

Logic and Set Theory

February 26, 2018

<i>CONTENTS</i>	2
-----------------	---

Contents

0	Miscellaneous	3
1	Propositional logic	4
2	Syntactic implication	6
3	Well-Orderings and Ordinals	10
4	Ordinals	14
4.1	Successors and limits	16
5	Posets and Zorn's lemma	18
5.1	Zorn's Lemma	19
5.2	Zorn's lemma and the axiom of choice	21
5.3	The Bourbaki-Witt theorem	21
6	Predicate Logic	23

0 Miscellaneous

Some introductory speech

1 Propositional logic

Let P denote a set of *primitive proposition*, unless otherwise stated, $P = \{p_1, p_2, \dots\}$.

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

- (1) $p \in L \ \forall p \in P$;
- (2) $\perp \in L$, where \perp is read as 'false';
- (3) If $p, q \in L$, then $(p \implies q) \in L$. For example, $(p_1 \implies L)$, $((p_1 \implies p_2) \implies (p_1 \implies p_3))$.

Note that at this point, each proposition is only a finite string of symbols from the alphabet $(,), \implies, \perp, p_1, p_2, \dots$ and do not really mean anything (until we define so).

By *inductively define*, we mean more precisely that we set $L_1 = P \cup \{\perp\}$, and $L_{n+1} = L_n \cup \{(p \implies q) : p, q \in L_n\}$, and then put $L = L_1 \cup L_2 \cup \dots$

Each proposition is built up *uniquely* from 1) and 2) using 3). For example, $((p_1 \implies p_2) \implies (p_1 \implies p_3))$ came from $(p_1 \implies p_2)$ and $(p_1 \implies p_3)$. We often omit outer brackets or use different brackets for clarity.

Now we can define some useful things:

- $\neg p$ (not p), as an abbreviation for $p \implies \perp$;
- $p \vee q$ (p or q), as an abbreviation for $(\neg p) \implies q$;
- $p \wedge q$ (p and q), as an abbreviation for $\neg(p \implies (\neg q))$.

These definitions 'make sense' in the way that we expect them to.

Definition. A *valuation* is a function $v : L \rightarrow \{0, 1\}$ s.t.

- (1) $v(\perp) = 0$; (2)

$$v(p \implies q) = \begin{cases} 0 & v(p) = 1, v(q) = 0 \\ 1 & \text{else} \end{cases} \quad \forall p, q \in L$$

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

Proposition. (1) If v, v' are valuations with $v(p) = v'(p) \ \forall p \in P$, then $v = v'$ (on L).

(2) For any $w : P \rightarrow \{0, 1\}$, there exists a valuation v with $v(p) = w(p) \ \forall p \in P$. In short, a valuation is defined by its value on P , and any values will do.

Proof. (1) We have $v(p) = v'(p) \ \forall p \in L_1$. However, 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 $L_n \ \forall n$.

(2) Set $v(p) = w(p) \ \forall p \in P$ and $v(\perp) = 0$: this defines v on L_1 . Having defined v on L_n , use the rules for valuation to inductively define v on L_{n+1} so we can extend v to L . \square

Definition. We say p is a *tautology*, written $\models p$, if $v(p) = 1 \forall$ valuations v .
Some examples:

(1) $p \implies (q \implies p)$: a true statement implies by anything. We can verify this by:

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

So we see that this is indeed a tautology;

(2) $(\neg\neg p) \implies p$, i.e. $((p \implies \perp) \implies \perp) \implies p$, called the "law of excluded middle";

(3) $[p \implies (q \implies r)] \implies [(p \implies q) \implies (p \implies r)]$.

Indeed, if not then we have some v with $v(p \implies (q \implies r)) = 1$, $v((p \implies q) \implies (p \implies r)) = 0$. So $v(p \implies q) = 1$, $v(p \implies r) = 0$. This happens when $v(p) = 1$, $v(r) = 0$, so also $v(q) = 1$. But then $v(q \implies r) = 0$, so $v(p \implies (q \implies r)) = 0$.

Definition. For $S \subset L$, $t \in L$, say S *entails* or *semantically implies* t , written $S \models t$ if $v(s) = 1 \forall s \in S \implies v(t) = 1$, for each valuation v .

("Whenever all of S is true, t is true as well.")

For example, $\{p \implies q, q \implies r\} \models (p \implies r)$. To prove this, suppose not: so we have v with $v(p \implies q) = v(q \implies r) = 1$ but $v(p \implies r) = 0$. So $v(p) = 1$, $v(r) = 0$, so $v(q) = 0$, but then $v(p \implies q) = 0$.

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

For $S \subset L$, v is a model of S if $v(s) = 1 \forall s \in S$. So $S \models t$ says that every model of S is a model of t . For example, in fact $\models t$ is the same as $\emptyset \models t$.

2 Syntactic implication

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

1. $p \implies (q \implies p) \forall p, q \in L$;
2. $[p \implies (q \implies r)] \implies [(p \implies q) \implies (p \implies r)] \forall p, q, r \in L$;
3. $(\neg\neg p) \implies p \forall p \in L$.

Note: these are all tautologies. Sometimes we say they are 3 axiom-schemes, as all of these are infinite sets of axioms.

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

For $S \subset L$, $t \in L$, a *proof* of t from S consists of a finite sequence t_1, \dots, t_n of propositions, with $t_n = t$, s.t. $\forall i$ 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*.

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

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

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

we deduce by the following:

- (1) $[p \implies (q \implies r)] \implies [(p \implies q) \implies (p \implies r)]$; (axiom 2)
- (2) $q \implies r$; (hypothesis)
- (3) $(q \implies r) \implies (p \implies (q \implies r))$; (axiom 1)
- (4) $p \implies (q \implies r)$; (mp on 2,3)
- (5) $(p \implies q) \implies (p \implies r)$ (mp on 1,4);
- (6) $p \implies q$; (hypothesis)
- (7) $p \implies r$. (mp on 5,6)

Example. Let's now try to prove $\vdash p \implies p$. Axiom 1 and 3 probably don't help so look at axiom 2; if we make $(p \implies q)$ and $p \implies (q \implies r)$ something that's a theorem, and make $p \implies r$ to be $p \implies p$ then we are done. So we need to take $p = p, q = (p \implies p), r = p$. Now:

- (1) $[p \implies ((p \implies p) \implies p)] \implies [(p \implies (p \implies p)) \implies (p \implies p)]$; (axiom 2)
- (2) $p \implies ((p \implies p) \implies p)$; (axiom 1)
- (3) $(p \implies (p \implies p)) \implies (p \implies p)$; (mp on 1,2)
- (4) $p \implies (p \implies p)$; (axiom 1)
- (5) $p \implies p$. (mp on 3,4)

Proofs are made easier by:

Proposition. (2, deduction theorem)

Let $S \subset L$, $p, q \in L$. Then $S \vdash (p \implies q)$ if and only if $(S \cup \{p\}) \vdash q$.

Proof. Forward: given a proof of $p \implies q$ from S , add the lines p (hypothesis), q (mp) to obtain a proof of q from $S \cup \{p\}$.

Backward: if we have proof $t_1, \dots, t_n = q$ of q from $S \cup \{p\}$. We'll show that $S \vdash (p \implies t_i) \forall i$, so $p \implies t_n = q$.

If t_i is an axiom, then we have $\vdash t_i \implies (p \implies t_i)$, so $\vdash p \implies t_i$;

If $t_i \in S$, write down $t_i, t_i \implies (p \implies t_i), p \implies t_i$ we get a proof of $p \implies t_i$ from S ;

If $t_i = p$: we know $\vdash (p \implies p)$, so done;

If t_i obtained by mp: in that case we have some earlier lines t_j and $t_j \implies t_i$.

By induction, we may assume $S \vdash (p \implies t_j)$ and $S \vdash (p \implies (t_j \implies t_i))$.

Now we can write down $[p \implies (t_j \implies t_i)] \implies [(p \implies t_j) \implies (t_i)]$ by axiom 2, $p \implies (t_j \implies t_i), p \implies t_j \implies (p \implies t_i)$ (mp), $p \implies t_j, p \implies t_i$ (mp) to obtain $S \vdash (p \implies t_i)$.

These are all of the cases. So $S \vdash (p \implies q)$. \square

This is why we chose axiom 2 as we did – to make this proof work.

Example. To show $\{p \implies q, q \implies r\} \vdash (p \implies r)$, it's enough to show that $\{p \implies q, q \implies r, p\} \vdash r$, which is trivial by mp.

Now, how are \vdash and \models related? We are going to prove the *completeness theorem*: $S \vdash t \iff S \models t$.

