MA3205

Set Theory

Jia Xiaodong

November 17, 2020

An introduction to set theory. The main goal is to introduce some basic facts about sets, cardinal and ordinal numbers, recursion, the axiom of choice, as well as to demonstrate how set theory can serve as a foundation for mathematics.

1 Preliminaries

This section will define some notions more concretely before we begin. We do not focus on the axioms, but will mention some of the more important ones. We will also touch on some of the later on, like the axiom of replacement and the axiom of choice.

Set existence axiom $\exists x\ (x = x)$.

Axiom of extensionality $\forall x \forall y\ [\forall z\ (z \in x \leftrightarrow z \in y) \to x = y]$.

Comprehension scheme For each formula $\phi(x)$, $\forall z \exists y \forall x\ [x \in y \leftrightarrow (x \in z \wedge \phi(x))]$.

Pairing axiom $\forall x \forall y \exists z\ (x \in Z \wedge y \in z)$.

Union axiom We can form unions of sets.

Power axiom We can form power sets.

We know the union of two sets $X \cup Y$. What about arbitrary unions? For any set A, define

$$\bigcup A = \{b \mid \exists a\ (a \in A \wedge b \in a)\}.$$

In other words, the (big) union of A collects the members of members of A. The axiom of union guarantees that $\bigcup A$ is a set. For instance if $A = \{X, Y\}$ then $\bigcup A = X \cup Y$. We define intersections the same way for non-empty A:

$$\bigcap A = \{b \mid \forall a\ (a \in A \implies b \in a)\}.$$

2 Relations and Functions

2.1 Ordered pairs

If you have two sets a and b, you can form a new set containing both a and b. This forms an unordered pair:

$$\{a, b\}$$

Definition 2.1. Let the *ordered pair* of a and b be denoted (a, b). It must satisfy the property $(a, b) = (c, d)$ iff $a = c$ and $b = d$. □

Theorem 2.1 (Kuratowski 1921)**.** *The following definition satisfies the defining property of ordered pairs:*

$$(a, b) = \{\{a\}, \{a, b\}\}.$$

Proof.
($\Longrightarrow$): Suppose that

$$\{\{a\}, \{a, b\}\} = \{\{c\}, \{c, d\}\}.$$

Any element on the left hand side must appear on the right hand side. Consider all possible cases:

- If $\{a\} = \{c\}$
  - and if $\{a, b\} = \{c\} = \{a\}$, then $b = a = c$. Then the left hand side becomes $\{\{a\}\}$ and $d = a$.
  - otherwise if $\{a, b\} = \{c, d\} = \{a, d\}$,
    - and if $b \neq a$ then $d = b$.
    - otherwise $b = a$ and $\{a, b\} = \{a\} = \{a, d\}$ which means $a = b = d = c$.
- $\{a\} = \{c, d\}$
  - then $a = c = d$. The right hand side reduces to $\{\{a\}\}$ and we gather $b = a$.

($\Longleftarrow$): By substitution. ■

Below are a few other definitions, though the one by Kuratowski is more widely used.

Theorem 2.2 (Hausdorff 1914)**.** *Choose two objects x, y that are not a or b. The following is a valid definition of an ordered pair:*

$$(a, b) = \{\{a, x\}, \{b, y\}\}$$

Proof.

($\Longrightarrow$): If

$$\{\{a, x\}, \{b, y\}\} = \{\{c, x\}, \{d, y\}\}$$

then it is quite obvious that $a = c$ and $b = d$.

($\Longleftarrow$): By substitution. ■

Theorem 2.3 (Wiener 1914)**.** *The following is a valid definition of an ordered pair:*

$$(a, b) = \{\{\{a\}, \emptyset\}, \{\{b\}\}\}$$

Proof. Similar to above. ■

Definition 2.2. The *Cartesian product* of A and B, denoted $A \times B$ is defined by

$$A \times B = \{(a, b) \mid a \in A \wedge b \in B\}$$

□

We can generalize the definitions. For example we can define $A \times B \times C = (A \times B) \times C$. We also use A^n to denote $A \times \cdots \times A$ n-times.

2.2 Relations and functions

Definition 2.3 (Relations)**.** A *binary relation* is a set of ordered pairs. Let R be a binary relation, then

- the *domain* of R is defined as $\text{dom}(R) = \{x \mid \exists y\ [(x, y) \in R]\}$
- and the *range* of R is defined as $\text{ran}(R) = \{y \mid \exists x\ [(x, y) \in R]\}$.

We write xRy or $R(x, y)$ instead of $(x, y) \in R$. □

Definition 2.4 (Functions)**.** A *function* f is a binary relation such that for every x there is at most one y for which $(x, y) \in f$:

$$(x, y) \in f \wedge (x, z) \in f \implies y = z.$$

□

Contrast this definition of functions with the naive definition of functions as rules. This view stem from the view of functions as analytic expressions. However the rules are irrelevant and we only care about what set it forms. For example $f(n) = n + 1 - 1$ is the same as $f(n) = n$, because it describes the same set.

The word function was first used by Leibniz. Later on Dirichlet defined functions in terms of an arbitrary correspondence between variables representing numerical sets. This was to separate the concept of function from its analytical representation to proof Fourier's claim regarding expansion of functions as sums of trigonometric functions.

The familiar definitions regarding functions are the same, and we quickly go through them:

Definition 2.5. Let $f : A \to B$ be a function.

- f is *injective* or one-one if $(f(x) = y) \wedge (f(x') = y) \implies x = x'$.
- f is *surjective* or onto if $\forall y \in B, \exists x, f(x) = y$.
- f is *bijective* if it is both injective and surjective. □

Definition 2.6. B^A denotes the set of all functions from A to B. □

Definition 2.7. Define the composition of $f : A \to B$ and $g : B \to C$ as

$$\forall x \in A, (f \circ g)(x) = g(f(x))$$

□

Definition 2.8. Given a function $f : X \to Y$, $X' \subseteq X$, $Y' \subseteq Y$, we define the *image* of X' under f to be

$$f[X'] = \{y \mid \exists x \in X' \ [f(x) = y]\}$$

Define the *inverse image* (pre-image) of Y' under f to be

$$f^{-1}[Y'] = \{x \mid f(x) \in Y'\}$$

□

Theorem 2.4. *Let A and B and A_i be sets and I an indexing set.*

i. $f[\bigcup\{A_i \mid i \in I\}] = \bigcup\{f[A_i] \mid i \in I\}$.

ii. $f[\bigcap\{A_i \mid i \in I\}] \subseteq \bigcap\{f[A_i] \mid i \in I\}$.

iii. $f[A] - f[B] \subseteq F[A - B]$.

