The aim is to show every subset of a well-ordered X is isomorphic to an initial segment.

Note. This is false for total orders. For example, $\{1, 5, 9\} \subseteq \mathbb{Z}$, or $\mathbb{Q} \subseteq \mathbb{R}$.

Given $Y \subseteq X$, intuitively we want to map the smallest element of Y to the smallest element of X and continue this way. But how do we show every element of Y is mapped somewhere? Instead, we should work backwards: given $y \in Y$, map y to the smallest element in X that is not mapped to.

Theorem 2.4 (Definition by recursion)**.** *Let X be well-ordering, Y any set and $G : \mathcal{P}(X \times Y) \to Y$. Then there exists $f : X \to Y$ such that $f(x) = G(f|_{I_x})$ for all $x \in X$. Moreover f is unique.*

Note.

1. For $f : A \to B$ and $C \subseteq A$, the *restriction* of f to C is
$$f|_C = \{(x, f(x)) : x \in C\}.$$
2. Slogan: to define $f(x)$, make use of $f|_{I_x}$, i.e. the values of $f(y)$ for $y < x$.

Proof. First we show existence. Define "h is an attempt" to mean $h : I \to Y$ where $I \subseteq X$ is some initial segment, and for all $x \in I$ we have $h(x) = G(h|_{I_x})$. Note that if h and h' are both attempts defined at x then $h(x) = h'(x)$: by induction on x, if $h(y) = h'(y)$ for all $y < x$, then $h(x) = h'(x)$.

Also for all $x \in X$ there exists an attempt defined at x by induction. Indeed we want an attempt defined at x, given that for all $y < x$ there exists an attempt defined at y. So for each $y < x$ we have a unique attempt h_y defined on $\{z : z \leq y\}$ (unique by what we just showed). Let
$$h = \bigcup_{y<x} h(y),$$
an attempt defined on I_x (which is single-valued by uniqueness) so
$$h' = h \cup \{(x, G(h))\}$$
is an attempt defined at x. Now set $f(x) = y$ if there exists an attempt h defined at x with $h(x) = y$ (also single-valued).

For uniqueness, if f and f' are both suitable then $f(x) = f'(x)$ for all $x \in X$ by induction — if $f(y) = f'(y)$ for all $y < x$ then $f(x) = f'(x)$. □

A typical application:

Proposition 2.5 (Subset collapse)**.** *Let X be well-ordered, $Y \subseteq X$. Then Y is isomorphic to an initial segment of X. Moreover, the initial segment is unique.*

Proof. To have an isomorphism $f : Y \to I \subseteq X$, we need precisely that for all $x \in Y$, $f(x) = \min X \setminus \{f(y) : y < x\}$. So done (existence and uniqueness) by the previous theorem. Note that $X \setminus \{f(y) : y < x\} \neq \emptyset$, because $f(y) \leq y$ for all y by induction so $x \notin \{f(y) : y < x\}$. □

A note to the pedantic: in proving the set $X \setminus \{f(y) : y < x\}$ is non-empty, we seem to use a circular argument by assuming f exists. But this is just a shorthand for the longer version: define

$$f(x) = \begin{cases} \min X \setminus \{f(y) : y < x\} & \text{if } X \setminus \{f(y) : y < x\} \neq \emptyset \\ \text{cabbage} & \text{otherwise} \end{cases}$$

and then proceed to show $f(x) \neq \text{cabbage}$ for all $x \in X$.

In particular, a well-ordered X cannot be isomorphic to a proper initial segment of X, by uniqueness in subset collapse.

So far we have proved that if two well-orderings are isomorphic there is a unique isomorphism, and that a subset of a well-ordering is isomorphic to a (unique) initial segment. The question now is, how do different general well-orderings relate to each other?

Definition. Say $X \leq Y$ if X is isomorphic to an initial segment of Y.

Example. Let $X = \mathbb{N}$, $Y = \{1 - 1/n : n = 1, 2, \dots\} \cup \{1\}$, then $X \leq Y$.

What we would hope is that there is a total order on the set of all well-orderings. Firstly we have

Theorem 2.6. *Let X, Y be well-orderings, then $X \leq Y$ or $Y \leq X$.*

Proof. Suppose $Y \not\leq X$. To obtain $f : X \to Y$ that is an isomorphism with an initial segment of Y, need for all $x \in X$,

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

We can't have $Y = \{f(y) : y < x\}$ as then Y is isomorphic to I_x. Done by the theorem. □

Proposition 2.7. *Let X, Y be well-orderings with $X \leq Y$ and $Y \leq X$ then X and Y are isomorphic.*

Proof. Let f be an isomorphism from X to an initial segment of Y and g from Y to X. Then $g \circ f : X \to X$ is an initial segment of X (as an initial segment of an initial segment is an initial segment). so $g \circ f = \text{id}$ by uniqueness in subset collapse. Similarly $f \circ g = \text{id}_Y$. Thus X is isomorphic to Y. □

2.2 Constructing well-orderings

So far we have very few examples of well-orderings, so we wish to build new well-orderings from old.

Notation. Write $X < Y$ if $X \leq Y$ but X is not isomorphic to Y. Equivalently, $X < Y$ if and only if X is isomorphic to a proper initial segment of Y.

Example. If $X = \mathbb{N}$, $Y = \{1 - 1/n : n = 1, 2, \dots\} \cup \{1\}$ then $X < Y$.

We can produce new well-orderings by

- add a bigger element: a simple yet bona fide way to make a bigger well-ordering is, given a well-ordering X, choose $x \notin X$ and set $x > y$ for all $y \in X$. This is a well-ordering on $X \cup \{x\}$, written X^+. Clearly $X < X^+$.
- put some together: let $(X, <_X)$ and $(Y, <_Y)$ be well-orderings. Say Y *extends* X if $X \subseteq Y$, $<_X, <_Y$ agree on X, and X is an initial segment of $(Y, <_Y)$.

A family of well-orderings $\{X_i : i \in I\}$ are *nested* if for all $i, j \in I$, X_i extends X_j or X_j extends X_i.

Proposition 2.8. *Let $\{X_i : i \in I\}$ be a nested family of well-orderings. Then there exists well-ordering X with $X \geq X_i$ for all i.*

Proof. Let $X = \bigcup_{i \in I} X_i$, with $x < y$ if there exists i such that $x, y \in X_i$ and $x <_i y$, where $<_i$ is the well-ordering on X_i. Then $<$ is a well-defined total order on X. Given $S \subseteq X$ non-empty, choose i with $S \cap X_i \neq \emptyset$. Then $S \cap X_i$ has a minimal element, which must also be a minimal element of S as X_i is an initial segment of X. Also $X \geq X_i$ for all i. □

2.3 Ordinals

We have shown that well-orderings can be compared, but are they totally ordered? This is a question that is not yet very meaningful, since we can have isomorphic well-orderings that are not equal. Now we employ a technique commonly used in studying collection of abstract mathematical objects — we identify well-orderings that are isomorphic as the same and work with equivalence classes of them.[1]

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

Definition (Order-type)**.** If X is a well-ordering corresponding to ordinal α, say X has *order-type* α.

Example. With slight abuse of notation, for each $k \in \mathbb{N}$, write k for the order-type of the (unique) well-ordering of a set of size k, and write ω for the order-type of $\mathbb{N}$. So in $\mathbb{R}$, $\{1, 3, 7\}$ has order-type 3, and $\{1 - 1/n : n = 2, 3, \dots\}$ has order-type ω.

[1] Technically, we are working with proper classes instead of sets, for example, by considering the collection of all singletons. See later.

Notation. For X of order-type α and Y of order-type β, write $\alpha \leq \beta$ if $X \leq Y$. This is well-defined. Similarly $\alpha < \beta$ and so on.

Equipped with these definitions, we now know for all α, β, $\alpha \leq \beta$ or $\beta \leq \alpha$ and if $\alpha \leq \beta, \beta \leq \alpha$ then $\alpha = \beta$, i.e. ordinals are totally ordered. But are they well-ordered?

Theorem 2.9. *Let α be an ordinal. Then the ordinals $< \alpha$ form a well-ordered set with order-type α.*

For example, the ordinals $< \omega$ are $0, 1, 2, ...$

Proof. Let X have order-type α. The well-orderings $< X$ are precisely (up to isomorphisms) the proper initial segments of X, i.e. the I_x for $x \in X$. But these are isomorphic to X itself via $I_x \mapsto x$. □

Notation. We often write $I_\alpha = \{\beta \text{ ordinal} : \beta < \alpha\}$ for this special well-ordered set with order-type α.

Proposition 2.10. *Let S be a non-empty set of ordinals. Then S has a least element.*

Proof. Choose $\alpha \in S$. If α is minimal in S then done. If not, then $S \cap I_\alpha \neq \emptyset$, so we have a minimal element of $S \cap I_\alpha$, which is therefore minimal in S. □

Given the proposition, it is very tempting to conclude that all well-orderings form a well-order. But there is one thing we haven't checked, namely well-orders are defined on a set. Unfortunately,

Theorem 2.11 (Burali-Forti paradox)**.** *The ordinals do not form a set.*

Proof. Suppose not. Let X be the set of all ordinals. Then X is a well-ordering, say of order-type α. So X is isomorphic to I_α, a *proper* initial segment of X. Absurd. □

This is saying that the collection of all well-orderings is too big to be a set, and thus to be a well-ordering. However, this does not prevent us from working locally with a set of well-orderings.

Recall the two ways of constructing well-orderings. Given α, we have $\alpha^+ > \alpha$. Also if $\{\alpha_i : i \in I\}$ is a set of ordinals, then there exists α with $\alpha \geq \alpha_i$ for all i, by applying Proposition 2.8 to the nested family $\{I_{\alpha_i} : i \in I\}$.

In fact, there is a least upper bound for $\{\alpha_i : i \in I\}$ — by applying Proposition 2.10 to the set

$$\{\beta \leq \alpha : \beta \text{ an upper bound of } \alpha_i\}.$$

This is denoted $\sup_{i \in I} \alpha_i$.

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

0	1	2	...				
ω	$\omega+1$	$\omega+2$	...				
$\omega 2$	$\omega 2+1$	$\omega 2+2$	...	$\omega 3$	...	$\omega 4$	...
ω^2	ω^2+1	...	$\omega^2+\omega$	...	$\omega^2+\omega 2$	...	
$\omega^2 2$	...	$\omega^2 3$	...	$\omega^2 4$	...		
ω^3	...	ω^4	...	ω^5	...		
ω^ω	$\omega^\omega+1$	...	$\omega^\omega+\omega$	...	$\omega^\omega 2$	...	$\omega^\omega 3$
$\omega^{\omega+1}$	...	$\omega^{\omega+2}$	...	$\omega^{\omega+3}$	...		
$\omega^{\omega 2}$	...	$\omega^{\omega 3}$	...				
ω^{ω^ω}	...	$\omega^{\omega^{\omega^\omega}}$	...				
ε_0	ε_0+1	...	$\varepsilon_0+\omega$	...	$\varepsilon_0+\omega^\omega$	...	
$\varepsilon_0 2$	...	$\varepsilon_0\omega$	$\varepsilon_0\omega^\omega$	...			
ε_0^2	...	ε_0^3	...	ε_0^4	...		
ε_0^ω	...	$\varepsilon_0^{\omega^\omega}$	...				
$\varepsilon_0^{\varepsilon_0}$	...	$\varepsilon_0^{\varepsilon_0^{\varepsilon_0}}$	...				
ε_1	...						

2.4 Some ordinals

Section 2.4 shows some ordinals in increasing order. In each row from left to right, adjacent values are successors to each other. The pattern in each row is the "obvious" one that the reader should be able to infer. The beginning entry of a row is the supremum of all entries in the previous row.

Some points to notice:

- we write $\omega 2 = \sup\{\omega+1, \omega+2, \dots\}$. The rationale for this unconventional

notation will soon be clear.

- everything in this table so far is countable, as they are built from operations such as union, subset, cartesian product on countable sets.

Is there an uncountable ordinal? In other words, is there an uncountable well-ordered set?

For example, we can well-order $\mathbb{N}$ and $\mathbb{Q}$, but what about $\mathbb{R}$? Unfortunately, no. We are always going to fail if we try to put a well-ordering on $\mathbb{R}$. But

Theorem 2.12. *There is an uncountable ordinal.*

Proof. The idea is to take the supremum of all countable ordinals, but first we have to check that it is a set, meaning that we can build it from existing sets using operations such as intersection, cartesian product, images of functions etc.

Let

$$R = \{A \in \mathcal{P}(\mathbb{N} \times \mathbb{N}) : A \text{ is a well-ordering of a subset of } \mathbb{N}\}.$$

Let S be the image of R under the function "order-type", i.e. S is the set of all order-types of well-orderings of $\mathbb{N}$ (and subsets thereof). It is the set of all countable ordinals.

Let $\omega_1 = \sup S$. Then ω_1 is uncountable: if not then $\omega_1 \in S$ so ω_1 would be the greatest member of S, which contradicts $\omega_1 < \omega_1^+$. Note that by construction ω_1 is the *least uncountable ordinal*. □

ω_1 has some strange properties, for example

1. ω_1 is uncountable, but for any $\alpha < \omega$, we have $\{\beta : \beta < \alpha\}$ countable.[1]
2. If $\alpha_1, \alpha_2, \cdots < \omega_1$ is a sequence, then it is *bounded* in ω_1: $\sup\{\alpha_1, \alpha_2, \dots\}$ is countable so $< \omega_1$.

