Part II - Logic and Set Theory

B. Couzens

Lent 2014

This is a condensed version of B. Couzens' notes on http://tartarus.org/gareth/maths/notes/ii/Logic_and_Set_Theory_2014.pdf

1 Posets and Zorn's Lemma

Definition (Partial/total order and posets)**.** Let A be a set. A *partial order* on A is a binary relation $\leq$ which is

(i) Reflexive $(\forall x \in A) x \leq x$

(ii) Transitive $(\forall x, y, z) x \leq y$ and $y \leq z \Rightarrow x \leq z$

(iii) Antisymmetric $(\forall x, y) x \leq y$ and $y \leq x \Rightarrow x = y$

$\leq$ is a *total order* if in addition $(\forall x, y \in A) x \leq y$ or $y \leq x$

A *poset* (or partially ordered set) is a pair $(A, \leq)$ where $\leq$ is a partial order on A.

Definition (cover)**.** y *covers* x in a poset $(P, \leq)$, written as $x \lhd y$, if $x < y$ (i.e. $x \leq y$ and $x \neq y$) but $\forall z \in P, x \leq z \leq y \Rightarrow z = x$ or $z = y$, i.e. y is strictly greater than x and there is nothing in between.

Definition (greatest member/(least) upper bound/completeness)**.** Let S be a subset of a poset $(P, \leq)$.

(i) x is a/the greatest member of S if $x \in S$ and $(\forall y \in S) y \leq x$.

(ii) x is an upper bound for S if $(\forall y \in S) y \leq x$

(iii) x is the *least upper bound/join/supremum* of S if it is the least member of $\{y \in P : y \text{ is an upper bound for } S\}$. Write $x = \bigvee S$.

Similarly we have the *greatest lower bound*, also known as *infimum* or *meet*, written as $x = \bigwedge S$.

(iv) $(P, \leq)$ is complete if every $S \subseteq P$ has a least upper bound.

Lemma. If $(P, \leq)$ is complete, then so is $(P, \geq)$

Proof. Let $S \subseteq P$. We are required to prove that S has a greatest lower bound.

Let $T = \{x \in P : x \text{ is a lower bound for S }\}$. and consider $y = \bigvee T$.

If $z \in S$ and $x \in T$, then $x \leq z$. So every $z \in S$ is an upper bound for T and hence satisfies $z \geq y$. So y is a lower bound for y, and hence the greatest member of T. □

Definition (chain)**.** A chain in a poset $(P, \leq$ is a non-empty subset $C \subseteq P$ which is totally ordered by $\leq$.

Definition (chain complete)**.** $(P, \leq)$ is chain complete if every chain $C \subseteq P$ has a least upper bound in P.

Theorem (Knaster-Tarski Theorem)**.** Let P be a complete poset. Let $f : P \to P$ be an order-preserving map (i.e. $x \leq y \Rightarrow f(x) \leq f(y)$). Then f has a fixed point.

Proof. Let $S = \{x \in P : x \leq f(x)\}$ be the set of prefixed points of f. Let $y = \bigvee S$.

For all $x \in S$, we have $x \leq y$ and hence $x \leq f(x) \leq f(y)$. So $f(y)$ is an upper bound for S.

So $y \leq f(y)$ and hence $f(y) \leq f(f(y))$ (since f is order-preserving). So $f(y) \in S$. Therefore $f(y) \leq y$. So $f(y) = y$ by antisymmetry. □

Corollary (Cantor-Bernstein Theorem)**.** Suppose we have sets A and B and injective functions $f : A \to B$, $g : B \to A$. Then there exists a bijection $h : A \to B$.

Proof. We seek a fixed point A_0 of the function $\Phi : \mathcal{P}A \to \mathcal{P}A$ defined by $\Phi(X) = A \setminus g(B \setminus f(X))$. Φ is order preserving since f and g are, and $A\setminus$ and $B\setminus$ reverse it.

So by Knaster-Tarski, Φ has a fixed point $A_0 \subseteq A$.

Define h by

$$h(x) = \begin{cases} f(x) & x \in A_0 \\ g^{-1}(x) & x \in A \setminus A_0 \end{cases}$$

□

Definition (inflationary map)**.** Given a poset P, f is inflationary if $x \leq f(x)$ for all x.

Theorem (Bourbaki-Witt Theorem)**.** Let P be a chain-complete poset and $f : P \to P$ an inflationary map. Then for every $x \in P$, $\exists y \in P$ with $x \leq y = f(y)$.

Proof. Define a subset $C \subseteq O$ to be closed if

(i) $(\forall y \in P) y \in C \Rightarrow f(y) \in C$

(ii) $\forall D \subseteq C$, D is a chain $\Rightarrow \bigvee D \in C$

Any intersection of closed sets is closed. In particular,

$$C(x) = \bigcap \{C \subseteq P : C \text{ is closed }, x \in C\}$$

is the smallest closed set containing x.

Suppose we can show that $C(x)$ is a chain. Then $y = \bigvee C(x) \in C(x)$ by (ii) and hence $f(y) \in C(x)$ by (i). So $f(y) \leq y$, and hence $y = f(y)$. So we are required to show that $C(x)$ is a chain.

Step 1: $\forall y \in C(x), x \leq y$

$\uparrow (x) = \{y \in P : y \geq x\}$ is a closed set containing x. So $C(x) \subseteq \uparrow (x)$

Say $y \in C(x)$ is normal if $(\forall z \in C(x)) z < y \Rightarrow f(z) \leq y$

Step 2: If y is normal, then $(\forall z \in C(x)) z \leq y$ or $f(y) \leq z$.

Consider $D = \{z \in C(x) : z \leq y \text{ or } f(y) \leq z\}$. Then $x \in D$ by Step 1. Suppose $z \in D$. Then either

(i) $z < y$, in which case $f(z) \leq y$ by normality; or

(ii) $z = y$, in which case $f(y) = f(z)$; or

(iii) $f(y) < f(z)$, in which case $f(y) \leq f(z)$ since f is inflationary.

So $f(z) \in D$.

If $E \subseteq D$ is a chain, then either

(i) $\forall z \in E, z \leq y$, in which case $\bigvee E \leq y$; or

(ii) $\exists z \in E$ such that $f(y) \leq z$, in which case $f(y) \leq \bigvee E$.

So $\bigvee E \in D$

So D is closed an contains x. Hence $D = C(x)$.

Step 3: $\forall y \in C(x)$, y is normal.

Consider $N = \{y \in C(x) : y \text{ is normal }\}$. Then $x \in N$, since $z < x$ is never satisfied for $z \in C(x)$ by Step 1.

Suppose $y \in N$ and $z \in C(x)$ satisfies $z < f(y)$. Then $z \not\geq f(y)$. So $z \leq y$ by Step 2. So either

(i) $z < y$, in which case $f(z) \leq y \leq f(y)$ (definition of normality); or

(ii) $z = y$, in which case $f(z) = f(y)$

So $f(y)$ is normal, and $f(y) \in N$.

Finally suppose $M \subseteq N$ is a chain and $z < \bigvee M$. Then $\exists y \in M$ such that $y \not\leq z$ and hence $f(z) \leq y$ by Step 2. So $f(z) \leq y \leq \bigvee M$.

So as before, we must have $N = C(x)$.

Hence by Step 2, for any $y, z \in C(x)$, either $z \leq y$ or $y \leq f(y) \leq z$. So $C(x)$ is a chain, and $y = \bigvee C(x)$ is the required fixed point. □

Corollary. Suppose $(P, \leq)$ is chain complete and $f : P \to P$ is order-preserving. Then for any $x \in P$ with $x \leq f(x)$, $\exists$ a least $y \geq x$ with $y = f(y)$. In particular, if P has a least element, then f has a least fixed point.

Proof. Let $Q = \{y \in P : y \leq f(y)\}$. Then $y \in Q \Rightarrow f(y) \leq f(f(y)) \Rightarrow f(y) \in Q$ since f is order-preserving.

And given a chain $C \subseteq Q$, then $z = \bigvee C$ satisfies $\forall y \in C, y \leq z$. So $\forall y \in C$, $y \leq f(y) \leq f(z)$. So $f(z)$ is an upper bound for C and $z \leq f(z)$. So $z \in Q$. Hence Q is chain-complete, and $f|_Q$ is an inflationary map $Q \to Q$. Hence $\forall x \in Q, \exists y \in Q$ with $x \leq y = f(y)$.

If $z \in Q$ is any other fixed point with $z \geq x$, then $\downarrow z = \{w \in Q : w \leq z\}$ is a closed set (if $w \in \downarrow z$, then $f(x) \leq f(z) = z \Rightarrow f(w) \in \Downarrow z$.

Hence z is an upper bound for the set $C(X)$ constructed in the proof above. So the fixed point $y = \bigvee C(x)$ satisfies $y \leq z$.

The final assertion follows from the fact that a least element 0 of P necessarily satisfies $0 \leq f(0)$. □

Axiom. Given a set $\{A_i : i \in I\}$ of sets with $A_i \neq \emptyset$ for all i, there exists a choice function $f : I \to \bigcup_{i \in I} A_i$ such that $f(i) \in A_i$ for all $i \in I$.

Zorn's Lemma. Given a chain complete poset $(P, \leq)$, every $x \in P$ lies below some maximal element.

Corollary. Axiom of choice $\Rightarrow$ Zorn's lemma

Proof. Let $(P, \leq)$ be chain-complete. Define a family of sets $\{A_x : x \in P\}$ as follows:

- If x is not maximal, $A_x = \{y \in P : x < y\}$

- If x is maximal, $A_X = \{x\}$.

By the Axiom of Choice, $\exists f : P \to P$ with $f(x) \in A_x$ for all x.

By construction $f(x) = x \Leftrightarrow x$ is maximal in P.

By Bourbaki-Witt, there exists a $y \geq x$ such that $f(y) = y$, i.e. y is maximal. □

Proposition. Zorn's Lemma ⇒ Axiom of Choice

Proof. Given a family $\{A_i : i \in I\}$ of non-empty sets, consider the set of partial functions $P \subseteq [I \rightharpoondown \bigcup_{i \in I} A_i]$ satisfying $f(i) \in A_i$ whenever $f(i)$ is defined.

It is easy to see that if $\{f_j : j \in J\}$ is a chain in P, then the join $\bigvee\{f_j : j \in J\}$ is a partial choice function. So P is chain-complete.

Also P is non-empty, as the everywhere undefined function is in P. So by Zorn's lemma, P has a maximal element f_0.

Suppose f_0 is not total. Then pick $i_o \in I \setminus \operatorname{dom} f_0$. Pick $x_0 \in A_{i_0}$. Now define

$$f_1(i) = \begin{cases} f_0(i) & \text{if } f_0(i) \text{ is defined} \\ x_0 & \text{if } i = i_0 \\ \text{undefined} & \text{otherwise} \end{cases} .$$

Then $f_1 \in P$ and $f_0 < f_1$. Contradiction. So f_0 is a total choice function. □

Theorem (Hamel's Theorem)**.** Every vector space has a basis

Definition. Let V be a vector space. Consider the set P of all linearly independent subsets of V, ordered by inclusion. If $\{S_i : i \in I\} \subseteq P$ is a chain, consider $\bigcup_{i \in I} S_i$. If we had a non-trivial linear relation $\sum_{j=1}^{n} \lambda_j x_j = 0$ on this set, then for each j, $\exists i_j$ such that $x_j \in S_{i_j}$. Since the S_i are totally ordered by $\subseteq$, $\exists i$ such that $x_1, \cdots, x_n$ all belong to S_i. So $\sum \lambda_j x_j = 0$ is a non-trivial linear relation on S_i. Contradiction. So $\bigcup S_i$ is linearly independent, and P is chain-complete.

Hence any linearly independent set (in particular) is contained in a maximal linearly independent set S_0.

Suppose S_0 does not span V. Pick $x \in V \subseteq \langle S_0 \rangle$ and define $S_1 = S_0 \cup \{x\}$.Then S_i is linearly independent, since a non-trivial linear relation on it would be either a linear relation on S_0 or an expression for x as a linear combination of members of S_0. Hence $S_0 \subseteq S_1$ but S_0 is maximal. Contradiction. So S_0 spans V and is hence a basis.

Theorem (Krull's theorem)**.** Every proper ideal in a ring R (with I) is contained in a maximal ideal.

Proof. Consider the poset P of proper ideals of R ordered by inclusion. So we need to show P is chain-complete. If $\{I_j : j \in J\}$ is a chain of proper ideals, then $\bigcup_{j \in J} I_J$ is an ideal. It is proper since $I \lhd R$ is proper $\Leftrightarrow 1 \in I$. So it is contained in a maximal ideal. □

Definition (Lattice)**.** A lattice is a poset $(L, \leq)$ in which every finite subset (including $\emptyset$) has both a join and a meet. In particular, L contains

$$0 = \bigvee \emptyset$$

$$1 = \bigwedge \emptyset$$

and it has binary operations $\vee, \wedge$ defined by

$$a \wedge b = \bigwedge \{a, b\}$$

$$a \vee b = \bigvee \{a, b\}$$

Note that the order relation is definable from either $\vee$or $\wedge$, since $a \leq b \Leftrightarrow a \vee b = b \Leftrightarrow a \wedge b = a$.

So a lattice homomorphism (i.e. function preserving $\vee, \wedge, 0$ and 1) is automatically order-preserving.

Definition (Distributive lattice)**.** A lattice L is distributive if it satisfies

$$a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c)$$

Definition (Complementary elements in lattice)**.** Two elements a, b are complementary in a lattice L if $a \vee b = 1$ and $a \wedge b = 0$.

