Sogic and Let Theory

March 2, 2020

1 Propositional Logic

Let P be a set of ***primitive propositions***, i.e. P is a set of symbols with $(,), \bot, \implies \notin P$. Unless stated otherwise (i.e. that P is uncountable), we may assume that $P = \{p_1, p_2, \ldots\}$.

The set of ***propositions***, denoted by $L(P)$ or simply just L, is defined inductively as follows:

1. $P \subset L$
2. $\bot \in L$, called FALSE
3. if $p, q \in L$, then $(p \implies q) \in L$

Each proposition is a string of symbols from $P \cup \{(,), \bot, \implies\}$, for instance we have the propositions $p_1, (p_1 \implies p_1), ((p_1 \implies p_2) \implies (p_2 \implies (\bot \implies p_3)))$. For readability, we often draw symbols $(,)$ in different ways, for instance as $[, (, \big($.

Sometimes we omit the outside pair of parentheses when writing down propositions, for instance $p_1 \implies p_2$ is shorthand for $(p_1 \implies p_2)$.

Also we use some abbreviations, e.g.:

NOT: $\neg p$ to mean $(p \implies \bot)$

OR: $p \vee q$ to mean $(\neg p \implies q)$

AND: $p \wedge q$ to mean $\neg(\neg p \vee \neg q)$

What do we mean by L "defined inductively"? Define $L_0 = P \cup \{\bot\}$. Then, given L_n, we can define $L_{n+1} = L_{n-1} \cup \{(p \implies q) : p, q \in L_{n-1}\}$. Then we set $L = \bigcup_{n=0}^{\infty} L_n$. Note: if $p \in L \setminus (P \cup \{\bot\})$, then it is easy to show that there are ***unique*** $q, r \in L$ with $p = (q \implies r)$.

1.1 Semantic Entailment

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

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

If $p \in L$ and $v(p) = 1$ for every valuation, we say that p is a ***tautology***, and write $\vDash p$.

Examples:

1. $\vDash (p \implies p)$

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

So this is a tautology.

2. $\vDash (p \implies (q \implies p))$

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

So this is a tautology.

3. Is $\vDash (p \implies (q \implies r)) \implies ((p \implies q) \implies (p \implies r))$?

Suppose not. Then for some p, q, r and valuation v we have:

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

So $v(p \implies q) = 1, v(p \implies r) = 0$. Hence $v(p) = 1, v(r) = 0, v(q) = 1$. But then $v(q \implies r) = 0$, and so $v(p \implies (q \implies r)) = 0 \lightning$.

4. $\vDash ((p \implies \bot) \implies \bot) \implies p$, i.e. $\neg\neg p \implies p$, i.e. $(\neg p \vee p)$. This is the Law of the Excluded Middle, and is also a tautology.

Note that a valuation is entirely determined by its values on the primitive propositions.

Proposition 1.1.

1. *Let $v, w : L \to \{0, 1\}$ be valuations with $v|_P = w|_P$. Then $v = w$.*
2. *Let $f : P \to \{0, 1\}$. Then there is a valuation $v : L \to \{0, 1\}$ with $v|_P = f$.*

Proof.

1. We prove this by induction on n, so that $v|_{L_n} = w|_{L_n}$. For the base case of $n = 0$, $v|_P = w|_P$, and $v(\bot) = 0 = w(\bot)$. Then for the induction step, $v|_{L_{n-1}} = w|_{L_{n-1}}$. Let $p \in L_n \setminus L_{n-1}$. Then $p = (q \implies r)$ for some $q, r \in L_{n-1}$. We know that $v(q) = w(q), v(r) = w(r)$, and so $v(p) = w(p)$.
2. We define v successively on $L_0, L_1, L_2, \ldots$

   L_0: Let $v|_P = f$ and let $v(\bot) = 0$

   L_n: If $p \in L_n \setminus L_{n-1}$, then $p = (q \implies r)$, and so set $v(p)$ to be 0 if $v(q) = 1, v(r) = 0$, and 1 otherwise. Since propositions are built up in a unique way, this is indeed a valuation.

□

Let $S \subset L$. We say that v is a ***model*** of S if v is a valuation with $v(x) = 1$ for all $x \in S$. If $S = \{p\}$, we say that v is a model of p. If every model of $S \subset L$ is a model of $p \in L$, we say that S ***semantically entails*** p, and write $S \vDash p$. Note that $\emptyset \vDash p$ is exactly the same as $\vDash p$.

For example, $\{p, p \implies q\} \vDash q$.

1.2 Syntactic Entailment (Provability)

Our proof system will have axioms as follows for all $p, q, r \in L$:

A1 $p \implies (q \implies p)$

A2 $(p \implies (q \implies r)) \implies ((p \implies q) \implies (p \implies r))$

A3 $((p \implies \bot) \implies \bot) \implies p$

Our proof system also has a ***deduction rule*** known as ***modus ponens*** (MP): for all $p, q \in L$, from p and $(p \implies q)$ we can deduce q.

Note that each axiom is a tautology. For MP, see the last example of **§1.1**

Let $S \subset L$ and $p \in L$. A ***proof*** of p from S is a sequence $t_1, t_2, \ldots, t_n \in L$ of finite length with $t_n = p$ such that, for each i, either t_i is an axiom, or $t_i \in S$ (a ***hypothesis***), or there exist $j, k < i$ with $t_k = (t_j \implies t_i)$.

If there exists a proof of p from S, we say that S ***syntactically entails*** p, or S ***proves*** p, and we write $S \vdash p$. If $S = \emptyset$, we say p is a ***theorem*** and write $\vdash p$.

Example: $\vdash (p \implies p)$
Use A2, with $r = p$, to get $(p \implies (q \implies p)) \implies ((p \implies q) \implies (p \implies p))$. Now the first bracket is a theorem by A1, and if we take $q = (p \implies p)$ in the second, we can use modus ponens twice with A1 to deduce the final bracket, that $(p \implies p)$. We will write this formally:

Lemma 1.2. *For all $p \in L, \vdash (p \implies p)$*

Proof.

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

□

Proposition 1.3 (The Deduction Theorem)**.** *Let $S \subset L$ and $p, q \in L$. Then $S \vdash (p \implies q)$ if and only if $S \cup \{p\} \vdash q$.*

Proof. Suppose $t_1, \ldots, t_n$ is a proof of $p \implies q$ from S. Then $t_1, \ldots, t_n, p, q$ is a proof of q from $S \cup \{p\}$. Suppose that $t_1, \ldots, t_n$ instead is a proof of q from $S \cup \{p\}$. We show by induction on i that $S \vdash (p \implies t_i)$ for each i, and then we will be done since $t_n = q$.

1. If $t_i \in S$:

- $t_i \implies (p \implies t_i)$ (A1)
- t_i (hypothesis)
- $(p \implies t_i)$ (MP)

2. If $t_i = p$, use Lemma **1.2**
3. If $t_k = (t_j \implies t_i)$ for some $j, k < i$, then write down proofs of $(p \implies t_j), (p \implies t_k)$ from S. Then append:
   - $(p \implies (t_j \implies t_i)) \implies ((p \implies t_j) \implies (p \implies t_i))$ (A2)
   - $(p \implies t_j) \implies (p \implies t_i)$ (MP)
   - $p \implies t_i$ (MP)

□

1.3 The Completeness Theorem and Applications

The key result of this section will be that $\vDash$ and $\vdash$ coincide. There will be two directions to prove:

1. ***Soundness:*** If $S \vdash p$ then $S \vDash p$.
2. ***Adequacy:*** If $S \vDash p$ then $S \vdash p$.

Proposition 1.4 (Soundness Theorem)**.** *Let $S \subset L$ and $p \in L$ with $S \vdash p$. Then $S \vDash p$.*

Proof. Let $t_1, \ldots, t_n$ be a proof of p from S. Let v be a model of S. We show by induction on i that $v(t_i) = 1$ for $1 \leq i \leq n$.

If $t_i \in S$ then $v(t_i) = 1$. If t_i is an axiom then $\vDash t_i$ so $v(t_i) = 1$. Otherwise, $t_k = (t_j \implies t_i)$ for some $j, k < i$. By the induction hypothesis, $v(t_j) = v(t_j \implies t_i) = 1$, so $v(t_i) = 1$. □

For adequacy, first consider the special case $p = \bot$, i.e. "If $S \vDash \bot$ them $S \vdash \bot$". We will prove the contrapositive: "If $S \nvdash \bot$ then $S \nvDash \bot$". If $S \nvdash \bot$ we say that S is ***consistent***. '$S \vDash \bot$" means "if v is a model of S then $v(\bot) = 1$". But $v(\bot) = 0$ for every valuation v, so this says "S has no model." Hence "$S \nvDash \bot$" says "S has a model".

Theorem 1.5 (Model Existence Lemma)**.** *Let $S \subset L$ be consistent. Then S has a model.*

Proof in the case P is countable. L is countable, as each $p \in L$ is a finite string of symbols from $P \cup \{(,), \bot, \implies\}$.

We write $L = \{x_1, x_2, \ldots\}$. We shall recursively construct sets $S_n \subset L$ with $S = S_0 \subset S_1 \subset \ldots$ and S_n consistent.

The base case is trivial, as $S_0 = S$ is consistent by hypothesis. Then for $n > 0$, we have S_{n-1} consistent. If $S_{n-1} \cup \{\neg x_n\}$ is consistent, let $S_n = S_n \cup \{\neg x_n\}$. Otherwise, $S_{n-1} \cup \{\neg x_n\} \vdash \bot$, and by the deduction theorem, $S_{n-1} \vdash (\neg x_n \implies \bot)$, i.e. that $S_{n-1} \vdash \neg\neg x_n$. But $S_{n-1} \vdash (\neg\neg x_n \implies x_n)$ by (A3), and so $S_{n-1} \vdash x_n$ by (MP). But S_{n-1} is consistent, so let $S_n = S_{n-1} \cup \{x_n\}$.

Then let $\bar{S} = \bigcup_{n=1}^{\infty} S_n$. Firstly, S_n is consistent - suppose $t_1, \ldots, t_n$ is a proof of $\bot$ from $\bar{S}$. Then there is some collection $i_1, \ldots, i_m \in \mathbb{N}$ such that the hypotheses used in the proof come

from $S_{i_1}, \ldots, S_{i_m}$. Let $I = \max\{i_1, \ldots, i_m\}$. Then every hypothesis comes from S_I, and so $t_1, \ldots, t_n$ is a proof of $\bot$ from S_I. $\lightning$

Also, for every $p \in L$ we have $p \in \bar{S}$ or $\neg p \in \bar{S}$. Moreover, $\bar{S}$ is ***deductively closed*** (d.c): if $\bar{S} \vdash p$ then $p \in \bar{S}$. Indeed, suppose that $\bar{S} \vdash p$ but $p \notin \bar{S}$. Then $\neg p \in \bar{S}$. Now $\bar{S} \vdash p$ and $\bar{S} \vdash \neg p$, i.e. $\bar{S} \vdash (p \implies \bot)$. So by (MP), $\bar{S} \vdash \bot$ $\lightning$.

Now let $v : L \to \{0, 1\}$ be the indicator function of $\bar{S}$. We must check that v is a valuation. As $\bar{S}$ is consistent, it is certainly true that $\bot \notin \bar{S}$, and so $v(\bot) = 0$.

Let $p, q \in L$. We want to think about $(p \implies q)$:

Case 1. Suppose $v(q) = 1$. Then $q \in \bar{S}$, so $\bar{S} \vdash (p \implies q)$, but $\bar{S}$ is deductively closed, and so $(p \implies q) \in \bar{S}$, and $v(p \implies q) = 1$.