This ensures that our proofs are sound, in the sense that everything it can prove is not absurd ($S \vdash t$ then $S \models t$), and are adequate, i.e. our axioms are powerful enough to define every semantic consequence of S , which is not obvious ($S \models t$ then $S \vdash t$).

Proposition. (3)

Let $S \subset L, t \in L$. Then $S \vdash t \implies S \models t$.

Proof. Given a valuation v with $v(s) = 1 \forall s \in S$, we want $v(t) = 1$.

We have $v(p) = 1 \forall p$ axiom as our axioms are all tautologies (proven earlier); $v(p) = 1 \forall p \in S$ by definition of v ; also if $v(p) = 1$ and $v(p \implies q) = 1$, then also $v(q) = 1$ (by definition of \implies). So $v(p) = 1$ for each line p of our proof of t from S . \square

We say $S \subset L$ consistent if $S \not\vdash \perp$. One special case of adequacy is: $S \models \perp \implies S \vdash \perp$, i.e. if S has no model then S inconsistent, i.e. if S is consistent then S has a model. This implies adequacy: given $S \models t$, we have $S \cup \{\neg t\} \models \perp$, so by our special case we have $S \cup \{\neg t\} \vdash \perp$, i.e. $S \vdash ((\neg t) \implies t)$ by deduction theorem, so $S \vdash \neg \neg t$. But $S \vdash ((\neg \neg t) \implies t)$ by axiom 3, so $S \vdash t$ (mp).

Theorem. (4)

Let $S \subset L$ be consistent, then S has a model.

The idea is that we would like to define valuation v by $v(p) = 1 \iff p \in S$, or more sensibly, $v(p) = 1 \iff S \vdash p$.

But maybe $S \not\vdash p_3, S \not\vdash \neg p_3$, but a valuation maps half of L to 1, so we want to 'grow' S to contain one of p or $\neg p$ for each $p \in L$, while keeping consistency.

Proof. Claim: for any consistent $S \subset L$, $p \in L$, $S \cup \{p\}$ or $S \cup \{\neg p\}$ consistent.
Proof of claim. If not, then $S \cup \{p\} \vdash \perp$ and $S \cup \{\neg p\} \vdash \perp$, then $S \vdash (p \implies \perp)$ (deduction theorem), i.e. $S \vdash \neg p$, so $S \vdash \perp$ contradiction.

Now L is countable as each L_n is countable, so we can list L as t_1, t_2, \dots . Put $S_0 = S$; set $S_1 = S_0 \cup \{t_1\}$ or $S_0 \cup \{\neg t_1\}$ so that S_1 is consistent. Then set $S_2 = S_1 \cup \{t_2\}$ or $S_1 \cup \{\neg t_2\}$ so that S_2 is consistent, and continue likewise. Set $\bar{S} = S_0 \cup S_1 \cup S_2 \cup \dots$. Then $\bar{S} \supset S$, and \bar{S} is consistent (as each S_n is, and each proof is finite). $\forall p \in L$, we have either $p \in \bar{S}$ or $(\neg p) \in \bar{S}$. Also, \bar{S} is *deductively closed*, meaning that is $\bar{S} \vdash p$ then $p \in \bar{S}$: if $p \notin \bar{S}$ then $(\neg p) \in \bar{S}$, so $\bar{S} \vdash p$, $\bar{S} \vdash (\neg p)$ so $\bar{S} \vdash \perp$ contradiction.

Define $v : L \rightarrow \{0, 1\}$ by $p \rightarrow 1$ if $p \in \bar{S}$, 0 otherwise. Then v is a valuation: $v(\perp) = 0$ as $\perp \notin \bar{S}$; for $v(p \implies q)$:

If $v(p) = 1$, $v(q) = 0$: We have $p \in \bar{S}$, $q \notin \bar{S}$, and want $v(p \implies q) = 0$, i.e. $(p \implies q) \notin \bar{S}$. But if $(p \implies q) \in \bar{S}$ then $\bar{S} \vdash q$ contradiction;

If $v(q) = 1$: have $q \in \bar{S}$, and want $v(p \implies q) = 1$, i.e. $(p \implies q) \in \bar{S}$. But $\vdash q \implies (p \implies q)$ so $\bar{S} \vdash (p \implies q)$;

If $v(p) = 0$: have $p \notin \bar{S}$, i.e. $(\neg p) \in \bar{S}$ and want $(p \implies q) \in \bar{S}$. So we need $(p \implies \perp) \vdash (p \implies q)$, i.e. $p \implies \perp, p \vdash q$ (deduction theorem). Thus it's enough to show that $\perp \vdash q$. But $(\neg \neg q) \implies q$, and $\vdash (\perp \implies (\neg \neg q))$ (axiom 3 and 1 – to see the second one, write \neg explicitly using \implies and \perp), so $\vdash (\perp \implies q)$, i.e. $\perp \vdash q$. \square

Remark. Sometimes this is called 'completeness theorem'. The proof used P being countable to get L countable; in fact, result still holds if P is uncountable (see chapter 3).

By remark before theorem 4, we have

Corollary. (5, adequacy)

Let $S \subset L$, $t \in L$. Then if $S \models t$ then $S \vdash t$.

And hence,

Theorem. (6, completeness theorem)

Let $S \subset L$, $t \in L$. Then $S \vdash t \iff S \models t$.

Some consequences:

Corollary. (7, compactness theorem)

Let $S \subset L$, $t \in L$ with $S \models t$. Then \exists finite $S' \subset S$ with $S' \models t$.

This is trivial if we replace \models by \vdash (as proofs are finite).

Special case for $t = \perp$: If S has no model then some finite $S' \subset S$ has no model. Equivalently,

Corollary. (7', compactness theorem, equivalent form)

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

This *isi* equivalent to corollary 7 because $S \models t \iff S \cup \{\neg t\}$ has no model and $S' \models t \iff S' \cup \{\neg t\}$ has no model.

Corollary. (8, decidability theorem)

There is an algorithm to determine (in finite time) whether or not, for a given finite $S \subset L$ and $t \in L$, we have $S \vdash t$.

This is highly non-obvious; however it's trivial to decide if $S \models t$ just by drawing a truth table, and $\models \iff \vdash$.

3 Well-Orderings and Ordinals

Definition. A *total order* or *linear order* on a set X is a relation $<$ on X , such that

- (1) Irreflexive: Not $x < x \forall x \in X$;
- (2) Transitive: $x < y, y < z \implies x < z \forall x, y, z \in X$;
- (3) Trichotomous: $x < y$ or $x = y$ or $y < x \forall x, y \in X$.

Note: two of (iii) cannot hold: if $x < y, y < x$ then $x < x$ by transitivity.

Write $x \leq y$ if $x < y$ or $x = y$, and $y > x$ if $x < y$.

We can also define total order in terms of \leq :

- (1) Reflexive: $x \leq x \forall x \in X$;
- (2) Transitive: $x \leq y, y \leq z \implies x \leq z \forall x, y, z \in X$;
- (3) Antisymmetric: $x \leq y, y \leq x \implies x = y \forall x, y \in X$;
- (4) 'Tri'chotomous (although it's only two): $x \leq y$ or $y \leq x \forall x, y \in X$.

Example. $\mathbb{N}, \mathbb{Q}, \mathbb{R}$ with the usual orders are all total orders.

\mathbb{N}^+ the relation 'divides' is not a total order: for example we don't have any of $2|3, 3|2$ or $2 = 3$.

$\mathcal{P}(S)$ for some S (with $|S| \geq 2$ to be rigorous), with $x \leq y$ if $x \subseteq y$ is not a total order for the same reason.

A total order is a *well-ordering* if every (non-empty) subset has a least element, i.e. $\forall S \subset X, S \neq \emptyset \implies \exists x \in S, x \leq y \forall y \in S$.

Example. 1. \mathbb{N} with the usual $<$ is a well ordering.

2. $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$ with the usual $<$ are not well orderings.

3. $\mathbb{Q}^+ \cup \{0\}$ with the usual $<$ is not a well ordering (e.g. $(0, \infty) \subset \mathbb{Q}^+ \cup \{0\}$).

4. The set $\{1 - \frac{1}{n} : n = 2, 3, \dots\}$ as a subset of \mathbb{R} with the usual ordering is a well ordering.

5. The set $\{1 - \frac{1}{n} : n = 2, 3, \dots\} \cup \{1\}$ as a subset of \mathbb{R} with the usual ordering is a well ordering.

6. The set $\{1 - \frac{1}{n} : n = 2, 3, \dots\} \cup \{2 - \frac{1}{n} : n = 2, 3, \dots\}$ (same assumption) is a well ordering.