Theorem 2.13 (Hartogs' lemma)**.** *For any set X, there is an ordinal that does not inject into X.*

Proof. Same proof as above, with $\mathcal{P}(X \times X)$ in place of $\mathcal{P}(\mathbb{N} \times \mathbb{N})$. □

Notation. We often write $\gamma(X)$ for least such ordinal. For example, $\gamma(\omega) = \omega_1$.

2.5 Successors and Limits

Given an ordinal α, does α has a greatest element?

If yes, say β is greatest. Then $\gamma < \beta$ or $\gamma = \beta$ implies $\gamma < \alpha$ and $\gamma < \alpha$ implies $\gamma < \beta$ or $\gamma = \beta$ (as we can't have $\gamma > \beta$). So $\alpha = \beta^+$. Call α a *successor*.

If no, then for every $\beta < \alpha$, then there exists $\gamma < \alpha$ such that $\gamma > \beta$. Thus $\alpha = \sup\{\beta : \beta < \alpha\}$ (note that this is false in general without the absence of greatest element hypothesis, e.g. $\omega + 5$). Call α a *limit*.

Example. 5 and $\omega + 5$ are successors. ω and $\omega + \omega$ are limits. 0 is a limit by definition.

[1] It would perhaps be less surprising if one considers the analogy that, given $\alpha < \omega$, $\{\beta : \beta < \alpha\}$ is finite.

2.6 Ordinal arithmetics

Definition (Ordinal addition (inductive))**.** Define $\alpha + \beta$ recursively by

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

Since the ordinals do not form a set, we cannot do recursion on ordinals. Instead we do it "locally": for each β we define $\alpha+\gamma$ for $\{\gamma : \gamma < \beta\}$ recursively. Then by uniqueness of recursion addition is well-defined.

Example.

$$\begin{aligned}
&\omega + 1 = (\omega + 0)^+ = \omega^+ \\
&\omega + 2 = (\omega + 1)^+ = \omega^{++} \\
&1 + \omega = \sup\{1 + \gamma : \gamma < \omega\} = \sup\{1, 2, 3, \dots\} = \omega
\end{aligned}$$

We can see that addition is not commutative. The is because in the definition of addition recursion is done on the second argument. However ordinal addition remains associative.

Proposition 2.14. *Ordinal addition is asssociative, i.e. for all α, β, γ,*

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

Proof. Since addition is defined by recursion, it is natural to consider an induction proof. Fix α and β and proceed by induction on γ.

- $\gamma = 0$: $\alpha + (\beta + 0) = \alpha + \beta = (\alpha + \beta) + 0$.
- $\gamma = \delta^+$ is a successor:

$$\begin{aligned}
\alpha + (\beta + \delta^+) &= \alpha + (\beta + \delta)^+ \\
&= (\alpha + (\beta + \delta))^+ \\
&= ((\alpha + \beta) + \delta)^+ \\
&= (\alpha + \beta) + \delta^+ \\
&= (\alpha + \beta) + \gamma
\end{aligned}$$

- γ is a limit:

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

On the other hand, we need to evaluate $\alpha + (\beta + \gamma)$. Claim $\beta + \gamma$ is a limit, i.e. $\beta + \gamma = \sup\{\beta + \lambda : \lambda < \gamma\}$: for any $\beta + \lambda$, since γ is a limit, there exists λ' such that $\lambda < \lambda' < \gamma$. Thus $\beta + \lambda < \beta + \lambda'$. Thus $\beta + \lambda$ is not the greatest element. Therefore

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

Now need to show that

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

Note that the two sets are not equal. For example, for $\beta = 3, \gamma = \omega$, $\alpha + 2$ is in LHS but not RHS.

- $\geq$: by set inclusion $\supseteq$.
- $\leq$: for $\lambda < \beta + \gamma$, we have $\lambda < \sup\{\beta + \lambda' : \lambda' < \gamma\}$. Thus $\lambda < \beta + \lambda'$ for some $\lambda' < \gamma$. Thus $\alpha + \lambda < \alpha + (\beta + \lambda)$.

□

Note that in the proof we assumed that $\beta < \gamma \implies \alpha + \beta < \alpha + \gamma$, which can be shown by induction on γ. Note that similar to the noncommutativity of ordinal addition, the does not hold for addition on the right: $1 < 2$ but $1 + \omega = 2 + \omega$.

The above definition is an inductive one, in which we used the recursive definition of ordinals to build addition bottom-up. Since ordinals can also be defined as order types of sets, there is an alternative definition of addition by constructing a set of the desired order type, and then declare it to be the sum of the two ordinals.

Definition (Ordinal addition (synthetic))**.** $\alpha + \beta$ is the order type of $\alpha \sqcup \beta$, the coproduct of the order α and β (i.e. product order on $\{0\} \times \alpha \cup \{1\} \times \beta$).

Example.

$$\omega + 1 = \omega^+$$
$$1 + \omega = \omega$$

With this definition, associativity is trivial by associativity of union.

Proposition 2.15. *The inductive and synthetic definition of addition coincide.*

Proof. Write $+$ and $+'$ for inductive and synthetic definition respectively. We want to show that $\alpha + \beta = \alpha +' \beta$. Induct on β.

- $\beta = 0$: $\alpha + 0 = \alpha = \alpha +' 0$.
- successor ordinal: $\alpha + \beta^+ = (\alpha + \beta)^+ = (\alpha +' \beta)^+ = \alpha +' \beta^+$.
- limit ordinal: $\alpha + \beta = \sup\{\alpha + \lambda : \lambda < \beta\} = \sup\{\alpha +' \lambda : \lambda < \beta\} = \alpha +' \beta$ where the second equality is because taking sup is the same as union.

□

The synthetic definition is usually easier to work with since they provide an encapsulation of information. For example, it was easy to show associativity, and also easy to see noncommutativity. However, the inductive definition is easier if we want to do induction.

Now we define multiplication.

Definition (Ordinal multiplication (inductive))**.** Define $\alpha\beta$ recursively by

- $\alpha 0 = 0$,
- $\alpha(\beta^+) = \alpha\beta + \alpha$,
- $\alpha\lambda = \sup\{\alpha\gamma : \gamma < \lambda\}$ for λ a non-zero limit,

Example.

$$\begin{aligned}
\omega 1 &= \omega 0 + \omega = 0 + \omega = \omega \\
\omega 2 &= \omega 1 + \omega = \omega + \omega \\
\omega\omega &= \sup\{\omega\gamma : \gamma < \omega\} = \sup\{0, \omega, \omega + \omega, \dots\} \\
2\omega &= \sup\{2\gamma : \gamma < \omega\} = \omega
\end{aligned}$$

In particular this shows multiplication is not commutative.

Definition (Ordinal multiplication (synthetic))**.** $\alpha\beta$ is the order-type of $\alpha \times \beta$, with $(x, y) < (z, w)$ if either $y < w$ or $y = w$ and $x < z$.

We can check that the definitions agree and associativity of multiplication etc.

Definition (Ordinal exponentiation)**.** Define α^β recursively by

- $\alpha^0 = 1$,
- $\alpha^{\beta^+} = \alpha^\beta \cdot \alpha$,
- $\alpha^\lambda = \sup\{\alpha^\gamma : \gamma < \lambda\}$ for λ a non-zero limit.

Example.

$$\begin{aligned}
\omega^1 &= \omega^0 \cdot \omega = 1 \cdot \omega = \omega \\
\omega^2 &= \omega^1 \cdot \omega = \omega \cdot \omega \\
2^\omega &= \sup\{2^\gamma : \gamma < \omega\} = \omega
\end{aligned}$$

Note that 2^ω is countable.

Similarly we can define towers and other arithmetic operations inductively. It is left as an exercise.

3 Posets and Zorn's Lemma

3.1 Partial Orders

Definition (Poset)**.** 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

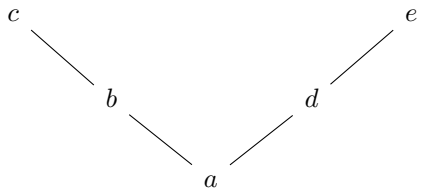
1. reflexive: for all x, $x \leq x$,
2. transitive: for all x, y, z, $x \leq y, y \leq z$ implies $x \leq z$,
3. antisymmetric: for all x, y, $x \leq y, y \leq x$ implies $x = y$.

Notation. Write $x < y$ if $x \leq y, x \neq y$.
In terms of $<$, a poset is

1. irreflexive: for all x, not $x < x$,
2. transitive: for all x, y, z, $x < y, y < z$ implies $x < z$.

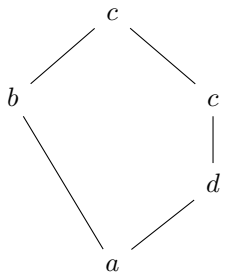
Example.

1. Any total order.
2. $\mathbb{N}^+$ with "divides".
3. For any set S, $\mathcal{P}(S)$ with $x \leq y$ if $x \subseteq y$.
4. Any $X \subseteq \mathcal{P}(S)$ with same relation as above. This specialises to, for example, all subspaces of a given vector space.
5. We can draw a *Hasse diagram* for a poset X: it consists of a drawing of elemnents of X, with an upward line from x to y if y covers x, meaning $y > x$ and no z such that $y > z > x$. For example



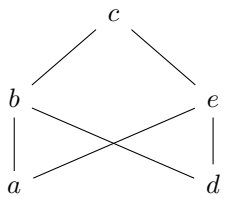
   Hasse diagrams can be useful to visualise a poset (e.g. $\mathbb{N}$), or useless (e.g. $\mathbb{Q}$).

6. In



   b and d are unrelated so there is no sense of "height" or "rank".

7.



8. A set in which no two elements are related is a poset.

Definition (Chain)**.** In a poset X, a *chain* is a set $S \subseteq X$ that is totally ordered: for all $x, y \in S$, $x \leq y$ or $y \leq x$.

Example.

1. Any subset of 1 above.
2. In 2, $\{1, 2, 4, 8, 16\}$.
3. In 5, $\{a, b, c\}$ or $\{a, c\}$ but not $\{b, d\}$.
4. In 8, only singletons and $\emptyset$.

Note. Chains can be uncountable, e.g. $(\mathbb{R}, \leq)$.

Definition (Anti-chain)**.** Give a poset X, $S \subseteq X$ is an *anti-chain* if no two elements are related: for all $x, y \in S$, $x \neq y$ implies that not $x < y$.

Example.

1. In 2: the set of primes.
2. In 5, $\{c, e\}$ or $\{c, d\}$, or $\{b\}$.
3. In 8, every subset.

Definition (Upper bound)**.** Give a poset X, $S \subseteq X$, an *upper bound* for S is any $x \in X$ such that $x \geq y$ for all $y \in S$.

Definition (Least upper bound)**.** Say X is a *least upper bound* or *supremum* for S if x is an upper bound for S and $x \leq y$ for every upper bound y for S. Write $x = \sup S$ or $x = \bigvee S$, the "join" of S.

Example.

1. In $\mathbb{R}$, $\{x : x^2 < 2\}$ has 7 as an upper bound and $\sqrt{2}$ as a supremum. This shows that $\sup S$ need not be in S.
2. In $\mathbb{R}$, the set $\mathbb{Z}$ has no upper bound.
3. In $\mathbb{Q}$, $\{x : x^2 < 2\}$ has 7 as an upper bound but no supremum.

4. In 5, $\{a, b\}$ has upper bounds b and c and supremum b.
5. In 5, $\{b, d\}$ has no upper bound.
6. In 7, $\{b, d\}$ has upper bounds c, b, e but no supremum.

Definition (Completeness)**.** A poset is *complete* if every subset has a supremum.

Example.

1. $(\mathbb{R}, \leq)$ is not complete as $\mathbb{Z}$ has no supremum. Thus completeness in posets is different from that in analysis.
2. $[0, 1]$ is complete.
3. $(0, 1)$ is not complete as $(0, 1)$ itself has no supremum.
4. $\mathcal{P}(S)$ is always complete — $\{A_i\}_{i \in I}$ has supremum $\bigcup_{i \in I} A_i$.

Note that in particular in any poset there is a greatest element, namely $\sup X$, and a least element, namely $\sup \emptyset$.

Definition (Order-preserving map)**.** A function $f : X \to X$ where X is a poset is *order-preserving* if $f(x) \leq f(y)$ for all $x \leq y$.

Example.

1. On $\mathbb{N}$, $f(x) = x + 1$.
2. On $[0, 1]$, $f(x) = \frac{1+x}{2}$, "halve the distance to 1".
3. On $\mathcal{P}(S)$, $f(A) = A \cup \{i\}$ for some fixed $i \in S$.

From above not every order-preserving function has a fixed point. But just as in Contraction Mapping Theorem, we can add condition to the space to make this happen. Unsurprisingly, this condition is completeness:

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

Proof. Let $E = \{x \in X : x \leq f(x)\}$ and $s = \sup E$. To show $f(s) = s$, we show both $s \leq f(s)$ and $s \geq f(x)$.