Case 2. Suppose $v(p) = 0$. Again, we must show that $v(p \implies q) = 1$, i.e. that $\bar{S} \vdash (p \implies q)$. By the Deduction Theorem, this is equivalent to $S \cup \{p\} \vdash q$, and $p \notin S$, so $\neg p \in S$ and it will be enough to show that $\{p, \neg p\} \vdash q$. We have:

1. $(p \implies \bot)$ (hyp)
2. p (hyp)
3. $\bot$ (MP on 1,2)
4. $((q \implies \bot) \implies \bot) \implies q$ (A3)
5. $\bot \implies ((q \implies \bot) \implies \bot)$ (A1)
6. $(q \implies \bot) \implies \bot$ (MP on 3,5)
7. q (MP on 4,6)

Case 3. $v(p) = 1, v(q) = 0$. We want to show that $v(p \implies q) = 0$. Suppose instead that $v(p \implies q) = 1$, so that $(p \implies q) \in \bar{S}, p \in \bar{S}$. But then by (MP) $\bar{S} \vdash q$, so $q \in \bar{S}$, so $v(q) = 1$. $\lightning$

We have now shown that v is a valuation. Moreover, $S \subset \bar{S}$ so $v(p) = 1$ for all $p \in S$. Hence v is a model of S. □

Corollary 1.6 (The Adequacy Theorem)**.** *Let $S \subset L$ and $p \in L$ with $S \vDash p$. Then $S \vdash p$.*

Proof. Suppose v is a model of $S \cup \{\neg p\}$. Then $v(p) = 1$ and $v(\neg p) = 1$, so $v(\bot) = 1$ $\lightning$. So $S \cup \{\neg p\}$ has no model, and so by the model existence lemma it is inconsistent. That is, $S \cup \{\neg p\} \vdash \bot$. Then by the deduction theorem, $S \vdash (\neg p \implies \bot)$, i.e. $S \vdash \neg\neg p$), and hence $S \vdash p$. □

Theorem 1.7 (The Completeness Theorem)**.** *Let $S \subset L$ and $p \in L$. Then $S \vDash p$ if and only if $S \vdash p$.*

Proof. Soundness and adequacy. □

Two important applications:

Corollary 1.8 (Compactness Theorem)**.** *Let $S \subset L$ such that every finite subset of S has a model. Then S has a model.*

Proof. Not at all obvious a priori, but obvious if we replace "has a model" by "is consistent". If S is not consistent, then $S \vdash \bot$, so, as proofs are finite, $T \vdash \bot$ for some finite $T \subset S$, so T is inconsistent. □

Corollary 1.9 (The Decidability Theorem)**.** *Let $S \subset L$ be finite and $p \in L$. Then there is an algorithm to determine in finite time whether or not $S \vdash p$.*

Proof. Obvious if we replace $\vdash$ by $\models$, and then do a truth table. □

1.4 What happens when P is uncountable?

This will be just a sketch - it will be a while before we can do this properly in chapter 3. We have only proved the completeness theorem under the assumption that P is countable. However, we only used this when showing that, if S is consistent, then there is a consistent $\bar{S} \supset S$ with $p \in \bar{S}$ or $\neg p \in \bar{S}$ for all $p \in L$.

We needed P to be countable so that $L = \{x_1, x_2, \ldots\}$ is countable and we can consider the x_is in turn, deciding if $x_i \in \bar{S}$ or $\neg x_i \in \bar{S}$.

Can we do without this assumption? Now allow P, and hence L, to be uncountable. Let $S \subset L$ be consistent and look for $\bar{S} \supset S$ consistent with $p \in \bar{S}$ or $\neg p \in \bar{S}$ for all $p \in L$. We could try $\bar{S} = S$, and if it works then we are done.

Otherwise, there is some $x_0 \in L$ with $x_0 \notin \bar{S}$ and $\neg x_0 \notin \bar{S}$. Exactly as in the countable case, either $\bar{S} \cup \{x_0\}$ or $\bar{S} \cup \{\neg x_0\}$ is consistent. So add x_0 or $\neg x_0$ to $\bar{S}$, keeping it consistent. If $\bar{S}$ works now, then we're done, otherwise there is some $x_1 \in L$ with $x_1 \notin \bar{S}$ and $\neg x_1 \notin \bar{S}$, and so on and so forth. If this never terminates, then we keep on going forever.

If after we've done this infinitely many times, if we are done the stop. Otherwise, we have x_ω with $x_\omega \notin \bar{S}$ and $\neg x_\omega \notin \bar{S}$, so either add x_ω or $\neg x_\omega$ to $\bar{S}$. If we're not done there is another $x_{\omega+1}$ and so on.

Either eventually we finish, or we have to go on forever, until we get to $x_{\omega \cdot 2}$, and then keep on going…

If this never terminates, we get loads and loads of x_is. What are they being indexed by? It looks to be some sort of extension of $\mathbb{N}$.:

$$0, 1, 2, 3, 4, \ldots, \omega, \omega+1, \omega+2, \ldots, \omega 2, \omega 2+1, \ldots, \omega 3, \ldots, \omega 4, \ldots,$$

$$\omega^2, \omega^2+1, \ldots, \omega^2+\omega, \ldots, \omega^2 2, \ldots, \omega^3, \ldots, \omega^4, \ldots, \omega^\omega, \ldots, \omega^{\omega^\omega}, \ldots, \omega^{\omega^{\omega^{\cdot^{\cdot^{\cdot}}}}} = \epsilon_0, \ldots$$

This is only countably many ordinals, but in fact if we keep on going, we can get to uncountably many numbers. We call these indices ***ordinals***, and we will define them in chapter 2 in such a way that:

- 0 is an ordinal
- For any ordinal α there is a least ordinal larger than it
- Given any set of ordinals, there is a least ordinal bigger than all of them

We will use Hartog's Lemma to show that in fact, eventually we do run out of stuff in L if we use this indexing, which says that, for any set X, there are more ordinals than there are elements of X.

Note that here we are using the axiom of choice, because this is part of the maths tripos which uses the axiom of choice. Later on we, will think about what might happen if we don't have the axiom of choice, but we won't worry about that for now.

2 Ordinals

2.1 Functions and Relations

A ***function*** $f : X \to Y$ from a set X to a set Y is a subset $f \subset X \times Y$ such that for all $x \in X$ there is a unique $y \in Y$ with $(x, y) \in f$. We write $f(x) = y$ to mean $(x, y) \in f$. If $Z \subset X$, the ***restriction*** of f to Z is the function $f|_Z : Z \to Y$ given by $f|_Z = f \cap (Z \times Y)$.

If $f^{-1}(\{y\}) = \{x \in X : f(x) = y\}$ has at most one element for every $y \in Y$, we say f is an ***injection*** (i.e. no two x's get mapped to the same y). If $f^{-1}(\{y\})$ has at least one element for every $y \in Y$, we say f is a ***surjection*** (i.e. every y gets mapped to by some x). If f is both an injection and a surjection, it is called a ***bijection***.

A ***relation*** R on a set X is a subset $R \subset X \times X$. We write xRy to mean $(x, y) \in R$. If R, S are relations on sets X, Y respectively, an ***isomorphism*** from (X, R) to (Y, S) is a bijection $f : X \to Y$ such that, for all $x, y \in X, xRy \iff f(x)Sf(y)$. If such an isomorphism exists, we say that (X, R) and (Y, S) are ***isomorphic***.

Note: this is not the right way to think about functions and relations, but it is convenient sometimes. Keep thinking of a function as "something that associates a unique element of Y with each element of X". Note that this does give us a way to define functions and relations in such a way that they live in the universe of sets.

2.2 Well-Ordering

A ***total order*** on a set X is a relation $<$ on X satisfying:

- For all $x, y \in X$, precisely one of $x = y, x < y, y < x$ holds. (trichotomy)
- Fro all $x, y, z \in X$, if $x < y$ and $y < z$ then $x < z$. (transitivity)

If $<$ is a total order on X, we can define a relation $\leq$ on X by $x \leq y$ if $x < y$ or $x = y$. This satisfies:

- For all $x, y \in X, x \leq y$ or $y \leq x$.
- For all $x, y \in X$, if $x \leq y$ and $y \leq x$, then $x = y$.
- For all $x, y, z \in X$, if $x \leq y, y \leq z$, then $x \leq z$.

Conversely, given a relation $\leq$ satisfying these conditions, we can define a total order $<$ on X by $x < y$ if $x \leq y$ and $x \neq y$. Some sources will define $\leq$ to be a total order, but here we will use the $<$ version.

A total order of X is a ***well-ordering*** of X if every non-empty subset $Z \subset X$ has a least element (i.e. an element $y \in Z$ such that for all $x \in Z, y \leq x$.

An ***ordinal*** is a well-ordered set with isomorphic sets considered to be the same. Given a well-ordered set X, the ***order-type*** of X is the corresponding ordinal.

A total order $<$ is a well-ordering if and only if there is no infinite descending sequence $x_1 > x_2 > \ldots$.

Examples:

1. $X = \{a, b, c, d\}$ with $a < b, c, d; b < c, d; c < d$ well orders X.
2. The usual order on $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}$ defines a total order. $\mathbb{N}$ is well ordered by this, $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$ are not, as $-1 > -2 > -3 > \ldots$.
3. $x < y$ if $|x| - \frac{1+\operatorname{sgn} x}{4} < |y| - \frac{1+\operatorname{sgn} y}{4}$ is however a well ordering on $\mathbb{Z}$, with $0 < 1 < -1 < 2 < -2 < \ldots$.
4. Take $\{\frac{1}{n} : n \in \mathbb{N} \setminus \{0\}\}$ with the usual ordering. This is not a well ordering, as $1 > \frac{1}{2} > \frac{1}{3} > \ldots$.
5. Take $\{-\frac{1}{n} : n \in \mathbb{N} \setminus \{0\}\}$ with the usual ordering. This is a well ordering, as there are only finitely many elements less than any given element. This has the same order type as $\mathbb{N}$ with the usual order.
6. What about $X = \left\{-\frac{1}{n} : n \in \mathbb{N} \setminus \{0\}\right\} \cup \left\{1 - \frac{1}{n} : n \in \mathbb{N} \setminus \{0\}\right\}$ with the usual order. Even though there are infinitely many elements less than say $\frac{2}{3}$, we can see that any descending sequence can have only finitely many positive terms, before a negative term, and then only finitely many negative terms. Hence this is in fact a well-ordering, but it is not the same order-type as $\mathbb{N}$. This should remind you of the discussion at the end of section 1, where we went through infinitely many elements, and then did it again.

Let X be a well-ordered set. An ***initial segment*** of X is a subset $Z \subset X$ such that if $z \in Z$ and $x \in X$ with $x \leq z$ then $x \in Z$. If Y is also a well-ordered set, then we say $Y \leq X$ if Y is isomorphic to some initial segment of X. We make some remarks about this notion:

1. If $Z \neq X$ is an initial segment of X then $X \setminus Z \neq 0$ so has some least element x. Then $Z = \{y \in X : y < x\} = X_{(<x)}$.
2. Clearly $X \leq X$.
3. This also gives a definition of $\alpha \leq \beta$ for ordinals α, β. (We should really check that this is well-defined: i.e. if $Y \leq X$ and $Y' \cong Y, X' \cong X$, then $Y' \leq X'$. This is obvious though because the definition is given in terms of isomorphism, which is transitive).
4. Transitivity is clear, as $\alpha \leq \beta, \beta \leq \gamma \implies \alpha \leq \gamma$. What about trichotomy?