Proof. Let $C = \bigcup\{A_i \mid i \in I\}$. Let $D = \bigcap\{A_i \mid i \in I\}$.

i.

$$\begin{aligned} y \in f[C] &\iff \exists x \in C \ (f(x) = y) \\ &\iff \exists i \in I, \exists x \in A_i \ [f(x) = y] \\ &\iff \exists i \in I \ [y \in f[A_i]] \\ &\iff y \in \bigcup\{f[A_i] \mid i \in I\} \end{aligned}$$

ii.

$$\begin{aligned} y \in f[D] &\iff \exists x \in D \ (f(x) = y) \\ &\implies \forall i \in I, \exists x \in A_i \ [f(x) = y] \\ &\iff \forall i \in I \ [y \in f[A_i]] \\ &\iff y \in \bigcap\{f[A_i] \mid i \in I\} \end{aligned}$$

There is no bidirectional on the third line since the existence of multiple x's does not mean that they are the same one. From this it can be seen that if f is injective then the equality will hold.

iii.

$$\begin{aligned} y \in f[A - B] &\iff \exists x \in A \ (f(x) = y) \wedge \neg\exists z \in B \ (f(z) = y) \\ &\iff \exists x \in A - B \ (f(x) = y) \\ &\iff y \in f(A - B) \end{aligned}$$

■

Corollary 2.4.1. *Let A and B and A_i be sets and I an indexing set.*

i. $f^{-1}[\bigcup\{A_i \mid i \in I\}] = \bigcup\{f^{-1}[A_i] \mid i \in I\}$.

ii. $f^{-1}[\bigcap\{A_i \mid i \in I\}] = \bigcap\{f^{-1}[A_i] \mid i \in I\}$.

iii. $f^{-1}[A - B] = f^{-1}[A] - f^{-1}[B]$.

Proof. Since the inverse function is always injective we get this result automatically. ■

A function f can be viewed as a set of pairs, which is also called its *graph*. We can also view a set as a function. For example, the indicator function $\mathbb{1}_A : A \to \{0, 1\}$:

$$\mathbb{1}_A = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{otherwise} \end{cases}$$

Using functions, we can also generalize the Cartesian product to infinitely many sets. Let $S = \langle S_i \mid i \in I\rangle$ be a function with domain I. We call it an *indexed system of sets*. We say that A is *indexed* by S if $A = \{S_i \mid i \in I\} = \text{ran}(S)$. In other words, every element of A is indexed. Then we can define the *product* of the indexed system S as the set

$$\prod S = \{f \mid (\text{dom}(f) = I) \wedge (\forall i \in I \ (f(i) \in S_i))\}$$

This is a reasonable definition since this essentially visits every element in each indexed sets. The image of f tells us which element to visit. We will revisit this after we discuss the Axiom of Choice.

Let R be a relation. If $\text{dom}(R) \cup \text{ran}(R) \subseteq X$, i.e. $R \subseteq X \times X$, then we say that R is a relation *in* X.

2.3 Equivalence relations and equivalence classes

Definition 2.9. Let R be a binary relation in A.

- R is *reflexive* in A if $\forall a \in A, aRa$.
- R is *symmetric* in A if $\forall a, b \in A, aRb \implies bRa$.
- R is *transitive* in A if $\forall a, b, c \in A, aRa \wedge bRc \implies aRc$.

□

Definition 2.10 (Equivalence relations)**.** R is called an *equivalence relation* on A if it is reflexive, symmetric, and transitive. □

The familiar equality ($=$) satisfies all three conditions, so equivalences are almost like equality, but more relaxed.

Definition 2.11 (Equivalence classes)**.** Let E be an equivalence on A and let $a \in A$. The *equivalence class of a modulo E* is the set

$$[a]_E = \{x \in A \mid xEa\}$$

A/E denotes the set of all equivalence classes:

$$A/E = \{[a]_E \mid a \in A\}.$$

□

Lemma 2.5. *Let $a, b \in A$, and E an equivalence on A.*

i. $aEb \iff [a]_E = [b]_E$.

ii. $\neg(aEb) \iff [a]_E \wedge [b]_E = \emptyset$.

Proof.

i. ($\implies$): Assume aEb. If $x \in [a]_E$, then xEa. By transitivity xEb, and so $x \in [b]_E$. If $x \in [b]_E$, then xEb. By symmetry bEa, and by transitivity xEa, and so $x \in [a]_E$. Therefore $[a]_E = [b]_E$.

($\impliedby$): Assume that $[a]_E = [b]_E$. Then $a \in [a]_E$, and so $a \in [b]_E$. By definition then aEb.

ii. ($\implies$): Suppose that there is some $x \in [a]_E \cap [b]_E$. Then xEa and xEb. By symmetry, aEx and by transitivity aEb. This is a contradiction.

($\impliedby$): Assume that $[a]_E \cap [b]_E = \emptyset$. Then $a \notin [b]_E$, since $a \in [a]_E$. Therefore $\neg(aEb)$.

■

Definition 2.12 (Partitions)**.** A system S of **non-empty** sets is called a *partition* of A if

- S is a collection of mutually disjoint sets, and
- the union of S is A.

□

Lemma 2.6. *Let E be an equivalence on A, then A/E is a partition of A.*

Proof. By lemma 2.5, if $[a]_E \neq [b]_E$, then $\neg(aEb)$, so $[a]_E \cap [b]_E = \emptyset$, so the first property holds. Since $a \in [a]_E$, $\bigcup A/E = A$, and the second property holds. ■

From equivalence classes we can induce a partition. From a partition we can also induce an equivalence class.