Definition (Boolean algebra)**.** A Boolean algebra is a distributive lattice in which every element has a complement.

Lemma. If $(L, \leq)$ is a distributive lattice, so is $(L, \geq)$

Proof. We have to show $a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c)$. But

$$\begin{aligned}(a \vee b) \wedge (a \vee c) &= ((a \vee b) \wedge a) \vee ((a \vee b) \wedge c) \\ &= a \vee (a \wedge c) \vee (b \wedge c) \\ &= a \vee (b \wedge c)\end{aligned}$$

□

Lemma. In a distributive lattice, any element has at most one complement.

Proof. Suppose b, c are both complements of a. Then $b \wedge (a \vee c) = b \wedge 1 = b$ but $b \wedge (a \vee c) = (b \wedge a) \vee (b \wedge c) = 0 \vee (b \wedge c) = b \wedge c$. So $b = b \wedge c$ and $b \leq c$. Similarly, $c \leq b$. So $b = c$. □

Lemma. Let L be a distributive lattice and a, b elements of L with $a \not\leq b$. Then there is a lattice homomorphism $f : L \to 2 = \{0, 1\}$ with $f(a) = 1, f(b) = 0$.

Proof. Let P be the set of all pairs (A, B) of subsets of L satisfying

(i) A is a filter:

  (a) $x \in A, x \leq y \Rightarrow y \in A$

  (b) $x, y \in A \Rightarrow x \wedge y \in A$

  (c) $1 \in A$

  i.e. it is closed upwards, includes the top element 1 and has a bottom element.

(ii) B is an ideal:

  (a) $x \in B, y \leq x \Rightarrow y \in B$

(b) $x, y \in B \Rightarrow x \vee y \in B$

(c) $O \in B$

i.e. it is closed downwards, includes the bottom element 0 and has a top element.

(iii) $A \cap B = \emptyset$.

Partially order P by $(A_1, B_1) \leq (A_2, B_2) \Rightarrow (a_1 \subseteq a_2$ and $b_1 \subseteq b_2)$.

P is chain complement with the least upper bound of $\{(A_i, B_i) : i \in I\}$ being $(\bigcup_{i \in I} A_i, \bigcup_{i \in IB_i})$

Now the pair $(\uparrow (a), \downarrow (b))$ is an element of P since $a \leq b$. So $\exists$ a maximal element (A_0, B_0) lying above it. Suppose $A_0 \cup B_0 \neq L$. Let $c \in L \setminus (A_0 \cup B_0)$. The set $A_1 = \{x \in L : x \geq (c \cap c)$ for some $y \in A_0\}$ is a filter strictly containing A_0. So $(A_1, B_0 \notin P$ by maximality of (A_0, B_0). And hence $A_1 \cap B_0 \neq \emptyset$. So $\exists x \in A_0$ such that $x \wedge c \in B_0$. Similarly, we can find $y \in B_0$ such that $y \vee c \in A_0$. Now $x \wedge (y \vee c) \in A_0$, since it's a meet of two elements of A_0. But $x \wedge (y \vee c) = (w \wedge y) \vee (x \wedge c) \in B_0$ since it is a join of two elements of B_0. So $A_0 \cap B_0 \neq \emptyset$. Contradiction.

So $A_0 \cup B_0 = L$ and we have a total function $f : L \to 2$ defined by

$$f(x) = \begin{cases} 1 & x \in A_0 \\ 0 & x \in B_0 \end{cases}$$

And f preserves $\wedge$ since $f(x \wedge y) = \Leftrightarrow x \wedge y \in A_0 \Leftrightarrow \{x, y\} \in A_0 \Leftrightarrow f(x) = f(y) = 1$, and similarly for $\vee$. □

Theorem (Birkhoff-Stone)**.** Any distributive lattice is isomorphic to a sublattice of a power set.

Proof. Given a distributive lattice L, let F be the set of all homomorphisms $L \to 2$. Define $\Phi : L \to \mathcal{P}f$ by $\Phi(x) = \{f \in F : f(x) = 1\}$. Then Φ is a lattice homomorphism:

$$\begin{aligned} f \in \Phi(x \wedge y) &\Leftrightarrow f(x \wedge y) = 0 \\ &\Leftrightarrow f(x) = f(y) = 1 \\ &= f \in \Phi(x) \text{ and } f \in \Phi(y) \\ &= f \in \Phi(x) \cap \Phi(y) \end{aligned}$$

We have Φ is injective since $a \not\leq b \Rightarrow \exists f \in \Phi(a) \setminus \Phi(b)$, i.e. $\Phi(a) \neq \Phi(b)$.

So Φ is an isomorphism from K to its image, which is a sublattice of $\mathcal{P}F$. □

2 Propositional calculus

In propositional calculus, we assume we are given a set P of "primitive propositions", which are abstract symbols capable of being "true" or "false".

Definition (Valuation and propositional formulae)**.** Given a set P of primitive propositions, a *valuation* of P is a function $f : P \to 2 = \{0, 1\}$. The set $\mathcal{L}(P)$ of *compound propositions* or *propositional formulae* is defined recursively by

(i) If $p \in P$, then $p \in \mathcal{L}(P)$.

(ii) If $s, t \in \mathcal{L}(P)$, so is $(s \Rightarrow t)$ (where $\Rightarrow$ is a meaningless symbol)

(iii) $\perp \in \mathcal{L}(P)$, which is "false", or "bottom", the statement that is always false.

Formally, if $\Sigma = P \cup \{(,), \Rightarrow, \perp\}$, then $\mathcal{L}(P)$ is the smallest subset of Σ^* closed under (i) to (iii), where Σ^* is the set of all words formed by Σ.

Given a formulation of P, $v : P \to 2$, we extend it to a function $\bar{v} : \mathcal{L}(P) \to 2$ by setting

(i) $\bar{v}(p) = \bar{v}(p)$ if $p \in P$

(ii) $\bar{v}(s \Rightarrow t) = 1 \Leftrightarrow$ either $\bar{v}(t) = 1$ or $\bar{v}(s) = 0$.

(iii) $\bar{v}(\perp) = 0$.

We defined the compound propositions

(i) $\neg p : (p \Rightarrow \perp)$

(ii) $\top$: $\neg\perp$

(iii) $p \vee q$: $(\neg p) \Rightarrow q$

(iv) $p \wedge q$: $\neg(p \Rightarrow \neg q)$

(v) $p \Leftrightarrow q$: $(p \Rightarrow q) \wedge (q \Rightarrow p)$

We have the following truth tables:

p	q	$p \vee q$	$p \wedge q$	$p \Leftrightarrow q$	$\neg q$
0	0	0	0	1	1
0	1	1	0	0	0
1	0	1	0	0	1
1	1	1	1	1	0

Lemma (Functional completeness)**.** For any function $f : 2^n \to 2$, $\exists$ a compound proposition in n primitive propositions whose truth table is f.

Proof. by induction on n. Then $n = 0$, then the constant maps 1 and 0 are the truth tables of $\perp$ and $\top$ respectively.

Suppose it is true for n. Given $f : 2^{n+1} \Rightarrow 2$, define $f_0, f_1 : 2^n \to 2$ by

$$f_0(x_1, \cdots, x_n) = f(x_1, \cdots, x_n, 0).$$

$$f_1(x_1, \cdots, x_n) = f(x_1, \cdots, x_n, 1).$$

Let t_0, t_1 have truth tables f_0, f_1 respectively. Consider $t = (\neg p_{n+1} \wedge t_0) \vee (p_{n+1} \wedge t_1)$. This has truth table f. □

Note: Two formulae s and t have the same truth-table iff $(s \Leftrightarrow t)$ has a truth table consisting entirely of 1's.

Definition (Semantic entailment)**.** Suppose $S \subseteq \mathcal{L}(P)$ and $t \in \mathcal{L}(P)$. S *semantically entails* t, denoted $S \models t$ if every valuation $v : P \to 2$ satisfying $\bar{v}(s) = 1$ for all $s \in S$ also satisfies $\bar{v}(t) = 1$.

In the particular case $S = \emptyset$, we write $\models t$, and call t a tautology.

For finite S, it is easy to determine whether $S \models t$. Write down the truth-tables of all members of S and of t, and cross out the rows in which $\bar{v}(s) = 0$ for some $s \in S$.

Example.

$$\{p, (p \Rightarrow q)\} \models q$$

p	q	$p \Rightarrow q$
~~0~~	~~0~~	~~1~~
~~0~~	~~1~~	~~1~~
~~1~~	~~0~~	~~0~~
1	1	1

Definition. We construct a deduction-system for the propositional calculus as follows:

As axioms, we take all formulae of the form

(i) $s \Rightarrow (t \Rightarrow s)$ (K)

(ii) $(s \Rightarrow (t \Rightarrow u)) \Rightarrow ((s \Rightarrow t) \Rightarrow (s \Rightarrow u))$ (S)

(iii) $\neg\neg s \Rightarrow s$ (T)

where s, t, u are formulae, as our rules of inference. We take modus ponens: from s and $s \Rightarrow t$, we can infer t.

Given a set $S \subseteq \mathcal{L}(P)$ of hypotheses, wed define a deduction from S to be a finite list $t_1, t_2, t_3, \cdots, t_n$, such that for each $i \leq n$, either

(i) $t_i \in S$

(ii) t_i is an axiom

(iii) $\exists j, k < i$ such that t_k is $t_j \Rightarrow t_i$.

S *syntactically entails* t, denoted $S \vdash t$, if there exists a deduction $(t_1, t_2, \cdots, t_n)$ from S with $t_n = t$. If $S = \emptyset$, we say t is a theorem, and write $\vdash t$.

Example. We show $\vdash (p \Rightarrow p)$.

1. $p \Rightarrow (p \Rightarrow p)$ by (K)
2. $p \Rightarrow ((p \Rightarrow p) \Rightarrow p)$ by (K)
3. $(p \Rightarrow ((p \Rightarrow p) \Rightarrow p)) \Rightarrow ((p \Rightarrow (p \Rightarrow p)) \Rightarrow (p \Rightarrow p))$ by (S)
4. $(p \Rightarrow (p \Rightarrow p)) \Rightarrow (p \Rightarrow p)$ MP (2 and 3)
5. $p \Rightarrow p$ MP (1 and 4)

Lemma (Soundness theorem)**.** If $S \vdash t$, then $S \models t$. In particular, every theorem is a tautology.

Proof. Let $t_1, t_2, \cdots, t_n = t$ be a deduction of t from S. We show by induction on i that $S \models t_i$ for all i.

- If $t_i \in S$, then trivially $S \models t_i$.
- If t_i is an axiom, then $\models t_i$. So $S \models t_i$.
- If $t_k = (t_j \Rightarrow t_i$ for some $j, k < i$, then by the induction hypothesis, any v making all of S true satisfies $\bar{v}(t_j) = 1$ and $\bar{v}(t_j \Rightarrow t_i) = 1$. This forces $\bar{v}(t_i) = 1$.

□

Theorem (Deduction theorem)**.** Let $S \subseteq \mathcal{L}(P)$ and $s, t \in \mathcal{L}(P)$. Then $S \vdash (s \Rightarrow t)$ iff $S \cup \{s\} \vdash t$.

Proof. ($\Rightarrow$) If we have a deduction of $s \Rightarrow t$ from S, we can write it down and add

- s hypothesis
- t MP

and obtain a deduction of t from $S \cup \{s\}$.

($\Leftarrow$) Suppose $(t_1, t_2, \cdots, t_n = t)$ is a deduction of t from $S \cup \{s\}$. We show for each i that $S \vdash (s \Rightarrow t_i)$.

(i) If $t_i = s$, we write down the proof that $s \Rightarrow s$ from above

(ii) If $t_i \in S$, we write down

   (a) t_i hypothesis
   (b) $t_i \Rightarrow (s \Rightarrow t_i)$ by (K)
   (c) $s \Rightarrow t_i$ MP

(iii) If t_i is an axiom, we write down

   (a) $t_i \Rightarrow (s \Rightarrow t_i)$ by (K)
   (b) t_i axiom
   (c) $s \Rightarrow t_i$ MP

(iv) If $t_k = t_j \Rightarrow t_i$ for some $j, k < i$, we write down our deductions

   (a) $s \Rightarrow t_j$
   (b) $s \Rightarrow (t_j \Rightarrow t_i)$ from S, and add
   (c) $(s \Rightarrow (t_j \Rightarrow t_i)) \Rightarrow ((s \Rightarrow t_j) \Rightarrow (s \Rightarrow t_i))$ by (S)
   (d) $(s \Rightarrow t_j) \Rightarrow (s \Rightarrow t_i)$ MP
   (e) $(s \Rightarrow t_i)$ MP

□

Definition. A proof in propositional calculus is a deduction from the empty set of hypothesis

Theorem (Completeness theorem)**.** If $S \models t$, then $S \vdash t$.

Proof. We may reduce to the case $t = \bot$: If $S \models t$, then $S \cup \{\neg t\} \models \bot$. If $S \cup \{\neg t\} \vdash \neg$, then $S \vdash \neg\neg t$ and hence $S \vdash t$. (recall $\neg t$ is defined as $t \Rightarrow \bot$) So if we can prove that $S \cup \{\neg t\} \models \bot$ implies $S \cup \{\neg t\} \vdash \neg$, then we know that $S \models t$ implies $S \vdash t$.