Remark. X is well-ordered iff there is no $x_1 > x_2 > x_3 > \dots$ in X .

Clearly if there is such a sequence then $S = \{x_1, x_2, \dots\}$ has no least element.

Conversely, if $S \subset X$ has no least element, then for each element $x \in S$ there exists a $x' \in S$ with $x' < x$, so we can just pick x, x', \dots inductively.

Definition. We say total orders X, Y are *isomorphic* if there exists a bijection $f : X \rightarrow Y$ that is order-preserving, i.e. $x < y \iff f(x) < f(y)$.

For example, 1 and 4 above are isomorphic; 5 and 6 are isomorphic; 4 and 5 are not isomorphic (one has a greatest element, and the other doesn't).

Here comes the first reason why well orderings are useful:

Proposition. (1, Proof by induction)

Let X be well-ordered, and let $S \subset X$ be s.t. if $y \in S \forall y < x$ then $x \in S$ (each $x \in X$). Then $S = X$.

Equivalently, if $p(x)$ is a property s.t. $\forall x: \text{if } p(y) \forall y < x \text{ then } p(x)$, then $p(x) \forall x$.

(I think we must assert S to be non-empty here, but the lecturer didn't agree with me; need to check later.)

Proof. If $S \neq X$ then let x be the least element of $X \setminus S$. Then $x \notin S$. But $y \in S \forall y < x$, contradiction. \square

A typical use:

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

Proof. Let f, g be isomorphisms. We'll show $f(x) = g(x) \forall x$ by induction. Thus we may assume $f(y) = g(y) \forall y < x$, and want $f(x) = g(x)$. Let a be the least element of $Y \setminus \{f(y) : y < x\}$. Then we must have $f(x) = a$: if $f(x) > a$, then some $x' > x$ has $f(x') = a$ by surjectivity, contradiction. The same shows $g(x)$ = least element of $Y \setminus \{g(y) : y < x\}$, but this is the same as a . So $f(x) = g(x)$. \square

Remark. This is false for total orders in general. One example is, consider from $\mathbb{Z} \rightarrow \mathbb{Z}$, we could either take identity, or $x \rightarrow x - 5$; or from \mathbb{R} to \mathbb{R} we could take identity or $x \rightarrow x - 5$ or $x \rightarrow x^3 \dots$

Definition. 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. For any $x \in X$, set $I(x) = \{y \in X : y < x\}$. Then this is an initial segment.

Obviously, not every initial segment is of this form: for example, in \mathbb{R} we can take $\{x : x \leq 3\}$; or in \mathbb{Q} , take $\{x : x^2 < 2\} \cup \{x < 0\}$ (this cannot be written as above form as $\sqrt{2} \notin \mathbb{Q}$).

Note: in a well-ordering, every proper initial segment *is* of the above form: let x be the least element of $X \setminus I$. Then $y < x \implies y \in I$. Conversely, if $y \in I$, then we must have $y < x$: otherwise $x \in I$, contradiction.

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

Note: this is very false for total orders: e.g. $\{1, 5, 9\} \subset \mathbb{Z}$, or $\mathbb{Q} \subset \mathbb{R}$. If we have $S \subset X$, we would like to define $f : S \rightarrow X$ that sends the smallest of S to the smallest of X , then remove them from both sets and send the smallest of the remaining to the smallest of the remaining, etc... But to do this we need a theorem.

Theorem. (3, definition by recursion)

Let X be well-ordered, Y be a set, and $G : \mathcal{P}(X \times Y) \rightarrow Y$. Then $\exists f : X \rightarrow Y$ s.t. $f(x) = G(f|_{I_x})$ for all $x \in X$. Moreover, such f is unique.

Here we define the restriction as: for $f : A \rightarrow B$, and $C \subset A$, the restriction of f to C is $f|_C = \{(x, f(x)) : x \in C\}$. (I think the lecturer is regarding a function as subset of a cartesian product)

In defining $f(x)$, make use of $f|_{I_x}$, i.e. the values of $f(y), y < x$.

Proof. Existence: define 'h is an attempt' to mean: $h : I \rightarrow Y$, some initial segment I of X , and $\forall x \in I$ we have $h(x) = G(h|_{I_x})$. Note that h, h' are

attempts, both defined at x , then $h(x) = h'(x)$ by induction on x . Since if $h(y) = h'(y) \forall y < x$ then $h(x) = h'(x)$.

Also, $\forall x \in X$ there exists an attempt defined at x by induction on x : we want attempt defined at x , given $\forall y < x$ there exists attempt defined at y . For each $y < x$, we have unique attempt h_y defined on $\{z : z \leq y\}$ (unique by what we just showed).

Let $h = \cup_{y < x} h_y$: an attempt defined on I_x . This is single-valued by uniqueness, so is indeed a function.

So $h' = h \cup \{(x, G(h))\}$ is an attempt defined at x .

Now set $f(x) = y$ if \exists attempt h , defined at x , with $h(x) = y$ (single-valued).

Uniqueness: if f, f' suitable then $f(x) = f'(x) \forall x \in X$ (induction on X) – since if $f(y) = f'(y) \forall y < x$ then $f(x) = f'(x)$. \square

A typical application:

Proposition. (4, subset collapse)

Let X be well-ordered, $Y \subset X$. Then Y is isomorphic to an initial segment of X . Moreover, such initial segment is unique.

Proof. To have f an isomorphism from Y to an initial segment of X , we need precisely that $\forall x \in Y : f(x) = \min X \setminus \{f(y) : y < x\}$. So done (existence and uniqueness) by theorem 3.

Note that $X \setminus \{f(y) : y < x\} \neq \emptyset$, e.g. because $f(y) \leq y \forall y$ (induction), so $x \notin \{f(y) : y < x\}$. \square

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

How do different well-orderings relate to each other?

We say $X \leq Y$ if X is isomorphic to an initial segment of Y . For example, $\mathbb{N} \leq \{1 - \frac{1}{n} : n = 2, 3, \dots\} \cup \{1\}$.

Theorem. (5)

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

Proof. Suppose $Y \not\leq X$. To obtain $f : X \rightarrow Y$ that is an isomorphism with an initial segment of Y , need $\forall x \in X : f(x) = \min Y \setminus \{f(y) : y < x\}$. So we are done by theorem 3.

Note that we cannot have $\{f(y) : y < x\} = X$, as then Y is isomorphic to I_x . \square

Proposition. (6)

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

Proof. We have isomorphism f from X to an isomorphism of Y , and g the other way round. Then $g \circ f : X \rightarrow X$ is an isomorphism from X to an initial segment of X (i.s. of i.s. is i.s.), but that is impossible unless the initial segment is X

itself. So $g \circ f$ is identity (by uniqueness in subset collapse). Similarly, $f \circ g$ is identity on Y . \square

New well-orderings from old:

Write $X < Y$ if $X \leq Y$ but X not isomorphic to Y . Equivalently, $X < Y$ iff X is isomorphic to a proper initial segment of Y . For example, if $X = \mathbb{N}$, $Y = \{1 - \frac{1}{n}\} \cup \{1\}$ then $X < Y$.

Make a bigger one: given well-ordered 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 $X \subset Y$, and $<_X, <_Y$ agree on X , and X an initial segment of $(Y, <_Y)$.

Well-orderings $(X_i : i \in I)$ are nested if $\forall i, j \in I : X_i$ extends X_j or X_j extends X_i .

Proposition. (7)

Let $(X_i : i \in I)$ be a nested family of well-orderings. Then there exist well-ordering X with $X \geq X_i \forall i$.

Proof. Let $X = \cup_{i \in I} X_i$, with $x < y$ if $\exists i$ with $x, y \in X_i$ and $x <_i y$. Then $<$ is a well-defined total order on X . given $S \subset X$, $S \neq \emptyset$, choose i with $S \cap X_i \neq \emptyset$. Then $S \cap X_i$ has a minimal element (as X_i is well-ordered), which must also be a minimal element of S (as X_i an i.s. of X). Also, $X \geq X_i \forall i$. \square

4 Ordinals

Are the well-orderings themselves well-ordered?

An ordinal is a well-ordered set, with two well-ordered sets regarded as the same if they are isomorphic. (Just as a rational is an expression $\frac{M}{N}$, with $\frac{M}{N}, \frac{M'}{N'}$ regarded as the same if $MN' = M'N$. But, unlike for \mathbb{Q} , we cannot formalise by equivalence classes – see later).

If X is a well-ordering corresponding to ordinal α , say X has order-type α .