Lemma 2.7. *Let S be a partition of A. The relation E_S in A is defined by*

$$E_S = \{(a, b) \in A \times A \mid \exists C \in S \ (a \in C) \wedge (b \in C)\}$$

E_S is an equivalence on A.

Proof. Every element of A must be in some set within S so E_S must be reflexive. It is quite obvious that E_S is symmetric and transitive. ■

Definition 2.13. A set $X \subseteq A$ is called a set of *representatives* for the equivalence E_S or for the partition S of A, if $\forall C \in S, X \cap C = \{a\}$ for some $a \in C$. □

Theorem 2.8.

i. If E is an equivalence on A, and $S = A/E$, then $E_S = E$.

ii. If S is a partition of A and E_S is the corresponding equivalence, then $A/E_S = S$.

Proof.

i. aE_Sb iff they are in the same equivalence class iff aEb.

ii. a and b are in the same partition iff aE_Sb iff they are in the same equivalence class induced by E_S.

■

2.4 Partial Orderings

Definition 2.14. Let R be a binary relation in A.

i. R is *antisymmetric* if $\forall a, b \in A, aRb \wedge bRa \implies a = b$.

ii. R is *asymmetric* if $aRb \implies \neg bRa$.

□

Definition 2.15 (Partial orderings)**.** A binary relation R in A which is reflexive, antisymmetric, and transitive (RAT) is called a *partial ordering* of A.

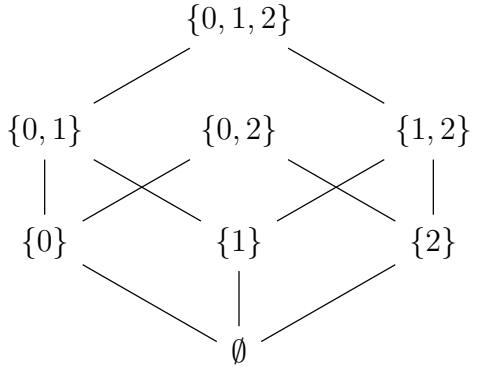
Then the pair (A, R) is called an partially ordered set (poset). We often use $\leq$ to denote a partial order. □

Definition 2.16 (Strict orders)**.** A binary relation R in A is a *strict* partial ordering if it is asymmetric and transitive. □

We often use $<$ to denote a strict partial ordering and $\leq$ to denote a non-strict partial ordering.

Example 2.1. $(\mathcal{P}(\{0,1,2\}), \subseteq)$ [1] is a partially ordered set.

Here is a hierarchy with some edges omitted for clarity.



[1] Here $\mathcal{P}(A)$ denotes the power set of A.

It is reflexive since $x \subseteq x$. It is antisymmetric since if $x \subseteq y$ and $y \subseteq x$ then $x = y$. It is also transitive because $x \subseteq y$ and $y \subseteq z$ then $x \subseteq z$.

This also highlights why it is called a *partial* order. Certain elements cannot be compared, for example $\{0\}$ and $\{1\}$. $\Diamond$

Example 2.2. The relation $D = \{(m, n) \mid m, n \in \mathbb{Z}^+ \wedge m \text{ divides } n\}$ is a partial ordering on $\mathbb{Z}^+$. This is also denoted $|$. The proof is omitted. $\Diamond$

Example 2.3. Let $X = \{f \mid f : \mathbb{R} \to \mathbb{R}\}$ and R be the relation such that

$$fRg \iff \forall x \, [f(x) \leq g(x)]$$

It inherits its properties from $\leq$, and we skip the proof. $\Diamond$

Lemma 2.9.

i. Let R be a partial ordering of A. Then the relation S defined in A by $aSb \iff (aRb) \wedge (a \neq b)$ is a strict partial ordering of A.

ii. Let S be a strict partial ordering of A. Then the relation R defined in A by $aRb \iff (aSb) \vee (a = b)$ is a partial ordering of A.

Proof.

i. S is asymmetric since $aSb \implies a \neq b$. But R is antisymmetric, so if aRb and bRa then $a = b$.

   S is transitive. Suppose aSb and bSc. Since aSb, aRb and $a \neq b$. Since bSc, bRc and $b \neq c$. Then aRc as R is transitive. $a \neq c$ because otherwise $aRb = cRb$ — together with bRc this implies $b = c$, a contradiction.

ii. R is reflexive since $a = a$ so aRa. If aSb then $\neg bSa$ so if both aRa and bRa, the only choice is that $a = b$ so R is antisymmetric. Next, since both S and $=$ are transitive, so is R.

■

Definition 2.17 (Linear orders)**.** An ordering $\leq$ (or $<$) of A is called *linear* or *total* if any two elements of A are comparable, i.e.

$$\forall x, y \, (x \leq y \vee y \leq x)$$
$$\forall x \neq y \, (x < y \vee y < x)$$

□

Definition 2.18. Let $\leq$ be a partial ordering on A, and let $B \subseteq A$.

- $b \in B$ is the *least* element of B if $b \leq x$ for all $x \in B$.
- $b \in B$ is the *minimal* element of B is there exists no $x \in B$ such that $x \leq b$ and $x \neq b$.
- $a \in A$ is a *lower bound* of B if $a \leq x$ for all $x \in B$.
- $a \in A$ is called the *infimum* of B if it is the greatest element of the set of all lower bounds of B (greatest lower bound).

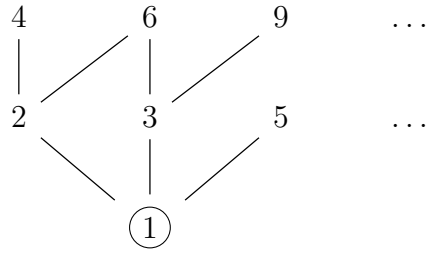
- The *greatest* element, *maximal* element, *upper bound*, *supremum* can be defined similarly.

□

Notice that the least element of a set has to be in the same set, but this is not necessarily so for the lower bound.

Since partial orders are antisymmetric, there can only be one least element. If the least element exists then it is also minimal. Furthermore, as long as a set is finite and non-empty then it must have a minimal and a maximal element [2].