- $s \leq f(s)$: suffices to show $f(s)$ is an upper bound for E (as s is the least upper bound). But
$$x \in E \implies x \leq s \implies f(x) \leq f(s) \implies x \leq f(x) \leq f(s).$$
- $s \geq f(s)$: suffices to show $f(s) \in E$ (as s is an upper bound). We know $s \leq f(s)$, so $f(s) \leq f(f(s))$ since f is order preserving.

□

Note. In any complete poset X, we have a greatest element, namely $\sup X$. We also have a least element, namely $\sup \emptyset$.

A typical application of Knaster-Tarski is

Theorem 3.2 (Schröder-Berstein theorem)**.** *Let A, B be sets such that there is an injection $f : A \to B$ and injection $g : B \to A$, then there exists a bijection from A to B.*

Proof. Seek partitions $A = P \cup Q, B = R \cup S$ such that $f(P) = R, g(S) = Q$. Then done by seting $h = \begin{cases} f & \text{on } P \\ g^{-1} & \text{on } Q \end{cases}$.

Note that we are done once we fix P: $R = f(P), S = B \setminus f(P), Q = g(B \setminus f(P))$. i.e. we seek $P \subseteq A$ such that

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

Define

$$\begin{aligned} \theta : \mathcal{P}(A) &\to \mathcal{P}(A) \\ P &\mapsto A \setminus (g(B \setminus f(P))) \end{aligned}$$

Then since $\mathcal{P}(A)$ is complete, θ is order-preserving (since it takes complement twice), there exists a fixed point by Knaster-Tarski. □

3.2 Zorn's Lemma

Definition. An element x in a poset X is *maximal* if there exists no $y \in X$ such that $y > x$.

Example. In example 5 before, c, e are both maximal.

Posets need not have a maximal element, for example $\mathbb{N}, \mathbb{Q}, \mathbb{R}$ with the ususal order. We notice something in common: in each of thoses cases, there exists a chain without an upper bound.

Theorem 3.3 (Zorn's lemma)**.** *Let X be a (non-empty) poset in which every chain has an upper bound, then X has a maximal element.*

Proof. Suppose not, then for each $x \in X$ there exists $x' \in X$ with $x' > x$. Also for any chain C we have an upper bound $u(C)$. Pick $x \in X$. Define $x_\alpha \in X$ for each $\alpha < \gamma(X)$ recursively by

$$\begin{aligned} x_0 &= x \\ x_{\alpha+1} &= x'_\alpha \\ x_\lambda &= u(\{x_\alpha : \alpha < \lambda\}) \text{ for } \lambda \text{ a nonzero limit} \end{aligned}$$

Then $\alpha \mapsto x_\alpha$ is an injection $\gamma(X) \to X$. Absurd. □

A typical application of Zorn's lemma: does every vector space V have a basis? Recall that a basis is a linearly independent (no non-trival finite relation) spanning (every element is a finite linear combination thereof) set.

Example.

1. Let $V = \mathbb{R}[X]$ be the space of all real polynomials. Then $\{X^i\}_{i \in \mathbb{N}}$ is a basis.

2. Let V be the set of all real sequences with pointwise addition. We might guess

$$\begin{aligned}\ell_1 &= (1, 0, 0, \dots)\\ \ell_2 &= (0, 1, 0, \dots)\\ &\vdots\end{aligned}$$

   is a basis. Unfortunately they are linearly independent but not spanning, e.g. $(1, 1, 1, \dots)$ is not in the span. It is actually easy to check that there is no countable basis. It also turns out that there is no *explicit* basis.

3. $\mathbb{R}$ as a $\mathbb{Q}$-vector space has a basis called a *Hamel basis*.

Theorem 3.4. *Every vector space has a basis.*

Proof. We seek a maximal linearly independent set. Let V be a vector space and

$$X = \{A \subseteq V : A \text{ linearly independent}\}$$

ordered by $\subseteq$. If we can find a maximal element of X, then done: if M is not spanning then choose $x \notin \langle M \rangle$ and then $M \cup \{x\}$ is linearly independent. Absurd.

We have $X \neq \emptyset$ as $\emptyset \in X$. Given a chain $\{A_i : i \in I\}$ in X, put $A = \bigcup_{i \in I} A_i$. Then $A \supseteq A_i$ for all i, so just need to check $A \in X$, i.e. A is linearly independent. Suppose not, so

$$\sum_{i=1}^{n} \lambda_i x_i = 0$$

for some $x_i \in A$, λ_i not all 0. We have $x_1 \in A_{i_1}, \dots, x_n \in A_{i_n}$ for some $i_1, \dots, i_n \in I$. But $A_{i_1}, \dots, A_{i_n} \subseteq A_{i_k}$ for some k (as $A_{i_1}, \dots, A_{i_n}$ are nested), contradicting A_{i_k} linearly independent. □

Note. The only "actual math" (i.e. linear algebra) in the proof was the "then done" part.

Another application of Zorn's lemma: completeness theorem when the primitive language is uncountable.

Theorem 3.5 (Completeness theorem)**.** *Let $S \subseteq L(P)$ where P is any set. Then S is consistent implies that S has a model.*

Proof. We seek a maximal consistent $\overline{S} \supseteq S$. Then done: for each $t \in L(P)$ have $\overline{S} \cup \{t\}$ or $\overline{S} \cup \{\neg t\}$ consistent, whence $t \in \overline{S}$ or $\neg t \in \overline{S}$. By maximality of $\overline{S}$, and now define

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

Let

$$X = \{T \subseteq L(P) : T \text{ consistent}, T \supseteq S\}$$

ordered by $\subseteq$. Then $X \neq \emptyset$ as $S \in X$. Given a non-empty chain $\{T_i : i \in I\}$ in X, take $T = \bigcup_{i\in I} T_i$, then $T \supseteq T_i$ for all i so just need $T \in X$. We have $S \subseteq T$ (as $I \neq \emptyset$) and T is consistent: suppose $T \vdash \bot$. Then $\{t_1, \dots, t_n\} \vdash \bot$ for some $t_1, \dots, t_n \in T$ (as proofs are finite). Since $t_1 \in T_{i_1}, \dots, t_n \in T_{i_n}$ for some $i_1, \dots, i_n \in I$, but $T_{i_1}, \dots, T_{i_n} \subseteq T_{i_k}$ for some k (as they are nested), $T_{i_k} \vdash \bot$. Absurd. □

One final application:

Theorem 3.6 (Well-ordering principle)**.** *Every set can be well-ordered.*

Remark. This is very surprising for, for example, $\mathbb{R}$, until you remember Hartogs' lemma.

Proof. Let S be the set. Let

$$X = \{(A, R) : A \subseteq S, R \text{ a well-ordering of } A\}$$

ordered by

$$(A, R) \leq (A', R') \text{ if } (A', R') \text{ extends } (A, R).$$

$X \neq \emptyset$ as $(\emptyset, \emptyset) \in X$. Given a chain $\{(A_i, R_i) : i \in I\}$, we have

$$(\bigcup_{i\in I} A_i, \bigcup_{i\in I} R_i) \in X$$

extending each (A_i, R_i) (from chapter 2). Thus by Zorn's lemma, X has a maximal element (A, R). Must have $A = S$: if not, choose $x \in S \setminus A$ and "take successor": well-order $A \cup \{x\}$ by setting $x > a$ for all $a \in A$, contradicting the maximality of (A, R). □

Remark. Proof of Zorn's lemma was easy because we knew ordinals, recursion and Hartogs' lemma.

3.3 Zorn's Lemma and Axiom of Choice

In the proof of Zorn's lemma, we chose for each $x \in X$ an $x' > x$ — i.e. we made infinitely many arbitrary choices (note that this has nothing to do with Hartogs' lemma. We made infinitely many choices even by the time we get to x_ω). We did the same in IA Numbers and Sets, in proving that the countable union of countable sets is countable: we chose for each set in the family an ordering whereof.

In terms of "rules for building sets", this is appealing to *axiom of choice*, which says that we may choose an element of each set in a family of non-empty sets. More precisely,

Axiom 3.7 (Axiom of choice)**.** *If $\{A_i : i \in I\}$ is a family of non-empty sets then it has a* choice function*, i.e. a function $f : I \to \bigcup_{i\in I} A_i$ such that $f(i) \in A_i$ for all i.*

This is of different character to the other set-building rules, such as union, power set etc in that the object whose existence is asserted is not uniquely specified by its properties, unlike, for example, union of sets. Thus often one points out when one has used Axiom of choice.

Remark. Axiom of choice is trivial if $|I| = 1$ ($A \neq \emptyset$ means by definition that there exists $x \in A$). By induction, it is true for I finite. However, it turns out that, for general I, axiom of choice *cannot* be deduced from the other set-theoretic rules.

In Zorn's lemma, we used axiom of choice. Is there a proof of Zorn's lemma without axiom of choice? No, because we can deduce axiom of choice from Zorn's lemma.

Proof of axiom of choice from Zorn's lemma. Given a family $\{A_i : i \in I\}$ of non-empty sets, a *partial choice function* is an $f : J \to \bigcup_{i\in I} A_i$ where $J \subseteq I$ such that $f(j) \in A_j$ for all $j \in J$. Let

$$(J, f) \leq (J', f') \text{ if } J \subseteq J' \text{ and } f'|_J = f.$$

This poset is non-empty as $(\emptyset, \emptyset)$ is an element. Given a chain $\{(J_q, f_q)\}_{q\in Q}$, we have $(\bigcup_{q\in Q} J_q, \bigcup_{q\in Q} f_q)$ as an upper bound. Thus by Zorn's lemma there exists a maximal element (J, f). We must have $J = I$, as if not we choose $i \in I \setminus J, x \in A_i$, and put $J' = J \cup \{i\}, f' = f \cup \{(i, x)\}$. Absurd. □

In conclusion, Zorn's lemma $\iff$ axiom of choice (in the presence of the set-building rules).

Actually, there is a three-way equivalence: we have shown Zorn's lemma implies well-ordering principle, and well-ordering principle implies axiom of choice trivially ($\bigcup_{i\in I} A_i$ is well-ordered and let $f(i)$ be least element of A_i). Therefore

$$\text{Zorn's lemma} \iff \text{axiom of choice} \iff \text{well-ordering principle.}$$

Exercise. Show that axiom of choice implies well-ordering principle directly.

Note. Zorn's lemma is hard to prove from first principles because we need theory of ordinals, recursions and Hartogs' lemma, not because of its equivalence with axiom of choice.

3.4 Bourbaki-Witt Theorem*

On one hand we have Zorn's lemma, which is a local (conditions on chains) to global (maximal element) principle, and on the other hand we have Knaster-Tarski fixed point theorem, a global fixed point theorem based on assumptions of the ambient space (completeness). The Bourbaki-Witt theorem is a "midpoint" between the two.

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

Example.

1. Any complete poset.
2. Any finite poset.
3. Given a vector space V, $\{A \subseteq V : A \text{ is linearly independent}\}$.

Definition (Inflationary)**.** A function $f : X \to X$ is *inflationary* if $f(x) \geq x$ for all x.

Theorem 3.8 (Bourbaki-Witt)**.** *Suppose X is chain-complete and $f : X \to X$ inflationary. Then f has a fixed point.*

Bourbaki-Witt follows immediately from Zorn's lemma: take maximal x and since $f(x) \geq x$, we must have equality.

However, intriguingly, we can prove Bourbaki-Witt *without* axiom of choice: injecting $\gamma(X)$ into X by explicitly set $x_{\alpha+1} = f(x_\alpha)$ and $x_\alpha = \sup\{x_\beta : \beta < \alpha\}$ for a non-zero limit α and derive a contradiction. We circumvent the issue of choice by exhibit an explicit upper bound.

Note. In chapter 2, we never used axiom of choice except in remark that well-ordering is equivalent to the absence of decreasing sequence, and that ω_1 does not have a countable supremum.

In fact, it is easy to deduce Zorn's lemma from Bourbaki-Witt (with axiom of choice) so we can view it as the "choice-free version of Zorn's lemma".

4 Predicate Logic

We studied propositional logic in chapter 1 but it is not powerful enough to express objects outside of primitive propositions. In this chapter we will introduce predicate logic, which is more intricate but more powerful. Before that we will have an overview of the theory we will develop.

A mathematical *structure* is a set with functions and relations defined on it. A function has an *arity* associated to it, which is the number of arguments it takes. For example, recall that a group is a set A equipped with functions $m : A^2 \to A$ (arity 2), and $i : A \to A$ (arity 1), and a constant $e \in A$, which could be seen as a function of arity 0, such that

$$(\forall x, y, z \in A)(m(x, m(y, z)) = m(m(x, y), z))$$
$$(\forall x \in A)(m(x, e) = x \wedge m(e, x) = x)$$
$$(\forall x \in A)(m(x, i(x)) = e \wedge m(i(x), x) = e)$$

As another example, a poset is a structure with relation: it is a set A equipped with a predicate (i.e. relation) $(\leq) \subseteq A^2$ such that

$$(\forall x \in A)(x \leq x)$$
$$(\forall x, y, z \in A)((x \leq y \wedge y \leq z) \implies (x \leq z))$$
$$(\forall x, y \in A)((x \leq y \wedge y \leq x) \implies (x = y))$$

Propostional logic	Predicate logic	Example in groups
language	language	group axioms
valuation	structure	m, i, e
model of S	same	structure in which each $s \in S$ holds
$S \vDash t$	same	group axioms $\vDash m(e, e) = e$
$S \vdash t$	same*	

Table 1: Comparison of concepts in propositional and predicate logic

The axioms and deduction rules in predicate logic is going to be more complicated. For example, it includes an axiom that says "if $p(x)$ holds for all x then we can substitute t for x".