Now we prove the contrapositive: if S is consistent, i.e. $S \nvdash \bot$, then S has a model, i.e. S has a valuation v such that $\bar{v}(s) = 1$ for all $s \in S$.

To do this, we first use Zorn's lemma to enlarge S to a maximal consistent set $\bar{S}$: the set $\mathcal{C} \subseteq \mathcal{L}(P)$ of consistent subsets of $\mathcal{L}(P)$ is chain-complete under $\subseteq$, since if $\{S_i : i \in I\}$ is a chain of consistent subsets and $\bigcup_{i\in I} S_i \vdash \bot$, then a deduction of $\bot$ form $\bigcup_{i\in I} S_i$ would use only finitely many hypotheses, say $s_1, s_2, \cdots, s_n$. Then $\exists i \in I$ such that $\{s_1, s_2, \cdots, s_n\} \subseteq S_i$ and $S_i \vdash \bot$. Contradiction. So $\mathcal{C}$ is chain-complete.

Note that

- $\bar{S}$ is deductively closed, i.e. $\bar{S} \vdash t$ implies $t \in \bar{S}$, since if $\bar{S} \vdash t$, then $\bar{S} \cup \{t\}$ is consistent and hence equals $\bar{S}$.
- For every t, either $t \in \bar{S}$, or $\neg t \in \bar{S}$, since if $t \notin \bar{S}$, then $\bar{S} \cup \{t\} \vdash \bot$ and so $\bar{S} \models \neg t$

We define $v : P \to 2$ by $v(p) = \begin{cases} 1 & p \in \bar{S} \\ 0 & p \notin \bar{S} \end{cases}$

Now claim that the canonical extension of v to $\bar{v} : \mathcal{L}(P) \to 2$ satisfies $v(s) = 1$ iff $s \in \bar{S}$.

This is true for $s \in P$ by definition. It is also true for $\bot$, since $\bot \notin \bar{S}$. Suppose it is true for s and t, and consider $s \to t$.

(i) If $\bar{v}(t) = 1$, then $t \in \bar{S}$. But $t \vdash (s \Rightarrow t)$. So $(s \Rightarrow t) \in \bar{S}$. So it holds for $s \Rightarrow t$.

(ii) If $\bar{v}(s) = 0$, then $\neg s \in \bar{S}$. But $\neg s \models (s \Rightarrow t)$. So $(s \Rightarrow t) \in \bar{S}$. So it is holds for $s \to t$.

(iii) If $\bar{v}(t) = 0$ and $\bar{v}(s) = 1$, then $s \in \bar{S}$ and $t \notin \bar{S}$. But $\{s, (s \Rightarrow t)\} \vdash t$. So $(s \Rightarrow t) \notin \bar{S}$. So it holds for $s \Rightarrow t$.

Since $S \subseteq \bar{S}$, this in particular implies $\bar{v}(s) = 1$ for all $s \in S$, i.e. v is a model of S. □

Corollary (Decidability theorem)**.** There exists an algorithm which, given a finite set S of propositions and a proposition t, determines whether $S \models t$.

Proof. This is obvious for $\models$ (write down the truth tables for the members of S and for t). We know that $\models$ coincides with $\vdash$ by soundness and completeness. □

Corollary (Compactness theorem)**.** If $S \models t$, then $\exists$ a finite $S' \subseteq S$ such that $S' \models t$. In particular, if every finite subset of S has a model, then S has a model.

Proof. Obvious for $\vdash$, since a deduction of t from S will only have finitely many steps and hence finitely many hypotheses from S. $\square$

Note: If we make the set V of all valuations $P \to 2$ into a topological space by taking basic open sets to be $U_t = \{v \in V : \bar{v}(t) = 1\}$ for all $t \in \mathcal{L}(P)$, then the assertion that $S \models \perp$ is equivalent to saying that $\{U_{\neg t} : t \in S\}$ covers V. So this is equivalent to the assertion that V is a compact space.

Example. A graph is n-colourable iff all its finite subgraphs are

A graph is a pair (V, E), where V is a set of vertices and E is a set of unordered pairs $\{v, w\}$ of distinct vertices.

An n-colouring of (V, E) is a partition of V into n subsets $V_1, V_2, \cdots, V_n$ such that for each $\{v, w\} \in E$, v and w belong to different subsets.

Given G, consider the set $P = \{p_{v,i} : v \in V, 1 \leq i \leq n\}$ of primitive propositions ($p_{v,i}$ corresponds to the statement "the "v"th vertex has colour i).

Now define the propositional theory S whose members are

$$\{\bigvee_{i=1}^{n} p_{v,i} : v \in V\} \cup \{(p_{v,i} \Rightarrow \neg p_{v,j}) : v \in V, i \neq j\} \cup \{(p_{v,i} \Rightarrow \neg p_{w,i}) : \{v, w\} \in E, 1 \leq i \leq n\}$$

Then a model for S 'is' an n-colouring of G. Result follows from compactness.

3 Predicate calculus

We want to formulate a language which enables us to talk about mathematical structures, e.g. groups or posets. We need to include both operation symbols and predicate (relation) symbols in our language.

Definition ((First-order) signature and structure)**.** A *(First-order) signature* $\Sigma = (\Omega, \Pi)$, consists of a set Ω of operation symbols, equipped with a function $\alpha : \Omega \to \mathbb{N}$, assigning to each $\omega \in \Omega$ its arity (number of inputs expected), together with a set of predicate symbols, again equipped with an arity function $\alpha : \Pi \to \mathbb{N}$.

An operation symbol of arity 0 is a *constant*

A predicate symbol of arity 0 is a *primitive proposition*

Given Σ, a *structure* for Σ is a set A equipped with a function $\omega_A : A^{\alpha(\omega)} \to A$ for each $\omega \in \Omega$ and a subset $[\![\pi_A]\!] \subseteq A^{\alpha(\pi)}$ for each $\pi \in \Pi$. (equivalently, a function $\pi_A : A^{\alpha(\pi)} \to 2$). A structure is an "instance" of the signature.

Definition (Term and interpretation)**.** Given a signature $\Sigma = (\Omega, \Pi)$, the terms of Σ (or over Ω) are defined recursively as follows:

(i) We have an infinite supply of variables $x, x', x'', x''', \cdots$ (or $x, y, z, \cdots$ or $x_1, x_2, x_3, \cdots$) which are terms

(ii) If $\omega \in \Omega$, $\alpha(\omega) = n$ and $t_1, t_2, \cdots, t_n$ are terms, then $\omega t_1 t_2 \cdots t_n$ is a term

i.e. if $X = \Omega \cup \{x,'\}$, then the set of terms is the smallest subset of X^* with these closure properties.

Given a term t of Σ, and a Σ-structure A with a list of variables $x_1, x_2, \cdots x_n$ (including all variables occurring in t), we define an *interpretation* of t in A as a function $t_A : A^n \to A$ as follows:

(i) If t is x_i for some $i \leq n$, then t_A maps $(a_1, a_2, \cdots, a_n) \mapsto a_i$.

(ii) If t is $\omega t_1 t_2 \cdots t_m$, where $\alpha(\omega) = m$, then t_A is the composite

$$A^n \xmapsto{((t_1)_A, (t_2)_A, \cdots, (t_m)_A)} A^m \xmapsto{\omega_A} A$$

Intuitively, it means "perform the substitution $x_1 = a_1, x_2 = a_2, \cdots$ for $a_i \in A$ in the term t"

Definition ((First-order) formulae)**.** Given Σ, the *(first-order) formulae* over Σ are defined as follows:

(i) If $\pi \in \Pi$, $\alpha(\pi) = n$, and $t_1, t_2, \cdots, t_n$ are terms, then $\pi(t_1, t_2, \cdots, t_n)$ is a formula

(ii) If s and t are terms, then $s = t$ is a formula

(iii) $\perp$ is a formula

(iv) If φ and ψ are formulae, then $\varphi \Rightarrow \psi$ is a formula. (Similarly, we introduce $\neg\varphi = (\varphi \Rightarrow \perp)$ etc.)

(v) If φ is a formula and x is a free variable of ϕ, then $(\forall x)\varphi$ is a formula in which all occurrences of x are bound, but all other free variables of φ are free.

Notation. $\mathcal{L}(\Sigma)$ is the set of all first-order formulae over Σ.

Notation. $(\exists x)\varphi$ is a shorthand for $\neg(\forall x)\neg\varphi$

Definition (Interpretation of formulae)**.** For each $\varphi \in \mathcal{L}(\Sigma)$ with free variables in the set $\{x_1, x_2, \cdots, x_n\}$ and each Σ-structure A, and interpretation of φ is a subset $[\![\varphi]\!]_A \subseteq A^n$, or equivalently a mapping $\varphi_A : A^n \to 2$ as follows:

(i) If φ is $\pi(t_1, \cdots, t_m)$, where $\pi \in \Pi$ and $\alpha(\pi) = m$, then φ_A is the composite

$$A^n \xmapsto{((t_1)_A, \cdots, (t_n)_A)} A^m \xmapsto{\pi_A} 2$$

(ii) If φ is $s = t$, then φ_A is the composite

$$A^n \xmapsto{s_A, t_A} A^2 \overset{\delta}{\mapsto} 2, \text{ where } \delta(a,b) = \begin{cases} 1 & a = b \\ 0 & a \neq b \end{cases}$$

(iii) If φ is $\perp$, then φ_A is the constant 0.

(iv) If φ is $\psi \Rightarrow \chi$, then φ_A is the composite

$$A^n \xmapsto{(\psi_A, \chi_A)} 2^2 \xmapsto{\Rightarrow_2} 2,$$

where $\Rightarrow_2$ is the truth table of $\Rightarrow$, i.e. $\Rightarrow_2 (x, y) = 0$ iff $x = 1, y = 0$.

(v) If φ is $\forall x_{n+1}\psi$, we interpret ψ relative to $\{x_1, x_2, \cdots, x_{n+1}\}$ as a subset $[\![\psi_A]\!]_A^{(n+1)}$ of A^{n+1}, and then put

$$[\![\varphi]\!]_A^{(n)} = \{(a_1, \cdots, a_n) : \forall a_{n+1} \in A, (a_1, \cdots, a_{n+1}) \in [\![\psi_A]\!]_A^{(n+1)}\}.$$

Definition. Two formulae φ, ψ are α-equivalent if we can obtain one from the other by renaming (some of) its bound variables. If φ and ψ are α-equivalent, then they have the same interpretation in any context $x_1, \cdots, x_n$.

Given a formulae φ, a variable x and a term t, define $\varphi[t|x]$ to be the formula obtained from φ by substituting a copy of t for each free occurrence of x in φ, provided no variable occurring in t is bound in φ (otherwise first replace φ by an α-equivalent formula).

Similarly, given a finite string $\vec{x} = (x_1, \cdots, x_n)$ and a string $\vec{t} = (t_1, \cdots, t_n)$, we define $\varphi[\vec{t}\,|\,\vec{x}]$ as the result of simultaneously substituting t_i for all free occurrences of x_i.

φ is satisfied in structure A, written $A \models \varphi$, if φ_A is the constant 1 (equivalently, if $[\![\varphi]\!]_A^{(n)} = A^n$).

Note: $A \models \varphi$ iff $A \models (\forall x)\varphi$ for any free variables x of φ. In general, $A \models \varphi$ iff $A \models (\forall \vec{x})\varphi$, where $\vec{x} = (x_1, \cdots, x_n)$ is the string of all free variables of φ.

A closed formula or sentence with no free variables. $(\forall \vec{x})\varphi$ is the universal closure

Definition (Theory)**.** A *theory* over a signature Σ is a set of sentences $\mathbb{T} \subseteq \mathcal{L}(\Sigma)$. A structure A is a model of $\mathbb{T}$, written as $A \models \mathbb{T}$, if it satisfies all members of $\mathbb{T}$, called the axioms of the theory.

Example. The group theory has signature with $\Sigma = \{m(2), i(1), e(0)\}$, where the arities are in brackets. They correspond to multiplication, inverse and the identity respectively. $\Pi = 0$.

The theory contains the axioms

- $(\forall x, y, z) m(x, m(y, z)) = m(m(x, y), z)$ (Associativity)
- $(\forall x) m(e, x) = x$ (Identity)
- $(\forall x) m(i(x), x) = e$ (Inverse)

Example. The theory of posets has signature with $\Sigma = \emptyset$, $\Pi = \{\leq\}$ and axioms

- $(\forall x) x \leq x$
- $(\forall x, y, z)(x \leq y) \Rightarrow ((y \leq z) \Rightarrow (x \leq z))$
- $(\forall x, y)(x \leq y) \Rightarrow ((y \leq x) \Rightarrow (x = y))$

Definition (Semantic entailment)**.** If S is a set of sentences in $\mathcal{L}(\Sigma)$ and φ is a sentence, we say S semantically entails φ, written as $S \models \varphi$, if for all Σ-structures A such that $A \models S$, we also have $A \models \varphi$.

For (sets of) formulae with free variables, we say $S \models \varphi$ if, given any Σ-structure A and an assignment of values in A to the free variables in φ making φ false, we can extend to an assignment of values to the free variables in all members of S making at least one member of S false.