Example 2.4. Let $A = \mathbb{Z}^+$ be the set of positive integers ordered by divisibility $|$. Let $B = \{2, 3, 4, \ldots\}$.



(1)

- 1 is the least element in A.
- There is no greatest element in A or B
- There is no maximal element in A or B
- B has no least element.
- The primes are the minimal elements of B.

◊

Definition 2.19 (Well-orderings)**.** A linear ordering $\prec$ of a set A is a *well-ordering* if every non-empty subset of A has a $\prec$-least element. The structure $(A, \prec)$ is then called a well-ordered set. □

Example 2.5. $(\mathbb{Z}, <)$ is not a well-ordered set. $(\mathbb{Z}^+, <)$ is a well-ordered set. ◊

2.5 Structures

Definition 2.20. A *binary operation* on S is a function mapping a subset of $S \times S$ to S.

A is said to be *closed* under the operation f if

$$\forall x, y \in \mathrm{dom}(A)\ (f(x, y) \in A)$$

□

Therefore condition only has to hold when f is defined. For example, the rational numbers are closed under division.

[2] The proof of this fact can be shown with induction, even though we have not introduced it yet, just like how we are using knowledge of integers right now without having really defined them first.

Definition 2.21. Generally a *structure* consists of a set A and several relations and operations on A. □

For simplicity we will only consider binary relations and unary functions. They can be generalized to higher arities.

Definition 2.22 (Isomorphism)**.** An isomorphism between structures (A, R_i, f_j) and (A', R'_i, f'_j) is a bijection h between the sets A and A' such that for any i and j,

- $f_j(x)$ is defined iff $f'_j(h(x))$ is defined
- $R_i(x, y) \iff R'_i(h(x), h(y))$
- $f_j(x) = y \iff f'_j(h(x)) = h(y)$.

□

Example 2.6. $h(x) = e^x$ is an isomorphism between $A = (\mathbb{R}, \leq, +)$ and $A' = (\mathbb{R}^+, \leq, \times)$.

h forms a bijection between $\mathbb{R}$ and $\mathbb{R}^+$ and preserves the ordering since it is monotone. Furthermore $e^{x+y} = e^x \cdot e^y$, so it also translates $+$ into $\times$. ◊

Lemma 2.10. *Let $(P, <)$ and $(Q, \prec)$ be linearly ordered sets. Let $h : P \to Q$ be a bijection that preserves the order: $p_1 < p_2 \implies h(p_1) \prec h(p_2)$. Then h is an isomorphism between $(P, <)$ and $(Q, \prec)$.*

Proof. What we need is the other direction, $h(p_1) \prec h(p_2) \implies p_1 < p_2$. Since $<$ is a linear order, one of the cases must be true:

- $p_1 < p_2$. In this case, we are done.
- $p_2 < p_1$. This also cannot occur because this would imply $h(p_1) \prec h(p_2)$.
- $p_1 = p_2$. This cannot occur because this would imply $h(p_1) = h(p_2)$.

■

3 Natural Numbers

Definition 3.1. Let 0 denote $\emptyset$, $1 = \{\emptyset\}$, $2 = \{\emptyset, \{\emptyset\}\} = \{0, 1\}$, and so on. For any set x, the *successor* of x, denoted by $S(x)$, is the set $x \cup \{x\}$. Some other common notations for the successor of x are x^+ and $x + 1$. □

Definition 3.2. A set I is called *inductive* if

- $0 = \emptyset \in I$, and
- it is closed under successor, i.e. $\forall a \in I \ (S(a) \in I)$.

□

The axiom of infinity basically says that an inductive set exists.

Definition 3.3. The set of all natural numbers is defined as

$$\mathbb{N} = \{x \mid x \in I \text{ for every inductive set } I\}$$

□

In other words, a set x a natural number iff it belongs to every inductive set.

Theorem 3.1. $\mathbb{N}$ *is an inductive set.*

Proof. First of all $\emptyset$ is in every inductive set so $\emptyset$ is also in $\mathbb{N}$. Next, if x is in any inductive set I then so is $S(x)$, which means $S(x) \in \mathbb{N}$ as well. ■

In fact, it can be seen that $\mathbb{N}$ is the smallest inductive set since it is effectively the intersection of all inductive sets.

Theorem 3.2 (Induction)**.** *Let $P(x)$ be a property. Assume that*

- *$P(0)$ holds.*
- *$\forall n \in \mathbb{N}\ [P(n) \implies P(n+1)]$. Then P holds for all natural numbers n.*

Proof. Consider the set $A = \{n \in \mathbb{N} \mid P(n)\}$. Since $P(0)$, $0 \in A$. The inductive step also means that A is closed under successor. So A is inductive. As $\mathbb{N}$ is the smallest inductive set, $\mathbb{N} \subseteq A$. But by definition $A \subseteq \mathbb{N}$. Therefore $A = \mathbb{N}$. ■

3.1 Ordering on $\mathbb{N}$

Definition 3.4 (von Neumann)**.** The relation $<$ on $\mathbb{N}$ is defined by $m < n \iff m \in n$. □

Lemma 3.3. *This definition preserves the familiar properties of $<$:*

i. $\forall n \in \mathbb{N}\ (0 \leq n)$ [3]

ii. $\forall k, n \in \mathbb{N}\ (k < n+1 \iff k \leq n)$.

Proof.

i. Induction on n. Let $P(x)$ be the statement $0 \leq x$. Since $0 \leq 0$. $P(0)$ holds trivially. Suppose that $P(n)$ holds. By the definition of $\leq$, either $0 \in n$ or $0 = n$. In either case $0 \in S(n)$. Thus $0 < S(n)$ and $P(S(n))$ holds. By the induction principle, $P(n)$ holds for all $n \in \mathbb{N}$.

ii. If $k < S(n)$ then $k \in n \cup \{n\}$. Therefore either $k \in n$ or $k = n$. Similarly so for the other direction.

■

Theorem 3.4. $(\mathbb{N}, <)$ *is a linearly ordered set.*

[3] Just to be clear here $0 \leq n$ means either $0 < n$ or $0 = n$.

Proof. We check that it is a strict partial order, followed by linearity.