We will generalize two important concepts from $\mathbb{N}$ - the ideas of ***induction*** and ***recursion***

Proposition 2.1 (Proof by Induction)**.** *Let X be a well ordered set and let $P(x)$ be a statement about $x \in X$. Suppose for all $x \in X$ that $(\forall y < x, P(y)) \implies P(x)$. Moreover, suppose for $z = \min X$ that $P(z)$. Then $\forall x \in X, P(x)$.*

Proof. Suppose not. Then the set $\{x \in X : \neg P(x)\}$ is non-empty, so has a least element x, which is strictly greater than z. Then $\forall y < x, P(y)$ but $\neg P(x)$. $\lightning$ □

Proposition 2.2. *Let X be a well-ordered set and $f : X \to Z$ be an isomorphism from X to an initial segment Z of X. Then f is the identity function on X, and in particular $X \not\leq X$.*

Proof. We prove this by induction on x that $\forall x \in X, f(x) = x$. Indeed, suppose that $x \in X$ and $f(y) = y$ for all $y < x$. As f is injective, $f(x) \geq x$; if $f(x) > x$ then, as Z is an initial segment

and f is surjective onto Z we have $f(z) = x$ for some $z > x$, so $f(z) = x < f(x)$, but $x < z$. $\lightning$ □

What about recursion? We want to generalise things like "define $f : \mathbb{N} \to \mathbb{N}$ recursively by $f(0) = 1$ and $f(n) = \sum_{i=0}^{n-1} f(i)$ for $n > 0$".

Proposition 2.3 (Definition by Recursion)**.** *Let X be a well-ordered set, let Y be a set, and let $G : \mathcal{P}(X \times Y) \to Y$. Then there is a unique function $f : X \to Y$ s.t. $\forall x \in X, f(x) = G(f|_{X_{(<x)}})$.*

Proof. For uniqueness, suppose functions f, f' both have this property. We show by induction that $\forall x \in X, f(x) = f'(x)$. Indeed, suppose $x \in X$ and $f(y) = f'(y)$ for all $y < x$. Then $f(x) = G(f|_{x_{(<x)}}) = G(f'|_{(X_{(<x)}}) = f'(x)$.

For existence, define h is an ***attempt*** at f to mean $h : Z \to Y$ for some initial segment Z of X with $\forall x \in Z, h(x) = G(h|_{Z_{(<x)}})$. Suppose that $h : Z \to Y$ and $h' : Z' \to Y$ are attempts. Then $Z \subset Z'$ or $Z' \subset Z$, WLOG take the first. Then $h'|_Z$ is an attempt with domain Z, and so by uniqueness $h'|_Z = h$. So attempts agree at any points where they are both defined.

Next we will prove by induction that, $\forall x \in X$, there is some attempt h_x with $x \in \mathrm{dom}(h_x)$. Indeed, let $x \in X$ and suppose that, for all $y < x$, h_y is an attempt with $y \in \mathrm{dom}(h_y)$. Let $h = \bigcup_{y<x} h_y$. Then by what we've' just done, h is a function. Indeed, h is an attempt and $X_{(<x)} \subset \mathrm{dom}(h)$.

If $x \in \mathrm{dom}(h)$, set $h_x = h$. If $x \notin \mathrm{dom}(h)$ then $\mathrm{dom}(h) = X_{(<x)}$. Then we can set $h_x = h \cup \{(x, G(h|_{X_{(<x)}})\}$. Finally, let $f = \bigcup_{x \in X} h_x$. □

Proposition 2.4. *Let X, Y be well-ordered sets with $Y \not\leq X$, then $X \leq Y$.*

Proof. Define $f : X \to Y$ recursively by letting $f(x)$ be the least element of $Y \setminus \{f(w) : w \in X, w < x\}$. This set is nonempty as $Y \not\leq X$. Then f is an isomorphism from X to an initial segment of Y. □

Proposition 2.5 (Trichotomy)**.** *Let α, β be ordinals. Then $\alpha \leq \beta$ or $\beta \leq \alpha$, and if both hold then $\alpha = \beta$.*

Proof. The first part follows immediately from **2.4**. For the second part, suppose we have isomorphisms $f : \alpha \to B, g : \beta \to A$ where A, B are initial segments of α, β respectively. Then $f \circ g$ is an isomorphism from β to some initial segment of itself. Then by **2.2** $f \circ g$ is the identity function on β. Hence $B = \beta$, and so $\alpha \cong \beta$, so $\alpha = \beta$. □

Theorem 2.6. *Let α be an ordinal. Then the ordinals less than α form a set well ordered by $<$.*

Proof. The ordinals less than α are precisely the order types of the proper initial segments $\alpha_{(<x)}$ for $x \in \alpha$ of α. Then $\bar{\alpha} = \{\alpha_{(<x)} : x \in \alpha\}$ is isomorphic to α via $\alpha \to \bar{\alpha}, x \to \alpha_{(<x)}$. Thus $\bar{\alpha}$ is well-ordered. □

Note that we can write I_α for the set of ordinals less than α. Then I_α has order type α.

Corollary 2.7. *Any non-empty set X of ordinals has a least element.*

Proof. $X \neq \emptyset$ so there is some $\alpha \in X$. If α is the least element of X then we're done. Otherwise $X \cap I_\alpha \neq \emptyset$. But I_α is well-ordered so $X \cap I_\alpha$ has a least element, which must also be the least element of X. □

This would lead us to want to say something along the lines of "the ordinals are well-ordered by $<$". However:

Theorem 2.8 (Burali-Forti Paradox)**.** *The ordinals do not form a set.*

Proof. Suppose X is the set of ordinals. Then X is well ordered by $<$ so has order type α for some $\alpha \in X$. But I_α also has order type α, and this is a proper initial segment of X. So X is isomorphic to a proper initial segment of itself, contradicting **2.2**.↯ □

2.3 Ordinal Arithmetic

Let α, β be ordinals. We define $\alpha + \beta$ to be the set α followed by β, and $\alpha\beta$ to be α followed by itself β times.

More precisely, $\alpha + \beta$ is the order type of the set $(\{0\} \times \alpha) \cup (\{1\} \times \beta)$ ordered by $(i, x) < (j, y)$ if and only if $i < j$ or $(i = j$ and $x < y)$, whilst $\alpha\beta$ is the order type of the set $\alpha \times \beta$ ordered by $(x, y) < (z, y)$ if and only if $y < w$ or $(y = w$ and $x < z)$.

As an exercise, check that these two orderings are indeed well-orderings, so that these definitions make sense.

Suppose that, for each $n \in \mathbb{N}$, we identify n with the ordinal well-ordering on an n-element set. Then these definitions generalise arithmetic in $\mathbb{N}$.

$2 + 3 =$  $= 5$

$2 \cdot 3 =$  $= 6$

$+$ is associative immediately from the definitions. However, if ω is the order type of $\mathbb{N}$, then $1 + \omega = \omega$ under the isomorphism "subtract 1", however $\omega + 1$ would have ω as an initial segment, and so $1 + \omega \neq \omega + 1$. As an exercise (see sheet 2), think about whether multiplication is associative, commutative, or distributive.

2.4 Ordinal Arithmetic II - Attack of the Ordinals

There is an alternative approach, using ideas of induction and recursion. First, we will need some preliminary ideas.

Proposition 2.9.

1. *Let α be an ordinal. Then there is a least ordinal $\beta > \alpha$.*
2. *Let X be a set of ordinals. then there is a least ordinal γ with $\gamma > \delta$ or all $\delta \in X$.*

Proof.

1. Take β to be α with a single additional element added greater than everything in α.

2. Let $Y = \{I_\delta : \delta \in X\}$. Then X is the set of order types of elements of Y and the elements of Y are nested sets of ordinals. So take *gamma* to be the order type of $\bigcup_{\delta \in X} I_\delta$.

□

In *1.*, we write $\beta = \alpha^+$, the ***successor*** of α. In *2*, we write $\gamma = \sup(X)$, the ***supremum*** of X. If $\alpha = \beta^+$ for some ordinal β, we say α is a ***successor ordinal***, and otherwise α is a ***limit ordinal***. For example, 0 is a limit, $n \in \mathbb{N}$ is a successor, ω is a limit. Note that α is a limit if and only if $\alpha = \sup(I_\alpha)$.

When doing induction or recursion for ordinals, we often deal with successors and limits in separate cases. We often treat 0 as different to other limits as well.

We can then define addition and multiplication for ordinals recursively as follows:

- $\alpha + 0 = \alpha$
- $\alpha + \beta^+ = (\alpha + \beta)^+$
- $\alpha + \gamma = \sup\{\alpha + \delta : \delta < \gamma\}$
- $\alpha \cdot 0 = 0$
- $\alpha \cdot \beta^+ = \alpha \cdot \beta + \alpha$
- $\alpha \cdot \gamma = \sup\{\alpha\delta : \delta < \gamma\}$.

However we have a problem - recursion doesn't work here since the ordinals do not form a set. We can get around this though.

Let α be an ordinal and suppose we want to define $\alpha + \beta$ for all ordinals β. Given an ordinal γ we can define $\alpha + \beta$ recursively for all $\beta \in I_\gamma$ as above, since I_γ is a set. Suppose we restrict this definition to $\beta \in I_{\gamma'}$ for some $\gamma' < \gamma$. Then this is the same as the direct definition on $I_{\gamma'}$. Hence we have uniquely defined $\alpha + \beta$ for all β, since $\beta \in I_{\beta^+}$.

What about induction? Suppose we have some statement $p(\alpha)$ about an ordinal α such that for all α, $(\forall \beta\alpha, p(\beta)) \implies p(\alpha)$. Take an ordinal α. By induction, for all $\gamma \in I_{\alpha^+}, p(\gamma)$. Hence $p(\alpha)$. So in fact $p(\alpha)$ holds for all ordinals α.

Lemma 2.10. *Let α, β, γ be ordinals. Then*

1. $\beta \leq \gamma \implies \alpha + \beta \leq \alpha + \gamma$
2. $\beta < \gamma \implies \alpha + \beta < \alpha + \gamma$
3. *γ a non zero limit ordinal* $\implies$ *$\alpha + \gamma$ a limit ordinal*

Proof.

1. We do this by induction on γ. If $\beta = \gamma$ then it is obvious, so assume $\beta < \gamma$. For $\gamma = 0$, there is no such β, so done. Otherwise, if $\gamma = \delta^+, \beta < \delta^+ \implies \beta \leq \delta \implies \alpha + \beta \leq \alpha + \delta < (\alpha + \delta)^+ = \alpha + \delta^+ = \alpha + \gamma$. The final case is γ is a non-zero limit. Then $\beta < \gamma \implies \alpha + \gamma = \sup\{\alpha + \delta : \delta < \gamma\} \geq \alpha + \beta$.
2. Suppose $\beta < \gamma$. Then $\beta^+ \leq \gamma$, so $\alpha + \beta < (\alpha + \beta)^+ = \alpha + \beta^+ \leq \alpha + \gamma$.

3. Suppose instead that $\alpha+\gamma = \delta^+$ for some δ. Then $\delta < \alpha+\gamma = \sup\{\alpha+\epsilon : \epsilon < \gamma\}$, so $\delta < \alpha+\epsilon$ for some $\epsilon < \gamma$. As γ is a limit, then $\epsilon^+ < \gamma$. Then $\alpha+\gamma = \delta^+ \leq (\alpha+\epsilon)^+ = \alpha+\epsilon^+ < \alpha+\gamma$ by *2.*. $\lightning$

□

Proposition 2.11. *Let α, β, γ be ordinals. Then $(\alpha+\beta)+\gamma = \alpha+(\beta+\gamma)$*

Proof. We use induction on γ. In the case $\gamma = 0, (\alpha+\beta)+0 = \alpha+\beta = \alpha+(\beta+0)$.

If $\gamma = \delta^+$, then $(\alpha+\beta)+\delta^+ = ((\alpha+\beta)+\delta)^+ = (\alpha+(\beta+\delta))^+ = \alpha+(\beta+\delta)^+ = \alpha+(\beta+\delta^+)$.