Definition ((First-order) predicate calculus)**.** The (first-order) predicate calculus has the following axioms:

- $\varphi \Rightarrow (\psi \Rightarrow \varphi)$ (K)
- $(\varphi \Rightarrow (\psi \Rightarrow \chi)) \Rightarrow ((\varphi \Rightarrow \psi) \Rightarrow (\varphi \Rightarrow \chi))$ (S)
- $\neg\neg\varphi \Rightarrow \varphi$ (for φ any formula of $\mathcal{L}(\Sigma)$) (T)
- $(\forall x)\varphi \Rightarrow \varphi[t|x]$ (φ any formula with x as free variable, t any term) (I)
- $(\forall x)(\varphi \Rightarrow \psi) \Rightarrow (\varphi \Rightarrow (\forall x)\psi)$ (x free in ψ, not in φ) (U)
- $(\forall x)(x = x)$ (R)
- $(\forall x, y)((x = y) \Rightarrow (\varphi \Rightarrow \varphi[y|x]))$ (x free in φ) (E)

and the following rules of inference

- (MP) from φ and $(\varphi \Rightarrow \psi)$, we may infer ψ, provided either ψ has a free variable or ϕ does not.
- (Gen) from φ we may infer $(\forall x)\varphi$, provided x appears free in φ but not in any hypothesis used in deducing φ.

By a deduction from a set S of hypotheses, we mean a finite list $\varphi_1, \varphi_2, \cdots, \varphi_n$ of formulae, each of which is either an axiom, a member of S, or obtainable from (an) earlier formula(e) by either (MP) or (Gen).

We say $S \vdash \varphi$ if there is a deduction from S whose last member is φ.

Proposition (Soundness theorem)**.** Suppose either $S \cup \{\varphi\}$ is a set of sentences or that φ has a free variable. If $S \vdash \varphi$, then $S \models \varphi$.

Proof. Like the soundness theorem for propositional logic - check that each axiom is a tautology and the rules of inference are sound. □

Proposition (Deduction theorem)**.** Suppose either that ψ has a free variable or that φ does not. Then $S \cup \{\varphi\} \vdash \psi$ iff $S \vdash (\varphi \Rightarrow \psi)$.

Theorem (Completeness theorem)**.** Suppose either that $S \cup \{\varphi\}$ consists of sentences or that φ has a free variable. Then $S \models \varphi$ implies $S \vdash \varphi$.

Proof. First reduce to the case of sentences, by replacing all free variables in $S \cup \{\varphi\}$ by new constants. If the resulting hypotheses and conclusion are S' and φ', then

(i) from $S \models \varphi$, we can deduce $S' \models \varphi'$

(ii) form $S' \vdash \varphi'$, we can deduce $S \vdash \varphi$

Second, reduce to the case $\varphi = \perp$ using the deduction theorem, just as in the propositional case. We prove the contrapositive: if S is a consistent set of sentences (i.e. $S \nvdash \perp$, then S has a model (i.e. $S \not\models \perp$)

Suppose we are given a consistent theory $\mathbb{T}_0$ in a language $\mathcal{L}(\Sigma_0)$.

We define increasing sequences of signatures Σ_n and theories $\mathbb{T}_n \subseteq \mathcal{L}(\Sigma_n)$ as:

- For even n, set $\Sigma_{n+1} = \Sigma_n$.

  Let $\mathbb{T}_{n+1}$ be a maximal consistent extension of $\mathbb{T}_n$.

- For odd n, let E_n be the set of formulae $\varphi \in \mathcal{L}(\Sigma_n)$ with one free variable x, such that $\mathbb{T}_n \vdash (\exists x)\varphi$.

  Set $\Sigma_{n+1} = \Sigma_n \cup \{c_\varphi : \varphi \in E_n\}$, where the c_φs are constant symbols not in Σ_n.

  Set $\mathbb{T}_{n+1} = T_n \cup \{\varphi[c_\varphi|x] : \varphi \in E_n\}$.

  We need to know $\mathbb{T}_{n+1}$ is consistent in this case. Consider adding a single row witness. So suppose $\mathbb{T} \models (\exists x)\varphi$, c is a constant not occurring in any member of $\mathbb{T}$, and $\mathbb{T} \cup \{\varphi[c_\varphi|x]\} \vdash \perp$.

  By the Deduction Theorem, we have $\mathbb{T} \models \neg\varphi[c_\varphi|x]$.

  We can rewrite this deduction, replacing all c_φs by xs, to get $\mathbb{T} \vdash \neg\varphi$.

  By (Gen), $\mathbb{T} \vdash (\forall x)\neg\varphi$.

  But we know $\mathbb{T} \vdash \neg(\varphi x)\neg\varphi$. So $\mathbb{T} \vdash \perp$.

  Hence by induction we can add witnesses for any finite number of existential formulae without destroying consistency.

  Since a deduction of $\perp$ from $\mathbb{T}_{n+1}$ would use only finitely many of the new axioms, $\mathbb{T}_{n+1}$ is consistent.

Now define $\Sigma_\infty = \bigcup_{n\geq 0} \Sigma_n$ and $\mathbb{T}_\infty = \bigcup_{n\geq 0} \mathbb{T}_n$.

Then $\mathbb{T}_\infty$ is consistent, since it is the union of a chain of consistent sets. $\mathbb{T}_\infty$ is maximally consistent, since for any $\varphi \in \mathcal{L}(\Sigma_\infty)$, $\varphi \in \mathcal{L}(\Sigma_{2n})$ for some n. So either φ or $\neg\varphi$ belongs to $\mathbb{T}_{2n+1}$.

Similarly, T_∞ is deductively closed.

And $\mathbb{T}_\infty$ has witnesses: if $\mathbb{T}_\infty \vdash (\exists x)\varphi$, then $\exists n$ such that $\varphi \in \mathcal{L}(\Sigma_{2n+1})$ and $\mathbb{T}_{2n+1} \vdash (\exists x)\varphi$. So $\mathbb{T}_{2n+2} \vdash \varphi[c_\varphi|x]$.

Now suppose $\mathbb{T} \subseteq \mathcal{L}(\Sigma)$ is a maximal consistent set of sentences, and has witnesses. Set $C = \{\text{closed terms over } \Sigma\}$ and factor C by $\sim$, where $s \sim t$ iff $\mathbb{T} \vdash (s = t)$.

On the set $A = C/\sim$, we interpret the operation and predicate symbols of Σ by $\omega_A([t_1], [t_2], \cdots, [t_n]) = [\omega t_1 t_2 \cdots t_n]$ and $([t_1], [t_2], \cdots [t_n]) \in [\![\pi]\!]_A$ iff $\mathbb{T} \vdash \pi(t_1, \cdots, t_n)$.

We can show that for any formula φ with free variables $x_1, \cdots, x_n$, we have $([t_1], \cdots, [t_n]) \in [\![\varphi]\!]_A$ iff $\mathbb{T} \vdash \varphi[t_1, \cdots, t_n|x_1, \cdots x_n]$.

Therefore, for any sentence φ, we have $[\![\varphi]\!]_A = 1$ iff $\mathbb{T} \vdash \varphi$ iff $\varphi \in \mathbb{T}$.

In particular, A is a model of $\mathbb{T}$. □

Corollary (Compactness theorem)**.** If $\mathbb{T}$ is a set of sentences and any finite subset of $\mathbb{T}$ has a model, then $\mathbb{T}$ has a model.

Proof. This is obvious if we replace 'has a model' by 'is consistent'. □

Corollary (Upward Löwenheim-Skolem Theorem)**.** If $\mathbb{T}$ has an infinite model, or $\mathbb{T}$ has finite models of arbitrarily large cardinality, then for any set I there is a $\mathbb{T}$ model A such that I injects into the underlying set of A. (i.e. there are models of arbitrarily large cardinality)

Proof. Add new constants $\{c_i : i \in I\}$ to the language, and let $\mathbb{T}' = \mathbb{T} \cup \{\neg(c_i = c_j) : i \neq j \in I\}$.

Any finite subset of $\mathbb{T}'$ has a model, since we can assign distinct values to the members of a finite subset of $\{c_i : i \in I\}$ in some $\mathbb{T}$-model.

So $\mathbb{T}'$ has a model A. Then we can inject $I \to A$ by $i \mapsto c_i$. □

Corollary (Downward Löwenheim-Skolem THeorem)**.** Suppose Σ is a countable signature, and that $\mathbb{T}$ is a theory in $\mathcal{L}(\Sigma)$ which has an infinite model. Then $\mathbb{T}$ has a countably infinite model.

Proof. Add constants $\{c_n : n \in \mathbb{N}$ to Σ and let $\mathbb{T}' = \mathbb{T} \cup \{\neg(c_m = c_n) : m \neq n\}$.

Then $\mathbb{T}'$ has a model. But the language of $\mathbb{T}'$ is still countable. So the construction in the proof of the completeness theorem produces a countable model of $\mathbb{T}'$, which must be countably infinite. □

In fact, the Downward Löwenheim-Skolem Theorem says that any infinite model of a first-order theory $\mathbb{T}$ has a countable structure which is still a model of $\mathbb{T}$.

The Löwernhei-Skolem Theorems tell us for any infinite structure A, we cannot have a first-order theory whose only model (up to isomorphism) is A.

Peano's Postulates for $\mathbb{N}$ (1899):

(i) 0 is a natural number

(ii) Every natural number has a successor

(iii) 0 is not a successor

(iv) Distinct natural numbers have distinct successors

(v) If P is a property of natural numbers which holds for 0, and holds for successor of n whenever it holds for n, then P holds $\forall \mathbb{N}$.

In modern language, Σ contains a constant 0 and a unary operation S. Then we have

(iii) $(\forall x)\neg(sx = 0)$

(iv) $(\forall x, y)((sx = sy) \Rightarrow (x = y))$

(v) $(\forall y1, \cdots, y_n)(\varphi[0/x] \Rightarrow ((\forall x)(\varphi \Rightarrow \varphi[sx|x]) \Rightarrow (\forall x)\varphi))$ for all $\varphi \in \mathcal{L}(\Sigma)$ with $\{x, y, \cdots, y_n\}$ the free variables of φ.

To get first-order Peano arithmetic, we add binary operation symbols a, m and axioms

- $(\forall x)(a(x, 0) = x)$
- $(\forall x, y)(a(x, s(y)) = s(a(x, y))$
- $(\forall x)(m(x, 0) = 0)$
- $(\forall x, y)(m(x, s(y)) = a(m(x, y), x))$

This theory has $\mathbb{N}$ as a model, but it also has an uncountable model.

Similarly, $\mathbb{R}$ is the unique (up to isomorphism) model of the theory of conditionally complete ordered fields (i.e. every non-empty bounded set has a least upper bound)

We can replace this by a scheme of axioms in the language of ordered rings, but the resulting theory has a countable models (e.g. the field of real algebraic numbers)

Definition (Categorical theory)**.** A first-order theory $\mathbb{T}$ is countably categorical, if any two countable models are isomorphic.

Similarly, a first-order theory $\mathbb{T}$ is κ-categorical for any infinite cardinal κ, if any two models of order κ are isomorphic.

Definition. Two models of a first-order theory are elementary equivalent, if they satisfy the same sentences. Clearly any two $\mathbb{T}$ models are elementary equivalent $\Leftrightarrow$ $\mathbb{T}$ is complete, i.e. $\forall\varphi$, either $\mathbb{T} \models \varphi$ or $\mathbb{T} \models \neg\varphi$.

4 Zermelo-Fraenkel Set Theory

Definition (Zermelo Set Theory)**.** Zermelo Set Theory is the first-order theory over a signature with one binary predicate $\in$ and the following axioms:

(i) Extensionality: $(\forall x, y)((\forall z)((z \in x) \Leftrightarrow (z \in y)) \Rightarrow (x = y))$. Intuitively, it says $x = y$ iff they have the same elements.

(ii) Separation scheme: $(\forall w_1, \cdots w_n)(\forall x)(\exists y)(\forall z)((z \in y \Leftrightarrow ((z \in x) \wedge \varphi))$ for any formula φ with free variables $z, w_1, \cdots, w_n$. Intuitively, it says $\{z \in x : \varphi\}$ exists.

(iii) Empty set: $(\exists x)(\forall y)\neg(y \in x)$. Intuitively, it says the empty set exists.

(iv) Pair set $(\forall x, y)(\exists z)(\forall w)((w \in z) \Rightarrow ((w = x) \vee (w = y))$. Intuitively, it says $\{x, y\}$ exists.

(v) Union set: $(\forall x)(\exists y)(\forall z)((z \in y) \Leftrightarrow (\exists y)((z \in w) \wedge (w \in x)))$. Here x is a collection of sets, and it says that the union of all the sets in x exists.

(vi) Power set: $(\forall x)(\exists y)(\forall z)((z \in y) \Leftrightarrow (\forall w)((w \in z) \Rightarrow (w \in x)))$. Intuitively, it says that the power set of x exists.

These axioms allow us to introduce a constant symbol φ and a binary operation symbol $(x, y) \mapsto \{x, y\}$, and two unary operations $\cup$ and $\mathcal{P}$. We abbreviate $\{x, x\}$ to $\{x\}$ and $\cup\{x, y\}$ to $(x \cup y)$.

For any $x \neq \emptyset$, we can define $\bigcap x$ to be $\{z \in y : (\forall w)(w \in x \Rightarrow z \in w)\}$

So we can define $x \cap y = \bigcap\{x, y\}$ and $x \setminus y = \{z \in x : \neg(z \in y)\}$.

Definition (Kuratowski ordered pair)**.** The Kuratowski ordered pair $\langle x, y\rangle$ is $\{\{x\}, \{x, y\}\}$.