- Transitive: Induction on n. Let $P(x)$ be the statement

$$\forall k, m \in \mathbb{N}\ [(k < m \wedge m < x) \implies k < x].$$

  $P(0)$ is vacuously true. Suppose $P(n)$ holds. Consider $P(n+1)$. By lemma 3.3, $m < n+1$ means $m < n$ or $m = n$. In the first case, $k < n$ by the inductive hypothesis and so $k < n+1$, and in the second case, $k < n$ directly and therefore $k < n+1$ as well.

- Antisymmetric: Let $P(a, b)$ be the statement

$$a < b \implies b \not< a.$$

  We first perform induction on n and show $P(n, n)$ is true for all $n \in \mathbb{N}$. This part is simple and can be skipped. Next, by transitivity, if $m < n$ and $n < m$ then $n < n$ which is impossible, hence $m < n \implies n \not< m$.

- Linear: Let $P(x)$ be the statement

$$\forall m\ (m < x) \vee (m = x) \vee (x < m).$$

  $P(0)$ is true from lemma 3.3. Assume $P(n)$ is true. If $m < n$ or $m = n$ then by transitivity $m < n+1$. If $n < m$, we perform induction on m to show that $n+1 \leq m$. This is skipped.

■

Theorem 3.5. $\forall n \in \mathbb{N}\ \neg(n < n)$.

Proof. In general $S(n) = \{0, 1, \ldots, n\}$. Perform induction on n. Let $P(x)$ be the statement that $\neg x < x$. It is easy to see that $P(0)$ is true. Now suppose $P(k)$ holds. Consider $P(S(k))$. Suppose instead $S(k) < S(k)$. Then from lemma 3.3, either $S(k) < k$ or $S(k) = k$. By definition $S(k) \neq k$. However if $k < S(k)$ and $S(k) < k$, by transitivity we conclude $k < k$ which contradicts the inductive hypothesis. Therefore $P(S(k))$ is true. ■

Theorem 3.6. $\forall m, n\ [m < n \iff m + 1 < n + 1]$.

Proof. We prove the forward direction with induction. Let $P(x)$ be the statement that $\forall m\ [m < x \implies m + 1 < x + 1]$. $P(0)$ is vacuously true. Suppose $P(n)$ is true for some $n \in \mathbb{N}$. Consider $P(n + 1)$. From lemma 3.3, $m < n + 1 \implies m \leq n$. If $m < n$ by the inductive hypothesis $m + 1 < n + 1 < n + 1 + 1$. If $m = n$ then $m + 1 = n + 1 < n + 1 + 1$. In both cases $P(n + 1)$ is true.

The converse is true by lemma 3.3. If $m + 1 < n + 1$, then either $m < m + 1 < n$, or $m < m + 1 = n$. ■

Theorem 3.7 (Strong induction)**.** *Let $P(x)$ be a property. Assume that for all $n \in \mathbb{N}$,*

$$\textit{If } P(k) \textit{ holds for all } k < n, \textit{ then } P(n).$$

Then P holds for all $n \in \mathbb{N}$.

Proof. Let $Q(x)$ be the statement $\forall k < x\ [P(k)]$. $Q(0)$ is vacuously true. Suppose $Q(n)$ is true for some n. So $P(k)$ holds for all $k < n$. By our supposition, this means that $P(n)$ is true. Therefore $P(k)$ holds for all $k < n + 1$, and $Q(n + 1)$ is true. ■

Theorem 3.8. *$(N, <)$ is a well-ordered set.*

Proof. Consider a non-empty set $X \subseteq \mathbb{N}$. We want to show that X has a least element. Suppose instead that X has no least element. Consider $\mathbb{N} \setminus X$. For any $k < n$, $k \in \mathbb{N} \setminus X \implies n \in \mathbb{N} \setminus X$, since otherwise n would be the least element of X. Then by the strong inductive principle, $\mathbb{N} \setminus X = \mathbb{N}$ and therefore $X \neq \emptyset$, a contradiction. ■

Corollary 3.8.1. *There is no infinite descending sequence of natural numbers. In other words, there is no function $f : \mathbb{N} \to \mathbb{N}$ such that for all $n \in \mathbb{N}$, $f(n) > f(n+1)$.*

Proof. Let $\text{ran}(f)$ is not empty, for instance $f(0)$ exists. So $\text{ran}(f)$ has a $<$-least element, a. Then there is some $n \in \mathbb{N}$ such that $f(n) = a$. Then $f(n+1) < f(n) = a$ contradicts the fact that a is the least element in $\text{ran}(f)$. ■

3.2 Recursion

Theorem 3.9 (Recursion on $\mathbb{N}$)**.** *Let A be a set, $a \in A$ and $g : A \times \mathbb{N} \to A$. Then there exists an unique function $f : \mathbb{N} \to A$ such that*

- *$f(0) = a$ and*
- *$\forall n \in \mathbb{N}\ [f(n+1) = g(f(n), n)]$*

Proof. A function $t : (m+1) \to A$ is called an m-step computation if $t(0) = a$ and for all k such that $0 \leq k < m$, $t(k+1) = g(t(k), k)$.

Let F be the set of all m-step computations for some $m \in \mathbb{N}$. Let $f = \bigcup F$.

We show that f satisfies all conditions by proving the following series of claims.

1. f is a function. This amounts to showing that for any $t_1, t_2 \in F$ are *compatible*, that is, if $x \in \text{dom}(t_1) \cap \text{dom}(t_2)$, then $t_1(x) = t_2(x)$. Assume $\text{dom}(t_1) = m \in \mathbb{N}$, $\text{dom}(t_2) = n \in \mathbb{N}$, and say $n \leq m$, then $n \subseteq m$ [4]. Therefore it suffices to show that $t_1(k) = t_2(k)$ for all $k < n$.

   Perform induction on k. When $k = 0$, $t_1(k) = t_2(k) = a$. Suppose $k + 1 < n$ and $t_1(k) = t_2(k)$. Then

   $$t_1(k+1) = g(t_1(k), k) = g(t_2(k), k) = t_2(k+1).$$

2. The second claim is that $\text{dom}(f) = \mathbb{N}$ and $\text{ran}(f) \subseteq A$. Firstly, $\text{dom}(f) \subseteq \mathbb{N}$ and $\text{ran}(f) \subseteq A$, since the same is true for all $t \in F$. To show that $\text{dom}(f) \supseteq \mathbb{N}$ it suffices to prove that for every $n \in \mathbb{N}$ there is an n-step computation t.