If γ is a non-zero limit, then $(\alpha+\beta)+\gamma = \sup(X)$ where $X = \{(\alpha+\beta)+\delta : \delta < \gamma\} = \{\alpha+(\beta+\delta) : \delta < \gamma\}$. Meanwhile, by **2.10** part 3, $\beta+\gamma$ is a limit, and so $\alpha+(\beta+\gamma) = \sup(Y)$ where $Y = \{\alpha+\epsilon : \epsilon < \beta+\gamma\}$. If $\delta < \gamma$ then by **2.10** part 2, $\beta+\delta < \beta+\gamma$ so $X \subset Y$, and so $(\alpha+\beta)+\gamma \leq \alpha+(\beta+\gamma$. On the other hand, if $\zeta \in Y$, then $\zeta = \alpha+\epsilon$ for some $\epsilon < \beta+\gamma = \sup\{\beta+\delta : \delta < \gamma\}$. Hence $\epsilon \leq \beta+\delta$ for some $\delta < \gamma$. Let $\eta = \alpha+(\beta+\delta)$. Then **2.10** part 1, $\eta \geq \zeta$ and also $\eta \in X$. Hence $\sup(X) \geq \sup(Y)$, so $(\alpha+\beta)+\gamma \geq \alpha+(\beta+\gamma)$, and so $(\alpha+\beta)+\gamma = \alpha+(\beta+\gamma)$. □

The definition in this section is the ***inductive*** definition of arithmetic, whilst the one in the previous section was the ***synthetic*** definition. It is often better to work with the synthetic definition than the inductive one. Fortunately however, they coincide.

Proposition 2.12. *The synthetic and inductive definitions of ordinal addition coincide.*

Proof. Let $+$ denote the synthetic definition. Then for all α, $\alpha+0 = \alpha$, and also $\alpha+\beta^+ = \leftarrow \alpha \rightarrow \leftarrow \beta \rightarrow \cdot = (\alpha+\beta)^+$

Finally, if α is an ordinal and γ is a non-zero limit ordinal, then $\alpha+\gamma = \leftarrow \alpha \rightarrow \leftarrow \sup\{\delta : \delta < \gamma\} \rightarrow = \sup(\alpha+\delta : \delta < \gamma\}$, by noting that δ is the order type of I_δ, and the I_δ are nested so $\sup \leftrightarrow \cup$.

Hence by induction on β the definitions agree on $\alpha+\beta$. □

Proposition 2.13. *The synthetic and inductive definitions of ordinal multiplication coincide.*

Proof. Left as an exercise for sheet 2. □

Can we define ordinal exponentiation? It's clear what to do inductively, but not so much synthetically. We define:

- $\alpha^0 = 0^+$
- $\alpha^{\delta^+} = \alpha^\delta \alpha$
- α^β for β a non-zero limit to be $\sup\{\alpha^\delta : \delta < \beta\}$.

2.5 Uncountable Ordinals

In **§1.4** our list of indices $0, 1, 2, \ldots, \omega, \ldots, \omega 2, \ldots, \epsilon_0$ are some ordinals. We've found lots of them, but in fact only countably many - $\epsilon_0 = \sup\{\omega, \omega^\omega, \ldots\}$. Since α is the order type of I_α, equivalently each of these ordinals is countable. However, there are more:

Proposition 2.14. *There exists an uncountable ordinal.*

Proof. Let $X \subset \mathcal{P}(\mathbb{N} \times \mathbb{N})$ be the set of all well-orderings of subsets of $\mathbb{N}$. Let $Y = \{\text{ord}(x) : x \in X\}$. Then Y is precisely the set o countable ordinals. Let $\omega_1 = \sup(Y)$. Suppose ω_1 is countable. Then $\omega_1 \in Y$ and the proper initial segment $Y_{(<\omega_1)}$ has the same order type as Y. $\lightning$ □

A couple of remarks:

- ω_1 is the least uncountable ordinal.
- The ordinal ω_1 is not the supremum of any countably set of countable ordinals. Indeed, if Z is a countable set of countable ordinals, then $\sup Z$ is the order-type of $\bigcup_{\alpha \in Z} I_\alpha$, a countable union of countable sets.

Theorem 2.15. *Let S be a set. Then there exists an ordinal γ with no injection $\gamma \to S$.*

Proof. Same as the previous proposition with $\mathbb{N}$ replaced by S. □

Theorem **2.15** is called ***Hartog's Lemma***. This makes sense of the statement in **§1.4** that we eventually run out of stuff, with a formal proof to follow in the next section.

3 Posets and Zorn's Lemma

A ***partially ordered set (poset)*** is a set P with a relation $\leq$ satisfying:

1. $(x \leq y, y \leq z) \implies x \leq z$
2. $\forall x \in P, x \leq x$
3. $(x \leq y, y \leq x) \implies y = x$

If $(P, \leq)$ is a poset and we define $x < y$ to mean $x \leq y$ and $x \neq y$, then

1. $(x < y, y < z) \implies x < z$
2. $x \not< x$

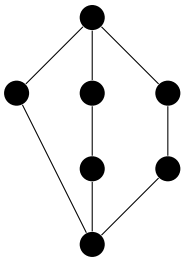
Conversely, we can check that if P is a set and $<$ is a relation on P satisfying the above then $\leq$ gives a partial order on P. We use $>, \geq$ in the obvious way.

A ***Hasse diagram*** of $(P, \leq)$ consists of a point for each $x \in P$ and line upwards from x to y if y ***covers*** x, i.e. that $y > x$ and there is no z with $y > z > x$.

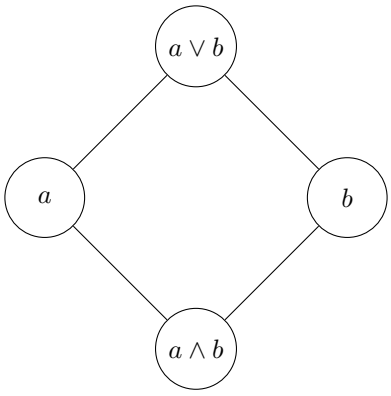
Examples

1. Any totally ordered set is a poset.
2. Any subset of a poset is a poset.
3. For any set X, $\mathcal{P}(X)$ is a poset under inclusion.

4. In a similar vein, take $\mathbb{N}$ with $|$ (divides). This is a poset.
5. We have the finite set with Hasse diagram given by:

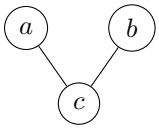


Let $(P, \leq)$ be a poset. We say that $x \in P$ is ***greatest*** if for all $y \in P, y \leq x$. We say that $x \in P$ is ***maximal*** if for all $y \in P$, $y \geq x \implies y = x$. If $S \subset P$, an ***upper bound*** for S is an $x \in P$ such that $y \in S \implies y \leq x$. We can similarly define ***least***, ***minimal***, and ***lower bound***. If S has a least upper bound we denote it by $\sup(S)$ or by $\bigvee S$, read "join S". Similarly, if we have a greatest lower bound or infimum of $S \subseteq P$, we denote it by $\inf(S)$ or $\bigwedge S$, read "meet S". We have the diagram:



Examples:

1. In $(\mathcal{P}S, \subseteq)$, S is greatest and $\emptyset$ is least.
2. In $(\mathbb{N}, \leq)$, 0 is least. There is no greatest or maximal element.
3. Take the finite set with diagram given by:



We have a is maximal, but a is not greatest since $a \not\geq b$. b is maximal as well. c is both minimal and least.

Note that a greatest implies that a is maximal, and if P has a greatest element then it must be unique. If every subset of P has a supremum then we say that P is ***complete***. If P is complete then P is non-empty, as $\bigvee \emptyset \in P$. In fact, $\bigvee \emptyset$ is the least element of P. P also has a greatest element, $\bigvee P$.

Examples:

1. $(\mathcal{P}S, \subseteq)$ is complete. $\bigvee \mathcal{A} = \bigcup_{A \in \mathcal{A}} A, \bigwedge \mathcal{A} = \bigcap_{A \in \mathcal{A}} A$.
2. $([0,1], \leq)$ is complete.

3. $(\mathbb{R}, \leq)$ is not complete.

If $(P, \subseteq)$ is a poset then a ***chain*** is a subset $C \subseteq P$ such that for all $x, y \in C$ either $x \leq y$ or $y \leq x$. If every chain has a supremum, we say that P is ***chain-complete.*** Note again that chain complete implies nonempty since the empty set is a chain.

Examples:

1. In $(\mathbb{N}, 1)$, the set $\{4^n : n \in \mathbb{N}\}$ is a chain.
2. In $(\mathbb{R}, \leq)$, $\mathbb{Q}$ is a chain, and so is $\mathbb{R}$.

A ***lattice*** L is a poset in which every finite subset has both a supremum and an infimum. For $a, b \in L$, we write $a \vee b = \bigvee\{a, b\}$ and $a \wedge b = \bigwedge\{a, b\}$. Note that L has a least element $\bigwedge \emptyset$ and a greatest element $\bigvee \emptyset$. We say that a lattice L is a ***Boolean algebra*** if:

1. For all $a, b, c \in L$, $a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c)$
2. For all $a, b, c \in L$, $a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c)$
3. For all $a \in L$ there is $b \in L$, $a \vee b = \bigwedge \emptyset, a \wedge b = \bigvee \emptyset$

These rules are essentially saying L behaves like a powerset, with $\wedge$ corresponding to $\cap$ and $\vee$ corresponding to $\cup$. Then the first two rules are de Morgan's laws, and the third is the notion of a complement.

3.1 Zorn's Lemma

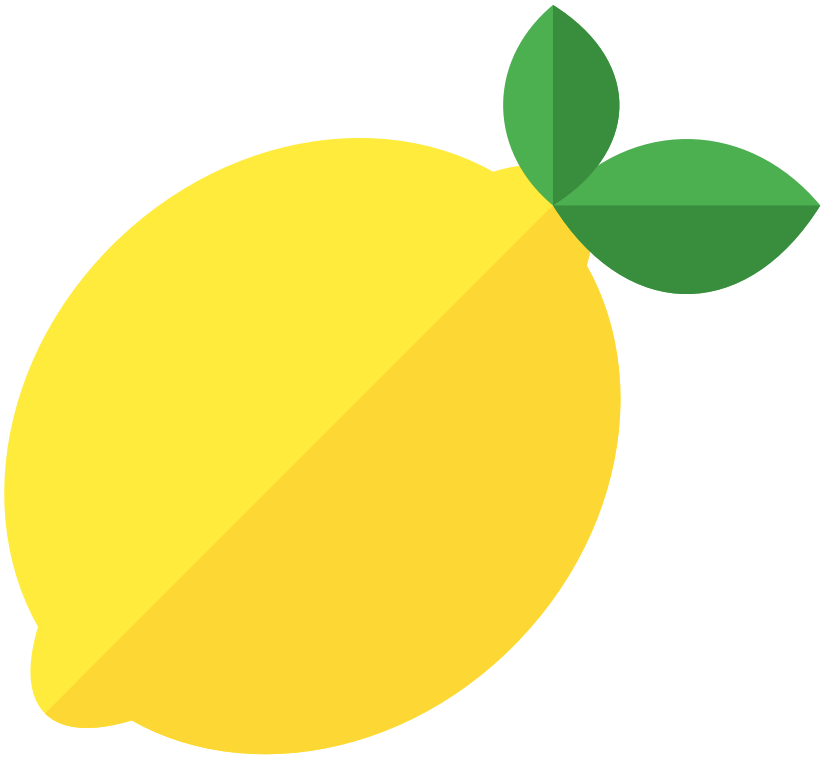


Figure 1: Zorn's Lemon

Theorem 3.1 (Zorn's Lemma)**.** *Let $(P, \leq)$ be a poset in which every chain has an upper bound. Then P has a maximal element.*

The idea is that, if not, we can take x_0 to be the upper bound of the empty set, then if x_0 is not maximal there is $x_1 > x_0$, and then we get another chain indexed by the ordinals, and eventually run out of stuff.

Proof. Suppose P has no maximal element. Then for any $x \in P$, we can find $x' \in P$ with $x' > x$. Moreover, for any chain $C \subseteq P$ we can find $C^* \in P$ such that for all $x \in C$, $C^* > x$, for instance take y to be an upper bound for C and let $C^* = y'$.