Definition.

$$\text{First}(t) = \begin{cases} \bigcup\bigcap t & t \neq \emptyset, \\ \emptyset & \text{otherwise} \end{cases}$$

$$\text{Second}(t) = \begin{cases} \bigcup(\bigcup t \setminus \bigcap t) & \bigcup t \setminus \bigcap t \neq \emptyset \\ \text{First}(t) & \text{otherwise} \end{cases}$$

't is an ordered pair' means '$t = \langle \text{First}(t), \text{Second}(t)\rangle$'

Definition (Cartesian product)**.**

$$x \times y = \{z \in \mathcal{PP}(x \cup y) : (z \text{ is an ordered pair}) \wedge (\text{First}(z) \in x) \wedge (\text{Second}(z) \in y)\}$$

To complete the axioms of Zermelo Set Theory, we need to add

Axiom (Axiom of infinity)**.** Infinity: $(\exists x)((\emptyset \in x) \wedge (\forall y)((y \in x) \Rightarrow (y^+ \in x)))$, where $y^+ = y \cup \{y\}$.

Given this, there is a unique smallest x with this property, which we denote by ω.

4.1 Classes

Given a first-order formula φ with one free variable x, we think of it as defining the class of all x's satisfying φ. If $(\forall x)(\varphi \Leftrightarrow \psi)$ holds, then we think of φ and ψ as defining the same class.

Write M for a typical class $t \in M$ for $\varphi[t|x]$, where φ is a formula defining M.

M is a set if $(\exists y)(\forall x)((x \in M) \Leftrightarrow (x \in y))$. M is a proper class if it is not a set.

Similarly, we think of a formula with n free variables extensionally, as a class of n-tuples.

A class of n-tuples is functional (or call it a function-class) if

$$(\forall x_1, \cdots x_{n-1}, y, z)(((\langle x_1, \cdots, x_{n-1}, y\rangle \in M) \wedge (\langle x_1, \cdots, x_{n-1}, z\rangle \in M)) \Rightarrow (y = z)).$$

i.e. it behaves like a function.

Definition (Zermelo-Fraenkel set theory)**.** Zermelo-Fraenkel set theory is obtained from Zermelo set theory by adding the axiom-scheme of replacement and the axiom of foundation:

(i) Replacement: for any formula φ with free variables x, y (and other free variables automatically universally generalized),

$$(\forall y, y')[(\varphi \wedge \varphi[y'|y]) \Rightarrow (y = y')] \Rightarrow \{(\forall u)(\exists v)[(\forall y)(y \in v) \Leftrightarrow (\exists x)((x \in u) \wedge \varphi)]\}$$

Intuitively, it states that if φ is a class-function and u is a set, then the image of φ is also a set.

(ii) Foundation:

$$(\forall x)((x \neq \emptyset) \Rightarrow (\exists y \in x)(y \cap x = \emptyset))$$

This shows that we cannot have $x \in x$. Otherwise, apply this axiom to $\{x\}$. It says that there is an element of $\{x\}$ which is disjoint from $\{x\}$. Since x is the only member of $\{x\}$, x is disjoint from $\{x\}$. Since $x \in \{x\}$, $x \notin x$.

Definition (Transitive set)**.** A set x is transitive if

$$(\forall y, z)\{[(z \in y) \wedge (y \in x)] \Rightarrow (z \in x)\}$$

i.e. membership is a transitive relation among members of x.

Equivalently. $x \subseteq \mathcal{P}x$ or $\bigcup x \subseteq x$.

Clearly any intersection of transitive sets is transitive. So if there is any transitive set containing x, there must be a smallest one. TC(x) is the smallest transitive set containing x.

Theorem. In the presence of the other axioms of Zermelo-Fraenkel, Foundation is equivalent to the scheme of $\in$-induction:

$$(\forall x)((\forall y)((y \in x) \Rightarrow \varphi[y|x]) \Rightarrow \varphi) \Rightarrow (\forall x)\varphi$$

where φ is any formula with free variable x (and other free variables automatically universally generalized).

Proof. Suppose $\in$-induction holds. Define 'x is a regular set' as

$$(\forall y)((x \in y) \Rightarrow (\exists z \in y)(z \cap y = \emptyset))$$

i.e. "sets that contain x satisfy the axiom of foundation".
Then Foundation is equivalent to the assertion $(\forall x)(x$ is regular$)$.

We can prove this by $\in$-induction. Consider x. Suppose $(\forall y)((y \in x) \Rightarrow$ $(y$ is regular $))$ and $x \in z$. If $x \cap z = \emptyset$, then we are done.

Otherwise, $(\exists y)((y \in x) \wedge (y \in z))$. But $y \in x$ implies y is regular. So $(\exists y' \in z)(y' \cap z = \emptyset)$. So x is regular.

Conversely, suppose that the induction hypothesis of $\in$-induction is true, but $\neg(\forall x)\varphi$. i.e. $(\exists x)\neg\varphi$. Consider the set

$$t = \{y \in \mathrm{TC}(\{x\}) : \neg\varphi[y|x]\}$$

This set is non-empty, because it contains at least x. So

$$(\exists y \in t)(\forall z \in y)(\neg(z \in t)).$$

But $z \in y$ implies $z \in \mathrm{TC}(\{x\})$ by transitivity. And $z \notin t$ implies $\varphi[z|x]$.

So we have $(\forall z \in y)\varphi[z|x]$. But not $\neg\varphi[y|x]$, contradicting the induction hypothesis. □

Definition (Locality and well-foundedness of classes)**.** Let R be a relation-class (i.e. a class of pairs) and M be a class.

R is *well-founded* on M if

$$(\forall x \subseteq M)((x \neq \emptyset) \Rightarrow (\exists y \in x \cap M)((\forall z \in M)((\langle z, y\rangle \in R) \Rightarrow \neg(z \in x))$$

Intuitively, seeing R as a partial order, given any subset x of M, there exists a minimal element y of x.

R is local if for any $x \subseteq M$, we can construct the $\downarrow R(x)$ of all those $y \in M$ such that $\langle y, z\rangle \in R$ for some $z \in x$ (i.e. the set of all things that are related to some member of x)

We can hence construct the set $RC(x) = \bigcup\{x, \downarrow R(x), \downarrow R(\downarrow R(x)), \cdots\}$ which is the smallest subset of M containing x which is R-closed, i.e. such that for any $z \in RC(x)$, $\langle y, z\rangle \in R$ implies $y \in RC(x)$.

Proposition (R-induction)**.** Suppose M is a class and R is a well-founded and local relation-class on M. Then

$$(\forall x \in M)((\forall y \in M)(\langle y, x\rangle \in R \Rightarrow \varphi[y|x]) \Rightarrow \varphi) \Rightarrow (\forall x \in M)\varphi$$

i.e. we can perform induction with R as the relation.

Lemma. Suppose R is a well-founded and local relation on a class M. Then there is a class of pairs $\bar{R} \supseteq R$ which is well-founded, local, and transitive on M.

Proof. We already saw that, for any set $x \subseteq M$, there is a set $RC_M(x)$ of all iterated R-predecessors of members of x lying on M.

So we define $\bar{R}$ as the set of all pairs $\langle x, y\rangle$ such that $x \in RC_M(\{y\})$. This is local by definition, and it is transitive since $x \in RC_M(\{y\})$ implies $RC_M(\{x\}) \subseteq RC_M(\{y\})$.

So we need to show it is well-founded.

Let $x \subseteq M$ be a non-empty set with no $\bar{R}$-minimal member.

Define $\bar{x} = x \cup \{y \in RC_M(x) : (\exists z \in RC_M(\{y\}))(z \in x)\}$.

$\bar{x}$ is non-empty, and if $\langle y, z\rangle \in R$ and $z \in \bar{x}$, then

- either $z \in x$, in which case it has an $\bar{R}$-predecessor in x, and has an R-predecessor either in x or in $\{y \in RC_M(x) : (\exists z \in RC_M(\{y\}))(z \in x)\}$;
- or $z \in \{y \in RC_M(x) : (\exists z \in RC_M(\{y\}))(z \in x)\}$, in which case it again has an R-predecessor in x.

So $\bar{x}$ has no R-minimal number. Contradiction. □

Theorem (R-Recursion theorem)**.** Let M be a class and R be a well-founded local relation-class.

Let G be a class of triples which is functional and satisfies

$$(\forall x, y)((x \in M) \Rightarrow (\exists! z)(\langle x, y, z\rangle \in G))$$

i.e. $G : M \times V \to V$. We write $G(x, y)$ for the unique z such that $\langle x, y, z\rangle \in G$.

Then there is a unique function class $F : M \to V$ satisfying

$$(\forall x \in M)(F(x) = G(x, \{F(y) : (y \in M) \wedge (\langle y, x\rangle \in R)\})) \qquad (*)$$

i.e. we can define functions recursively.

Proof. Suppose both F and F' both satisfy $(*)$. We can show uniqueness by proving $(\forall x \in M)(F(x) = F'(x))$ by R-induction over M.

To show existence, we consider attempts on F, i.e. sets f such that f is a function, $\operatorname{dom} f$ is a R-closed subset of M, and f satisfies the recursion relation.

If f and f' are attempts, then they agree in regions where their domains overlap by the uniqueness argument. So

$$F = \{\langle x, y\rangle : (\exists f)(f \text{ is an attempt}) \wedge (\langle x, y) \in f)\}$$

is a function class.

Suppose F is undefined at some $x \in M$. Then there is an $\bar{R}$ minimal member x_0 of $\{x \in M : F(x) \text{ undefined}\}$.

Define

$$f_0 = \{\langle x, y\rangle : (x \in RC_M(\{x_0\})) \wedge (\exists f)((f \text{ is an attempt}) \wedge (\langle x, y\rangle \in f)))\}$$

Then f_0 is an attempt, with domain $RC_M(\{x_0\})$, and we may extend it to

$$f_1 = f_0 \cup \{\langle x_0, G(x_0, \{F(y) : (y \in M) \wedge (\langle y, x_0\rangle \in R)\})\}$$

which is an attempt with $x_0 \in \operatorname{dom} f_1$. So F is defined on M. □

Definition (Extensional class)**.** A relation-class R is extensional on a class M if

$$(\forall x, y \in M)((\forall z \in M)((\langle z, x\rangle \in R) \Leftrightarrow (\langle z, y\rangle \in R)) \Rightarrow (x = y))$$

i.e. if x and y are related to the same things, then they are equal.

Theorem (Mostowski's Isomorphism Theorem)**.** Let a be as set and $r \subseteq a \times a$ an extensional, well-founded relation on a. Then there is a unique pair (b, f) where b is a transitive set, and $a \to b$ is a bijection, and

$$(\forall x, y \in a)(\langle x, y\rangle \in r \Leftrightarrow (f(x) \in f(y)))$$

Proof. Uniqueness: Suppose (b', f') also satisfy the conditions. Let g be the composite $b \overset{f^{-1}}{\longmapsto} a \overset{f'}{\longmapsto} b'$. Then $(\forall x, y \in b)((x \in y) \Leftrightarrow (g(x) \in g(y)))$. So $(\forall x \in b)(g(x) = x)$ by $\in$-induction over b. So $b' = b$, and $f' = f$.

Existence: We define f by r-recursion over a:

$$f(x) = \{f(y) : (y \in a) \wedge (\langle y, x\rangle \in r)\}$$

(i.e. we take $G(x, y) = y$ in the statement of the recursion theorem) and we define b to be the image $\{f(x) : x \in a\}$.

By definition, f is surjective and $((\langle x, y\rangle \in r) \Rightarrow (f(x) \in f(y))$.

To show $(f(x) \in f(y)) \Rightarrow (\langle x, y\rangle \in r)$, we need to show f is injective. Consider the formula φ with one free variable x:

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

We prove $(\forall x \in a)\varphi$ by r-induction:

Assume $(\forall z \in a)((\langle z, x\rangle \in r) \Rightarrow \varphi[z|x])$ and $f(x) = f(y)$. From $f(x) = f(y)$, we can deduce

$$(\forall z \in a)(((\langle z, x\rangle) \in r) \Rightarrow (\exists t \in a)(((\langle t, y\rangle \in r) \wedge (f(z) = f(t))))$$

But for any such z we have $\varphi[z|x]$. So we can deduce $z = t$.i.e.

$$(\forall z \in a)((\langle z, x\rangle \in r) \Rightarrow (\langle z, y\rangle \in r))$$

Similarly, we have

$$(\forall z \in a)((\langle z, y\rangle \in r) \Rightarrow (\langle z, x\rangle \in r))$$

By extensionality of r, we have $x = y$. □

Definition (Trichotomous relation)**.** A binary relation $r \subseteq a \times a$ is *trichotomous* if

$$(\forall x, y \in a)((\langle x, y\rangle \in r) \wedge (\langle y, x\rangle \in r) \wedge (x = y))$$

Note that a well-founded relation is necessarily irreflexive (i.e. $(\forall x)\neg(\langle x, x\rangle \in r)$), since $(\langle x, x\rangle \in r)$ would imply that $\{x\}$ has no r-minimal member.

And a well-founded trichotomous relation is necessarily transitive, since if we have $(\langle x, y\rangle \in r) \wedge (\langle y, z\rangle \in r)$ but not $(\langle x, z\rangle \in r)$, then we have $\langle(z, x) \in r)$ or $x = z$, and in either case $\{x, y, z\}$ has no r-minimal member.

In this case, we will normally write $(\langle x, y\rangle \in r)$ as $x < y$ and think of $(x < y) \vee (x = y)$ as a total ordering on a.

We call such a relation $<$ a well-ordering of a. Equivalently, a well-ordering is a strict total ordering in which every non-empty $b \subseteq a$ has a least member.