   Perform induction on n. Clearly the 0-step computation exists. Assume that t is an n-step computation, then define $t^+ : (n+1)+1 \to A$ by

   $$t^+(k) = \begin{cases} t(k) & \text{if } k \leq n \\ g(t(n), n) & \text{if } k = n+1 \end{cases}$$

   Then t^+ is an $(n+1)$-step computation.

[4] It is easy to see that this is true from the definition of the successor function

3. f satisfies recursive equations (i) and (ii). First of all $f(0) = a$ since $t(0) = a$ for all $t \in F$. For every $n \in \mathbb{N}$, let t be a $(n+1)$-step computation. Then $f(n) = t(n)$ and $f(n+1) = t(n+1)$. Therefore

$$f(n+1) = t(n+1) = g(t(n), n) = g(f(n))$$

4. Finally, we show the uniqueness of f. Let f_1 and f_2 both satisfy the theorem. We will perform induction on n to show that $f_1(n) = f_2(n)$ for all $n \in \mathbb{N}$. $f_1(0) = f_2(0) = a$ by definition. Suppose $f_1(n) = f_2(n)$, then

$$f_1(n+1) = g(f_1(n), n) = g(f_2(n), n) = f_2(n+1).$$

■

Theorem 3.10 (Recursion with parameters)**.** *Let $a : P \to A$ and $g : A \times A \times \mathbb{N} \to A$ be functions. There exists an unique function $f : P \times \mathbb{N} \to A$ such that*

- $\forall p \in P\ [f(p, 0) = a(p)]$
- $\forall n \in \mathbb{N}, \forall p \in P\ [f(p, n+1) = g(p, f(p, n), n)]$

Proof. The proof is similar as for the case without parameters. ■

Theorem 3.11 (Addition on $\mathbb{N}$)**.** *There is an unique binary operation $A : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ such that*

- $\forall m \in \mathbb{N}\ [A(m, 0) = m]$.
- $\forall m, n \in \mathbb{N}\ [A(m, S(n)) = S(A(m, n))]$.

We usually write $m + n$ instead of $A(m, n)$. We will also try to convince ourselves that this definition of addition is just like the addition familiar to us.

Lemma 3.12. $\forall n, m \in \mathbb{N}\ [n + S(m) = S(n) + m]$

Proof. Perform induction on m. When $m = 0$,

$$n + S(0) = S(n + 0) = S(n) = S(n) + 0$$

Assume the claim is true for some m. Then

$$\begin{aligned} n + S(S(m)) &= S(n + S(m)) \\ &= S(S(n) + m) \\ &= S(n) + S(m) \end{aligned}$$

using the definition of addition. This establishes the claim. ■

Theorem 3.13. *The operation $+$ is commutative.*

Proof. Let $P(x)$ be the proposition

$$\forall m \in \mathbb{N}\ (x + m = m + x).$$

$P(0)$ holds since by definition $0 + m = m$. We just have to show that $m + 0 = m$ as well.

Let $Q(y)$ be the proposition,

$$0 + y = y.$$

$Q(0)$ holds by definition. Assume $Q(m)$. Then

$$\begin{aligned} 0 + S(m) &= S(0 + m) \\ &= S(m) \\ &= S(m) + 0 \end{aligned}$$

Therefore $Q(m)$ holds for all $m \in \mathbb{N}$ by the induction principle.

Now, let us continue. Assume $P(n)$. Then

$$\begin{aligned} m + S(n) &= S(m + n) \\ &= S(n + m) \\ &= n + S(m) \\ &= S(n) + m \end{aligned}$$

using lemma 3.12 for the last step. Thus $P(S(n))$ holds, and so by the inductive principle P holds for all $n \in \mathbb{N}$. ■

Theorem 3.14. *The operation $+$ is associative.*

Proof. Let $P(x)$ be the proposition

$$\forall k, m\ [(k + m) + x = k + (m + x)].$$

$P(0)$ is clearly true. Assume the $P(n)$ holds for some $n \in \mathbb{N}$. Consider $P(S(n))$.

$$\begin{aligned} (k + m) + S(n) &= S((k + m) + n) && \text{(Definition of addition)} \\ &= S(k + (m + n)) && \text{(Inductive hypothesis)} \\ &= k + S(m + n). && \text{(Definition of addition)} \end{aligned}$$

Thus $P(S(n))$ also holds. ■

Theorem 3.15 (Multiplication in $\mathbb{N}$)**.** *There is an unique binary operation $M : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ such that*

- $\forall n \in \mathbb{N}\ [M(n, 0) = 0]$.
- $\forall m, n \in \mathbb{N}\ [M(m, n + 1)] = M(m, n) + m]$.

Again, we will write $m \cdot n$ instead of $M(m, n)$.

Theorem 3.16. *The operation $\cdot$ is commutative.*

Theorem 3.17. *The operation $\cdot$ is associative.*

Theorem 3.18. *The operation $\cdot$ is distributive over addition.*

Proof. Let $P(x)$ be the proposition

$$\forall m, n\ [x \cdot (m+n) = x \cdot m + x \cdot n].$$

$P(0)$ is obviously true. Suppose $P(k)$ holds for some $k \in \mathbb{N}$. Then consider $P(S(k))$.

$$\begin{aligned} S(k) \cdot (m+n) &= k \cdot (m+n) + (m+n) && \text{(Defn. of mult., associativity)} \\ &= k \cdot m + k \cdot n + m + n && \text{(Inductive hypothesis)} \\ &= S(k) \cdot m + S(k) \cdot n && \text{(Defn. of mult., associativity).} \end{aligned}$$

Thus $P(k+1)$ holds as well. ■

There is no division operation because we run into trouble with the rational numbers. However we can use multiplication to perform something similar to division.