By Hartog's Lemma, we can find an ordinal γ such that there is no injection from $\gamma \to P$.

Recursively for $\alpha < \gamma$ define x_α as follows:

$\alpha = 0$: $x_0 = \emptyset^*$ - we can use any element of P

$\alpha = \beta^+$: $x_{\beta^+} = x'_\beta$

Otherwise: α is a non-zero limit, so let $x_\alpha = \{x_\delta : \delta < \alpha\}^*$.

We can do this because $\{x_\delta : \delta < \alpha\}$ is a chain as $x_\zeta > x_\xi \iff \zeta > \xi$, and the ordinals are totally ordered.

Then $\alpha \mapsto x_\alpha$ is an injection $I_\gamma \to P$, and $I_\gamma \cong \gamma$. $\lightning$ $\square$

But can we actually let $x_\alpha = \{x_\delta : \delta < \alpha\}^*$? Even though we know the thing on the right is a chain, we don't know that until we're halfway through the recursion - when we define what x_α will be, we are saying its the star of something that may or may not be a chain. It is only once we find out what this something is that we know it is indeed a chain and we can take a star of it, and this first assignment is slightly dodgy. To get around this, we use the technical fix of cabbages:

Cabbages. Take something not in P, for instance a cabbage that we happen to have lying around. For $\alpha < \gamma$, define $x_\alpha \in P \cup \{\text{cabbage}\}$ exactly as above, except for α a non-zero limit.

Then set for α a non-zero limit, let $x_\alpha = \begin{cases} \{x_\delta : \delta < \alpha\}^* & \text{this set is a chain} \\ \text{cabbage} & \text{otherwise} \end{cases}$, and also define cabbage$'$ = cabbage. By induction, we don't get any cabbage. In these proofs, cabbages are implicitly assumed.

3.2 Applications of Zorn's Lemma

The general method of applying Zorn's Lemma is as follows:

1. Define a suitable poset.
2. Check every chain has an upper bound.
3. Use Zorn to find a maximal element x.
4. Check x is the thing we are looking for.

It is not unusual for most of the work to come in step 4.

Corollary 3.2. *Let P be a set of primitive propositions, and let $S \subset L = L(P)$ be consistent. Then there is a consistent $\bar{S} \subset L$ with $S \subset \bar{S}$ and for all $p \in L$, $p \in \bar{S}$ or $\neg p \in \bar{S}$.*

Proof. Let X be the poset $\{R \subset L | R \text{ consistent and } S \subset R\}$, ordered by $\subset$. Let $\mathscr{C} \subset X$ be a chain. If $\mathscr{C} = \emptyset$ then $S \in X$ is an upper bound for $\mathscr{C}$.

Suppose $\mathscr{C} \neq \emptyset$. Let $Q = \bigcup_{R \in \mathscr{C}} R$. Then for all $R \in \mathscr{C}, R \subset Q$. As $\mathscr{C} \neq \emptyset$, there exists $R \in \mathscr{C}$ and $S \subset R \subset Q$. Suppose that Q is inconsistent, i.e. that $Q \vdash \bot$. As proofs are finite, there are some $q_1, \ldots, q_n \in Q$ such that $\{q_1, \ldots, q_n\} \vdash \bot$. For each i, pick $R_i \in \mathscr{C}$ with $q_i \in R_i$. As $\mathscr{C}$ is a chain and there are only finitely many of $R_1, \ldots, R_n$, there exists j such that $R_i \subseteq R_j$ for all i. Then $q_1, \ldots, q_n \in R_j$, and so $R_j \vdash \bot$. $\lightning$

Hence Q is consistent. Thus $Q \in X$ is an upper bound for $\mathscr{C}$. By Zorn's lemma, X has a maximal element $\bar{S}$. Let $p \in L$ and suppose $p \notin \bar{S}$. Then $\bar{S} \cup \{p\}$ is inconsistent, as otherwise $\bar{S} \cup \{p\} \in X$ with $\bar{S} \subsetneq \bar{S} \cup \{p\}$. That is, $\bar{S} \cup \{p\} \vdash \bot$, and so by the deduction theorem $\bar{S} \vdash (p \implies \bot)$, i.e. $\bar{S} \vdash \neg p$. So, similarly to the above, $\neq p \in \bar{S}$. $\square$

This fills the gap in the proof of the Model Existence Lemma, thus finishing the proof of the completeness theorem. It is very common when using Zorn to consider a poset of the form $(X, \subseteq)$ where $X \subseteq \mathcal{P}(A)$ for some set A. If $\mathscr{C}$ is a chain in X, an obvious guess for an upper bound for $\mathscr{C}$ is $Q = \bigcup_{R \in \mathscr{C}} R$. Clearly for all $R \in \mathscr{C}$, $R \subseteq Q$, but often work is needed to show that $Q \in X$,

so that Q is genuinely an upper bound for $\mathscr{C}$ in X. Also, care must be taken to remember that $\emptyset$ si a chain, and often needs t be considered separately.

Corollary 3.3. *Every vector space has a basis.*

Proof. Let V be a vector space over a field k. Let $P = \{S \subseteq V | S \text{ is linearly independent}\}$, ordered by $\subseteq$. Let $\mathscr{C} \subseteq P$ be a chain, and let $R = \bigcup_{S \in \mathscr{C}} S$. Suppose that R is linearly dependent. Then there are some $n \in \mathbb{N} \setminus \{0\}$, and some distinct $v_1, \dots, v_n \in R$ and some $\lambda_1, \dots, \lambda_n \in k$ with $\sum_{i=1}^n \lambda_i v_i = 0$. For each i, pick $S_i \in \mathscr{C}$ with $v_i \in S_i$. As $\mathscr{C}$ is a chain, there is some j for which $S_i \subseteq S_j$ for all $i = 1, \dots, n$. Then $v_1, \dots, v_n \in S_j$, so S_j is linearly dependent. $\lightning$

Hence R is linearly independent, and so R is an upper bound for $\mathscr{C}$. By Zorn, P has a maximal element B. Certainly B is linearly independent. Suppose there is some $v \in V \setminus \operatorname{span} B$. We show that $B \cup \{v\}$ is linearly independent. Indeed, if not then we have a linear combination $\mu v - \sum_{i=1}^n \lambda_i b_i = 0$, where $b_i \in B$, but then $v = \sum \mu^{-1} \lambda_i b_i \in \operatorname{span} B \lightning$. But then $B \cup \{v\} \in P$, and $B \subsetneq B \cup \{v\}$, $\lightning$ maximality of B.

So B is a linearly independent set spanning V, and so B is a basis for V. □

Corollary 3.4 (Bourbaki-Witt Theorem)**.** *Let $(X, \leq)$ be a poset, and let X be chain-complete, and $f : X \to X$ be such that for all $x \in X, f(x) \geq x$. Then f has a fixed point.*

Proof. By Zorn X has a maximal element x. Then $f(x) \in X, f(x) \geq x$, so by maximality $f(x) = x$. □

Corollary 3.5 (Well-Ordering Principle)**.** *Every set can be well-ordered.*

Proof. Let X be a set. Let $P = \{(Y, R) | Y \subset X \text{ and } R \text{ is a well-ordering of } Y\}$. Define $(Y, R) \leq (Z, S)$ if $Y \subset Z$ and R, S agree on Y, and Y is an initial segment of Z in S. Then $(P, \leq)$ is a poset. Let $\mathscr{C} \subset P$ be a chain, say $\mathscr{C} = \{(Y_i, R_i) | i \in I\}$. Let $Y = \bigcup_{I \in I} Y_i$. Define a relation R on Y as follows: if $x, y \in Y$ then choose $j, k \in I$ such that $x \in Y_j$ and $y \in Y_k$. If $Y_j \subset Y_K$ then $xRy \iff xR_k y$, otherwise $Y_k \subset Y_j$ in which case $xRy \iff xR_j y$.

It is easy to show that R is a total ordering of Y. We must show further that it is a well ordering. Let $A \subset Y$ with $A \neq \emptyset$. Then there exists $x \in A$, so $x \in Y_\ell$ for some $\ell \in I$, so $A \cap Y_\ell \neq \emptyset$. But we know Y_ℓ is well ordered by R_ℓ, so $A \cap Y_\ell$ has an R_ℓ-least element, z, say. So z is also the R-least element of $A \cap Y_\ell$, and hence R is a well ordering of Y.

Thus, $(Y, R) \in P$ and is an upper bound for $\mathscr{C}$. Then by Zorn, P has a maximal element (Z, S). Suppose there is some $x \in X \setminus Z$. Then $S \cup \{(z, x) | z \in Z \cup \{x\}\}$ is a well ordering of $Z \cup \{x\} \supsetneq Z$, and Z is an initial segment. $\lightning$

Hence $Z = X$ and S is a well ordering of X. □

3.3 The Axiom of Choice

In the proof of Zorn's Lemma (ZL), we had to make infinitely many arbitrary choices, e.g. for all $x \in X$, pick some $x' > x$. We've done this quite a bit throughout tripos, for instance in IA Numbers and Sets, when we proved that a countable union of countable sets is countable by picking an ordering of each set so that we can move along the list diagonally.

Axiom 3.6 (The Axiom of Choice)**.** *Let $\{X_i | i \in I\}$ be a set of non-empty sets. Then there exists a function $f : I \to \bigcup_{i \in I} X_i$ such that, for all $i \in I$, $f(i) \in X_i$.*

The function f is sometimes called a ***choice function*** for the collection $\{X_i | i \in I\}$.

The axiom of choice (AC) has a different feel from our other usual assumptions about sets, e.g. it doesn't explicitly construct anything, so it can be of interest to see if we've used it. For example we used AC when proving Zorn's Lemma, but it might be an interesting question to ask if we needed to. What does this even mean? To give a sensible answer, we'd need to axiomatise set theory, and think about proof systems, and then show that if we only assume axioms other than AC then Zorn's Lemma would no longer be a theorem. This would take a lot of work (c.f. Russell and Whitehead, Principia Mathematica). For now, we will do a similar thing - assume Zorn's Lemma and use it to prove AC.

Proof. Let $\{X_i | i \in I\}$ be a collection of non-empty sets, and let

$$P = \left\{ g \subset \mathcal{P}\left(I \times \bigcup_{i \in I} X_i \right) \middle| g : J \to \bigcup_{i \in I} X_i, J \subseteq I; \forall i \in \mathrm{dom}(g), g(i) \in X_i \right\}$$

ordered by inclusion. Not $g \subset g'$ means $\mathrm{dom}\, g \subset \mathrm{dom}\, g'$ and $g'|_{\mathrm{dom}\, g} = g$. Let $\mathscr{C}$ be a chain in P. Let $h = \bigcup_{g \in \mathscr{C}} g$. Then h is a function with $\mathrm{dom}\, h = \bigcup_{g \in \mathscr{C}} \mathrm{dom}\, g$.

Moreover, $h \in P$ and is an upper bound for $\mathscr{C}$. So by Zorn P has a maximal element f. Suppose $i \in I \setminus \mathrm{dom}\, f$. Pick $x \in X_i$[1]. Then $f \subsetneq f \cup \{(i, x)\} \in P$. $\lightning$ So f is our required choice function. □

However, in the case where the index set I is finite, AC *does* follow from our assumptions. For instance, if $I = \{a, b, c\}$, then if we have nonempty sets X_a, X_b, X_c, we can, by definition write $x_a \in X_a, x_b \in X_b, x_c \in X_c$. Then we can write down our choice function $= \{(a, x_a), (b, x_b), (c, x_c)\}$.

Another way of thinking of the axiom of choice is that, if $X_i \neq \emptyset$ for all $i \in I$, then the Cartesian product $\prod_{i \in I} X_i \neq \emptyset$.