Example. 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 order-type of \mathbb{N} . So, in \mathbb{R} , $\{1, 3, 7\}$ has order-type 3. $\{1 - \frac{1}{n} : n = 2, 3, \dots\}$ has order-type ω . For X of o-t α and Y of o-t β , write $\alpha \leq \beta$ if $X \leq Y$ (this is independent of choice of X, Y). Similarly for $\alpha < \beta$ etc.

We know: $\forall \alpha, \beta, \alpha \leq \beta$ or $\beta \leq \alpha$, and if $\alpha \leq \beta, \beta \leq \alpha$ then $\alpha = \beta$.

Theorem. Let α be an ordinal. Then the ordinals $< \alpha$ form a well-ordered set of order-type α . e.g. the ordinals $< \omega$ are $0, 1, 2, 3, \dots$

Proof. Let X have o-t α . the well-orderings $< X$ are precisely (up to isomorphism) the proper initial segments of X , i.e. the $I_x, x \in X$.

But these are isomorphic to X itself, via $x \rightarrow I_x$. □

We often write I_α to be the set of ordinals less than α .

Proposition. (9)

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

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

Theorem. (10, Burali-Forti paradox):

The ordinals do not form a set.

Proof. Suppose not, let X be set of all ordinals. Then X is a well-ordering, say order-type α . So X is isomorphic to I_α . But I_α is a proper i.s. of X . □

Given α , we have $\alpha^+ > \alpha$. Also, if $\{\alpha_i : i \in I\}$ is a set of ordinals, then there exists α with $\alpha \geq \alpha_i \forall i$ (by applying prop 7 to the nested family of $I_{\alpha_i}; i \in I$).

In fact, there is therefore a least upper bound for $\{\alpha_i : i \in I\}$ by applying prop 9 to the set $\{\beta \leq \alpha : \beta \text{ an upper bound for the } \alpha_i\}$. This is written $\sup\{\alpha_i : i \in I\}$, e.g. $\sup\{2, 4, 6, 8, \dots\} = \omega$.

Some ordinals: $0, 1, 2, \dots, \omega, \omega + 1$ (officially ω^+), $\omega + 2, \dots$,
 $\omega + \omega = \omega \cdot 2 = \sup\{\omega + 1, \omega + 2, \dots\}$, $\omega^2 + 1, \omega^2 + 2, \dots$,

$\omega 3, \dots, \omega 4, \dots, \dots, \omega \omega = \omega^2 = \sup\{\omega, \omega 2, \omega 3, \dots\},$
 $\omega^2 + 1, \dots, \omega^2 + \omega, \omega^2 + \omega + 1, \dots, \omega^2 + \omega 2, \dots, \omega^2 + \omega^2 = \omega^2 2, \dots, \omega^2 3, \dots, \omega^2 4, \dots, \omega^2 5, \dots, \omega^2 \omega =$
 $\omega^3, \dots, \omega^3 2, \dots, \omega^4, \dots, \omega^\omega = \sup\{\omega, \omega^2, \omega^3, \dots\},$
 $\omega^\omega + 1, \dots, \omega^\omega 2, \dots, \omega^\omega \omega = \omega^{\omega+1},$
 $\omega^{\omega+2}, \dots, \omega^{\omega+3}, \dots, \omega^{\omega^2}, \dots, \omega^{\omega^3}, \dots, \omega^{\omega^\omega}, \dots$
 And as expected we have $\omega^{\omega^{\omega^{\omega^{\dots}}}} = \sup\{\omega, \omega^2, \omega^3, \dots\} := \varepsilon_0$, and then $\varepsilon_0 + 1, \dots$,
 and then the whole thing again until $\varepsilon_1 = \varepsilon_0^{\varepsilon_0}$.

However, although this thing looks quite magnificent, they are all just countable (as we have just done it). Is there an uncountable ordinal? In other words, is there an uncountable well-ordered set?

Theorem. (11)

There is an uncountable ordinal.

Proof.

IDEA : takes up all countable ordinals. However, this might not be a set.

Let $R = \{A \in \mathcal{P}(\mathbb{N} \times \mathbb{N})\}$ s.t. A is a well-ordering of a subset of \mathbb{N} . Let S be image of R under 'order-type', i.e. S is the set of all order-types of well-orderings of some subset of \mathbb{N} . Then S is the set of all countable ordinals. Let ω_1 be $\sup S$. Then ω_1 is uncountable: otherwise, then $\omega_1 \in S$, so ω_1 would be the greatest member of S . But then $\omega_1 + 1$ is also in S . \square

Note that, by contradiction, ω_1 is the *least* uncountable ordinal. ω_1 has some strange properties, e.g.

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

Similarly we have

Theorem. (11', Hartogs' lemma)

For any set X , there is an ordinal that does not inject into X .

To see that, just replace $\mathcal{P}(\mathbb{N} \times \mathbb{N})$ by $\mathcal{P}(X \times X)$ in the previous proof.

Write $\gamma(X)$ for the least such ordinal – e.g. $\gamma(\omega) = \omega_1$.

4.1 Successors and limits

Given ordinal α , does α (any set of order-type α , e.g. I_α) have a greatest element?

If yes: say β is that greatest element. Then $\gamma < \beta$ or $\gamma = \beta \implies \gamma < \alpha$, and $\gamma < \alpha \implies \gamma < \beta$ or $\gamma = \beta$ (as we can't have $\gamma > \beta$). In other words, $\alpha = \beta^+$. In that case, we call α a *successor*;

If not: then $\forall \beta < \alpha, \exists \gamma < \alpha$ s.t. $\gamma > \beta$. So $\alpha = \sup\{\beta : \beta < \alpha\}$. (this is false in general, e.g. $\omega + 5$). We call α a *limit*.

For example, 5 is a successor, $\omega + 5$ is a successor, ω is a limit, $\omega + \omega$ is a limit. (0 is a limit as well).

For ordinals α, β , define $\alpha + \beta$ by recursion on β (α fixed) by: $\alpha + 0 = \alpha$, $\alpha + \beta^+ = (\alpha + \beta)^+$, $\alpha + \lambda = \sup\{\alpha + \gamma : \gamma < \lambda\}$ for λ a non-zero limit.

For example, $\omega + 1 = (\omega + 0)^+ = \omega^+$, $\omega + 2 = \omega^{++}$, $1 + \omega = \sup\{1 + \gamma : \gamma < \omega\} = \omega$ – so addition is not commutative.

Officially, by 'recursion on the ordinals', we mean: define $\alpha + \gamma$ on $\{\gamma : \gamma \leq \beta\}$ (a set) recursively, plus uniqueness. Similarly for induction: if know $p(\beta) \forall \beta < \alpha \implies p(\alpha)$ (for each α), then must have $p(\alpha) \forall \alpha$. If not, say $p(\alpha)$ false: then look at $\{\beta \leq \alpha : p(\beta) \text{ false}\}$.

Note that $\beta \leq \gamma \implies \alpha + \beta \leq \alpha + \gamma$ (induction on γ). Also, $\beta < \gamma \implies \alpha + \beta < \alpha + \gamma$. Indeed, $\gamma \geq \beta^+$, so $\alpha + \gamma \geq \alpha + \beta^+ = (\alpha + \beta)^+ > \alpha + \beta$. However, $1 < 2$, but $1 + \omega = 2 + \omega$.

Proposition. (12)

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

Proof. Induction on γ :

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

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

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

Claim: $\beta + \lambda$ is a limit.

Proof of claim: We have $\beta + \gamma = \sup\{\beta + \gamma' : \gamma' < \gamma\}$. But $\gamma < \lambda \implies \exists \gamma' < \lambda$ with $\gamma < \gamma' \implies \beta + \gamma < \beta + \gamma'$. So $\{\beta + \gamma : \gamma < \lambda\}$ does not have a greatest element.

Back to the main proof, now $\alpha + (\beta + \gamma) = \sup\{\alpha + \delta : \delta < \beta + \lambda\}$. So want $\sup\{\alpha + (\beta + \gamma) : \gamma < \lambda\} = \sup\{\alpha + \delta : \delta < \beta + \lambda\}$.

\leq : $\gamma < \lambda \implies \beta + \gamma < \beta + \lambda$, so LHS \subset RHS;

\geq : $\delta < \beta + \lambda \implies \delta < \beta + \gamma$, some $\gamma < \lambda$ (definition of $\beta + \lambda$). So $\alpha + \delta \leq \alpha + (\beta + \gamma)$. \square

Alternative viewpoint:

Above is the 'inductive' definition of $+$. There is also a synthetic definition: $\alpha + \beta$ is the order-type of $\alpha \sqcup \beta$ (α disjoint union β), with all of α coming before all of β .