Theorem 3.19 (Euclidean division)**.** *If m and d are non-zero natural numbers then there exist $q, r \in \mathbb{N}$ such that*

$$m = (d \cdot q) + r \qquad r < d.$$

Proof. We perform induction on $m \in \mathbb{N}$. Let P be the statement as above. Then $P(0)$ is true because $0 = 0 \cdot 0 + 0$. Suppose $P(n)$ is true for some $n \in \mathbb{N}$. Consider $P(n+1)$. We can express $n+1$ in the required form:

$$n + 1 = \begin{cases} d \cdot q + r + 1, & \text{if } r + 1 < d \\ d \cdot (q+1), & \text{if } r + 1 = d \end{cases}$$

These are the only possible cases since $r < d$ by the inductive hypothesis, and using theorem 3.6 we get $r + 1 < d + 1$ and lemma 3.3 gives us the two cases.

In the first case both sides are clearly equivalent. In the second case, expanding and substituting d shows that it is equal as well. So $P(n+1)$ is also true. ■

3.3 Closures

Definition 3.5. Consider a structure (A, f, g) where f is a binary operation on A and g is a unary operation on A. Let $C \subseteq A$. Define $\overline{C}$, the *closure* of C under f and g to be the smallest set containing C which is also closed under f and g:

$$\overline{C} = \bigcap \{B \subseteq A \mid B \supseteq C \text{ and } B \text{ is closed under } f \text{ and } g\}.$$

□

This definition is difficult and perhaps impossible to use if one were to try to find the closure. For that we need the concept of sequences.

Definition 3.6. A sequence is a function whose domain is either a natural number or $\mathbb{N}$. A sequence whose domain is some natural number is called a *finite sequence* of length n and is denoted

$$\langle a_i : i < n \rangle = \langle a_0, a_1, \ldots, a_{n-1} \rangle$$

$\text{seq}(A) = \bigcup_{n \in \mathbb{N}} A^n$ denotes the set of all finite sequence of elements of A.

If the domain of a sequence if $\mathbb{N}$, we call it an *infinite sequence* and it is denoted as

$$\langle a_i : i \in \mathbb{N} \rangle$$

□

Theorem 3.20. *Define a sequence of sets* $\{C_n \mid n \in \mathbb{N}\}$ *by*

- $C_0 = C$,
- $C_{n+1} = C_n \cup f(C_n \times C_n) \cup g(C_n)$

Let $\underline{C} = \bigcup\{C_n \mid n \in \mathbb{N}\}$. *Then* $\underline{C} = \overline{C}$.

Proof. ($\overline{C} \subseteq \underline{C}$): It suffices to show that $\underline{C}$ is closed under f and g. Assume that $a, b \in \underline{C}$. Then $a \in C_m$ and $b \in C_n$ for some $m, n \in \mathbb{N}$. Assume $m \leq n$. Then both $f(a, b) \in C_{n+1} \ni g(a)$. Hence $f(a, b) \in \underline{C} \ni g(a)$.

($\underline{C} \subseteq \overline{C}$): We perform induction on n to show that $C_n \subseteq \overline{C}$. First of all $C_0 = C \subseteq \overline{C}$. Assume that $C_n \subseteq \overline{C}$. Then since $\overline{C}$ is closed under f and g, $f[C_n \times C_n]$ and $g[C_n]$ are contained in $\overline{C}$. Therefore $C_{n+1} \subseteq \overline{C}$. By the inductive principle $C_n \subseteq \overline{C}$ for all $n \in \mathbb{N}$, so $\underline{C} \subseteq \overline{C}$. ■

Notice that $C_n \subseteq C_{n+1}$. Intuitively every C_{n+1} captures images of f and g applied on C_n that might be outside of C_n. Applied

Theorem 3.21 (Structural induction)**.** *Let* $P(x)$ *be a property, and* (A, f, g) *a structure. Assume that*

- $\forall a \in C \ [P(a)]$
- $\forall a, b \in A \ [P(a) \wedge P(b) \implies P(f(a, b)) \wedge P(g(a))]$.

Then $P(x)$ *holds for all* $x \in \overline{C}$.

Proof. This follows from theorem 3.20. Every $x \in \overline{C}$ can be obtained through repeated applications of f and g on some element in C. The suppositions then imply that $P(x)$ is true for all $x \in \overline{C}$. ■

3.4 Peano Arithmetic

1. $S(n) = S(m) \implies n = m$. The success function is one-to-one.
2. $S(n) \neq 0$. 0 is not a successor.
3. $\exists k \ [n \neq 0 \implies n = S(k)]$ Non-zero numbers have a predecessor.

4. $n + 0 = n$. Defining equations for addition
5. $n + S(m) = S(n + m)$.
6. $n \cdot 0 = 0$. Defining equations for multiplication.
7. $n \cdot S(m) = (n \cdot m) + n$.
8. The induction schema: Let A be an arithmetical property. If 0 has property A and if $A(k) \implies A(S(k))$ for every k then every number has property A.

4 Cardinality

Definition 4.1. Let A and B be sets. We say that the *cardinality* of A is less than or equal to the cardinality of B if there is an one-to-one mapping of A into B (i.e. injection). We write $|A| \leq |B|$.

We say that A are *equipotent* (same cardinality) if there is an one-to-one from A onto B (i.e. a bijection). We write $|A| = |B|$. □

Note that at this moment there is no definition for $|A|$ or $|B|$.