Do we need AC to prove the well ordering principle? Note that we can prove the axiom of choice from the well ordering principle by well ordering $\bigcup_i X_i$ and choosing the least element of X_i.

Bourbaki-Witt does not need the axiom of choice though.

4 Cardinals

Let X and Y be sets. If there is a bijection $X \to Y$ we say X, Y ***have the same cardinality***, and write $|X| = |Y|$.

Note that $|X| = |X|; |X| = |Y| \implies |Y| = |X|$, and $(|X| = |Y|, |Y| = |Z|) \implies |X| = |Z|$.

We want to define cardinals in such a way that every set X has the same cardinality as precisely one cardinal. By WOP every set X can be well-ordered, so X has the same cardinality as some ordinal.

[1] We can do this without using AC since we are only making one choice, and that is precisely what it means for a set to be nonempty - we can pick an element from it

We then define the ***cardinality*** $\text{card}(X)$ to be the least ordinal with the same cardinality as X.

If X, Y are sets, then we write $|X| \leq |Y|$ to mean there is an injection from X to Y. Clearly $|X| \leq |X|$, and $(|X| \leq |Y|, |Y| \leq |Z|) \implies |X| \leq |Z|$. Moreover, by Cantor-Schröder-Bernstein, if $|X| \leq |Y|$ and $|Y| \leq |X|$, then $|X| = |Y|$.

But then the question is what are the cardinals? We define an ordinal α to be ***initial*** if for all $\beta\alpha$ there is no bijection $\beta \to \alpha$. For example, 0,1,2,3,...are all initial ordinals. ω, ω_1 are also initial ordinals. However, no ordinal α with $\omega < \alpha < \omega_1$ is initial.

For α an ordinal, we can define ω_α recursively:

$\alpha = 0$: $\omega_0 = \omega$

$\alpha = \beta^+$: ω_{β^+} is the least ordinal with no injection to ω_β which exists by Hartog's lemma.

α limit: $\omega_\alpha = \sup\{\omega_\delta : \delta < \alpha\}$.

For every α, ω_α is an infinite initial ordinal. We can also prove the converse: it is easy to show by induction that, for all $\alpha, \omega_\alpha \geq \alpha$.

Now let α be an infinite initial ordinal. Then it is easy to check that, if δ is least with $\omega_\delta \geq \alpha$, then $\alpha = \omega_\delta$.

We then write $\aleph_\alpha = \text{card}(\omega_\alpha)$, so that the cardinals are $0, 1, 2, \ldots$ and the infinite cardinals, the $\aleph_\alpha$ for α ordinals. We have two different notations for the same thing because sometimes we want to think about it as a cardinal and sometimes as an ordinal. The important thing about cardinals is that there is precisely one of each cardinality - the fact that we've defined them as ordinals to make this work is not important.

4.1 Cardinal Arithmetic

Let κ, λ be cardinals. We define:

- $\kappa + \lambda = \text{card}(\{0\} \times \kappa \cup \{1\} \times \lambda\}$
- $\kappa\lambda = \text{card}(\kappa \times \lambda)$
- $\kappa^\lambda = \text{card}(\{f | f : \lambda \to \kappa\})$

Warning: this is not the same definition of exponentiation as for ordinals, as for instance $2^{\omega_0} = \omega_0$, but $2^{\aleph_0} = \text{card}(\mathcal{P}(\mathbb{N})) \neq \text{card}(\mathbb{N}) = \aleph_0$.

Proposition 4.1.

1. $(\kappa + \lambda) + \mu = \kappa + (\lambda + \mu)$
2. $\kappa + \lambda = \lambda + \kappa$
3. $(\kappa\lambda)\mu = \kappa(\lambda\mu)$
4. $\kappa\lambda = \lambda\kappa$
5. $\kappa(\lambda + \mu) = \kappa\lambda + \kappa\mu$
6. $(\kappa\lambda)^\mu = \kappa^\mu\lambda^\mu$
7. $\kappa^{\lambda+\mu} = \kappa^\lambda\kappa^\mu$

8. $(\kappa^\lambda)^\mu = \kappa^{\lambda\mu}$.

This is left as an exercise, but we'll illustrate an example here. For, say, *5.*, we need to find a bijection:

$$f : \kappa \times ((\{0\} \times \lambda) \cup (\{1\} \times \mu)) \to (\{0\} \times (\kappa \times \lambda)) \cup (\{1\} \times (\kappa \times \mu))$$

So take $f(x,(n,y)) = (n,(x,y))$ - this does the job.

In fact, addition and multiplication are easy because:

Proposition 4.2. *Let κ be an infinite cardinal. Then $\kappa\kappa = \kappa$.*

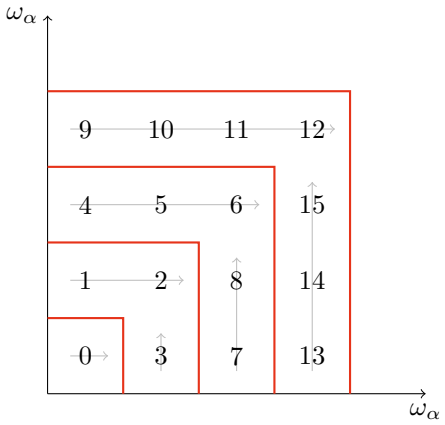
Proof. We have $\kappa = \aleph_\alpha$ for some ordinal α. Think of κ as the initial ordinal ω_α, so it is well-ordered.

We then need to show that $\omega_\alpha \times \omega_\alpha$.

Well order $\omega_\alpha \times \omega_\alpha$ by $(x,y) < (z,w)$ if:

- $\max(x,y) < \max(z,w)$, or
- $y = w \geq x, z$ and $x < z$, or
- $x = z > y, w$ and $y < w$, or
- $y = z, y \geq x, z > w$.

This looks like a mess, but we can think of it more clearly with a picture:



In this well-ordering, what can we say about the order-type of $\omega_\alpha \times \omega_\alpha$?

Consider a proper initial segment $(\omega_\alpha \times \omega_\alpha)_{(<\delta)}$. As ω_α is initial and hence a limit, $\delta = \beta \times \beta$ for some $\beta < \omega_\alpha$. Then, writing $I = (\omega_\alpha \times \omega_\alpha)_{(<\delta)}$, in fact $I \subset \beta \times \beta$. As ω_α is initial, $\mathrm{card}(\beta) < \mathrm{card}(\omega_\alpha)$, so either β is finite, in which case $\beta \times \beta$ is finite, or β is infinite, in which case by the inductive hypothesis, $\mathrm{card}(\beta \times \beta) = \mathrm{card}(\beta)$.

In either case, $\mathrm{card}(\beta \times \beta) < \aleph_\alpha$, and so $\mathrm{ord}(I) < \omega_\alpha$. Hence $\mathrm{card}(\omega_\alpha \times \omega_\alpha) \leq \mathrm{card}(\omega_\alpha)$, so $\kappa\kappa \leq \kappa$. But clearly $\kappa \leq \kappa\kappa$, and so $\kappa\kappa = \kappa$ as claimed. □

Corollary 4.3. *Let κ, λ be infinite cardinals. Then:*

$$\kappa + \lambda = \kappa\lambda = \max(\kappa, \lambda)$$

Proof. WLOG $\kappa \geq \lambda$. Then:

$$\kappa \leq \kappa + \lambda \leq \kappa + \kappa = 2\kappa \leq \kappa\lambda \leq \kappa\kappa = \kappa$$

□

Examples:

1. $\operatorname{card} \mathbb{R} = \operatorname{card} \mathcal{P}(\mathbb{N}) = 2^{\alpha_0}$.
2. Let X be the set of real sequences. What is $\operatorname{card}(X)$? Well, $\operatorname{card}(X) = \operatorname{card}(\mathbb{R}^{\mathbb{N}}) = (2^{\aleph_0})^{\aleph_0} = 2^{\aleph_0 \aleph_0} = 2^{\aleph_0}$, i.e. there is a bijection from the set of the reals to the set of real sequences.

What about exponentiation? This turns out to be HARD. What is $2^{\aleph_0}$? We know that $2^{\aleph_0} = \aleph_\alpha$ for some $\alpha \geq 1$. We could guess that $2^{\aleph_0} = \alpha_1$, but it is impossible to prove one way or the other, even using the axiom of choice as we have been. We cannot prove that $\mathbb{R}$ has an uncountable subset that doesn't biject with $\mathbb{R}$.

One thing that is true: for any $\kappa, 2^\kappa > \kappa$. This is left as an exercise - show there is no bijection $\mathcal{P}(X) \to X$ for any X.

4.2 A Slight Remark on the Axiom of Choice

This whole section has been heavily dependent on AC. We even needed AC to define what the cardinals are! In fact, without AC, we can still define a nice definition of cardinals and define arithmetic, with the basic properties still holding. However, we can no longer prove:

- Every infinite cardinal is $\aleph_\alpha$ for some ordinal α
- **4.2** and **4.3** only work for the $\aleph$ cardinals, but not necessarily for the other cardinals that might exist.
- For cardinals κ, λ we have either $\kappa \leq \lambda$ or $\lambda \leq \kappa$.

We can still prove that $(\kappa \leq \lambda, \lambda \leq \kappa) \implies \kappa = \lambda$.

5 First Order Predicate Logic

Keep the following two examples in mind:

- The Theory Of Groups (GT): A group is a set G with multiplication m, inverse i, and identity e satisfying certain axioms. We can think of m, i, e as functions. $m : G^2 \to G, i : G \to G, e : G^0 \to G$.
- The Theory Of Posets (PT): A poset is a set P with a relation $\leq$ satisfying certain axioms. We have $\leq \subset P^2$.

A ***first order signature*** is an ordered triple $\Sigma = (\Omega, \Pi, \alpha)$ where Ω and Π are disjoint sets of symbols with $w, ', (,), =, \implies, \bot, \forall \notin \Omega \cup \Pi$ and $\alpha : \Omega \cup \Pi \to \mathbb{N}$. We call Ω the set of ***function symbols***, Π the set of ***relation symbols***, and α the ***arity function***.

Examples:

GT: $\Omega = \{m, i, e\}, \Pi = \emptyset, \alpha(m) = 2, \alpha(i) = 1, \alpha(e) = 0$.

PT: $\Omega = \emptyset, \Pi = \{\leq\}, \alpha(\leq) = 2$.

We then define the ***first-order language*** $L = L(\Sigma) = L(\Omega, \Pi, \alpha)$ by a sequence of inductive definitions.

The ***variables*** are defined inductively by:

- w is a variable.
- x is a variable, so x' is a variable.

So the variables are $w, w', w'', w''', \ldots$. Informally, we often use $w_0, w_1, \ldots$, or $x, y, z, \ldots$.

The ***terms*** are defined inductively by:

- Every variable is a term.
- If $f \in \Omega, \alpha(f) = n, t_1, \ldots, t_n$ are terms, then $ft_1t_2 \ldots t_n$ is a term.

Informally, we'll sometimes add brackets and commas, i.e. $f(t_1, \ldots, t_n)$.

Examples:

GT: Some terms are $mxmmyzu$ or $mxiy$ or e[2]

PT: The only terms are the variables - we have no function symbols.

The ***atomic formulae*** are defined inductively by:

- $\bot$ is an atomic formula.
- If s, t are terms then $s = t$ is an atomic formula.
- If $\varphi \in \Pi$ and $\alpha(\varphi) = n$ and $t_1, \ldots, t_n$ are terms, then $\varphi t_1 \ldots t_n$ is an atomic formula.

Again, we often insert brackets and commas.

The ***formulae*** are defined inductively by:

- Every atomic formula is a formula.
- If p, q are formulae then $(p \implies q)$ is a formula.
- If x is a variable and p is a formula then $(\forall x)p$ is a formula.

Examples:

GT: $((\forall x)(mxx = e) \implies (\forall x)(x = y))$ or $(\forall x)(y = z)$ or $mxy = e$.