4.1 Definitions

Let Ω, Π be disjoint sets and set $\alpha : \Omega \cup \Pi \to \mathbb{N}$.

Definition (Language)**.** Let Ω, Π be disjoint and $\alpha : \Omega \cup \Pi \to \mathbb{N}$. A *language* $L(\Omega, \Pi, \alpha)$ is the set of *formulæ*, defined by

- *variables*: $x_1, x_2, ...$ We sometimes use $x, y, ...,$
- *terms*: defined inductively by
  1. each variable is an a term,
  2. if $f \in \Omega, \alpha(f) = n$ and $t_1, ..., t_n$ are terms then $ft_1, ..., t_n$ is a term. We usually write $f(t_1, ..., t_n)$.

Example. In the language of groups, we have $\Omega = (m, i, e)$, with arities $2, 1, 0$ respectively. $\Pi = \emptyset$.

Some terms: $x_1, m(x_1, x_1), e, m(e, e), m(x_1, i(x_1))$.

- *atomic formulæ*: consists of
  1. $\bot$,
  2. $(s = t)$ for any terms s, t,
  3. $\phi(t_1, \ldots, t_n)$ for any $\phi \in \Pi, \alpha(\phi) = n$ and any terms $t_1, \ldots, t_n$.
- *formulæ*: defined inductively by
  1. each atomic formula is a formula,
  2. if p, q are formulæ then so is $(p \implies q)$,
  3. if p is a formula, x is a variable, then $(\forall x)p$ is a formula,

     Example.

     (a) group:

     $$(\forall x)(m(x, x) = e),$$
     $$(\forall x)(m(x, x) = e) \implies (\exists y)(m(y, y) = x).$$

     (b) poset: $(\forall x)(x \leq x)$.

Note.

1. A formula is a string of symbols.
2. Just as in chapter 1 we defined symbols such as $\neg$, $\wedge$ and $\vee$ in terms of $\bot$, we write $(\exists x)p$ for $\neg(\forall x)(\neg p)$.

Definition (Closed)**.** A term is *closed* if it contains no variables.

Example. $e, m(e, e), m(e, m(e, e))$ are closed terms, but $m(x, i(x))$ isn't.

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

Example. In $m(x, x) = e \implies (\exists y)(m(y, y) = e)$, x is free and y is bound.

Note. We say an "occurrence" and not a "variable" is bound or free because officially, it is possible to have a variable that is both free and bound in a formula. For example,

$$(m(x, x) = e) \implies (\forall x)(\forall y)(m(x, y) = m(y, x)).$$

But promise to *never* use it!

Definition (Sentence)**.** A *sentence* is a formula without free variables.

Example. $(\forall x)(m(x, e) = x)$.

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

Do not worry too much about the word "free" in the above definition, it is there to prevent us from doing stupid things in stupid formula such as the bad example above where a variable is both free and bound. Just follow your common sense!

Example. If p is the statement $(\exists y)(m(y, y) = x)$ then $p[t/x]$ is

$$(\exists y)(m(y, y) = t).$$

4.2 Semantic Entailment

Definition (Structure)**.** Let $L = L(\Omega, \Pi, \alpha)$. An *$L$-structure* is a non-empty set A equipped with

1. for each $f \in \Omega$ with $\alpha(f) = n$, a function $f_A : A^n \to A$,
2. for each $\phi \in \Pi$ with $\alpha(\phi) = n$, a relation $\phi_A \subseteq A^n$.

See note on page 33 for why "non-empty".

Example.

1. Suppose L is the language of groups, then an L-structure is a set A with functions $m_A : A^2 \to A, i_A : A \to A, e_A \in A$. Note that A need not be a group.
2. Suppose L is the language of posets, then an L-structure is a set A with a relation $(\leq_A) \subseteq A^2$.

We want to define the *interpretation* $p_A \in \{0, 1\}$ of a sentence p in an L-structure A as "p holds in A". For example, $(\forall x)(m(x, x) = e)$ should be "true in A" if $\forall a \in A, m_A(a, a) = e_A$. Informally this can be done by inserting "$\in A$" and subscripting A, and saying it aloud. This recipe captures the essence of the definition of interpretation and is perfectly valid, except that it precludes, for example, the entire French community from studying predicate logic. Thus we rephrase it in the language common to all mathematicians:

Definition (Interpretation)**.** For an L-structure A, define *interpretation* $p_A \in \{0, 1\}$ of a sentence p recursively by

1. closed term: define t_A recursively by $(ft_1, \dots, t_n)_A = f_A((t_1)_A, \dots, (t_n)_A)$ for any $f \in \Omega, \alpha(f) = n$ and closed terms $t_1, \dots, t_n$.

   Example. $m(e, i(e))_A = m_A(e_A, i_A(e_A))$. Note that e_A is already defined.

2. atomic formulæ: define $p_A \in \{0, 1\}$ for p atomic by

(a) $\bot_A = 0$,

(b) $(s = t)_A = \begin{cases} 1 & \text{if } s_A = t_A \\ 0 & \text{otherwise} \end{cases}$ for closed terms s, t.

(c) $(\phi t_1, \dots, t_n)_A = \begin{cases} 1 & \text{if } ((t_1)_A, \dots, (t_n)_A) \in \phi_A \\ 0 & \text{otherwise} \end{cases}$ for each $\phi \in \Pi, \alpha(\phi) = n$ and closed terms $t_1, \dots, t_n$.

3. sentence: p_A defined inductively by

(a) $(p \implies q)_A = \begin{cases} 0 & \text{if } p_A = 1, q_A = 0 \\ 1 & \text{otherwise} \end{cases}$

(b) $((\forall x)p)_A = \begin{cases} 1 & \text{if } p[\overline{a}/x]_{\overline{A}} = 1 \text{ for all } a \in A \\ 0 & \text{otherwise} \end{cases}$ where, for each $a \in A$, add constant symbol $\overline{a}$ to L obtaining L', and made an L'-structure $\overline{A}$ by setting $\overline{a}_{\overline{A}} = a$.

If p has free vairables, we can define $p_A \subseteq A^{\#\text{free variables of } p}$. For example, if p is the formula $(\exists y)(m(y, y) = x)$, then

$$p_A = \{a \in A : \exists b \in A \text{ with } m_A(b, b) = a\}.$$

Definition (Model)**.** If $p_A = 1$, say p is *true* is A, or p *holds* in A, or A is a *model* of p.

Definition (Semantic entailment)**.** For T a *theory* (set of sentences), say T *semantically entails* p, written $T \vDash p$, if every model of T is a model of p.

Definition. p is a *tautology* if $\emptyset \vDash p$ (or written $\vDash p$), i.e. p holds in every L-structure.

Example. $\vDash (\forall x)(x = x)$.

Example.

1. Theory of groups: $\Omega = \{m, i, e\}, \Pi = \emptyset$. Let T be the usual group axioms, i.e.

$$\{(\forall x, y \in A)(m(x, m(y, z)) = m(m(x, y), z))$$
$$(\forall x \in A)(m(x, e) = x \wedge m(e, x) = x)$$
$$(\forall x \in A)(m(x, i(x)) = e \wedge m(i(x), x) = e)\}$$

Then an L-structure is a model of T if and only if T is a group. Note that there are two implications here (compare with, for example, a group is a model of the sentence representing associativity). We say T *axiomatises the class of groups* or *axiomatises the theory of groups*. Sometimes call the elements of T the *axioms* of T.

2. Theory of fields: $\Omega = \{+, \times, -, 0, 1\}, \Pi = \emptyset$. Note that multiplicative inverse is not among them. T is

$$\begin{aligned}&\{\text{abelian group under } (+, -, 0)\\ &\quad \times \text{ is commutative, associative, and distributive over } +\\ &(\forall x)(1x = x)\\ &\neg(1 = 0)\\ &(\forall x)(\neg(x = 0)) \implies (\exists y)(xy = 1)\}\end{aligned}$$

   Then T axiomatises the class of fields.

   T entails "inverses are unique", i.e.

$$T \vDash (\forall x)(\neg(x = 0)) \implies ((\forall y)(\forall z)((yx = 1) \wedge (zx = 1) \implies (y = z))).$$

   T does not entail "characteristic equals 2", i.e. $T \nvDash 1 + 1 = 0$.

3. Theory of posets: $\Omega = \emptyset, \Pi = \{\leq\}$. T is

$$\begin{aligned}&\{(\forall x)(x \leq x)\\ &\quad (\forall x)(\forall y)(\forall z)((x \leq y \wedge y \leq z) \implies x \leq z)\\ &\quad (\forall x)(\forall y)((x \leq y) \wedge (y \leq x) \implies x = y)\}\end{aligned}$$

4. Theory of graphs: $\Omega = \emptyset, \Pi = \{a\}$ where a means "is adjacent to". T is

$$\begin{aligned}&\{(\forall x)(\neg a(x, x))\\ &\quad (\forall x)(\forall y)(a(x, y) \implies a(y, x))\}\end{aligned}$$

4.3 Syntactic Implication

As before, we need some logical axioms. There are 7 of them: 3 from propositional logic, 2 for $=$, 2 for $\forall$.

1. $p \implies (q \implies p)$ for any formulæ p, q.
2. $(p \implies (q \implies r)) \implies ((p \implies q) \implies (p \implies r))$ for any formulæ p, q, r.
3. $(\neg\neg p) \implies p$ for any formula p.
4. $(\forall x)(x = x)$ for any variable x.
5. $(\forall x)(\forall y)((x = y) \implies (p \implies p[y/x]))$ for any variables x, y and formula p in which y does not occur bound.
6. $(\forall x)(p) \implies p[t/x]$ for any variable x, term t and formula p with no variable in t occurring bound in p.
7. $((\forall x)(p \implies q)) \implies (p \implies (\forall x)q)$ for any variable x and formulæ p, q with x not occurring free in p.

As deduction rules we have

1. modus ponens: from $p, p \implies q$ can deduce q.
2. generalisation: from p can deduce $(\forall x)p$, if x does not occur free in any premise used to prove p.

Definition (Proof)**.** For $S \subseteq L, p \in L$, a *proof* of p from S is a finite sentence of formulæ, ending with p, such that each line is one of

- a logical axiom,
- a sentence of S,
- following from earlier lines by modus ponens or generalisation.

Write $S \vdash p$ if there exists a proof of p from S.

Note.

1. Each logical axiom is a tautology.
2. If we allow the empty structure A (for a language with no constants), then $(\forall x)\bot$ holds in A, and $\bot$ deos not hold in A, so

$$((\forall x)\bot) \implies \bot$$

   does not hold in A, which is an instance of axiom 6. To resolve the issue, we can either obfuscate our axioms by adding more technical restrictions (of the form, for example, "with x occurring bound"), or simply ban empty structure.

Example. $\{x = y, x = z\} \vdash y = z$. The idea is to use axiom 5, with p being $x = x$. But there are two quantifiers so we use axiom 6 to kill them.

1. $(\forall x)(\forall y)(x = y \implies (x = z \implies y = z))$, A5
2. $(\forall x)(\forall y)(x = y \implies (x = z \implies y = z)) \implies (\forall y)(x = y \implies (x = z \implies y = z))$, A6
3. $(\forall y)(x = y \implies (x = z \implies y = z))$, MP
4. $(\forall y)(x = y \implies (x = z \implies y = z)) \implies (x = y \implies (x = z \implies y = z))$, A6
5. $x = y \implies (x = z \implies y = z)$, MP
6. $x = y$, hypothesis
7. $x = z \implies y = z$, MP
8. $x = z$, hypothesis
9. $y = z$, MP

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

Proof.

- $\implies$: write down p and apply MP to obtain $S \cup \{p\} \vdash q$.

- $\Longleftarrow$: as before, show $S \vdash p \implies t_i$ for each t_i in the proof of $S \cup \{p\} \vdash q$. The only new case is generalisation. So in proof of q from $S \cup \{p\}$ we have line

  1. r
  2. $(\forall x)r$

  and have a proof of $(p \implies r)$ from S, and we want $S \vdash (p \implies (\forall x)r)$. But in proof of r from $S \cup \{p\}$, no premise had x free. Thus in proof of $(p \implies r)$ from S, no proof had x free either. Hence $S \vdash (\forall x)(p \implies r)$ by generalisation. Now

  – if x does not occur free in p, have $S \vdash (p \implies (\forall x)r)$ by A7 and MP,
  – if x does occur free in p, proof of r from $S \cup \{p\}$ cannot have used p. So in fact $S \vdash (\forall x)r$, whence $S \vdash (p \implies (\forall x)r)$ by A1 and MP.