Clearly we have $\alpha + (\beta + \gamma) = (\alpha + \beta) + \gamma$ with this definition (same order-type). We need:

Proposition. (13)

The synthetic and inductive definition of $+$ coincide.

Proof. Write $\alpha + \beta$ for inductive, $\alpha +' \beta$ for synthetic. Do induction on β (α fixed).

0: $\alpha + 0 = \alpha = \alpha +' 0$:

Successors: $\alpha +' \beta^+ = (\alpha +' \beta)^+ = (\alpha + \beta)^+ = \alpha + \beta^+$;

λ a non-zero limit: $\alpha +' \gamma = \text{order-type of } \alpha \sqcup \gamma = \sup \text{ of order-type of } \alpha \sqcup \gamma, \gamma < \lambda$ (nest union, so order-type of union = sup – this was proved before) = $\sup(\alpha +' \gamma : \gamma < \lambda) = \sup(\alpha + \gamma : \gamma < \lambda) = \alpha + \lambda$. \square

Normally we prefer to use synthetic than inductive, *if* we do have a synthetic definition available.

Ordinal multiplication:

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. e.g:

$\omega 1 = \omega 0 + \omega = 0 + \omega = \omega$;

$\omega 2 = \omega 1 + \omega = \omega + \omega$;

$\omega\omega = \sup\{0, \omega, \omega + \omega, \omega + \omega + \omega, \dots\}$ (as in our big picture)

$2\omega = \sup\{2\gamma : \gamma < \omega\} = \omega$, so multiplication is not commutative.

Similarly, this also has a synthetic definition: $\alpha\beta$ is the order-type of $\alpha \times \beta$, with $(x, y) < (z, t)$ if either $y < t$ or $y = t$ and $x < z$. We can check that these coincide on the previous examples. Also we can see $\alpha(\beta\gamma) = (\alpha\beta)\gamma$ etc.

We can define ordinal exponentiation, powers, etc. Similarly. For example, let's define exponentiation:

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

Note that $\omega^1 = \omega$, $\omega^2 = \omega \cdot \omega$, and $2^\omega = \sup\{2^\gamma : \gamma < \omega\} = \omega$ (and is countable). This is different to what we expect from cardinality, but the notation in cardinality and here is different.

5 Posets and Zorn's lemma

A *Partially ordered* set or poset is a pair (X, \leq) where X is a set and \leq is a relation on X that is reflexive, transitive and antisymmetric. Write $x < y$ if $x \leq y, x \neq y$. In terms of $<$, a poset is irreflexive and transitive.

For example, any total order is a partial order; \mathbb{N}^+ with divides; for any set S , $\mathcal{P}(S)$, with $x \leq y$ if $x \subset y$; for any $X \subset \mathcal{P}(S)$, with same relation of $x \leq y$ if $x \subset y$ (e.g. all subspaces of a given vector space).

In general, a hasse diagram for a poset X consists of a drawing of the posets of X , with an upward line from x to y if y *covers* x , i.e. $y > x$, but no z that $y > z > x$.

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

In a poset X , a *chain* is a set $S \subset X$ that is totally ordered ($\forall x, y \in S : x \leq y$ or $y \leq x$).

Note: chains can be uncountable, e.g. in (\mathbb{R}, \leq) take \mathbb{R} .

We say $S \subset X$ is an *antichain* if no two element are related.

For $S \subset X$, an *upper bound* for S is an $x \in X$ s.t. $x \geq y \forall y \in S$.

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 of S .

Write $x = \sup S$ or $x = \vee S$.

e.g. In \mathbb{R} , $\{x : x^2 < 2\}$ has $\sqrt{2}$ as least upper bound, and $\sup = \sqrt{2}$ (so $\sup S$ need not be in S). In \mathbb{R} , \mathbb{Z} has no upper bound. In \mathbb{Q} , $\{x : x^2 < 2\}$ has $\sqrt{2}$ as an upper bound, but no least upper bound.

We say a poset is *complete* if every subset has a sup.

e.g. (\mathbb{R}, \leq) is not complete: \mathbb{Z} has no sup (so different to notion of 'completeness' from analysis);

$[0, 1]$ is complete; $(0, 1)$ is not complete: itself has no sup;

$\mathbb{P}(S)$ is always complete: $\{A_i : i \in I\}$ has $\sup \cup_{i \in I} A_i$.

A function $f : X \rightarrow X$, where X is any poset, is order-preserving if $f(x) \leq f(y) \forall x \leq y$.

e.g. on \mathbb{N} : $f(x) = x + 1$; on $[0, 1]$: $f(x) = \frac{1+x}{2}$ (halve the distance to 1); on $\mathbb{P}(S)$: $f(A) = A \cup \{i\}$ for some fixed $i \in S$.

not every order-preserving f has a fixed point ($f(x) = x$), e.g. $f(x) = x + 1$ on \mathbb{N} .

Theorem. (1, Knaster-Tarski fixed point theorem):

Let X be a complete poset. Then every order-preserving function $f : X \rightarrow X$ has a fixed point.

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

$s \leq f(s)$: Enough to show $f(s)$ is an upper bound for E (as s the *least* upper bound). But $x \in E \implies x \leq s \implies f(x) \leq f(s) \implies x \leq f(x) \leq f(s)$.

$s \geq f(s)$: Enough to show $f(s) \in E$ (as s an upper bound). We know $s \leq f(s)$, and want $f(s) \leq f(f(s))$. But that's true because f is order preserving. \square

Note: in any complete poset X , we have a greatest element ($x.s.t.x \geq y \forall y$), namely $\sup X$. A typical application of knaster-tarski:

Theorem. (2, schröder-bernstein theorem)

Let A, B be sets s.t. there exists injection $f : A \rightarrow B$ and an injection $g : B \rightarrow A$. Then there exists an bijection from A to B .

Proof. Seek partition $A = P \sqcup Q, B = R \sqcup S$ s.t. $f(P) = R$ and $g(S) = Q$. Then we are done: set h to be f on P , y^{-1} on Q , then $h : A \rightarrow B$ is a bijection.

i.e. we seek $P \subset A$ s.t. $A \setminus g(B \setminus f(P)) = P$. Define $\theta : \mathcal{P}(A) \rightarrow \mathcal{P}(A)$ via $P \rightarrow A \setminus g(B \setminus f(P))$. Then since $\mathcal{P}(A)$ is complete, θ order-preserving, there is a fixed point by K-T theorem. \square

5.1 Zorn's Lemma

An element x in poset X is *Maximal* if no $y \in X$ has $y > x$.

Posets need not have a maximal element, for example $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$.

Theorem. (3, Zorn's lemma)

Let X be a non-empty poset in which every chain has an u.b.. Then X has a maximal element.

Proof. Suppose not. Then for each $x \in X$ there is some $x' \in X$ with $x' > x$. Also, for any chain C we have an upper bound $u(C)$. Pick $x \in X$. Define $x_\alpha \in X$, each $\alpha < \gamma(x)$ ($\gamma(x)$ is the u.b.?) recursively by: $x_0 = x$, $x_{\alpha+1} = x'_\alpha$, $x_\lambda = u(\{x_\alpha : \alpha < \lambda\})$ for λ a non-zero limit (this is a chain by induction). Then $\alpha \rightarrow x_\alpha$ is an injection from $\gamma(X)$ to X . \square

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

e.g. $V =$ space of all real polynomials. We can take $1, x, x^2, \dots$

Let V now be all real sequences. But $l_1 = (1, 0, 0, 0, \dots)$, $l_2 = (0, 1, 0, 0, \dots)$, then l_1, l_2 LI but not spanning! (recall span must be a finite linear combination!) It's easy to check that there is no countable basis. Also, it turns out that there is no

explicit basis.

\mathbb{R} as a vector space over \mathbb{Q} . Basis is called a Hamel basis.

Theorem. (4) Every vector space V has a basis.

Proof. Let $X = \{A \subset V : A \text{ is LI}\}$, ordered by \subset . We seek a maximal element M of X (then we are done: if M does not span then choose $x \notin \langle M \rangle$, and now $M \cup \{x\}$ is LI, contradiction).

We have $X \neq \emptyset$, as $\emptyset \in X$.

Given a chain $\{A_i : i \in I\}$ in X , put $A = \cup_{i \in I} A_i$, then $A \supset A_i \forall i$, so just need $A \in X$, i.e. A LI. Suppose A is not LI, then $\sum_{i=1}^n \lambda_i x_i = 0$ for some $x_1, \dots, x_n \in A$, and λ_i scalars not all zero. We have $x_i \in A_{i_1}, \dots, x_n \in A_{i_n}$ for some $i_1, \dots, i_n \in I$. But $A_{i_1}, \dots, A_{i_n} \in A_{i_k}$, some k (as they are nested), contradicting A_{i_k} being LI. \square

Note: the only actual maths (i.e. linear algebra) in the proof was the 'then done' part.