Suppose that p is a formula and x is a variable that appears in p. We say an occurrence of x in p is ***bound*** if it is within a subformula of the form $(\forall x)q$; it is ***free*** otherwise. A formula with no free variables is called a ***sentence***.

[2]These are more commonly recognised as $(x \cdot ((y \cdot z) \cdot u))$ or $(x \cdot y^{-1})$ or e

Examples:

GT: $(\forall x)(\forall y)(\forall z)(mmxyz = mxmyz)$ is a sentence.

GT: $(\forall x)(x = y)$ - x is bound and y is free.

GT: We can do stupid things like $((x = y) \implies (\forall x)(x = z))$. The first x is free, the second is bound.

The ***first order language*** $L(\Sigma)$ is the set of formulae. A ***theory*** is a set T of sentences.

For instance, the ***language of groups*** is generated by $\Omega = \{m, i, e\}, \Pi = \emptyset, \alpha(m) = 2, \alpha(i) = 1, \alpha(e) = 0$. The ***theory of groups*** is:

$$\begin{aligned} T = \big\{ & (\forall x)(\forall y)(\forall z)(mxmyz = mmxyz), && \text{(associativity)} \\ & (\forall x)(mxe = x), && \text{(identity)} \\ & (\forall x)(mxix = e) \big\} && \text{(inverses)} \end{aligned}$$

We sometimes call the sentences in T the ***axioms*** of the theory.

In a similar way, we have the ***theory of posets*** given by:

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

We've used the informal notations $x \leq y$ for $\leq xy$, and the abbreviations $\wedge, \vee, \neg$ as in chapter 1. We also use the abbreviation $(\exists\, x)p$ meaning $\neg(\forall x)(\neg p)$.

We also have the ***theory of total orders***, using the same language as posets, but including an extra sentence in the theory: $(\forall x)(\forall y)(x \leq y \vee y \leq x)$.

The ***language of ordered fields*** has $\Omega = \{a, m0, i, 0, 1\}, \Pi = \{\leq\}, \alpha(a) = \alpha(m) = \alpha(\leq) = 2, \alpha(-) = \alpha(i) = 1, \alpha(0) = \alpha(1) = 0$. As an exercise, write down the theory of ordered fields.

The ***language of graphs*** has $\Omega = \emptyset, \Pi = \{a\}, \alpha(a) = 2$. The ***theory of graphs*** is:

$$\begin{aligned} T = \{ & (\forall x) \neq axx, \\ & (\forall x)(\forall y)(axy \implies ayx) \} \end{aligned}$$

5.1 Semantic Entailments

Let $L = L(\Sigma) = L(\Omega, \Pi, \alpha)$ be a first order language. An ***L-structure*** is a non-empty set A endowed with, for each $f \in \Omega$, a function $f_A : A^{\alpha(f)} \to A$ and for each $\varphi \in \Pi$, a subset $\varphi_A \subseteq A^{\alpha(\varphi)}$.

An important thing to note is that, if L is say the language of groups, an L-structure need not be a group.

A term t is ***closed*** if it has no free variables. A symbol $c \in \Omega$ is a ***constant*** if $\alpha(c) = 0$. If p is a formula and x a variable, t is a term, then we write $p[t/x]$ for the formula obtained by replacing every free occurrence of x in p by t.

Let A be an L-structure. For each closed term t, we define the ***interpretation*** t_A inductively as follows:

- For c a constant, we have defined $c_A : A^0 \to A$ - think of $c_A \in A$ by identifying it with the unique element of $\mathfrak{Im}(c_A)$.
- $(ft_1 \dots t_n)_A = f_A((t_1)_A, \dots, (t_n)_A)$ for $f \in \Omega$, $\alpha(f) = n, t_1, \dots, t_n$ all closed terms.

For p a sentence, we define the interpretation $p_A \in \{0, 1\}$ of p in A as follows:

We start with the atomic formulae, so that:

- $p = \bot$. Then $\bot_A = 0$.
- $p = (s = t)$ for s, t closed terms. Then $(s = t)_A = \begin{cases} 1 & s_A = t_A \\ 0 & \text{otherwise} \end{cases}$
- $p = \phi t_1 \dots t_n$ where $\phi \in \Pi, \alpha(\phi) = n$, $t_1, \dots, t_n$ closed terms. Then $(\phi t_1 \dots t_n)_A = \begin{cases} 1 & ((t_1)_A, \dots, (t_n)_A) \in \phi_A \\ 0 & \text{otherwise} \end{cases}$.

Then for non-atomic formulae we proceed inductively:

- $p = (q \implies r)$. Then q, r are sentences, so define $(q \implies r)_A = \begin{cases} 0 & q_A = 1, r_A = 0 \\ 1 & \text{otherwise} \end{cases}$.
- $p = (\forall x)q$ for some formula q. As p is a sentence, the only possible free variable in q is x. We want something like "substitute every $a \in A$ into q in place of x, and if what we get is always true then so is p".

We introduce a new constant for each $a \in A$. That is, let $\bar{A} = \{\bar{a} : a \in A\}$ be a set of symbols, where $a \mapsto \bar{a}$ is a bijection $A \to \bar{A}$. Assume WLOG $\bar{A} \cap (\Omega \cup \Pi) = \emptyset$. Let $\Omega' = \Omega \cup \bar{A}$ and extend α to α' by setting $\alpha'(\bar{a}) = 0$ for all $a \in A$. Let $L' = L(\Omega', \Pi, \alpha')$. Then extend our interpretation of A as an L-structure to being an L'-structure by interpreting $\bar{a}_A = a$.

Then, for each $a \in A, q[\bar{a}/x]$ is a sentence. So define:

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

Let L be a language and A an L-structure. If p is a sentence with $p_A = 1$, we say that p is ***satisfied*** or ***true*** in A, or that A is a ***model*** of p. If T is a theory, we say A is a ***model*** of T if for all $p \in T$, A is a model of p.

If T is a theory and p is a sentence such that every model of T is a model of p we say that T ***semantically entails*** p, and we write $T \vDash p$. If $\emptyset \vDash p$, we write $\vDash p$ and say p is a ***tautology***. So p is a tautology if and only if it is true in every L-structure.

For example, if T is the theory of groups in the language L of groups then an L-structure is a group if and only if it is a model of T. We have:

$$T \vDash (\forall x)((\forall y)(mxy = y \wedge myx = y) \implies (x = e))$$

Sometimes we need to interpret a formula p with free variables. Suppose the set of free variables of p, $FV(p) \subset \{w_0, \dots, w_{n-1}\}$. We can then define the interpretation of p as $p_A \subset A^n$ by:

$$p_A = \{(a_0, \dots, a_{n-1} \in A^n : (p[\bar{a}_0/w_0] \dots [\bar{a_{n-1}}/w_{n-1}])_A = 1\}$$

We say A is a model of p if $p_A = A^n$

Note that this final definition is independent of n, and is consistent with our earlier definition in the case where p is a sentence.

5.2 Syntactic Entailment

There are seven logical axioms:

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

Note that every axiom is a tautology. Note also that one instance of axiom 6 is $(\forall x)\bot \implies \bot$. If we allow $\emptyset$ as an L-structure, then in $\emptyset$, $(\forall x)\bot$ is true, and so that would imply $\bot$ is true, making axiom 6 not a tautology. This is why we insisted that L-structures had to be non-empty.

We also have two deduction rules:

- Modus ponens: from p and $p \implies q$ we can infer q.
- Generalisation: From p we can infer $(\forall x)p$ as long as x didn't appear free in any premise used in the proof of p.

Let S be a set of formulae and p a formula. A ***proof*** of p from S is a sequence $\ell_1, \ell_2, \ldots, \ell_n = p$ of formulae, sometimes called ***lines***, such that each line is an axiom or an element of S or can be inferred from earlier lines by a deduction rule. If there is a proof of p from S, we say S ***syntactically entails*** or ***proves*** p and write $S \vdash p$. If $S = \emptyset$, we write $\vdash p$.

Proposition 5.1 (Deduction Theorem)**.** *Let S be a set of formulae and p, q be formulae. Then $S \vdash (p \implies q)$ if and only if $S \cup \{p\} \vdash q$.*

Proof. Exactly the same as in chapter 1 with one new case: when assuming $S \cup \{p\} \vdash q$ and carrying out our induction on the lines of the proof, we may come across a line of the proof of q from $S \cup \{p\}$ of the form $(\forall x)r$ that is inferred from an earlier line r by (Gen).

Case 1 x does not appear free in p. By induction hypothesis, $S \vdash (p \implies r)$. Then by (Gen), $S \vdash (\forall x)(p \implies r)$. So $S \vdash (\forall x)(p \implies r)$, and by (A7) and (MP), $S \vdash (p \implies (\forall x)r)$.

Case 2 x does appear free in p. Then as we inferred $(\forall x)r$ from r by (Gen), we cannot have used p in the proof of r from $S \cup \{p\}$. So in fact $S \vdash r$, and by (Gen), $S \vdash (\forall x)r$. Then by (A1) and (MP), $S \vdash (p \implies (\forall x)r)$.

□

5.3 Completeness Theorem

Examinable: statements of results. Non-examinable: everything else.

Proposition 5.2 (Soundness Theorem)**.** *Let S be a set of sentences and p a sentence with $S \vdash p$. Then $S \vDash p$.*

Proof. Induction on the lines of the proof. □

The next step and heart of the proof, is to show that if S is consistent (i.e. $S \nvdash \bot$), then S has a model ($S \nvDash \bot$).

Let S be a consistent set of sentences in a language $L = (\Omega, \Pi, \alpha)$. Then we let A be the collection of closed terms of L made into an L-structure via:

1. $f \in \Omega, \alpha(f) = n : f_A(t_1, \ldots, t_n) = (ft_1 \ldots t_n)_A$.
2. $\phi \in \Pi, \alpha(\phi) = n : \phi_A = \{(t_1, \ldots, t_n) \in A^n : S \vdash \phi t_1 \ldots t_n\}$.

In general, this is not yet a model of S:

Problem 1: E.g. (GT): $e, mee \in A$, and we know that $S \vdash (e = mee)$, but $(e = mee)$ is not true in A.

Solution: Quotient out by the equivalence relation $t \sim u$ if $S \vdash (t = u)$.

Problem 2: E.g. L has constants a, b, c and $S = \{(a = b) \vee (a = c)\}$. Then $S \nvdash (a = b)$ and $S \nvdash (a = c)$. So $(a = b) \vee (a = c)$ not true even in $A/\sim$.

Solution: Use Zorn's lemma on S to complete it.

Problem 3: E.g $\Omega = \{a, f\}, \Pi = \emptyset, \alpha(a) = 0, \alpha(f) = 1$. If $S = \{(\exists\ x)(fx = a)\}$. Then $A = \{a, fa, ffa, fffa, \ldots\}$. We could extend S to a complete theory with included $\neg(fa = a), \neg(ffa = a), \ldots$.

Solution: Add a constant b to the language and add the sentence $fb = a$ to S. It is easy to check that if S is consistent then so is $S \cup \{(fb = a)\}$. And indeed, S is still consistent if we do this simultaneously for every formula in S of the form $(\exists\ x)p$, adding a new constant for each. We say that S ***has witnesses*** if whenever $(\exists\ x)p \in S$ there is a constant b with $p[b/x] \in S$.

However, if S has witnesses and then we extend it to a complete theory, it may no longer have witnesses. If S is complete and we add witnesses, we may make S incomplete. The solution is to just repeatedly solve problems 2 and 3.

Proposition 5.3 (Model Existence Lemma)**.** *Let S be a consistent theory. Then S has a model.*

Sketch proof. Let L_0 be the language of the theory S. By Zorn, we can extend S to a complete consistent theory S_1. Next, extend S_1 to a consistent theory T_1 that has witnesses in a language $L_1 \supset L_0$. Now by Zorn again, we can extend T_1 to a complete consistent theory S_2. Then add witnesses to give a consistent theory T_2 in language $L_2 \supset L_2$.