□

4.4 Gödel Completeness Theorem*

Proposition 4.2 (Soundness)**.** *Let $S \subseteq L, p \in L$. Then if $S \vdash p$ then $S \vDash p$.*

Proof. Have proof of p from S and a model A of S, want $p_A = 1$. This is an easy induction down the lines of the proof. □

For adequacy, want if $S \vDash p$ then $S \vdash p$, i.e. if $S \cup \{\neg p\} \vDash \bot$ then $S \cup (\neg p) \vdash \bot$.

Theorem 4.3 (Model existence lemma)**.** *Let $S \subseteq L$ be a set of sentences. Then if S is consistent then it has a model.*

Note that some people call it the completeness theorem since it contains the majority of the work.

Unlike most other proofs we have met in tripos, this proof has not one, not two, but five key ideas:

1. To build a model out of a language, we first need a candidate structure. If you think about it carefully, we have no choice but to let A be a set of closed terms of L, with "obvious" operations like $(1+1) +_A (1+1) = (1+1)+(1+1)$ in the example of fields.

2. Say S is the theory of fields, taking our above definition, $(1+1)+1 \neq 1+(1+1)$, but $S \vdash ((1+1)+1 = 1+(1+1))$. The solution is to quotient by
$$s \sim t \text{ if } S \vdash (s = t).$$

3. Suppose S is the theory of fields of characteristic 2 or 3, i.e. usual field axioms and $(1+1=0) \vee (1+1+1=0)$. Then $S \nvdash (1+1=0)$, so $[1+1] \neq [0]$. Also $S \nvdash (1+1+1=0)$, so $[1+1+1] \neq [0]$. Thus our structure does not satisfy $(1+1=0) \vee (1+1+1=0)$. Actually this is similar to the failure in attempting to define a valuation of a set S when we are proving model existence lemma for propositional logic: there are propositions implied by S but not in S. Thus we need to extend S to be maximal consistent.

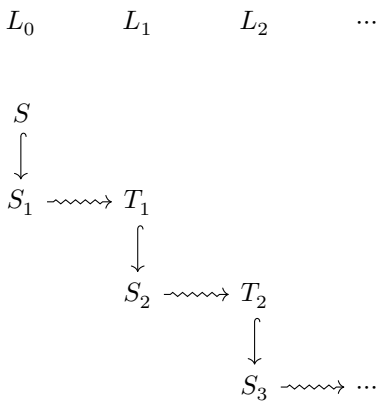
4. Suppose S is the theory of fields with a square root of 2, i.e. usual field axioms and $(\exists x)(xx = 1 + 1)$. In our constrution, maybe no closed term has $[tt] = [1 + 1]$. Thus S lacks "witnesses". Solution? For each $(\exists x)p$ in S add new constant c to language and add $p[c/x]$ to S.

5. Now our set may be no longer consistent, so loop back to step 3. But are we certain that the process will terminate?

Proof. Suppose we have S consistent contained in $L_0 = L(\Omega, \Pi)$. Extend to maximal consistent S_1 by Zorn's lemma. So for each sentence $p \in L$, we have either $p \in S_1$ or $(\neg p) \in S_1$. Thus S_1 is complete (for every p, either $S_1 \vdash p$ or $S_1 \vdash (\neg p)$).

Now add witnesses: for each $(\exists x)p \in S$, add new constant c and axiom $p[c/x]$. We obtain T_1, in language $L_1 = L(\Omega \cup C_1, \Pi)$, where C_1 is the set of all the c's, that "has witnesses" for S_1 (if $(\exists x)p \in S$, then some closed term t has $p[t/x] \in T_1$). Easy to check T_1 is consistent.

Now extend T_1 to maximally consistent S_2 in L_1. Add witnesses, obtaining T_2 in language $L_2 = L(\Omega \cup C_1 \cup C_2, \Pi)$. Continue inductively.



Put $\overline{S} = S_1 \cup S_2 \cup \cdots$ in language $\overline{L} = L(\overline{\Omega}, \Pi)$ where $\overline{\Omega} = \Omega \cup C_1 \cup \cdots$. Check

- $\overline{S}$ is consistent: if $\overline{S} \vdash \bot$ then some $S_n \vdash \bot$ as proofs are finite. Absurd.
- $\overline{S}$ is complete: given sentence $p \in \overline{L}$, have $p \in L_n$ for some n as p contains only finitely many constants. So $S_{n+1} \vdash p$ or $S_{n+1} \vdash (\neg p)$.
- $\overline{S}$ has witnesses (for itself): given $(\exists x)p \in \overline{S}$, have $(\exists x)p \in S_n$ for some n. So $p[t/x] \in T_n$ for some closed term t whence $p[t/x] \in \overline{S}$.

On set of closed terms of $\overline{L}$, define $s \sim t$ if $\overline{S} \vdash (s = t)$, clearly an equivalence relation. Let the set of equivalence classes be A. Make A into an $\overline{L}$-structure by setting

- $f_A([t_1], \ldots, [t_n]) = [ft_1, \ldots, t_n]$ for each $f \in \overline{\Omega}, \alpha(f) = n$, closed terms $t_1, \ldots, t_n$,
- $\phi_A = \{([t_1], \ldots, [t_n]) : \overline{S} \vdash \phi(t_1, \ldots, t_n)\}$ for each $\phi \in \Pi, \alpha(\phi) = n$, closed terms $t_1, \ldots, t_n$.

Claim that $p_A = 1$ if and only if $\overline{S} \vdash p$ for each sentence $p \in \overline{L}$ (then done since if A is a model of $\overline{S}$ then it a model of S).

Proof. An easy induction:

1. atomic sentences:
   (a) $\bot$: $\bot_A = 0$ and $\overline{S} \nvdash \bot$.
   (b) $s = t$: $\overline{S} \vdash (s = t)$ if and only if $[s] = [t]$ by definition of $\sim$, if and only if $s_A = t_A$ by definition of operation on A, if and only if $(s = t)_A = 1$.
   (c) $\phi(t_1, \ldots, t_n)$: exactly the same.
2. induction step:
   (a) $p \implies q$: $\overline{S} \vdash (p \implies q)$ if and only if $\overline{S} \vdash (\neg p)$ or $\overline{S} \vdash q$ (only if: if not then $\overline{S} \vdash p, \overline{S} \vdash (\neg q)$ (as $\overline{S}$ is complete), whence $\overline{S} \vdash \neg(p \implies q)$), if and only if $p_A = 0$ or $q_A = 1$ by induction, if and only if $(p \implies q)_A = 1$.
   (b) $(\exists x)p$: $\overline{S} \vdash (\exists x)p$ if and only if $\overline{S} \vdash p[t/x]$ for some closed term t (only if: $\overline{S}$ has witnesses), if and only if $p[t/x]_A = 1$ for some closed term t by induction, if and only if $((\exists x)p)_A = 1$ (if: A is the set of equivalence classes of closed terms).

□

□

By remark before, have

Corollary 4.4 (Adequacy)**.** *Let $S \subseteq L, p \in L$. Then if $S \vDash p$ then $S \vdash p$.*

Theorem 4.5 (Gödel completeness theorem for first-order logic)**.** *Let S be a set of sentences and p a sentence in language L, then $S \vDash p$ if and only if $S \vdash p$.*

Proof.

- $\Longleftarrow$: soundness.
- $\Longrightarrow$: adequacy.

□

Remark.

1. If L is countable (i.e. Ω and Π are countable) then Zorn's lemma is not needed in the first step.
2. "First-order" means that variables range over elements of our structure, not, for example, subsets thereof.

Theorem 4.6 (Compactness)**.** *Let $S \subseteq L$ be a set of sentences. Then if every finite subset of S has a model then S has a model.*

Proof. Trivial if we replace $\vDash$ with $\vdash$ as proof are finite. □

Note. Unlike in propositional logic, there is no decidability theorem, since we don't know how to check if $S \vDash t$.

Some consequences of compactness/completeness: let's think about axiomatisability of theories. We axiomatised the theory for groups, fields and graphs easily. Can we axiomatise the class of finite groups? In other words, we want some sentences S (in language of groups) such that a structure is a model for S if and only it is a finite group.

We may attempt to find some theorem that holds only for finite groups. For example, from IA Groups we may say that finiteness of conjugacy classes is a property of finite groups, but it also holds for the infinite group $\mathbb{Z}$. We may go to IB Groups, Rings and Modules and even IID Representation Theory to find some more theorems, but they also holds for some infinite groups. This gives us the inkling that maybe we should look in the other direction. In fact,

Corollary 4.7. *The class of finite groups cannot be axiomatised (in language of groups).*

Remark. It is amazing that we can actually *prove* this, as opposed to "believing it might be true".

Proof. Suppose S axiomatises finite groups. Add to S the sentences

$(\exists x_1)(\exists x_1)(\neg(x_1 = x_2))$ ("order at least 2")
$(\exists x_1)(\exists x_2)(\exists x_3)(\neg(x_1 = x_2) \wedge \neg(x_2 = x_3) \wedge \neg(x_3 = x_1))$ ("order at least 3")
$\vdots$

Then every finite subset has a model (e.g. $\mathbb{Z}/n\mathbb{Z}$ for n sufficiently large), but the set itself has no model, contradicting compactness. □

Similarly,

Corollary 4.8. *Let S be a theory in a language L. Then if S has arbitrarily large finite models, then it has an infinite model.*

Proof. Add sentences as in the previous proof and apply compactness. □

The takeaway is: finiteness is not a first-order property.

Corollary 4.9 (Upward Löwenheim-Skolem)**.** *If a theory S has an infinite model then it has an uncountable model.*

Proof. Add uncountably many constants $\{c_i\}_{i \in I}$ to the language and add to S the set of sentences "$\neg(c_i = c_j)$" for each distinct $i, j \in I$. Then any finite subset has a model (e.g. any infinite model of S), so the whole set has a model by compactness. □

Remark. Similarly for any set X, can find a model that doesn't inject into X. For example add $\gamma(X)$ of $\mathcal{P}(X)$ constants.

For example, we could find a model into which $\mathcal{P}(\mathcal{P}(\mathbb{R}))$ injects, by letting it be the indexing set. Thus we cannot axiomatise $\mathbb{R}$.

Example. We know there is an infinite field $\mathbb{Q}$, thus there exists a field of the same size as $\mathcal{P}(\mathcal{P}(\mathbb{R}))$.

Corollary 4.10 (Downward Löwenheim-Skolem)**.** *Let S be a theory in a countable language L. Then if S has a model then it has a countable model.*

Proof. The model constructed in Model existence lemma of $\overline{S}$ is countable. □

Note that this theorem and its proof *are* examinable.

4.5 Peano Arithmetic

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

Definition (Peano arithmetic)**.** *Peano arithmetic* (PA) or *formal number theory* has language L consisting of $\Omega = \{0, s, +, \times\}$ where $\alpha(0) = 0$, $\alpha(s) = 1$ is the *successor*, $\alpha(+) = \alpha(\times) = 2$. $\Pi = \emptyset$. The axioms are

1. $(\forall x)\neg(s(x) = 0)$.
2. $(\forall x)(\forall y)(s(x) = s(y) \implies x = y)$.
3. $(\forall y_1)\cdots(\forall y_n)((p[0/x] \wedge (\forall x)(p \implies p[s(x)/x])) \implies (\forall x)p)$ where the quantifiers in front are parameters, for each formula p and free variables $x, y_1, \ldots, y_n$.
4. $(\forall x)(x + 0 = x)$.
5. $(\forall x)(\forall y)(x + s(y) = s(x + y))$.
6. $(\forall x)(x \times 0 = 0)$.
7. $(\forall x)(\forall y)(x \times s(y) = s \times y + x)$.

Note. The 3rd axiom is the induction axiom. Our first guess would have been

$$(p[0/x] \wedge (\forall x)(p \implies p[s(x)/x])) \implies (\forall x)p,$$

but then we miss properties like $x \geq y$ where y is chosen earlier.

PA has an infinite model $\mathbb{N}$ so by upward Löwenheim-Skolem, it has an uncountable model, which is not isomorphic to $\mathbb{N}$. Doesn't this contradict the fact that the usual axioms characterise $\mathbb{N}$ uniquely?

The answer is no: axiom 3 is only "first-order" induction. Even in $\mathbb{N}$ itself, it refers to only countably many subsets, as opposed to true induction, which talks about the uncountably many subsets of $\mathbb{N}$.

Note that in PA the only constant is 0. We write $1 = s(0), 2 = 1 + 1 = s(1)$ etc.

Definition (Definable)**.** A subset $S \subseteq \mathbb{N}$ is called *definable* if there exists a formula $p \in L$ and free variable x such that for all $m \in \mathbb{N}$ we have $m \in S$ if and only if $p[m/x]$ holds in $\mathbb{N}$.

Example.

1. The set of square numbers. $p(x)$ is $(\exists y)(y \times y = x)$.
2. The set of primes. $p(x)$ is $\neg(x = 0) \wedge \neg(x = 1) \wedge (\forall y)(y \mid x \implies y = 1 \vee y = x)$ where $y \mid x$ means $(\exists z)(y \times z = x)$.
3. Powers of 2. $p(x)$ is $(\forall x)((y \mid x \wedge y \text{ prime}) \implies y = 2)$.

Note that only countably many subsets are definable.