Corollary. For any well-ordered set $(a, <)$, there is a unique transitive set b such that $\in \cap b \times b$ is trichotomous, together with an isomorphism of ordered sets $(a, <) \overset{f}{\rightarrowtail} (b, \in \cap(b \times b))$.

5 Ordinals

Definition. An ordinal is a transitive set α that is trichotomous, i.e.

$$(\forall x, y \in \alpha)((x \in y) \wedge (y \in x) \wedge (x = y))$$

Example. $0 = \emptyset, 1 = \{\emptyset\}, 2 = \{\emptyset, \{\emptyset\}\}, 3 = \{0, 1, 2\}$ etc. are ordinals.

Lemma. If α is an ordinal, then so is $\alpha^+ = \alpha \cup \{\alpha\}$.

Proof. Transitivity: if $x \in y$, $y \in \alpha^+$, then either $y \in \alpha$, in which case $x \in \alpha \subseteq \alpha^+$ by transitivity of α, or $y = \alpha$, in which case $x \in \alpha \subseteq \alpha^+$. Trichotomy: if $x, y \in \alpha^+$, then either

- x, y in α, in which case trichotomy of α is inherited; or
- None of them is in α, in which case $x = y = \alpha$; or
- $x = \alpha$ and $y \in \alpha$, in which case $y \in x$ (or vice versa)

□

Lemma. Every member of an ordinal is an ordinal.

Proof. Suppose α is an ordinal and $x \in \alpha$.

Transitivity: if $y \in z \in x$, then $y, z \in \alpha$ by transitivity of α. Hence we have $(y \in x) \wedge (x \in y) \wedge (x = y)$ by trichotomy of α. But $(x \in y) \wedge (x = y)$ would contradict foundation. So $y \in x$.

Trichotomy: If $y, z \in x$, then $y, z \in \alpha$ by transitivity of α. So $(y \in z) \wedge (z \in y) \wedge (y = z)$ by trichotomy of α. □

Notation. On is the class of all ordinals (which is a transitive class by above).

Lemma. If $\alpha, \beta \in \mathrm{On}$, then either $\alpha \subseteq \beta$ or $\beta \subseteq \alpha$.

Proof. Suppose $\alpha \not\subseteq \beta$. Then $\alpha \setminus \beta$ is non-empty, so it has an ϵ-least member γ.

Now if $\delta \in \gamma$, then $\delta \in \alpha$ by transitivity. By minimality of γ, $\delta \notin \alpha \setminus \beta$. So $\delta \in \beta$. Hence $\delta \in \alpha \cap \beta$.

On the other hand, if $\delta \in \alpha \cap \beta$, then $(\delta \in \gamma) \wedge (\gamma \in \delta) \wedge (\gamma = \delta)$ by trichotomy of α. And either $\gamma \in \delta$ or $\gamma = \delta$ would imply $(\gamma \in \beta)$. So $\delta \in \gamma$.

Therefore $\delta \in \alpha \cap \beta \Leftrightarrow \delta \in \gamma$. So $\gamma = \alpha \cap \beta$. In particular, $\alpha \cap \beta \in \alpha$.

Similarly, if $\beta \not\subseteq \alpha$, then $\alpha \cap \beta \in \beta$.

So if both $\alpha \subseteq \beta$ and $\beta \subseteq \alpha$, then $\alpha \cap \beta \in \alpha \cap \beta$, contradicting Foundation. □

Corollary. For ordinals α, β, we have one of $(\alpha \in \beta), (\beta \in \alpha)$ or $(\alpha = \beta$.

$(\alpha \subseteq \beta)$ is equivalent to $(\alpha \in \beta) \vee (\alpha = \beta)$.

Proof. By above, we have one of $\alpha = \beta$, we have $\alpha = \beta$, $\alpha \subsetneq \beta$ or $\beta \subsetneq \alpha$. And $\alpha \subsetneq \beta$ implies $\alpha = \alpha \cap \beta \in \beta$ by the proof of above. □

Corollary. On is a proper class.

Proof. If On were a sets, then we would have $\mathrm{On} \in \mathrm{On}$, contradicting foundation. □

Corollary. If α is a subset of On, then $\bigcup \alpha \in \mathrm{On}$.

Proof. $\cup\alpha$ is transitive, since it is a union of transitive sets. The members of $\cup\alpha$ are ordinals and thus satisfy trichotomy. □

Theorem. Let M be any class satisfying $(\forall x)((x \in M) \Rightarrow (x^+ \in M))$ and $(\forall x)((x \subseteq M) \Rightarrow (\bigcup x \in M))$. Then $\mathrm{On} \subseteq M$.

Proof. By $\in$-induction: Suppose $\alpha \in \mathrm{On}$ and $(\forall \beta \in \alpha)(\beta \in M)$.

If α has an $\in$-greatest member, say β, then $(\forall \gamma)((\gamma \in \alpha) \Rightarrow ((\gamma \in \beta) \vee (\gamma = \beta)))$ by maximality of β. But we also have $(\forall \gamma)(((\gamma \in \beta) \vee (\gamma = \beta)) \Rightarrow (\gamma \in \alpha))$ by transitivity of α. So $\alpha = \beta^+$ by extensionality. But $\beta \in M$. So $\alpha \in M$.

If α has no greatest member, then $(\forall \beta)((\beta \in \alpha) \Rightarrow (\exists \gamma \in \alpha)(\beta \in \gamma))$, i.e. $\alpha \subseteq \bigcup \alpha$. But we also have $\bigcup \alpha \subseteq \alpha$ by transitivity. Hence $\alpha = \bigcup \alpha$. But $\alpha \subseteq M$ by induction hypothesis. So $\alpha \in M$. □

Definition (Successor and limit ordinals)**.** If $\alpha = \beta^+$ for some β, then α is a *successor ordinal.* Otherwise, α is a *limit ordinal.*

Note: 0 is a limit ordinal. The above theorem says that we can prove things by $\in$-induction over On, or define them by $\in$-recursion, by considering separately the cases of successor and limit ordinals.

For example, we define the function-class $(\alpha \mapsto V_\alpha) : \mathrm{On} \to V$ by $\in$-recursion:

(i) If $\alpha = 0$, then $V_\alpha = \emptyset$.

(ii) If $\alpha = \beta^+$, the $V_\alpha = \mathcal{P}(V_\beta)$.

(iii) If α is a limit, $V_\alpha = \cup\{V_\beta : \beta < \alpha\}$.

The sets V_α are called the von Neumann hierarchy of sets. We define the function-class rank: $V \to \mathrm{On}$ by $\in$-recursion.

Definition. $\mathrm{rank}(x) = \bigcup\{\mathrm{rank}(y)^+ : y \in x\}$. We can prove that the rank of any set is an ordinal, and the rank of an ordinal is itself by $\in$-induction.

Theorem. For all sets x and ordinals α, we have

- $(x \in V_\alpha) \Leftrightarrow (\mathrm{rank}(x) < \alpha)$
- $(x \subseteq V_\alpha) \Leftrightarrow (\mathrm{rank}(x) \leq \alpha)$

Proof. The second assertion follows from the first since $x \subseteq V_\alpha \Leftrightarrow x \in V_{\alpha^+}$ and $\mathrm{rank}(x) \leq \alpha \Leftrightarrow \mathrm{rank}(x) < \alpha^+$.

To prove the first assertion,

($\Rightarrow$) We use $\in$-induction on α. Suppose $(\forall x)(\forall \beta < \alpha)((x \in V_\beta) \Rightarrow \mathrm{rank}(x) < \beta$, and $x \in V_\alpha$.

If α is a limit, then $x \in V_\beta$ for some $\beta < \alpha$. So $\mathrm{rank}(x) < \beta < \alpha$.

If $\alpha = \beta^+$ is a successor, then $(\forall y \in x)(y \in V_\beta)$. So $(\forall y \in x)(\mathrm{rank}(y) < \beta)$. So $(\forall y \in x)(\mathrm{rank}(y)^+ \leq \beta)$. So $\mathrm{rank}(x) \leq \beta < \alpha$.

($\Leftarrow$). We use $\in$-induction on x. Suppose $(\forall y \in x)(\forall \alpha \in \mathrm{On})((\mathrm{rank}(y) < \alpha) \Rightarrow (y \in V_\alpha))$ and $\mathrm{rank}(x) \leq \beta < \alpha$.

Then $(\forall y \in x)(\mathrm{rank}(y) < \beta)$. So $(\forall y \in x)(y \in V_\beta)$. So $x \subseteq V_\beta$. So $x \in V_{\beta^+} \subseteq V_\alpha$. □

Proposition. $\alpha \in \mathrm{On}$ is equivalent to α is a transitive set, all of whose members are transitive.

Proof. ($\Rightarrow$) Members of ordinals are ordinals.

($\Leftarrow$) Proved by $\in$-induction. Suppose x is a transitive set, all of whose members are transitive. Suppose $(\forall y \in x)((y \text{ is transitive}) \Rightarrow (y \in \text{On}))$.

So x is a transitive set of ordinals. But the either $x = \beta^+$ for some $\beta \in x$, or $x = \bigcup x$, and so in either case, x is an ordinal. □

Definition (Ordinal addition and multiplication)**.** We can define them synthetically: we take $\alpha + \beta$ to be the order-type of the well-ordered set $\alpha \perp\!\!\!\perp \beta = \alpha \times \{0\} \cup \beta \times \{1\}$ ordered by $\langle \gamma, i\rangle < \langle \delta, j\rangle \Leftrightarrow (i < j)$ or ($i = j$ and $\gamma < \delta$), i.e. reverse lexicographic order. (the order type of a well-ordered set is the ordinal it is isomorphic to)

We define $\alpha\beta$ to be the order type of $\alpha \times \beta$, again under the reverse lexicographic order.

Lemma.

(i) Ordinal addition satisfies the recursive definition

(a) $\alpha + 0 = \alpha$

(b) $\alpha + \beta^+ = (\alpha + \beta)^+$

(c) $\alpha + \lambda = \bigcup\{\alpha + \gamma : \gamma < \lambda\}$ if λ is a non-zero limit.

(ii) Ordinal multiplication satisfies the recursive definition

(a) $\alpha \cdot 0 = 0$

(b) $\alpha \cdot \beta^+ = \alpha\beta + \alpha$

(c) $\alpha \cdot \lambda = \bigcup\{\alpha \cdot \gamma : \gamma < \lambda\}$ if λ is a non-zero limit.

Proof.

(i)

$$\begin{aligned}
\alpha + 0 &= \text{otp}(\alpha \times \{0\}) \\
&= \alpha \\
\alpha + (\beta^+) &= \text{otp}(\alpha \times \{0\} \cup \beta \times \{1\} \cup \{\langle \beta, 1\rangle\}) \\
&= \text{otp}(\alpha \times \{0\} \cup \beta \times \{1\})^+ \\
&= (\alpha + \beta)^+ \\
\alpha + \lambda &= \text{otp}(\alpha \times \{0\} \cup \lambda \times \{1\}) \\
&= \text{otp}(\alpha \times \{0\} \cup (\bigcup\{\gamma : \gamma < \lambda\} \times \{1\}) \\
&= \text{otp}(\bigcup\{\alpha \times \{0\} \cup \gamma \times \{1\} : \gamma < \lambda\}) \\
&= \text{otp}(\bigcup\{\alpha + \gamma : \gamma < \lambda\}) \\
&= \bigcup\{\alpha + \gamma : \gamma < \lambda\}
\end{aligned}$$

(ii) $\alpha \times 0 = \emptyset$. So $\text{otp}\{\alpha \times 0\} = 0$.

$\alpha \times (\beta^+) = \alpha \times \beta \cup \alpha \times \{\beta\}$. So

$$\begin{aligned}
\text{otp}(\alpha \times (\beta^+)) &= \text{otp}((\alpha \times \beta) \perp\!\!\!\perp \alpha) \\
&= \text{otp}(\alpha\beta \perp\!\!\!\perp \alpha) \\
&= \alpha\beta + \alpha
\end{aligned}$$

For a limit λ, we have $\alpha \times \lambda = \bigcup\{\alpha \times \gamma : \gamma < \lambda\}$.

By induction, we have order isomorphisms $\alpha \times \gamma \xmapsto{f_\gamma} \alpha\gamma$ for each $\gamma < \lambda$. And if $\gamma < \delta$, then $\alpha \times \gamma$ is an initial segment of $\alpha \times \gamma$. So f_γ and f_δ agree where both are defined. So the f_γ can be patched together to produce an order-isomorphism $\alpha \times \gamma \mapsto \cup\{\alpha\gamma : \gamma < \lambda\}$.

□

Lemma.

(i) If $\beta < \gamma$, then $\alpha + \beta < \alpha + \gamma$.

(ii) If $\alpha \leq \beta$, then $\alpha + \gamma \leq \beta + \gamma$.

(iii) If $\alpha > 0$ and $\beta < \gamma$, then $\alpha\beta < \alpha\gamma$.

(iv) If $\alpha \leq \beta$, then $\alpha\gamma \leq \beta\gamma$.

Proof. We prove (i) and (iii) using the synthetic approach: If $\beta < \gamma$, then $\alpha \perp\!\!\!\perp \beta$ is a proper initial segment of $\alpha \perp\!\!\!\perp \gamma$. So $\mathrm{otp}(\alpha \perp\!\!\!\perp \beta) < \mathrm{otp}(\alpha \perp\!\!\!\perp \gamma)$.

Similarly, if $\alpha \neq 0$, then $\alpha \times \beta$ is a proper initial segment of $\alpha \times \gamma$. So $\mathrm{otp}(\alpha \times \beta) < \mathrm{otp}(\alpha \times \gamma)$.