Another application: completeness theorem when proposition language uncountable.

Theorem. (5)

Let $S \subset L(P)$, where P is any set. Then S consistent implies that S has a model.

Proof. We seek a maximal consistent $\bar{S} \supset S$. Then done: for each $t \in L(P)$ we have $\bar{S} \cup \{t\}$ or $\bar{S} \cup \{\neg t\}$ consistent (see chapter 1), hence $t \in \bar{S}$ or $\neg t \in \bar{S}$ by maximality of \bar{S} . Now define $v(t) = 1$ if $t \in \bar{S}$, 0 otherwise (as in chapter 1). Let X be the set of all consistent subsets of $L(P)$, ordered by \subset . Then $X \neq \emptyset$, as $S \in X$. Given a non-empty chain $(T_i : i \in I)$ in X , put $T = \cup_{i \in I} T_i$. Then $T \supset T_i$ for each i , so we just need $T \in X$. We have $S \subset T$ as $T \neq \emptyset$. Also T is consistent: if $T \vdash \perp$, then $\{t_1, \dots, t_n\} \vdash \perp$ for some $t_1, \dots, t_n \in T$. We have $t_1 \in T_{i_1}, \dots, t_n \in T_{i_n}$ for some $i_1, \dots, i_n \in I$. But $T_{i_1}, \dots, T_{i_n} \subset T_{i_k}$ for some k (nested), contradicting T_{i_k} being consistent. \square

One more:

Theorem. (6, well-ordering principle)

Every set S can be well-ordered.

Note that this is very surprising for e.g. $S = \mathbb{R}$.

Proof. Let $X = \{(A, R) : A \subset S \text{ and } R \text{ is a well-ordering of } A\}$. We order this by: $(A, R) \leq (A', R')$ if (A', R') extends (A, R) . Then $X \neq \emptyset$, as $(\emptyset, \emptyset) \in X$. Given a chain $((A_i, R_i) : i \in I)$, we have $(\cup_{i \in I} A_i, \cup_{i \in I} R_i) \in X$, and extends each (A_i, R_i) from chapter 2. So by Zorn's lemma, X has a maximal element (A, R) . We must have $A = S$: otherwise choose $x \in S \setminus A$ and take 'successor': well-order $A \cup \{x\}$ by putting $x > a \forall a \in A$, contradicting maximality of (A, R) . \square

Remark. Proof of zorn was easy, but we used a lot of machinery there (ordinals, recursion, hartog's lemma).

5.2 Zorn's lemma and the axiom of choice

In proof of Zorn's lemma, we chose, for each $x \in X$, and $x' \supset x$, i.e. we made infinitely many arbitrary choices, even by time we get to x_ω . We did the same in part IA, to prove that a countable union of countable sets is countable. This is appealing to the axiom of choice, saying that we may choose an element of each set in a family of non-empty sets.

More precisely, the axiom of choice states that, if $(A_i : i \in I)$ is a family of sets, we have a choice function, meaning a function $f : I \rightarrow \cup_{i \in I} A_i$ s.t. $f(i) \in A_i \forall i$. This is of a different character to the other set-building rules in that the object whose existence is asserted is not uniquely specified by its properties (unlike, e.g., $A \cup B$).

So often one points out when one has used axiom of choice.

Note that AC is trivial $|I| = 1$ ($A \neq \emptyset$ means $\exists x \in A$). Similarly for I finite by induction. However, there is no derivation of AC from the other set-building rules for general I .

Also, we cannot prove ZL without AC because we can deduce AC from ZL: Given family $(A_i : i \in I)$ of non-empty sets, a partial choice function is an $f : J \rightarrow \cup_{i \in I} A_i$ for some $J \subset I$, s.t. $f(j) \in A_j \forall j \in J$. Put $(J, f) \leq (J', f')$ if $J \subset J'$ and $f'|_J = f$. This poset is not empty. Also, given a chain we have an upper bound being the union of them. So by ZL, there is a maximal of such. We must have $J = I$ in that case, as if not we can choose (???) $i \in I \setminus J$, $x \in A_i$ and put $J' = J \cup \{i\}$, $f' = f \cup \{(i, x)\}$. Contradiction.

Conclusion: $ZL \iff AC$ (in presence of the other set-building rules).

Also, we had $ZL \implies WO$, and $WO \implies AC$ trivially (well order $\cup_{i \in I} A_i$ and let $f(i)$ be the least element of A_i). So we get $ZL \iff AC \iff WO$.

5.3 The Bourbaki-Witt theorem

Poset X is *chain-complete* if $X \neq \emptyset$ and every non-empty chain has a sup. For example, any complete poset is chain-complete; any finite poset is chain-complete; and $\{A \subset V : A \text{ is LI}\}$, for a vector space V is also.

We say $f : X \rightarrow X$ is *inflationary* if $f(x) \geq x \forall x$.

Theorem. (Bourbaki-Witt)

X chain-complete, $f : X \rightarrow X$ inflationary. Then f has a fixed point.

Note that BW follows instantly from ZL: take maximal x , and now $f(x) \geq x \implies f(x) = x$.

However, we can prove BW without AC: we pick some $x_0 \in X$, then let $x_1 = f(x_0)$, $x_2 = f(x_1)$, ..., and let x_ω be the sup of them.

In chapter 2, we did not use AC, except in remark that well-ordering \iff no decreasing sequence, and that ω_1 is not a countable sup.

In fact, it's easy to deduce ZL from BW (using AC). So we can view BW as the choice-free version of ZL.

6 Predicate Logic

Recall that a group is a set equipped with functions:

$M : A^2 \rightarrow A$ ('arity' (slots) 2) and inverse $iA \rightarrow A$ ('arity' 1), and a constant $e \in A$ (kind of 'arity' 0), s.t.

$$\begin{aligned} & (\forall x, y, z \in A)(M(x, M(y, z)) = M(M(x, y), z)), \\ & (\forall x \in A)(M(x, e) = x \wedge M(e, x) = x), \\ & (\forall x \in A)(M(x, i(x)) = e \wedge M(i(x), x) = e) \end{aligned}$$

And a poset is a set A equipped with a predicate (relation) \leq (arity 2) $\subset A^2$ s.t

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

We try to establish these correspondence between propositional logic and predicate logic: Language \rightarrow e.g. language of groups (thinks like the definitions above);

Valuation \rightarrow structure (set equipped with functions and relations of given arities);

Model of S (valuation making each $s \in S$ true) \rightarrow model of S (structure in which each $s \in S$ holds);

$S \models t \rightarrow$ same (e.g. In language of groups, should have the above 3 definitions $\models M(e, e) = e$ etc);

$S \vdash t \rightarrow$ same (but a bit more complicated).

Let Ω (function symbols) and Π (relation symbols) be disjoint sets, and α (arity) : $\Omega \cup \Pi \rightarrow \mathbb{N}$. The *language* $L = L(\Omega, \Pi, \alpha)$ is the set of *formulae*, defined by:

- variables: x_1, x_2, x_3, \dots (can use x, y , etc);

- terms: defined inductively by:

- (i) each variable is a term;

- (ii) If $f \in \Omega$, $\alpha(f) = n$, and t_1, \dots, t_n are terms, then $ft_1 \dots t_n$ is a term (and as always, we can add brackets, commas, etc). For example, in the language of groups: $\Omega = \{m, i, e\}$ of arities 2, 1, 0, $\Pi = \emptyset$. Some terms: $x_1, m(x_1, x_2), e, m(e, e), m(x_1, i(x_1))$, etc.

- Atomic formulae, consists of:

- (i) \perp ;

- (ii) $(s = t)$, any terms s, t ;

- (iii) $\phi(t_1, \dots, t_n)$, any $\phi \in \Pi$, $\alpha(\phi) = n$, and terms t_1, \dots, t_n .

Again use the language of groups as example: $m(x, y) = m(y, x)$, $m(x, i(x)) = e$;

In language of posets: $\Omega = \emptyset$, $\Pi = \{\leq\}$ of arity 2. We could take $x = y, x \leq y, x \leq x$.