Exercise. Write down the defining formula for the following sets:

1. powers of 4.
2. powers of 6.

The key question we are concerned with is: is PA a complete theory? In other words, for each sentence p, is it true that $\text{PA} \vdash p$ or $\text{PA} \vdash \neg p$?

Theorem 4.11 (Gödel incompleteness theorem)**.** *PA is not complete.*

Take p with $\text{PA} \nvdash p$ and $\text{PA} \nvdash \neg p$. Then we have p or $\neg p$ holds in $\mathbb{N}$, although we cannot prove it. Thus there the conclusion is there exists a sentence p such that p is true in $\mathbb{N}$ but $\text{PA} \nvdash p$.

Note that this *doesn't* contradict Gödel completeness theorem for first-order logic which says that if p is true in *all* models of PA then $\text{PA} \vdash p$.

5 Set Theory

The aim of this chapter is to answer the question, what does "the universe of set" look like? The key starting point is to view set theory as just another first order theory. There are many formulations of set theory and we will take Zermelo-Fraenkel set theory.

5.1 Zermelo-Fraenkel Set Theory

Definition (Zermelo-Fraenkel set theory)**.** *Zermelo-Fraenkel set theory* (ZF) has language L consisting of $\Omega = \emptyset, \Pi = \{\in\}, \alpha(\in) = 2$ with the ZF axioms (to be stated below).

A universe of sets will be a model $(V, \in)$ of the ZF axioms, so a worked example from chapter 4, except that every model $(V, \in)$ should be incredibly complicated (contains the entire world of maths, or a copy of it).

There are 9 ZF axioms in total: 2 to get started, 4 to build things and the last 3 might not be the things we would have thought of at first.

Axiom (Axiom of extension)**.**

$$(\forall x)(\forall y)((\forall z)(z \in x \iff z \in y) \implies x = y).$$

"If two sets have the same member, then they are equal."

Note. The converse is also true, which is an instance of a logical axiom.

Axiom (Axiom of separation)**.**

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

for each formula p with free variables $z, t_1, \ldots, t_n$.

"We can form a subset of a set." More precisely, given a set x and a property p, we can form $\{z \in x : p(z)\}$.

Note. We do want parameters, e.g. to have $\{z \in x : t \in z\}$ where t is chosen earlier.

Axiom (Axiom of empty set)**.**

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

"There is a set with no member."

We write $\emptyset$ for the unique (by axiom of extension) such set x. This is just an abbreviation: $p(\emptyset)$ means $(\exists x)((\forall y)(\neg y \in x) \wedge p(x))$.

Similarly, we write $\{z \in x : p(z)\}$ for sets formed by axiom of separation.

Axiom (Axiom of pair set)**.**

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

"We can form $\{x, y\}$."

We write $\{x, y\}$ for this set, and singleton $\{x\}$ for $\{x, x\}$. Now we can defined an ordered pair:

Definition (Ordered pair)**.** An *ordered pair* (x, y) is $\{\{x\}, \{x, y\}\}$.
x is an ordered pair if $(\exists y)(\exists z)(x = (y, z))$.

It is easy to check that $(x, y) = (t, u) \iff (x = t \wedge y = u)$ using axioms we have so far.

Definition (Function)**.** f is a *function* if

$$\begin{aligned}&(\forall x)(x \in f \implies x \text{ an ordered pair})\\ \wedge&(\forall x)(\forall y)(\forall z)((x, y) \in f \wedge (x, z) \in f \implies y = z).\end{aligned}$$

The *domain* of a function f, written $x = \operatorname{dom} f$, is such that

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

Write $f : x \to y$ for

$$(f \text{ a function}) \wedge (x = \operatorname{dom} f) \wedge (\forall t)[(\exists z)((z, t) \in f) \implies t \in y].$$

Axiom (Axiom of union)**.**

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

"We can form unions."

Axiom (Axiom of power set)**.**

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

where $z \subseteq x$ *means* $(\forall t)(t \in z \implies t \in x)$.

"We can form power sets."

Note.

1. We write $\bigcup x$ and $\mathcal{P}(x)$ for the two sets. We also write $x \cup y$ for $\bigcup\{x, y\}$ etc.
2. No extra axiom is needed for intersections: we can form $\bigcap x$ (with $x \neq \emptyset$) as a subset of y for any $y \in x$, which can be done by axiom of separation.
3. We can now form Cartesian product of sets $x \times y$ as a suitable subset of $\mathcal{P}(\mathcal{P}(x \cup y))$: since if $t \in x, u \in y$ then $(t, u) = \{\{t\}, \{t, u\}\} \in \mathcal{P}(\mathcal{P}(x \cup y))$.

4. Similarly we can form the set of all functions from x to y, as a subset of $\mathcal{P}(x \times y)$.

The axioms so far should be quite intuitive. The next three are more subtle. Note that so far a model V of ZF must be infinite. For example, write $x^+ = x \cup \{x\}$, then it is easy to check $\emptyset, \emptyset^+, \emptyset^{++}, ...$ are all distinct. We often write 0 for $\emptyset$, 1 for $\emptyset^+$, 2 for $\emptyset^{++}$ etc, so

$$\begin{aligned} 1 &= \{0\} \\ 2 &= \{0, 1\} \\ 3 &= \{0, 1, 2\} \\ &\vdots \end{aligned}$$

But this shows that V is infinite, which is not the question we are interested in. We want to know whether V has an infinite set, e.g. an x such that $\emptyset \in x, \emptyset^+ \in x$ etc.

In the real world of maths, yes, we do have V (which is defined to be a set in the first place). However, in V we cannot take V because of Russell's paradox.

Definition (Successor set)**.** x is a *successor set* if

$$(\emptyset \in x) \wedge (\forall y)(y \in x \implies y^+ \in x).$$

Axiom (Axiom of infinity)**.**

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

"There is an infinite set."

Note that any intersection of successor sets is a successor set, so there exists a least one. Call it ω. This will be our version in V of $\mathbb{N}$. Therefore

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

Note that if $x \subseteq \omega$ is a successor set then $x = \omega$ by definition, i.e.

$$(\forall x)(x \subseteq \omega \wedge \emptyset \in x \wedge (\forall y)(y \in x \implies y^+ \in x)) \implies x = \omega.$$

This is ω-*induction*. Note this is genuine induction in V over all subsets $x \subseteq \omega$, as opposed to first order induction in PA.

It is also easy to check that

$$\begin{aligned} &(\forall x)(x \in \omega) \implies \neg(x^+ = \emptyset) \\ &(\forall x)(\forall y)(x \in \omega \wedge y \in \omega \wedge x^+ = y^+) \implies x = y \end{aligned}$$

Thus ω satisfies (in V) all the usual axioms for the natural numbers.

Subsequently, we say x is *finite* if $(\exists y)(y \in \omega \wedge x \text{ bijects with } y)$. Then x is *countable* if x is finite or x bijects with ω. x is *infinite* if it is not finite.

Axiom (Axiom of foundation)**.**

$$(\forall x)(x \neq \emptyset \implies (\exists y)(y \in x \wedge (\forall z)(z \in x \implies \neg(z \in y)))).$$

"Sets are built up from simpler sets", or every (non-empty) set has an $\in$-minimal member.

The intuition behind is like this: we want to disallow $x \in x$ to avoid possible contradiction, $x \in y \wedge y \in x$ to agree with our intuition that "sets have a hierarchy", and also infinite chains $\cdots \in x_3 \in x_2 \in x_1 \in x_0$. What is common to all of them is that they do not have a $\in$-minimal element: $\{x\}$, $\{x, y\}$ and $\{x_0, x_1, \dots\}$ respectively do not have such an element in the above examples.

For our next axiom, we want if for each $i \in I$ we have A_i, then we can take $\{A_i : i \in I\}$. But how do we know that $\{A_i : i \in I\}$ is a set? One may say that "$i \mapsto A_i$" looks like a function, so the image is a set. But recall that functions are also sets. Is this rule a set?

This one is different from previous axioms we have. So far every axiom allows us to build a new set "near" the one in the universe we starting with, e.g. power set, union set. However, this one goes far out to V from I.

Thus what we really want to say is "the image of a set, under something that looks like a function, is a set." See Appendix A for a discussion on how to formalise this idea using classes.

Axiom (Axiom of replacement)**.**

$$\begin{aligned}&(\forall t_1)\cdots(\forall t_n)\underbrace{((\forall x)(\forall y)(\forall z)((p \wedge p[z/y]) \implies y = z))}_{p \text{ a function-class}}\\ \implies &\underbrace{((\forall x)(\exists y)(\forall z)(z \in y \iff (\exists t)(t \in x \wedge p[t/x, z/y])))}_{\text{image of } x \text{ under } p \text{ is a set}}\end{aligned}$$

for each formula p, free variables $x, y, t_1, \dots, t_n$.

"The image of a set under a function-class is a set."

Intuitively we think this axiom exactly as what the slogan says, but since classes are not part of the structure, formally we have to substitute the function-class with the (first order) rule.

Note that this holds also for partial functions.

Example. For any set x, can form $\{\{t\} : t \in x\}$, with function-class $t \mapsto \{t\}$. This is a "bad" example as we don't actually need axiom of replacement to know this is a set (it is a subset of the power set of x). See later for "good" examples.

Those are all the ZF axioms.

Note.

1. Sometimes axiom of separation is called "axiom of comprehension" and axiom of foundation is called "axiom of regularity".

2. ZF axioms do not include axiom of choice. ZF + AC is called ZFC where

Axiom (Axiom of choice)**.**

$$\begin{aligned}&(\forall f)(f \text{ a function} \wedge (\forall x)(x \in \operatorname{dom} f \implies \neg(f(x) = \emptyset))\\ \implies &(\exists g)(g \text{ a function} \wedge \operatorname{dom} g = \operatorname{dom} f \wedge (\forall x)(x \in \operatorname{dom} f \implies g(x) \in f(x)))).\end{aligned}$$

"Every family of (non-empty) sets has a choice function."

Remark. We have not proven ZF is consistent, i.e. there exists a model of ZF. Sadly, by one of Gödel incomplete theorem ZF $\nvdash$ "ZF has a model", so no proof in ordinary maths (including ZF, ZFC etc).

In this course, every theorem we prove about ZF will be preceeded with the premise that, either explicitly or implicitly, it holds in a model of ZF. Thus incompleteness does not pose a problem to our theory, although to make some practical sense out of the theory, we better have some faith in the existence of such a model!

5.2 Properties of ZF

Definition (Transitive)**.** x is *transitive* if every member of a member of x is itself a member of x:

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

i.e. $\bigcup x \subseteq x$.

Example.

1. Every member of ω, e.g. $2 = \{\emptyset, \{\emptyset\}\}$, is transitive.
2. ω is transitive as $n = \{0, 1, \dots, n-1\}$ for all $n \in \omega$.

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

Remark.

1. It officially says: let $(V, \in)$ be a model of ZF, then in V this statement holds, or equivalently, ZF $\vdash$ the statement (by completeness).
2. Any intersection of transitive sets is transitive, so we will know that there exists a *least* transitive set containing x, called the *transitive closure* of x, written $TC(x)$.

Proof. Consider $x \cup (\bigcup x) \cup (\bigcup\bigcup x) \cup \dots$, which is the obvious step to take. This is a set by axiom of union applied to

$$\{x, \bigcup x, \bigcup\bigcup x, \dots\},$$

which is a set by axiom of replacement applied to the function-class

$$\begin{aligned} 0 &\mapsto x \\ 1 &\mapsto \bigcup x \\ 2 &\mapsto \bigcup\bigcup x \\ &\vdots \end{aligned}$$

But why is this a function-class? We can't use recursion since that would be circular logic. Again we use the clever idea of attempts.

Define "f is an attempt" to mean

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

Then attempts exist, i.e.

$$(\forall n)(\exists f)(f \text{ attempt} \wedge n \in \operatorname{dom} f)$$

and are unqiue wherever they are defined, i.e.

$$(\forall n)(\forall f)(\forall f')((f, f' \text{ attempts} \wedge n \in \operatorname{dom} f \wedge n \in \operatorname{dom} f')$$
$$\implies f(n) = f'(n))$$

by ω-induction. So take function-class $p(y, z)$ to be

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

□

Remark. This is a "good" use of axiom of replacement.

Now let's take a look at axiom of foundation. The slogan says "sets are built out of simpler set", but does it achieve what it claims to do, or are we just banning things randomly for no reason? Actually there exists a good test: if the slogan is true, then suppose $p(y) \forall y \in x$ implies $p(x)$, we have $p(x) \forall x$,

Theorem 5.2 (Principle of $\in$-induction)**.** *Let p be a formula, free variables $x, t_1, \ldots, t_n$. Then*

$$(\forall t_1) \cdots (\forall t_n)((\forall x)((\forall y)(y \in x \implies p(y)) \implies p(x))) \implies (\forall x) p(x)$$

where $p(y)$ means $p[y/x]$, $p(x)$ means p.

Proof. Given $t_1, \ldots, t_n$, have $p(y) \forall y \in x \implies p(x)$, and suppose $(\forall x) p(x)$ is not true, so $(\exists x)(\neg p(x))$. We want to say: "choose $\in$-minimal y of $\{x : \neg p(x)\}$", then $\neg p(t)$, but $p(s) \forall s \in t$, contradiction. But this might not be a set, e.g. if $p(x)$ is $x \neq x$. This is where transitive closure comes in.