We prove (ii) and (iv) by induction on γ. If $\gamma = 0$, then $\alpha + \gamma = \alpha \leq \beta = \beta + \gamma$. If $\gamma = \delta^+$, then $\alpha + \delta \leq \beta + \delta$ by the induction hypothesis. So $\alpha + \gamma = (\alpha + \delta)^+ \leq (\beta + \gamma)^+ = \beta + \delta$.

If γ is a limit, then by the induction hypothesis, we have $\alpha + \delta \leq \beta + \delta$ for all $\delta < \gamma$. So $\alpha + \delta = \bigcup\{\alpha + \delta : \delta < \gamma\} \leq \bigcup\{\beta + \delta : \delta < \gamma\} = \beta + \gamma$.

Similarly, if $\gamma = 0$, then $\alpha \cdot 0 = 0 = \beta \cdot 0$.

If $\gamma = \delta^+$, then $\alpha \cdot \delta \leq \beta \cdot \delta$ by the induction hypothesis, and $\alpha \leq \beta$. So $\alpha\gamma = \alpha\delta + \alpha \leq \beta \cdot \delta + \beta = \beta \cdot \gamma$.

If γ is a limit, then $\alpha\delta \leq \beta\delta$ for all $\delta < \gamma$. So $\alpha\delta = \cup\{\alpha\delta : \delta < \gamma\} \leq \bigcup\{\beta\delta : \delta : \gamma\} = \beta\gamma$. □

Lemma. (i) $0 + \alpha = \alpha$ and $0 \cdot \alpha = 0$

(ii) $1 \cdot \alpha = \alpha \cdot 1 = \alpha$

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

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

(v) $(\alpha\beta)\gamma = \alpha(\beta\gamma)$

Proof. We can prove (iv) with the synthetic approach: $\alpha \cdot (\beta + \gamma = \mathrm{otp}(\alpha \times (\beta \perp\!\!\!\perp \gamma))$, $\alpha \times (\beta \perp\!\!\!\perp \gamma)$ contains elements of the form $\langle \delta, \langle \epsilon, i \rangle$.

$\alpha\beta + \alpha\gamma = \mathrm{otp}((\alpha \times \beta) \perp\!\!\!\perp (\alpha \times \gamma))$, and $(\alpha \times \beta) \perp\!\!\!\perp (\alpha \times \gamma)$ contains elements of the form $\langle\langle \delta, \epsilon \rangle, i \rangle$. The obvious bijection $\langle \delta, \langle \epsilon, i \rangle\rangle \mapsto \langle\langle \delta, \epsilon \rangle, i \rangle$ is an order isomorphism. So $\mathrm{otp}(\alpha \times (\beta \perp\!\!\!\perp \gamma)) = \mathrm{otp}((\alpha \times \beta) \perp\!\!\!\perp (\alpha + \gamma))$ and the result follows.

We can prove (v) by induction on γ. □

Note: $\omega + 1 = \omega^+ > \omega$ but $1 + \omega = \bigcup\{1 + n : n < \omega\} = \omega$, and $\omega \cdot 2 = \omega + \omega > \omega$, while $2 \cdot \omega = \omega$. Also $(1+1)\omega = \omega \neq 1 \cdot \omega + 1 \cdot \omega$.

Lemma (Division Algorithm)**.** If $\alpha, \beta \in \mathrm{On}$ and $\beta \neq 0$, then $\exists$ a unique $\gamma, \delta \in \mathrm{On}$ with $\delta < \beta$ and $\alpha = \beta\gamma + \delta$.

Proof. First, since $\beta \geq 1$, we have $\beta \cdot \epsilon \geq \epsilon$. So $\exists \epsilon$ such that $\beta\epsilon > \alpha$. Hence there is a least such ϵ. Analysis shows that this must be a successor. Say $\epsilon = \gamma^+$. Then $\beta \cdot \gamma \leq \alpha < \beta \cdot \gamma^+ = \beta \cdot \gamma + \beta$.

Set $\delta = \mathrm{otp}(\alpha \setminus \beta \cdot \gamma)$. Then $\mathrm{otp}(\beta \cdot \gamma \perp\!\!\!\perp \delta) = \alpha$. i.e. $\alpha = \beta \cdot \gamma + \delta$. And $\beta\gamma + \beta > \alpha = \beta\gamma + \delta$.

Conversely, to show uniqueness, if $\alpha = \beta\gamma + \delta$ and $\delta < \beta$, then $\beta\gamma \leq \alpha < \beta\gamma^+$. So γ^+ is the least ϵ such that $\beta\epsilon > \alpha$. Then γ is uniquely determined as $\mathrm{otp}(\alpha \setminus \beta \cdot \gamma)$. □

Definition. $[\beta, \alpha]_f = \{f : \beta \to \alpha : \{\gamma : f(\gamma) \neq 0\} \text{ is finite}\}$.

Lemma. $[\beta, \alpha]_f$ is well-ordered by reverse lexicographic ordering, i.e. $f < g \Leftrightarrow f(\gamma) < g(\gamma)$ for the largest γ such that $f(\gamma) \neq g(\gamma)$.

Proof. By induction on β. Let $S \subseteq [\beta, \alpha]_f$ be a non-empty subset. Pick $f \in S$. If f is identically 0, then it is the least element of S.

Otherwise, there is a largest γ such that $f(\gamma) \neq 0$. Let $S_1 = \{g \in S : g(\gamma') = 0 \forall \gamma' > \gamma\}$. Then $S_1 \neq \emptyset$ since $f \in S_1$, and it is an initial segment of S. So if it has a least member, then that will be the least member of S.

Now consider $\{g(\gamma) : g \in S_1\}$. This is a non-empty subset of α, so has a least member δ. Set $S_2 = \{g \in S_1 : g(\gamma) = \delta\}$. Again, $S_2 \neq \emptyset$ and it is an initial segment of S_1. Now the mapping $g \mapsto g|_\gamma$ is an order-isomorphism from S_2 to a non-empty subset of $[\gamma, \alpha]_f$. So by the induction hypothesis, S_2 has a least element, which is the least element of S. □

Definition (Ordinal exponentiation)**.** $\alpha^\beta = \mathrm{otp}([\beta, \alpha]_f)$.

Lemma. Ordinal exponentiation satisfies the recursive definition

- $\alpha^0 = 1$
- $\alpha^{\beta^+} = \alpha^\beta \cdot \alpha$
- $\alpha^\lambda = \bigcup\{\alpha^\gamma : \gamma < \lambda\}$

Proof.

(i) There is a unique function $\emptyset \to \alpha$, and it has finite support. So $\alpha^0 = \mathrm{otp}([\emptyset, \alpha]_f) = 1$.

(ii) We have an order-isomorphism $[\beta^+, \alpha]_f \to [\beta, \alpha]_f \times \alpha$ by $f \mapsto \langle f|_\beta, f(\beta)\rangle$. So $\alpha^{\beta^+} = \mathrm{otp}([\beta^+, \alpha]_f) = \mathrm{otp}([\beta, \alpha]_f \times \alpha) = \alpha^\beta \times \alpha$.

(iii) Any function $\lambda \to \alpha$ of finite support has support $\subseteq \gamma$ for some $\gamma < \lambda$. So $[\gamma, \alpha]_f = \bigcup\{S_\gamma : \gamma < \lambda\}$, where S_γ is order-isomorphic to $[\gamma, \alpha]_f$ and the S_γ are all initial segments of $[\gamma, \alpha]_f$. So $\alpha^\gamma = \bigcup\{\alpha^\gamma : \gamma < \lambda\}$ by patching together the Mostowski isomorphism of the subsets S_γ.

□

Lemma. Ordinal exponentiation satisfies the identities $\alpha^{\beta+\gamma} = \alpha^\beta \cdot \alpha^\gamma$ and $\alpha^{\beta\gamma} = (\alpha^\beta)^\gamma$.

Proof. Induction for the first:

$$\begin{aligned}
\alpha^{\beta+0} &= \alpha^{\beta} \\
&= \alpha^{\beta} \cdot 1 \\
&= \alpha^{\beta} \cdot \alpha^{0} \\
\alpha^{\beta+\gamma^{+}} &= \alpha^{(\beta+\gamma)^{+}} \\
&= \alpha^{\beta+\gamma} \cdot \alpha \\
&= (\alpha^{\beta} \cdot \alpha^{\gamma}) \cdot \alpha \text{ (by induction hypothesis)} \\
&= \alpha^{\beta} \cdot (\alpha^{\gamma} \cdot \alpha) \\
&= \alpha^{\beta} \cdot \alpha^{\gamma^{+}} \\
\alpha^{\beta+\lambda} &= \alpha^{\bigcup\{\beta+\gamma : 0<\gamma<\lambda\}} \\
&= \alpha^{\bigcup\{\delta : 0<\delta<\beta+\lambda\}} \\
&= \bigcup\{\alpha^{\delta} : 0 < \delta < \beta+\lambda\} \\
&= \bigcup\{\alpha^{\beta+\gamma} : 0 < \gamma < \lambda\} \\
&= \bigcup\{\alpha^{\beta}\alpha^{\gamma} : 0 < \gamma < \lambda\} \\
&= \alpha^{\beta} \cdot \bigcup\{\alpha^{\gamma} : 0 < \gamma < \lambda\} \\
&= \alpha^{\beta} \cdot \alpha^{\lambda}
\end{aligned}$$

We prove synthetically for the second. There is a bijection $[\gamma, [\beta, \alpha]_f]_f \mapsto [\beta \times \alpha, \alpha]_f$ with $f \mapsto \hat{f} = (\langle \delta, \epsilon \rangle \mapsto f(\epsilon)(\delta))$, since $\hat{f}$ has finite support $\Leftrightarrow$ only finitely many $f(\epsilon)$ are not identically 0, and all $f(\epsilon$ have finite support. This is a bijection when both sets are ordered by reverse lexicographic ordering. So $(\alpha^{\beta})^{\gamma} = \alpha^{(}\beta\gamma)$. □

Note: We do not have $(\alpha \cdot \beta)^{\gamma} = \alpha^{\gamma}\beta^{\gamma}$ in general, e..g $(\omega \cdot 2)^2 = \omega \cdot (2 \cdot \omega) \cdot 2 = \omega \cdot \omega \cdot 2 = \omega^2 \cdot 2 \neq \omega^2 \cdot 2^2$.

6 Choice, Well-Ordering and Cardinal Arithmetic

Lemma (Hartog's Lemma)**.** For any set a, there is an ordinal which cannot be mapped injectively to a.

Proof. Consider the set $S = \{r \subseteq a \times a : r \text{ is a well-ordering of a subset of } a\}$. Mostowski yields a surjective function-class from S to a class of ordinals which can be injected into a. By replacement, the latter is a set $\gamma(a)$. It is an ordinal since $\alpha \leq \beta \in \gamma(a)$ implies $\alpha \in \gamma(a)$ Since $\gamma(a) \notin \gamma(a)$, it cannot be injected into a. □

Corollary (Bourbaki-Witt Theorem)**.** Let P be a chain-complete poset and $f : P \to P$ be an inflationary map. Then for every $x \in P$, $\exists y \in P$ with $x \leq y = f(y)$.

Proof. Let $\gamma(P)$ be the Hartogs ordinal of P. Define $g : \gamma(P) \to P$ by the following recursion:

- $g(0) = x$
- $g(\alpha^+) = f(g(\alpha))$
- $g(\lambda) = \bigvee\{g(\alpha : \alpha < \lambda\}$.

g is order-preserving by induction. It is not injective. So $\exists \alpha < \beta < \gamma(P)$ with $g(\alpha) = g(\beta)$. Then $f(g(\alpha)) = g(\alpha^+) \leq g(\beta) = g(\alpha)$. So $g(\alpha)$ is a fixed-point of f. □

Notation. For a set a, $\mathcal{P}^+a$ is $\mathcal{P}a \setminus \{\emptyset\}$.

Definition (Choice function)**.** A choice function for a is a function $g : \mathcal{P}^+a \to a$ such that $g(b) \in b$ for all b.

Note: If $\{a_i : i \in I\}$ is a family of non-empty sets, then each $a_i \in \mathcal{P}^+ \bigcup\{a_i : i \in I\}$. So a choice function (in this sense) for $\bigcup\{a_i : i \in I\}$ yields a choice function (in the sense of Chapter 1) for $\{a_i : i \in I\}$. So the assertion $(\forall x)(\exists g)(g \text{ is a choice function for } x)$ becomes our formulation of AC.

Theorem (Well-ordering theorem)**.** A set a can be well-ordered $\Leftrightarrow$ it has a choice function.

Proof. Suppose a is non-empty.

($\Rightarrow$) Given a well-ordering of a, we have a choice function $g : \mathcal{P}^+a \to a$ sending each $b \in \mathcal{P}^+a$ to its least element.

Suppose we have a choice function $g : \mathcal{P}^+a \mapsto a$. We define the partial function $f : \gamma(a) \rightharpoondown a$ by $f(a) = g(a \setminus \{f(\beta) : \beta < \alpha\})$ for $\{f(\beta) : \beta < \alpha\} \neq a$. By construction, f is injective. We know that f is not total, since there is no injection from $\gamma(a) \to a$. So there exists an α such that $\{f(\beta) : \beta < \alpha) = a$. So f is surjective. So f is a bijection between a and a subset of $\gamma(a)$. Define $x < y \Leftrightarrow f^{-1}(x) < f^{-1}(y)$. □

Note: Well-ordering can be deduced from Zorn's Lemma. We can prove well-ordering by considering the well-orders of subsets of a ordered by inclusion. It is clearly chain-complete and has a maximal element. Now the maximal element

must be a well-order of the whole of a, or else we can add a member of a to create a larger well-ordered subset.
Note: Many applications of Zorn's lemma can be proved with well-ordering. e.g. Hamel's theorem: given a vector space V, well-order its vectors as x_α. The transfinitely recurse through the vectors, adding each vector to the basis if the is not already in the span of the basis. Then this is by construction linearly independent and each vector must be in some basis.