We get a sequence of consistent theories $S \subset S_1 \subset T_1 \subset S_2 \subset T_2 \subset \ldots$, and languages $L_0 \subset L_1 \subset L_2 \subset \ldots$, such that each S_i is complete and each T_i has witnesses and S_i, T_{i+1} are in language L_{i+1}.

Let $\bar{L} = \bigcup_{i=0}^{\infty} L_i$ and $\bar{S} = \bigcup_{i=1}^{\infty} S_i$.

Then $\bar{S}$ is a complete consistent theory with witnesses in the language $\bar{L}$.

Let A be the set of closed terms in the language $\bar{L}$, and let $\sim$ be the relation on A defined by $s \sim t$ if and only if $\bar{S} \vdash (s = t)$. Then we can check that $\sim$ is an equivalence relation on A,

and makes $A/\sim$ into an $\bar{L}$-structure as described previously. Now it is easy to check inductively that for any sentence p in language $\bar{L}$ then p is true in $A/\sim$ if and only if $\bar{S} \vdash p$. □

The other results now follow exactly as in chapter 1.

Corollary 5.4 (Adequacy Theorem)**.** *Let S be a theory, p a sentence with $S \vDash p$. Then $S \vdash p$.*

Theorem 5.5 (Gödel's Completeness Theorem For First-Order Logic)**.** $S \vdash p \iff S \vDash p$.

5.4 Applications

Back to everything examinable.

Corollary 5.6 (The Compactness Theorem)**.** *Let S be a set of sentences such that every finite subset of S has a model. Then S has a model.*

Proof. This is obvious if we use completeness to replace "has a model" by "is consistent", since proofs are finite. □

Unfortunately, we no longer have a decidability theorem as in this system it is not obvious how to check if $S \vDash p$.

A typical application of compactness is to try to axiomatize the theory of finite groups in the language of groups. Suppose T is the theory of finite groups.

We enrich our language by adding new constants $c_0, c_1, c_2, \ldots$, and for all $i, j \in \mathbb{N}$ with $i \neq j$ add $\neg(c_i = c_j)$ to T to create a new theory T'. Now any finite subset of T' has a model, as there are arbitrarily large finite groups, so by compactness T' has a model, which must be an infinite model of T. ↯

The same method gives:

Proposition 5.7. *If T is a theory with arbitrarily large finite models then T has an infinite model.*

Even more:

Theorem 5.8 (Upward Lowenheim-Skolem Theorem)**.** *Let T be a theory with an infinite model and κ a cardinal. Then T has a model of cardinality $\geq \kappa$.*

Proof. Do the same thing but with κ new constants. □

Theorem 5.9 (Downward Lowenheim-Skolem Theorem)**.** *A consistent theory in a countable language has a countable model.*

Proof. Take the model used in the proof of the Model Existence Lemma. □

These two theorems say that a first-order theory cannot uniquely specify an infinite structure. What's going on?

5.5 Peano Arithmetic

It is well known that $\mathbb{N}$ is uniquely specified by the ***Peano axioms***:

- $0 \in \mathbb{N}$
- $n \in \mathbb{N} \implies s(n) \in \mathbb{N}$, the ***successor*** of $\mathbb{N}$
- $n \in \mathbb{N} \implies s(n) \neq 0$
- $m \neq n \implies s(m) \neq s(n)$
- If $p(0)$ and for all n, $p(n) \implies p(s(n))$, then for all $n, p(n)$.

We try to make this a first-order theory. We'll also include $+, \times$.

The ***language of Peano arithmetic*** has $\Omega = \{0, s, a, m\}, \Pi = \emptyset, \alpha(0) = 0, \alpha(s) = 1, \alpha(a) = \alpha(m) = 2$. Then the theory PA has axioms:

1. $(\forall x) \neq (0 = sx)$
2. $(\forall x)(\forall y)((sx = sy) \implies (x = y))$
3. $(\forall y_1) \dots (\forall y_n)(\forall x)(p[0/x] \wedge (\forall z)p[z/x] \implies p[sz/x]) \implies p)$ where p is a formula with free variables $\{y_1, \dots, y_n, x\}$.
4. $(\forall x)(ax0 = x)$
5. $(\forall x)(\forall y)(axsy = saxy)$
6. $(\forall x)(mx0 = 0)$
7. $(\forall x)(\forall y)(mxsy = amxyx)$

Now $\mathbb{N}$ is a model of PA with $0_{\mathbb{N}} = 0, s_{\mathbb{N}}(b) = b + 1, a_{\mathbb{N}}(b, c) = b + c, m_{\mathbb{N}}(b, c) = bc$. So by Upward Lowenheim-Skolem, PA has an uncountable model. Doesn't this contradict that the Peano axioms uniquely specify $\mathbb{N}$.

No: genuine induction says that if $X \subset \mathbb{N}$ with $0 \in X$ and for all $n \in \mathbb{N}, n \in X \implies n+1 \in X$, then $X = \mathbb{N}$. We know $\mathbb{N}$ has uncountably many subsets.

We say a subset $S \subset \mathbb{N}$ is ***definable*** if there is a PA-formula p with $FV(p) = \{w\}$ such that $p_{\mathbb{N}} = S$. First-order induction only tells us about the definable subsets, of which there are only countably many. So first order induction is weaker than genuine induction.

Why do we have parameters $y_1, \dots, y_n$ in axiom 3? We wan tot prove things like for all ℓ, m, n that $\ell + (m + n) = (\ell + m) + n$. We can do a lot with definable subsets:
Example: s

1. Set of squares: $(\exists\, y)(myy = w)$
2. Set of primes: $\neg(w = 1) \wedge (\forall y)((\exists\, z)(w = myz) \implies ((y = w) \vee (y = s0))$
3. Set of powers of 2: $(\forall y)(((y \text{ prime}) \wedge (y|w)) \implies y = 2)$.

Harder problems/exercises are to show that the powers of 4 or of 10 are definable.

Our next hope is that maybe PA is a complete theory, so that anything that is true in $\mathbb{N}$ is provable in PA. Not even this is true: we can find a sentence p in the language of PA with PA $\nvdash p$ and PA$\nvdash \neq p$:

We will find a statement p that says '$PA \nvdash p$'. Then suppose that $PA \vdash p$. Then p is true in a model e.g. $\mathbb{N}$, so $PA \nvdash p$ ↯. Hence by completeness $PA \vdash \neg p$. Then $\neg p$ is true in $\mathbb{N}$. So $PA \vdash p$, so $PA \vdash \bot$, but PA has a model. ↯.

We can extend the definition of definable to subsets of $\mathbb{N}^k$ and partial function $\mathbb{N}^k \to \mathbb{N}$. It can be shown that every recursive function is definable. Church's Thesis says that for any sane model of computation, computable function are precisely the recursive functions. So we can assume any function or set given by an algorithm is definable in PA.

A formula in PA is a finite strin from $(,), =, \bot, \forall, \implies, w, ', 0, s, a, m$. Number these symbols $1, \ldots, 12$. A formula p corresponds to a finite sequence $(n_1, \ldots, n_k)$ of numbers between 1 and 12. So p can be represented by the single number $2^{n_1} 3^{n_2} \ldots p_k^{n_k}$ where p_k is the k^{th} prime. We call this the ***Gödel number*** (or just ***code***) of p. Then coding and decoding are recursive.

Write $c(p)$ for the code of p. If n codes a formula, call it S_n. Given a proof of $S_{n_1}, S_{n_2}, \ldots, S_{n_k}$, we can code it by the single number $2^{n_1} 3^{n_2} \ldots p_k^{n_k}$. Now let $U = \{(m, n) \in \mathbb{N}^2 : n \text{ codes a formula and } m \text{ codes a proof of } S_n\}$. Then U is recursive, so there is some PA-formula θ with $FV(\theta) = \{x, y\}$, saying "y codes a formula and x codes a proof of S_y."

Let φ be $(\exists\, x)\theta$, i.e. "y codes a provable formula". For n a natural number, write $\bar{n}$ for the term $\underbrace{ss\ldots s}_{n}0$ of PA. Let p be a formula with $c(p) = n$. Suppose that $PA \vdash p$. Then let m be the code of some proof of p. Then $PA \vdash \theta[\bar{m}/x][\bar{n}/y]$, so $PA \vdash \varphi[\bar{n}/y]$. Conversely, suppose $PA \vdash \varphi[\bar{n}/y]$. Then $\mathbb{N}$ is a model of PA, so $\varphi[\bar{n}/y]$ is true in $\mathbb{N}$, so $\exists\, m \in \mathbb{N}$ with $(m, n) \in U$. So m codes a proof of p. So $PA \vdash p$.

Hence PA proves p if and only if PA proves p is provable.

Let $V = \{m \in \mathbb{N} : m \text{ codes a formula with } FV(S_m) = \{w\}\}$. Let $f : V \to \mathbb{N}$ be defined by $f(m) = c(S_m[\bar{m}/w])$. Now V is recursive, so there is a formula ψ of PA with $FV(\psi) = \{w\}$, saying "w codes a formula with free variables $\{w\}$." Moreover, the function f is recursive and so definable in PA. Thus we have a formula $\xi = \psi\wedge \neq \varphi[f(w)/w]$. Finally, let h be the code for ξ and p be $\xi[\bar{h}/w]$. What does p say?

"h codes a formula with $FV(S_h) = \{w\}$, and $S_h[\bar{h}/w]$ is unprovable", i.e. "p is unprovable." Hence by the previous discussion, we have:

Theorem 5.10 (Gödel's First Incompleteness Theorem)**.** *The theory PA is incomplete.*

The incompleteness of PA is essential. Note that p must be true in $\mathbb{N}$, as if p is false then p is provable. But maybe $PA \cup \{p\}$ is complete? - No. The same proof works as $\mathbb{N}$ is still a model. But what if we let T be the set of all true formulae in $\mathbb{N}$? This theory is clearly complete, so the incompleteness theorem won't work. Hence the only possibility is that T is not recursive, and so truth is not recursive.

6 Set Theory

This chapter is essentially a worked example of the previous chapter. We want to know what the universe of sets looks like. The idea is to start from $\emptyset$ and build up from there.

6.1 The Axioms of Zermelo-Fraenkel Set Theory

The ***language of sets*** has $\Omega = \emptyset, \Pi = \{\in\}, \alpha(\epsilon) = 2$. For readability, we write $x \in y$ for $\in xy$.

The ***theory of sets***, ZF, consists of axioms as follows:

1. Extensionality. "If two sets have the same elements, they are equal."

$$(\forall x)(\forall y)((\forall z)(z \in x) \iff (z \in y)) \implies (x = y)) \tag{Ext}$$

   Note that the converse is a logical theorem.

2. Separation. We want something like "if p is a formula with free variable x then $\{x|p\}$ is a set." However, we don't want something like $p = \neg(x \in x)$. Instead, we say "we can form a subset of all elements with property p."

$$(\forall t_1) \dots (\forall t_n)(\forall y)(\exists\, z)(\forall x)((x \in z) \iff ((x \in y) \wedge p)) \tag{Sep}$$

   where p is a formula with free variables $\{t_1, \dots, t_n, x\}$. We introduce the notation $\{x \in y|p\}$ for the set z in the axiom. This is more of an ***axiom scheme*** - we have one instance of it for each formula p.

3. Empty set. "The empty set exists."

$$(\exists\, x)(\forall y) \neq (y \in x) \tag{Emp}$$

   By (Ext) the empty set is unique, and we denote it by $\emptyset$. That is, $z \in \emptyset$ is an abbreviation for $((\exists\, x)(\forall y) \neq (y \in x)) \wedge (z \in x)$. Similarly for the notation introduced after (Sep).