Let $t = TC(\{x\})$ so $x \in t$ and $\neg p(x)$. Let $u = \{y \in t : \neg p(y)\}$. Obviously $u \neq \emptyset$ so let y be an $\in$-minimal element of u. Then $\neg p(y)$. But $(\forall z \in y) p(z)$ (as $z \in y \implies z \in t$ and y is $\in$-minimal in n), so $p(y)$. Contradiction. □

Remark.

1. We used existence of transitive closure in the proof, i.e. the lemma above.
2. In fact, $\in$-induction is equivalent to the axiom of foundation, as we can deduce axiom of foundation from $\in$-induction (in the presence of the other axioms): say "x is regular" if

   $$(\forall y)(x \in y \implies y \text{ has an } \in\text{-minimal element}).$$

   Then axiom of foundation says every set is regular. To prove this by $\in$-induction, given y regular for all $y \in x$, want x to be regular. For $x \in z$, if x is minimal then done. If not, some $y \in x$ has $y \in z$. But y is regular so z has a minimal element.

Now we have induction, what about recursion? We want "$f(x)$ defined in terms of the $f(y)$ where $y \in x$".

Theorem 5.3 ($\in$-recursion)**.** *Let G be a function-class everywhere defined. Then there is a function class F (i.e. $(x,y) \in F \iff q(x,y)$ for some formula q) such that*

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

Moreover, F is unique.

Note. $F|_x = \{(z, F(z)) : z \in x\}$ is a set by axiom of replacement.

Proof. Say "f is an attempt" if

$$(f \text{ a function}) \wedge (\operatorname{dom} f \text{ transitive}) \wedge (\forall x)(x \in \operatorname{dom} f \implies f(x) = G(f|_x)).$$

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

$$(\forall x)(f, f' \text{ attempts defined at } x \implies f(x) = f'(x))$$

by $\in$-induction since if f and f' agree at all $y \in x$ then they agree at x. Also

$$(\forall x)(\exists \text{ an attempt } f \text{ defined at } x)$$

by $\in$-induction. Indeed suppose $(\forall y \in x)(\exists$ an attempt defined at $x)$. So $\forall y \in x$ there exists a unique attempt f_y defined on $TC(\{y\})$. Put $f = \bigcup_{y \in x} f_y$ and now set

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

So done by taking $q(x,y)$ to be

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

Uniqueness follows from $\in$-induction. □

Note. $\in$-induction and $\in$-recursion proofs look very similar to induction and recursion from chapter 2.

What properties of the "relation-class" $\in$ (i.e. the formula $p(x,y) = x \in y$) have we used?

1. p is well-founded: every non-empty set has a p-minimal element. We used it to make everything work.

2. p is local: $\{y : p(y,x)\}$ is a set for each x. We used this to build p-closure, i.e. transitive closure. By contrast $\ni$ and "superset of" is not local.

So in fact we have p-induction and p-recursion for any $p(x,y)$ that is well-founded and local. In particular, for a relation r on a set a, trivially r is local (as a is a set), so to have r-induction and r-recursion, just need r to be well-founded. Thus with this view in mind, induction and recursion from chapter 2 are special cases of this.

Recall that in chapter 2, we spend effort on induction so as to prove subset collapse. Here something similar happens: can we "model" a relation by $\in$?

For example, let $a = \{a_1, a_2, a_3\}$ and $r = \{(a_1, a_2), (a_2, a_3)\}$. Can we build a set b with the same relations but using $\in$? i.e. can we find $b = \{b_1, b_2, b_3\}$ and relation s defined by $b_i s b_j \iff b_i \in b_j$, and (a, r) is isomorphic to (b, s)? Certainly. We put $b_1 = \emptyset, b_2 = \{\emptyset\}, b_3 = \{\{\emptyset\}\}$. Then $a_i r a_j \iff b_i s b_j$ for all i, j. Moreover b is transitive.

Can we do this for all relations? Well not for all, since for example, axiom of foundation forbids the relation xrx.

Definition (Extensionality)**.** A relation r on a set a is *extensional* if

$$(\forall x, y \in a)((\forall z \in a)(zrx \iff zry) \implies x = y).$$

Example. a in the example above the definition, the relation $\in$ on any transitive set.

The analogue of subset collapse is

Theorem 5.4 (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 bijection $f : a \to b$ such that*

$$(\forall x, y \in a)(xry \iff f(x) \in f(y)).$$

Moreover b and f are unique.

Proof. This is basically r-recursion: once the images of all elements relate to a_n are fixed, we have no choice for $f(a_n)$ but let it be the set of images of all those things.

Define $f(x) = \{f(y) : yrx\}$, which is a definition by r-recursion on the set a, which should be the only sensible thing to try. Note that f is a function, not just a function-class, as it is an image of the set a.

Let $b = \{f(x) : x \in a\}$, which is a set by axiom of replacement. Then b is transitive by definition of f, and f is surjective by definition of b. If we can show f injective, then we also have $xry \iff f(x) \in f(y)$. We will show that

$$(\forall x \in a)(\forall y)(f(y) = f(x) \implies y = x)$$

by r-induction on x. So given y with $f(y) = f(x)$, want $y = x$, with the assumption that

$$(\forall t)(\forall u)((t, u \in a \wedge trx \wedge f(u) = f(t)) \implies u = t).$$

From $f(y) = f(x)$, we have

$$\{f(u) : ury\} = \{f(t) : trx\}$$

whence $\{u : ury\} = \{t : trx\}$ by induction assumption. Thus $x = y$ as r is extensional.

For uniqueness, if f and f' are both suitable then $(\forall x \in a)(f(x) = f'(x))$ by r-induction. □

Now we can do something that is owed from chapter 2. We defined ordinals to be equivalent classes of well-orderings, with two well-orderings regarded the same if there is an order-isomorphisms between them. But a hiccup is that the set of all well-orderings do not form a set so "equivalence class" does not make sense. Now we can instead define it formally in the language of ZF.

Definition (von Neumann ordinal)**.** An *ordinal* or *von Neumann ordinal* is a transitive set that is well-ordered by $\in$.

Note that by by axiom of foundation, we can say instead "totally ordered by $\in$".

Example. $\emptyset$, $\{\emptyset\}$, any $n \in \omega$, ω itself.

Mostowski's collapsing theorem tells us that any well-ordered X is order-isomorphic to a unique ordinal α. Say X has *order-type* α.

Remark (Irrelevant remark)**.** We know that for every ordinal α, have $\{\beta : \beta < \alpha\}$ is a well-order of order-type α. Hence by definition of f in Mostowski's collapsing theorem,

$$\alpha < \beta \iff \alpha \in \beta$$

so $\alpha = \{\beta : \beta < \alpha\}$, For example, $\omega = \{0, 1, 2, \dots\}$.

Thus the successor for ordinal is the same as the successor for set. For example,

$$\alpha^+ = \alpha \cup \{\alpha\}$$
$$\sup\{\alpha_i : i \in I\} = \bigcup\{\alpha_i : i \in I\}$$

although this might not be the most useful way to view things.

5.3 Picture of the Universe

In this section we build the universe of sets, where everything in mathematics takes place[1][2], starting with $\emptyset$ and taking power set many times.

Definition (von Neumann hierarchy)**.** For each ordinal α, define set V_α by recursion:

$$\begin{aligned} V_0 &= \emptyset \\ V_{\alpha+1} &= \mathcal{P}(V_\alpha) \\ V_\lambda &= \bigcup_{\alpha<\lambda} V_\alpha \text{ for } \lambda \text{ a non-zero limit} \end{aligned}$$

How do we know this is the whole universe? We want to show every set x belongs to some V_α.

Lemma 5.5. *Each V_α is transitive.*

Proof. Induction on α:

- 0: done.

[1] One might be tempted to think of this as a model of ZFC, or at least modulo some set vs class technicality, but there are many reasons not to do so, one of them being the violation of axiom of foundation. Curiously, "universe" may also in other context refer to the opposite, namely a model of ZFC (of course assuming consistency whereof).

[2] If you don't understand the previous footnote then don't worry and move on, since it's probably not intended for you!

- successor: Power set of a transitive set is transitive. More specifically, given $x \in y \in V_{\alpha+1}$, have $y \in \mathcal{P}(V_\alpha)$ so $x \in V_\alpha$ so $x \subseteq V_\alpha$, i.e. $x \in \mathcal{P}(V_\alpha) = V_{\alpha+1}$.
- limit: any union of transitive sets is transitive.

□

Lemma 5.6. *$V_\alpha \subseteq V_\beta$ whenever $\alpha \leq \beta$.*

Proof. Induction on β with α fixed:

- $\beta = \alpha$: done.
- successors: given $V_\alpha \subseteq V_\beta$, want $V_\alpha \subseteq \mathcal{P}(V_\beta)$. But $V_\beta \subseteq \mathcal{P}(V_\beta)$: $x \in V_\beta$ implies $x \subseteq V_\beta$ by transitivity of V_β.
- limits: obvious.

□

Theorem 5.7.

$$(\forall x)(\exists \alpha)(x \in V_\alpha).$$

The slogan is "the universe is the union of sets", or the suggestive identity $V = \bigcup_{\alpha \in ON} V_\alpha$ where ON is the class of ordinals. Note the subtlety hidden in the notation: you cannot take union of a family indexed by a proper class! (or more pedantically, nothing prevents you except that the result is not a set)

Note.

- $x \subseteq V_\alpha \iff x \in V_{\alpha+1}$.
- If $x \subseteq V_\alpha$, then there exists *least* such α, which we call the *rank* of x.

Example.

$$\begin{aligned} \text{rank}(\emptyset) &= 0 \\ \text{rank}(\{\emptyset\}) &= 1 \\ \text{rank}(\omega) &= \omega \end{aligned}$$

and $\text{rank}(\alpha) = \alpha$ for all ordinal α by induction.

Proof. We will show that $(\forall x)(\exists \alpha)(x \subseteq V_\alpha)$ by $\in$-induction. Given x, for each $y \in x$, we have $y \subseteq V_\alpha$ for some α, so $y \subseteq V_{\text{rank}(y)}$, i.e. $y \in V_{\text{rank}(y)^+}$. Let $\alpha = \sup\{\text{rank}(y)^+ : y \in x\}$, then $x \subseteq V_\alpha$. □

Remark.

1. What the proof says essentially is

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

   which is the right way to think about rank. For example, $\text{rank}(\{6\})$ is 7 as $\text{rank}(6) = 6$ since it is an ordinal.

2. Most of maths takes place in $V_{\omega+10}$, except in this course when we discussed order-types!

6 Cardinals

We will discuss "sizes" of sets in this chapter. We will work predomiantly under ZFC, insofar we do not refrain ourselves from comparing and contrasting results under ZF.

6.1 Definitions

We want to define $\operatorname{card} x$ so that $\operatorname{card} x = \operatorname{card} y$ if and only if $x \leftrightarrow x$. We cannot define it naïvely by

$$\operatorname{card} x = \{y : y \leftrightarrow x\}$$

as this may not be a set. But we do know $x \leftrightarrow \alpha$ for some ordinal α so we can define $\operatorname{card} x$ to be the least such α. It follows that $\operatorname{card} x = \operatorname{card} y$ if and only if $x \leftrightarrow y$.

If we choose to work without axiom of choice, among all y that bijects with x, we need to pick one. This seems impossible without choice, but we have the clever *Scott trick*: define the *essential rank* of x $\operatorname{essrank}(x)$ to be the least rank such that there exists y of this rank that bijects with x, and then define $\operatorname{card} x = \{y \subseteq V_{\operatorname{essrank}(x)} : y \leftrightarrow x\}$.

Definition (Cardinal)**.** m is a *cardinal* or a *cardinality* if $m = \operatorname{card} x$ for some x.

For $x \in \omega$, we write "x" for $\operatorname{card} x$. For example $\operatorname{card} 3 = 3$.

For cardinals m and n, say $m \leq n$ if M injects into N for some M and N with $\operatorname{card} M = m, \operatorname{card} N = n$ (which does not depend on the choice of M and N). Similarly, write $m < n$ if $m \leq n$ and $m \neq n$. For example, $\operatorname{card} \omega < \operatorname{card} \mathcal{P}(\omega)$.

Note that if $m \leq n, n \leq m$ then $m = n$ by Schröder-Berstein. So $\leq$ is a partial order, and even a total order (by well-ordering). Note that in ZF $\leq$ need not be a total order.

What do the cardinals look like? Of course there are the finite (i.e. boring) ones. Note that not all ordinals are cardinals as for example $\omega \leftrightarrow \omega + 1$.

Definition (Initial ordinal)**.** An ordinal α is *initial* if for all $\beta < \alpha$, β does not biject with α.

Example. $0, 1, 2, \dots, \omega, \omega_1, \gamma(X)$ for any set X. However ω^2 is not initial as $\omega^2 \leftrightarrow \omega$.

How do we get all the initial ordinals then? Obviously we have $\omega = \omega_0, \omega_1 = \gamma(\omega_0), \omega_2 = \gamma(\omega_1), \dots$. Think for a little longer and we find their supremum $\omega_\omega = \sup\{\omega_n : n = 0, 1, \dots\}$ is also initial as otherwise some smaller cardinal α would bijects with some $\beta < \alpha$.