- Formulae: defined inductively by:

- (i) Each atomic formula is a formula;

- (ii) If p, q are formulae, then so is $(p \implies q)$;

- (iii) If p is a formulae, x is a variable, then $(\forall x)p$ is a formula.

e.g. in language of groups $(\forall x)(m(x, x) = e)$, $(\forall x)((m(x, x) = e) \implies$

$(\exists y)(m(y, y) = x)$ (note that we have not talked about \exists yet; we'll do that later).

In language of posets: $(\forall x)(x \leq x)$.

Notes:

1. A formula is just a string of symbols.
2. We can now write $\neg p$ for $p \implies \perp$, and similarly for $p \wedge q$, $p \vee q$ etc, and $(\exists x)p$ for $\neg(\forall x)(\neg p)$.

A term is *closed* if it contains no variables. For example, $e, m(e, e), m(e, m(e, e))$. However, $m(x, i(x))$ is *not* closed.

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

For example, in $m(x, x) = e \implies (\exists y)(m(y, y) = x)$, each x is free and each y is bound.

Note that in some cases we can make a variable both free and bound: $(m(x, x) = e) \implies (\forall x)(\forall y)(m(x, y) = m(y, x))$. We see that x in LHS is free, but in RHS is bound (although it's not a very helpful expression).

A *sentence* is a formula without free variables: e.g., $(\forall x)(m(x, e) = x)$. For formula p , variable x , term t , the *substitution* $p[t/x]$ is obtained by replacing each free occurrence of x with t .

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

Semantic entailment: An L -structure consists of a non-empty (see later wfor why) set A equipped with, for each $f \in \Omega$ with $\alpha(f) = m$, a function $f_A : A^m \rightarrow A$, and for each $\phi \in \Pi$, with $\alpha(\phi) = n$, a relation $\phi_A \subset A^n$.

For example, let L be the language of groups: an L -structure is a set A with functions $m_A : A^2 \rightarrow A$, $i_A : A \rightarrow A$, e_A an element of A (need not be a group! These have no 'meaning' yet).

Another example: L be the language of posets: an L -structure is a set A with a relation $\leq_A \subset A^2$.

We want to define the *interpretation* $p_A \in \{0, 1\}$ of a sentence p in structure A , e.g. $(\forall x)(m(x, x) = e)$ should be 'true in A ' if $\forall a \in A : m_A(a, a) = e_A$.

So: 'insert $\in A$ subscript A and say it aloud'.

Formal bit: For L -structure A , define *interpretation* of a closed term t to be $t_A \in A$, defined inductively by:

$(ft_1 \dots t_n)_A = f_A(t_{1A}, \dots, t_{nA})$ for any $f \in \Omega$, $\alpha(f) = n$, closed terms t_1, \dots, t_n .
e.g. $m(e, i(e))_A = m_A(e_A, i_A(e_A))$ (and e_A already defined).

Atomic formulae: define $p_A \in \{0, 1\}$ for p atomic by:

- (i) $\perp_A = 0$;
- (ii)

$$(s = t)_A = \begin{cases} 1 & s_A = t_A \\ 0 & \text{else} \end{cases}$$

for s, t closed terms;

(iii)

$$\phi(t_1 \dots t_n)_A = \begin{cases} 1 & (t_{1A}, \dots, t_{nA}) \in \phi_A \\ 0 & \text{else} \end{cases}$$

for $\phi \in \Pi$, $\alpha(\phi) = n$, closed terms t_1, \dots, t_n .

Sentences: p_A defined inductively by:

(i)

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

(ii)

$$((\forall i)_p)_A = \begin{cases} 1 & p[\bar{a}/x]_A = 1 \text{ for all } a \in A \\ 0 & \text{else} \end{cases}$$

where, for any $a \in A$, add constant symbol \bar{a} to L , obtaining L' , and make A an L' -structure by setting $\bar{a}_A = a$.

If p has free variables, we can define $p_A \subset A^{\text{number of free variables of } p}$.

e.g. if p is $(\exists y)(m(y, y) = x)$, then $p_A = \{a \in A : \exists b \in A \text{ with } m_A(b, b) = a\}$.

If $p_A = 1$, say p *true* in A , or p holds in A , or A is a *model* of p . For T a theory (set of sentences), say T semantically entails p , written $T \models p$, if every model of T is a model of p .

p is a *tautology* if $\phi \models p$ (or just $\models p$), i.e. p holds in every L -structure. For example, $\models (\forall x)(x = x)$.

Examples: theory of groups: $\Omega = (m, i, e)$, $\Pi = \phi$. Let

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

Then an L -structure is a model of $T \iff$ it is a group.

Say T 'axiomatises' the class of groups or 'axiomatises the theory of groups'.

Sometimes call the elements of T the 'axioms' of T .

Theory of fields: $\Omega = \{+, \times, -, 0, 1\}$. T is: abelian group under $(+, -, 0)$; X is commutative, associative, distributive under $+$; $(\forall x)(1x = x)$, $\neg(1 = 0)$, $(\forall x)((\neg(x = 0)) \implies (\exists y)(xy = 1))$. Then T axiomatises the class of fields. E.g., $T \models$ inverses are unique: $(\forall x)((\neg(x \neq 0)) \implies ((\forall y)(\forall x)((yx = 1 \wedge zx = 1) \implies y = z))$.

Theory of posets: $\Omega = \phi$, $\Pi = \{\leq\}$.

T is: $(\forall x)(x \leq x)$, $(\forall x)(\forall y)(\forall z)((x \leq y \wedge y \leq z) \implies x \leq z)$, $(\forall x)(\forall y)((x \leq y \wedge y \leq x) \implies x = y)$.

Theory of graphs: $\Omega = \phi$, $\Pi = \{a\}$ ('is adjacent to').

T is $(\forall x)(\neg a(x, x))$, $(\forall x)(\forall y)(a(x, y) \implies a(y, x))$.

Proofs:

Logical axioms:

- (1) $p \implies (q \implies p)$ (any formulae p, q);
- (2) $p \implies (q \implies r) \implies ((p \implies q) \implies (p \implies r))$ (any formulae p, q, r);
- (3) $(\neg\neg p) \implies p$ (any formula p);
- (4) $(\forall x)(x = x)$; (any variable x);
- (5) $(\forall x)(\forall y)(x = y) \implies (p \implies p[y/x])$ (any variables x, y , formula p where y is a bound);
- (6) $(\forall x)p \implies p[t/x]$ (any variable x , term t , formula p with no variable in t occurring bound in p);
- (7) $(\forall x)(p \implies q) \implies (p \implies (\forall x)q)$ (any variable x , formulae p, q with x not occurring free in p).

As rules of deduction, we take:

Modus Ponens: From $p, p \implies q$ can deduce q ;

Generalisation: From p can deduce $(\forall x)p$, if x does not occur free in any premise used to prove p .

For $S \subset L$, $p \in L$, a proof of p from S is a finite sequence of formulae, ending with p , s.t. each line is a logical axiom, or a member of S , or follows from earlier lines by MP or GEN. Write $S \vdash p$ (' S proves P ') if there exists a proof of p from S .

Example: $\{x = y, x = z\} \vdash \{y = z\}$ (use axiom 5, with p being ' $x = z$ ').

1. $(\forall x)(\forall y)(x = y \implies (x = z \implies y = z))$ (axiom 5);
2. $(\forall x)(\forall y)(x = y \implies (x = z \implies y = z)) \implies (\forall y)(x = y \implies (x = z \implies y = z))$ (axiom 6, $t = 'x'$);
3. $(\forall y)(x = y \implies (x = z \implies y = z))$ (MP on 1,2);
4. $(\forall y)(x = y \implies (x = z \implies y = z)) \implies (x = y \implies (x = z \implies y = z))$ (axiom 6);
5. $x = y \implies (x = z \implies y = z)$ (MP on 3,4);
6. $x = y$ (hypothesis)
7. $x = y \implies y = z$ (mp on 5,6)
8. $x \implies z$ (hypothesis)
9. $y = z$ (mp on 7,8).

Aim: $T \vdash p \iff T \models p$.

e.g. if p holds in every group then p can be proved from the three group axioms (completely obvious).

Proposition. (1, deduction theorem)

Let $S \subset L$, $p, q \in L$. Then $S \vdash (p \implies q) \iff S \cup \{p\} \vdash q$.

Proof. Forward: as for propositional logic, from $p \implies q$ write down p and apply MP to obtain $S \cup \{p\} \vdash q$;

Backward: as for propositional logic: the only new case is 'generalisation'. So in proof of q from $S \cup \{p\}$ we have something like r then $(\forall x)r$ (Gen), and have a proof of $p \implies r$ from S (induction), and we want $S \vdash p \implies (\forall x)r$. In proof of r from $S \cup \{p\}$, no premise had x free. So in proof of $p \implies r$ from S , no premise had x free. Hence $S \vdash (\forall x)(p \implies r)$ (gen).