Definition. An ordinal is initial if there is no bijection $\alpha \to \beta$ for any $\beta < \alpha$. Intuitively, it is the smallest ordinal of the given cardinality.

According to this definition, every finite ordinal is initial, and ω is initial.

We can enumerate the infinite ordinals as $\{\omega_\alpha : \alpha \in \mathrm{On}\}$ by the recursive definition

(i) $\omega_0 = \omega$

(ii) $\omega_{\alpha^+} = \gamma(\omega_\alpha)$

(iii) $\omega_\lambda = \bigcup\{\omega_\alpha : \alpha < \lambda\}$.

Lemma. The infinite initial ordinals are exactly the ω_α for $\alpha \in \mathrm{On}$.

Proof. ω_0 is initial. ω_{α^+} is initial since a bijection $\omega_{a^+} \to \gamma$ for $\gamma < \omega_{\alpha^+}$ would yield an injection $\omega_{\alpha^+} \to \omega_\alpha$. ω_λ is initial since a bijection $\omega_\lambda \to \gamma$ for some $\gamma < \omega_\lambda$ would yield an injection $\omega_{\beta^+} \to \omega_\beta$, where $\gamma \leq \omega_\beta < \omega_\lambda$.

Conversely, suppose β is an infinite initial ordinal. By Sheet 4 Q. 3, we have $\alpha \leq \omega_\alpha$ for all α. So there is a least α such that $\omega_\alpha > \beta$. This α must be a successor δ^+. So we have $\omega_\delta \leq \beta < \omega_{\delta^+} = \gamma(\omega_\delta$. So we have injections $\omega \to \beta$ and $\beta \to \omega_\delta$. By Cantor-Bernstein, we have a bijection $\beta \to \omega_\delta$. So by initiality of β, we have $\beta = \omega_\delta$. □

Informally, a cardinal is an equivalence class of sets under the relation $a \sim b \Leftrightarrow \exists$ bijection $a \to b$. Except for $\{\emptyset\}$, the equivalence classes of this relation are all proper classes, so we seek a function class card : $V \to V$ such that $(\forall x, y)((\operatorname{card} x = \operatorname{card} y) \Leftrightarrow (x \sim y))$.

Definition.

(a) If we assume Choice, then every $\sim$-equivalence class contains an ordinal by the Well-Ordering Theorem (since there is a bijection between any (well-ordered) set and some ordinal). So it contains a unique initial ordinal. We can define $\operatorname{card} x$ to be the unique initial ordinal α satisfying $x \sim \alpha$.

(b) If we do not assume Choice, we define the essential rank of x to be the least ordinal α such that $(\exists y)((\operatorname{rank} y = \alpha) y \sim x)$. Then define

$$\operatorname{card} x = \{y \in V^+_{\text{ess. rk}(x)} : x \sim y\}$$

(Intuitively, since we cannot include the set of all sets of same size as x, we pick a von-Neumann universe (which is a set) that is "large enough" to contain sets as big as x, and take the cardinality to be the set of all sets in that universe that are as big as x)

Clearly, $\operatorname{card} x = \operatorname{card} y$ implies $(\exists z)(z \in \operatorname{card} x \cap \operatorname{card} y)$, and hence $x \sim z \sim y$.

Notation. $\operatorname{card}\omega_\alpha = \aleph_\alpha$.

Note: The class $\{\aleph_\alpha : \alpha \in \mathrm{On}\}$ may or may not be all of the cardinals, depending on whether Choice is true.

Definition (Order of cardinals)**.** For cardinals m, n, $m \leq n$ iff $\exists$ injection $x \to y$, where $\operatorname{card} x = m$, $\operatorname{card} y = n$. This is obviously (well-defined), reflexive and transitive. It is also anti-symmetric by Cantor-Bernstein.

By Sheet 4 Q. 7, it is a total order on cardinals iff the Axiom of Choice holds.

Definition (Sum, product and exponentiation for cardinals)**.** If $\operatorname{card} x = m$, $\operatorname{card} y = n$, then

(i) $m + n = \operatorname{card}(x \perp\!\!\!\perp y)$

(ii) $m \cdot n = \operatorname{card}(x \times y)$

(iii) $m^n = \operatorname{card}(x^y)$, where x^y is the set of all functions $y \to x$.

It is easy to verify that these are well-defined.

Lemma.

(i) Addition is associative and commutative, with $0 = \operatorname{card}\emptyset$ as the identity element.

(ii) Multiplication is associative and commutative, with $1 = \operatorname{card}\{\emptyset\}$ as the identity element.

(iii) $m(n + p) = mn + mp$

(iv) $m^{n+p} = m^n m^p$

(v) $m^{np} = (m^n)^p$

(vi) $(mn)^p = m^p n^p$

Proof is by constructing the obvious bijections.

Lemma. For any ordinal α, $\aleph_\alpha \alpha_\alpha = \alpha_\alpha$

Proof. Perform induction on α.

Define a well-ordering of $\omega_\alpha \times \omega_\alpha$ as follows: $\langle \gamma_1, \delta_1 \rangle < \langle \gamma_2, \delta_2 \rangle$ iff

- $(\gamma_1 \cup \delta_1) < (\gamma_2 \cup \delta_2)$; or
- $(\gamma_1 \cup \delta_1) = (\gamma_2 \cup \delta_2)$ and $(\gamma_1 < \gamma_2)$; or
- $\gamma_1 = \gamma_2 \geq (\delta_1 \cup \delta_2)$ and $\delta_1 < \delta_2$.

We need to check this is a well-ordering: given a non-empty $S \subseteq \omega_\alpha \times \omega_\alpha$, first set

$$S_1 = \{\langle \gamma, \delta \rangle \in S : (\forall \langle \gamma', \delta' \rangle \in S)(\gamma \cup \delta < \gamma' \cup \delta')\}$$
$$S_2 = \{\langle \gamma, \delta \rangle \in S_1 : (\forall \langle \gamma', \delta' \rangle \in S_1)(\gamma \leq \gamma')\}$$

If S_2 is not a singleton,

$$S_3 = \{\langle \gamma, \delta \rangle \in S_2 : (\forall \langle \gamma', \delta' \rangle \in S_2)(\delta \leq \delta')\}$$

Then the member of S_3 or S_2 is the minimum.

Now let β be the order type this well-ordering. For any $\theta < \beta$, there exists $\gamma < \omega_\alpha$ such that θ injects into $\gamma \times \gamma$.

Then either γ is finite, in which case $\gamma \times \gamma$ is finite, so $\operatorname{card} \theta < \aleph_\alpha \leq \aleph_\alpha$;

Or γ is infinite, in which case $\operatorname{card} \gamma = \alpha_\delta$ for some $\delta < \alpha$. So $\operatorname{card}(\gamma \times \gamma) = \aleph_\delta \times \aleph_\delta = \alpha_\delta < \alpha_\alpha$. So $\operatorname{card} \theta < \alpha_\alpha$.

In either case, we deduce $\theta < \omega_\alpha$ for all such θ. So $\beta \leq \omega_\alpha$. But we also have $\omega_\alpha \leq \beta$, since ω_α injects into $\omega_\alpha \times \omega_\alpha$ trivially and ω_α is initial. So $\beta = \omega_\alpha$.

So $\aleph_\alpha \times \aleph_\alpha = \aleph_\alpha$. □

Corollary. For any ordinals α and β, we have $\aleph_\alpha + \aleph_\beta = \aleph_\alpha = \aleph_{\alpha \cup \beta}$.

Proof. wlog assume $\beta \leq \alpha$. By constructing suitable injections, we have $\aleph_\alpha \leq \aleph_\alpha + \aleph_\beta \leq \aleph_\alpha \cdot \aleph_\beta \leq \aleph_\alpha^2 = \aleph_\alpha$. So they must be all equal. (Cantor-Bernstein) □

Lemma. Suppose $m = \operatorname{card}(a)$ and $n = \operatorname{card}(b)$ satisfy $m + n = m \cdot n$. Then $\exists$ either an injection $a \to b$ or a surjection $a \to b$.

Proof. By assumption, we have a bijection $a \perp\!\!\!\perp b \overset{f}{\mapsto} a \times b$. Consider the composite $g : a \to a \perp\!\!\!\perp b \to a \times b \to b$. If this is a surjection, then we are done. If not, pick $y_0 \in b \setminus \text{img}$ and consider the map $a \to a \times b \to a \perp\!\!\!\perp b$ by $x \mapsto \langle x, y_0 \rangle \xmapsto{f^{-1}} a \perp\!\!\!\perp b$.This has image contained in $b \times \{1\}$. So its composite with the mapping $\langle y, 1 \rangle \mapsto y$ is an injection $a \to b$. □

Corollary. Suppose every cardinal $m \geq \aleph_0$ satisfies $m^2 = m$. Then the Axiom of Choice holds.

Proof. We show that any set a can be well-ordered. Given a, consider $\gamma(a)$. If a is finite, then any order is a well-order.

Otherwise, $\gamma(a) \geq \omega$. Let $m = \operatorname{card}(\gamma(a))$ and $n = \operatorname{card}(a)$. Then $m + n \geq \aleph_0$. But we have $m + n \leq m \cdot n \leq (m + n)^2 = m + n$. By Cantor-Bernstein, we have $m + n = m \cdot n$.

Since there is no injection $\gamma(a) \to a$, there must be a surjection $f : \gamma(a) \to a$.

Now define $g : a \to \gamma(a)$ by $g(x) =$ least element of $f^{-1}(x)$. Then g is injective, and we can well-order a by $x < y \Leftrightarrow g(x) < g(y)$. □

If $m = \operatorname{card}(a)$, then $2^m = \operatorname{card}(\mathcal{P}a)$. Cantor's diagonal argument tells us that $m < 2^m$ for all m.

From now we assume Choice, so that all infinite cardinals are $\aleph$s.

Lemma. If $\beta \leq \alpha^+$, then $\aleph_\beta^{\aleph_\alpha} = 2^{\aleph_\alpha}$.

Proof. Since $2 \leq \aleph_\beta$, we have $2^{\aleph_\alpha} \leq \aleph_\beta^{\aleph_\alpha}$.

But $2^{\aleph_\alpha} > \aleph_\alpha$. So $2^{\aleph_\alpha} \geq \aleph_{\alpha^+} \geq \aleph_\beta$. So $\aleph_\beta^{\aleph_\alpha} \leq (2^{\aleph_\alpha})^{\aleph_\alpha} = 2^{\aleph_\alpha^2} = 2^{\aleph_\alpha}$. By Cantor-Bernstein, the result follows. □

Definition. Given a family of sets $\{a_i : i \in I\}$ with $\operatorname{card}(a_i) = m_i$, write

$$\sum_{i\in I} m_i = \operatorname{card}(\amalg_{i\in I} a_i)$$
$$\prod_{i\in I} m_i = \operatorname{card}(\prod_{i\in I} a_i)$$

where $\amalg_{i\in I} = \cup\{a_i \times \{i\} : i \in I\}$ and $\prod_{i\in I} a_i$ is the set of all choice functions for $\{a_i : i \in I\}$ (generalization of $i \in I$).

Lemma (König's Lemma)**.** If $(\forall i) m_i < n_i$, then $\sum_{i\in I} m_i < \prod_{i\in I} n_i$.

Proof. We will only show that they are not equal. Suppose $\operatorname{card}(a_i) = m_i$ and $\operatorname{card}(b_i) = n_i$ for each i. Given a function $f : \amalg_{i\in I} \to \prod_{i\in I} b_i$, we show f cannot be surjective.

For each i, the composite $f_i : a_i \to \amalg_{i\in I} a_i \xrightarrow{f} \prod_{i\in I} b_i \to b_i$ is not surjective. So we can choose $y_i \in b_i \setminus \operatorname{im} f_i$.

Then the function $i \mapsto y_i$ is an element of $\prod_{i\in I} b_i$, but not in the image of f. □

Note: If we set $m_i = 0$, then König's Lemma becomes the statement of the Axiom of Choice. If we set $m_i = 1$ and $n_i = 2$, then it reduces to Cantor's Theorem.

Corollary. $2^{\aleph_0} \neq \aleph_\omega$

Proof. Take $I = \omega$ and set $a_0 = \omega$, $a_i = \omega_i \setminus \omega_{i-1}$ and $b_i = \omega_\omega$ for all i.

Then $\operatorname{card} a_i = \aleph_i \leq \aleph_\omega = \operatorname{card} b_i$ for all i. So $\aleph_\omega = \operatorname{card} \omega_\omega = \operatorname{card}(\amalg_{i\in I} a_i) < \operatorname{card}(\prod_{i\in I} b_i) = \aleph_\omega^{\aleph_0}$. But $2^{\aleph_0} = 2^{\aleph_0 \cdot \aleph^0} = (2^{\aleph_0})^{\aleph_0}$. So $2^{\aleph_0} \neq \aleph_\omega$.

Similarly, if λ is any limit ordinal of cardinality ω, then $2^{\aleph_0} \neq \aleph_\lambda$. □