Define ω_α, for each ordinal α, recursively by

$$\begin{aligned} \omega_0 &= \omega \\ \omega_{\alpha+1} &= \gamma(\omega_\alpha) \\ \omega_\lambda &= \sup\{\omega_\alpha : \alpha < \lambda\} \text{ for } \lambda \text{ a non-zero limit} \end{aligned}$$

Then every ω_α is initial by induction.

Also every infinite initial ordinal δ is an ω_α. Indeed, the ω_α's are unbounded in the ordinals (e.g. $\omega_\alpha \geq \alpha$ by induction) so there exists least α with $\omega_\alpha \geq \delta$, so $\omega_\alpha = \delta$ by definition of ω_α.

Definition (Aleph number)**.** Define the *aleph number* for each ordinal α

$$\aleph_\alpha = \mathrm{card}(\omega_\alpha).$$

Thus $\aleph_\alpha$'s are the cardinalities of all infinite sets (in ZF: of all infinite well-orderable sets).

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

6.2 Cardinal Arithmetics

Definition (Cardinal arithmetic)**.** For cardinals m and n, let $m = \mathrm{card}\, M, n = \mathrm{card}\, N$ for some M and N. Define

$$\begin{aligned} m + n &= \mathrm{card}(M \sqcup N) \\ mn &= \mathrm{card}(M \times N) \\ m^n &= \mathrm{card}(M^N) \end{aligned}$$

where $M \sqcup N$ is the disjoint union and

$$M^N = \{f : f \text{ a function from } N \text{ to } M\}.$$

Note. They are independent of choice of M and N and thus well-defined.

We can also define the sum over an indexed family

$$\sum_{i \in I} m_i = \mathrm{card}(\bigsqcup_{i \in I} M_i).$$

Note that axiom of choice is needed for this to be well-defined.

Example.

1. $\mathbb{R} \leftrightarrow \mathcal{P}(\omega) \leftrightarrow \{0,1\}^\omega$ so $\mathrm{card}(\mathbb{R}) = 2^{\aleph_0}$.
2. How many real sequences are there? In IA Numbers and Sets we have to fiddle around but this course provides a slick proof:

   $$\mathrm{card}(\mathbb{R}^\omega) = (2^{\aleph_0})^{\aleph_0} = 2^{\aleph_0 \aleph_0} = 2^{\aleph_0}.$$

   We are using obvious facts like:

   (a) $m + n = n + m$ as $M \sqcup N \leftrightarrow N \sqcup M$.
   (b) $mn = nm$ as $M \times N \leftrightarrow N \times M$.
   (c) $(m^n)^p = m^{np}$ as $(M^N)^P \leftrightarrow M^{N \times P}$.
   (d) $\aleph_0 \aleph_0 = \aleph_0$ as $\omega \times \omega \leftrightarrow \omega$.

We know $\aleph_0 \aleph_0 = \aleph_0$ but what about $\aleph_1 \aleph_1$? It turns out addition and multiplication of cardinals are trivially easy, thanks to

Theorem 6.1. *For all α,*

$$\aleph_\alpha \aleph_\alpha = \aleph_\alpha.$$

Proof. We will show that $\omega_\alpha \times \omega_\alpha \leftrightarrow \omega_\alpha$ by induction. Since ω_α is an ordinal we naturally want to equip $\omega_\alpha \times \omega_\alpha$ with a well-ordering. Product order doesn't work since it doesn't embed into ω_α. Recall that proof of $\mathbb{N} \times \mathbb{N} \leftrightarrow \mathbb{N}$ in IA Numbers and Sets: we traverse the half-lattice by following successive anti-diagonals. We use the same idea here, with anti-diagonals replaced by squares for easier of presentation.

Define a well-ordering on $\omega_\alpha \times \omega_\alpha$ by "going up in squares": put $(x, y) < (z, t)$ if

- either $\max\{x, y\} < \max\{z, t\}$ (bigger square beats smaller square),
- or $\max\{x, y\} = \max\{z, t\} = \beta$ (within a square) and one of
  - $y = t = \beta, x < z$,
  - or $x = z = \beta, y < t$,
  - or $t = \beta, y < \beta$.

This is obviously a well-ordering.

Given $r \in \omega_\alpha \times \omega_\alpha$, we have $r \in \beta \times \beta$ for some $\beta < \omega_\alpha$ (since ω_α is a limit), so $I_r \subseteq \beta \times \beta$ by definition of $<$. But $\beta \times \beta \leftrightarrow \beta$ (or β is finite) by induction hypothesis. Thus I_r has order-type $< \omega_\alpha$. Thus every proper initial segment of $(\omega, <)$ has order-type $< \omega_\alpha$. Thus $(\omega, <)$ has order-type $\leq \omega_\alpha$. Take cardinality, $\aleph_\alpha \aleph_\alpha \leq \aleph_\alpha$. The other direction is trivial by, for example, the diagonal embedding. □

Corollary 6.2. *Let $\alpha \leq \beta$. Then*

$$\aleph_\alpha + \aleph_\beta = \aleph_\alpha \aleph_\beta = \aleph_\beta.$$

So simply "take the bigger one". Cardinal addition and multiplication are boring!

Proof.

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

□

Example. For any infinite set X we have $X \leftrightarrow X \sqcup X$ (in ZFC).

However, cardinal exponentiation is much harder. Just a warning: exponentiation is different for cardinals and ordinals.

Example. For ordinals, ω^ω is countable as by definition,

$$\omega^\omega = \sup\{\omega, \omega^2, \dots\}$$

On the other hand, for cardinals, by Cantor's diagonal argument

$$\aleph_0^{\aleph_0} \geq 2^{\aleph_0} > \aleph_0.$$

To get an idea of how hard cardinal exponentiation is, $2^{\aleph_0}$ might not even be an aleph in ZF. In ZFC, we can still ask the question if $2^{\aleph_0} = \aleph_1$. Equivalently, if every $S \subseteq \mathbb{R}$ is either countable or bijects with $\mathbb{R}$.

This is the *continuum hypothesis*, and has proven to be independent of ZFC. Depending on your philosophical view, this is not so intuitively obvious as the other axioms in ZF or ZFC.

Even today, not all implications about values of $2^{\aleph_\alpha}$ are known. For example, if we are given that $2^{\aleph_\alpha} = \aleph_{\alpha+1}$ for $\alpha = 0, 1, 2, ...$, the best we can say, based on the results so far, is that $2^{\aleph_\omega} \leq \aleph_{\omega_4}$.

7 Gödel Incompleteness Theorem*

The aim of this non-examinable section is to show PA is incomplete, i.e. there exists p such that PA $\nvdash p$ and PA $\nvdash \neg p$. It suffices to show that there exists p, true in $\mathbb{N}$, such that PA $\nvdash p$. We are going to abbreviate "true in $\mathbb{N}$" as "true" and "PA $\nvdash p$" as "not provable".

We try to find a p saying "I am not provable", i.e. p such that p is true if and only if p is not provable. Then done: if p is false then PA $\vdash p$, so p holds in every model of PA, in particular in $\mathbb{N}$, absurd. Thus p is true so not provable.

The idea is to "code up" formulæ, proofs etc in PA, i.e. as natural numbers. At first glance it seems that we are doomed to fail since however we do it, "p is not provable" must be longer/more complicated than "p".

Recall that a subset $S \subseteq \mathbb{N}$ is *definable* if there exists formula $p(x)$ (which means a formula p with free variable x) such that $m \in S$ if and only if $p(m)$ is true. Similarly, $f : \mathbb{N} \to \mathbb{N}$ is *definable* if there exists a formula $p(x, y)$ such that $f(m) = n$ if and only if $p(m, n)$.

Example. $f(n) = 2n$ is definable: take $p(x, y)$ to be "$y = x + x$".

We should take as fact

Proposition 7.1. *Any function given by an algorithm is definable.*

Example. $f(n) = 2^n$ is definable.

Proof. A reference can be found in P. T. Johnstone Chapter 4 and 9. □

The language of PA has symbols $s, 0, +, \times, =, \bot, \Longrightarrow, \forall$, the two parentheses, as well as countably many variables $\{x_1, x_2, \dots\}$. It is possbile to reduce them just two symbols x and $\cdot'$, so x_1 is x, x_2 is x', x_3 is x'' etc. There is a total of 12 of symbols.

We can now code a formula by raising successive primes to the power of the successive symbols in p. For example, if p is

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

then its code is

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

Note that not every number codes a formula, for example $2^7 \cdot 3^5$ is non sense. "m codes a formula" is definable as there exists an algorithm.

Notation. Write S_m for the formula coded by m, and $S_m = \bot$ if m does not code a formula.

Note that "m codes an axiom" (logical or PA axiom) is definable. Also "ℓ, m, n code formula, with S_n following from S_ℓ and S_m by modus ponens" is definable, same for generalisation.

Now move on to proofs. We code a sequence of statements by

$$S(p_1, \dots, p_n) = 2^{c(p_1)} \cdot 3^{c(p_2)} \cdots (n\text{th prime})^{c(p_n)}.$$

Thus $\theta(m,n)$ = "m codes a proof of S_n" is definable. Then $\phi(n)$ = "n codes a provable statement" is definable since $\phi(n) \iff (\exists m)\theta(m,n)$. Note that this is *not* by algorithm, but because we can find a formula!

Here comes in the clever part. Consider $\chi(n)$ = "n codes formula S_n with one free variable, and $S_n(n)$ is not provable". This is clearly definable, say by formula $p(x)$ ($p(n)$ is true if and only if $\chi(n)$ where p is a formula of PA).

Let $N = c(p)$. Then $\chi(N)$ asserts that: N codes a formula with one free variable (so far this is true, since N codes $p(x)$), and this formula, with variable set to N (namely $p[N/x]$) is not provable. So $p(N)$ is true if and only if $p(N)$ is not provable. Done!

We have thus shown

Theorem 7.2 (Gödel incompleteness theorem)**.** *PA is not complete.*

But maybe PA is too weak. Could we add some clever sentences p to PA to make it complete? Maybe, for example, the p used in the proof above. The answer is no: just run the same proof with PA replaced by PA $\cup\{p\}$.

However, we can certainly extend PA to a complete theory, in an almost trivial way by adding to the axioms

$$T = \{p : p \text{ true in } \mathbb{N}\}.$$

Why does proof of the theorem fails?

We have to rewind all the way to the beginning — it can only be the case that T is not definable, i.e.

Theorem 7.3.

$$\{m : m \textit{ codes all true statements}\}$$

is not definable.

The slogan is "truth is not definable".

Another objection is why doesn't our proof of the incompleteness theorem (in particular, that p is true) formalise into a proof in PA that p is true? The answer is that we assumed the existence of a model of PA (namely $\mathbb{N}$), i.e. PA is consistent, which by completeness is con(PA) = "$\bot$ is not provable". Thus PA $\cup$ con(PA) $\vdash p$. So by deduction theorem, Gödel incompleteness theorem can be reformulated as

Theorem 7.4.

$$PA \nvdash \text{con}(PA).$$

How about ZF? Certainly ZF $\vdash$ con(PA) (note that con(PA) means slightly different things than before: it now means that for all $n \in \omega$, n does not code a proof of $\bot$). This is because ZF $\vdash$ "PA has a model" (namely ω).

But copying proof of incompleteness theorem gives

Theorem 7.5. *ZF is not complete.*

And by the same reasoning above,

Theorem 7.6.

$$ZF \nvdash con(ZF).$$

A Classes

$x \mapsto \{x\}$ for all x looks like a function, but isn't because the "domain" is too big: every function f has a domain $\operatorname{dom} f$ (defined as a suitable subset of $V \cup f$) and this "function" would have domain V, absurd as there is no universal set, i.e. $\neg(\exists x)(\forall y)(y \in x)$ (Russell's paradox).

Definition (Class)**.** For an L-structure $(V, \in)$, a collection C of points of V is called a *class* if there is a formula p, free variables x (parameterised), such that x belongs to C if and only if $p(x)$ holds in V.

Here "collection" is simply a set, in the true maths world. But to avoid confusion with subset in the sense of ZF we give it an alias. Same for "points" and "belongs to".

Example.

1. V is a class. Take $p(x)$ to be $x = x$.
2. For any t, $\{x : t \in x\}$ is a class. Take $p(x)$ to be $t \in x$. This shows that we need parameter t.
3. Every set y is a class. Take $p(x)$ to be $x \in y$.

Definition (Proper class)**.** If C is not a set (in V), i.e. $\neg(\exists y)(\forall x)(x \in y \iff p(x))$, say C is a *proper class*.

Example. V is a proper class, as is $\{x : x \text{ infinite}\}$.

Definition (Function-class)**.** A *function-class* is a collection F of ordered pairs from V such that there is a formula p, free vaiables x, y (parameterised), such that

1. $(x, y) \in F$ if and only if $p(x, y)$,
2. if $(x, y) \in F, (x, z) \in F$ then $y = z$.

Example. $x \mapsto \{x\}$ is a function-class. Take $p(x, y)$ to be $y = \{x\}$.

Index