- If x does not occur free in p : we have $S \vdash p \implies (\forall x)r$ by axiom 6 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 axiom 1. \square

Proposition. (2, soundness)

Let S be a set of sentences, p a sentence. Then if $S \vdash p$ then $S \models p$.

Proof. We have proof of p from S , and a model A of S , and we want $p_x = 1$. This is an induction down the lines of the proof. \square

For adequacy, we want if $S \models p$, i.e. that if $S \cup \{\neg p\} \models \perp$, then $S \cup \{\neg p\} \vdash \perp$.

Theorem. (3, model existence lemma, or completeness theorem)

Let $S \subset L$ be a set of sentences. Then S consistent implies that S have a model.

Ideas:

- 1. Build model out of language: let A be the set of closed terms of L , with operation line $(1 + 1) +_A (1 + 1) = (1 + 1) + (1 + 1)$;
- 2. Say for S be the theory of fields: $(1 + 1) + 1 \neq 1 + (1 + 1)$, but $S \vdash (1 + 1) + 1 = 1 + (1 + 1)$. So quotient out by $s \sim t$ if $S \vdash s = t$;
- 3. Suppose s is the fields of characteristic 2 or 3, i.e. field axioms, and the statement $1 + 1 = 0 \vee 1 + 1 + 1 = 0$. Then $S \not\vdash 1 + 1 = 0$. So $[1 + 1] \neq [0]$, where $[\cdot]$ denotes the equivalent class under \sim . Also, $S \not\vdash 1 + 1 + 1 = 0$, so $[1 + 1 + 1] \neq [0]$.

So our structure does not satisfy $1 + 1 = 0 \vee 1 + 1 + 1 = 0$. Then we need to extend S to maximal consistent.

- 4. If S is 'fields with a square root of 2': field axioms + $(\exists x)(xx = 1 + 1)$. Maybe no closed term t has $[tt] = [1 + 1]$. So s lacks 'witnesses'.

Solution: for each $(\exists x)p$ in S , add new constant c to language, and add $p[c/x]$ to S . (e.g. $cc = 1 + 1$).

Now no longer maximal consistent, so go back to step 3.

Problem: this might not terminate.

Proof. We have consistent S in language $L_0 = L(\Omega, \Pi)$. Extend to maximal consistent S_1 (zorn), so for each sentence $p \in L$, we have $p \in S_1$, or $(\neg p) \in S_1$. Thus S_1 is complete (for every p , $S_1 \vdash p$ or $S_1 \vdash (\neg p)$). Add witnesses: for each $(\exists x)p$ in S_1 , add new constant c and axiom $p[c/x]$. We obtain T_1 in language $L_1 = L(\Omega \cup C_1, \Pi)$ that has *witnesses* for S_1 (if $(\exists x)p \in S$, then some closed term t has $p[t/x] \in T_1$). It's easy to check T_1 consistent. Now extend T_1 to maximal consistent S_2 (in L). Add witnesses, obtaining T_2 in language $L_2 = L(\Omega \cup C_1 \cup C_2, \Pi)$.

Continue inductively.

Put $\bar{S} = S_1 \cup S_2 \cup \dots$. In language $\bar{L} = L(\Omega \cup C_1 \cup C_2 \cup \dots)$.

- \bar{S} is consistent: If $\bar{S} \vdash \perp$, then some $S_n \vdash \perp$ (as proofs are finite), contradiction;
- \bar{S} is complete: given sentence $p \in \bar{L}$, we have $p \in L_n$ for some n (as p mentions only finitely many constants), so $S_{n+1} \vdash p$ or $S_{n+1} \vdash (\neg p)$ (choice of S_{n+1}).
- \bar{S} has witnesses (for itself): given $(\exists x)p \in \bar{S}$, we have $(\exists x)p \in S_n$ for some n . So $p[t/x] \in T_n$ for some closed term t (choice of T_n), whence $p[t/x] \in \bar{S}$. \square

On set of closed terms of \bar{L} , define $s \sim t$ if $\bar{S} \vdash (s = t)$.

This is clearly an equivalent relationship. let A be the set of equivalent classes. Make A into an \bar{L} -structure by setting $f_A([t_1], \dots, [t_n]) = [ft_1 \dots t_n]$ (each $f \in \bar{\Omega}$, $\alpha(f) = n$, closed terms $t_1 \dots t_n$), $\varphi_A = \{([t_1], \dots, [t_n]) : \bar{S} \vdash \phi(t_1, \dots, t_n)\}$ (each $\phi \in \Pi$, $\alpha(\phi) = n$, closed terms $t_1 \dots t_n$).

Claim: $\phi_A = 1 \iff \bar{S} \vdash p$ for each sentence $p \in \bar{L}$. (Then done: A is a model of \bar{S} , so A is a model of S .)

Proof. An easy induction:

Atomic sentences:

\perp : $\perp_A = 0$ and $\bar{S} \not\vdash \perp$.

$s = t$:

$$\begin{aligned} \bar{S} \vdash (s = t) &\iff [s] = [t] \\ &\iff s_A = t_A \\ &\iff (s = t)_A = 1 \end{aligned}$$

$\phi(t_1 \dots t_n)$: same.

Induction step:

$p \implies q$:

$$\begin{aligned} \bar{S} \vdash (p \implies q) &\iff \bar{S} \vdash (\neg p) \text{ or } \bar{S} \vdash q \\ &\iff p_A = 0 \text{ or } q_A = 1 (\text{induction}) \\ &\iff (p \implies q)_A = 1 \end{aligned}$$

where the second step is because, say if the forward direction doesn't hold, then $\bar{S} \vdash p$, $\bar{S} \vdash (\neg q)$ (since \bar{S} is complete), but then $\bar{S} \vdash \neg(p \implies q)$, contradiction).

$(\exists x)p$:

$$\begin{aligned} \bar{S} \vdash (\exists x)p &\iff \bar{S} \vdash p[t/x] \\ &\iff p[t/x]_A = 1 \\ &\iff ((\exists x)p)_A = 1 \end{aligned}$$

for some closed term t . The last line is because A is the set of equivalent classes of closed terms. \square

By remark before theorem 3 we have

Corollary. (4, adequacy)

If $S \models p$, then $S \vdash p$.

Hence:

Theorem. (5, Gödel's completeness theorem for first-order logic)

Let S be a set of sentences and p a sentence (in language L). Then $S \models p \iff S \vdash p$.

The proof is just soundness + adequacy.

Note:

- If L is countable (i.e. Ω, Π countable), then we don't need Zorn's lemma;
- 'First-order' means variables range over elements of our structure (not, e.g., subsets).

Theorem. (6, compactness)

Let $S \subset L$ be a set of sentences. Then if every finite subset of S has a model, then S has a model.

Proof. This is trivial if we replace \models with \vdash (as proofs are finite). \square

Note: we have no decidability theorem – how to check if $S \models t$?

Some consequences of completeness/compactness:

Can we axiomatise the class of finite groups? In other words, we want some sentences S (in language of groups) s.t. a structure is a model for $S \iff$ it is a finite group.

However, this is not possible.

Corollary. (7)

the class of finite groups cannot be axiomatised (in language of groups).

Proof. Suppose S axiomatises finite groups. We add to S the sentences:

$$\begin{aligned} &(\exists x_1)(\exists x_2)(\neg(x_1 = x_2)) \\ &(\exists x_1)(\exists x_2)(\exists x_3)(\neg(x_1 = x_2) \wedge \neg(x_1 = x_3) \wedge \neg(x_2 = x_3)) \\ &\dots \end{aligned}$$

which stands for $|G| \geq 2$, $|G| \geq 3$, etc.

Then ever finite subset has a model (e.g. \mathbb{Z}_n , n large). However, the set itself has no model – contradicting compactness. \square

Similarly,

Corollary. (7')

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 corollary 7, and apply compactness theorem. \square

So we know *finiteness is not a first-order property*.

Corollary. (8, upward Löwenheim-Skolem theorem)

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 $c_i \neq c_j$ (for each distinct $i, j \in I$). Then any finite subset has a model. So the whole set has a model by compactness. \square

Similarly, we couldn't find a model into which $P(P(R))$ injects (choose $I = P(P(R))$). E.g., there exists an infinite field (\mathbb{Q}), so there exists a field as big as $P(P(R))$.

Corollary. (9, downward Löwenheim-Skolem theorem):

Let S be a theory in countable language L . If S has a model, then it has a countable model.

Proof. The model constructed in theorem 3 is countable. □