Lemma 4.1. *$|A| = |B|$ is reflexive, symmetric, and transitive. $|A| \leq |B|$ is reflexive and transitive.*

Proof. Reflexivity can be satisfied with the identity function. The inverse of the bijection shows symmetry. Composition shows transitivity. ■

Lemma 4.2 (Dedekind)**.** *If $A_1 \subseteq B \subseteq A$ and $|A_1| = |A|$ then $|B| = |A|$.*

Proof. Let $f : A \to A_1$ be the one-to-one mapping. We define two sequences of sets, $\{A_n\}$ and $\{B_n\}$ recursively.

$$\begin{aligned} A_0 &= A, & B_0 &= B \\ A_{n+1} &= f[A_n], & B_{n+1} &= f[B_n] \end{aligned}$$

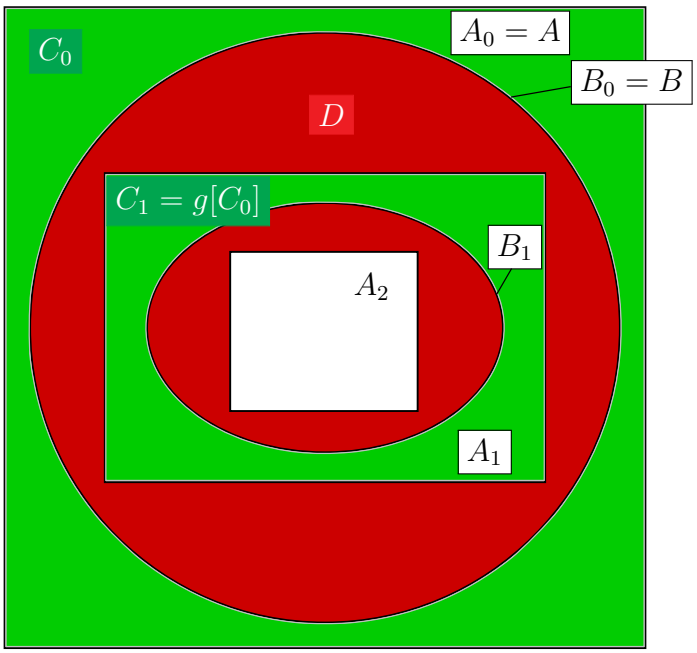
Since $A_0 \supseteq B_0 \supseteq A_1$, we have that $A_n \supseteq B_n \supseteq A_{n+1}$. Now let

$$C_n = A_n - B_n, \qquad C = \bigcup_n C_n, \qquad D = A - C$$

Notice $f[C_n] = f[A_n] - f[B_n] = A_{n+1} - B_{n+1} = C_{n+1}$. This makes use of theorem 2.4. Then let

$$g(x) = \begin{cases} f(x), & \text{if } x \in C \\ x, & \text{if } x \in D \end{cases}$$

See fig. 1 for an illustration of how g "compresses" A into B. Perhaps we may also remark on the similarity between g and Hilbert's Hotel. g is one-to-one since $g \restriction C$ and $g \restriction D$ are one-to-one and their ranges are disjoint. Furthermore, g is onto, since $f[C] \cup D = B$. ■



A_1

Figure 1: Illustration of C and the action of g.

Theorem 4.3 (Cantor-Schröder-Bernstein)**.** *If* $|X| \leq |Y|$ *and* $|Y| \leq |X|$ *then* $|X| = |Y|$.

Proof. Again we must refrain from treating the cardinalities as numbers. In full this states that if we have one-to-one functions $f : X \to Y$ and $g : Y \to X$, then we have a bijection $h : X \to Y$.

$|X| = |f[X]|$ and $|Y| = |g[Y]|$ since f and g are one-to-one. Then $|g[f[X]]| = |f[X]| = |X|$ since g is one-to-one. $g[f[X]] \subseteq g[Y] \subseteq X$. Using lemma 4.2 we have that $|Y| = |g[Y]| = |X|$.

There is also another proof by König. Let $x \in X$. An X-orbit of x is a finite or infinite sequence of the form

$$g^{-1}(x), f^{-1}(g^{-1}(x)), g^{-1}(f^{-1}(g^{-1}(x))), \ldots$$

If the orbit is infinite, we let $x \in X_0$. If the orbit is finite and has an even number of elements (so it stops on the X side), we let $x \in X_1$. If the orbit is finite with an odd number of elements (so it stops on the Y side) we let $x \in X_2$. Now define

$$h(x) = \begin{cases} f(x), & \text{if } x \in X_0 \cup X_1 \\ g^{-1}(x), & \text{if } x \in X_2 \end{cases}$$

h is a function since X_0, X_1, X_2 are disjoint. The domain of h is X, since if $x \in X_0 \cup X_1$ then $f(x)$ is defined, and if $x \in X_2$, then the orbit of x is at least length 1 so $g^{-1}(x)$ is defined.

h is injective. Let $x_1 \neq x_2$. If they are both in $X_0 \cup X_1$ then $h(x_1) \neq h(x_2)$ since f is injective. Similarly for the case when they are both in X_2. Now suppose that $x_1 \in X_0 \cup X_1$ and $x_2 \in X_2$ and $f(x_1) = h(x_1) = h(x_2) = g^{-1}(x_2)$. Then $x_1 = f^{-1}(g^{-1}(x_2))$ and so the orbit of x_2 is either infinite or even length which contradicts the fact that $x_2 \in X_2$.

h is surjective. Let $y \in Y$. Either $y \in \text{ran}(f)$ or not. If $y \in \text{ran}(f)$, then let $y = f(x)$. If $x \in X_0 \cup X_1$ then $h(x) = y$. If $x \in X_2$ then let $z = g(y)$. Observe that $z \in X_2$ so $h(z) = y$. Therefore $y \in \text{ran}(h)$. If $y \notin \text{ran}(f)$ then $z = g(y) \in X_2$ so $h(z) \in y$. ∎

Definition 4.2. A set S is *finite* if it is equipotent to some natural number n. We then define $|S| = n$ and say that S has n elements. A set is *infinite* if it is not finite. □

Theorem 4.4 (Pigeonhole principle)**.** *If $n \in \mathbb{N}$ then there is no one-to-one mapping of n onto a proper subset $X \subset n$.*

Proof. We perform induction on n. Let $P(n)$ be the statement above. Then $P(0)$ is vacuously true since there is no proper subset of 0.

Assuming $P(n)$ is true, we want to show that $P(n+1)$ is also true. Suppose on the contrary that there is a one-to-one mapping $f : n+1 \to X$ where $X \subset n+1$. Either $n \in X$ or not.

If $n \notin X$, then $X \subseteq n$ and $f \restriction n$ maps n onto a proper subset $X - \{f(n)\}$ of n.

If $n \in X$, then $n = f(k)$ for some $k \leq n$. Define

$$g(i) = \begin{cases} f(i), & \text{if } i \neq k, i < n \\ f(n), & \text{if } i = k < n \end{cases}$$

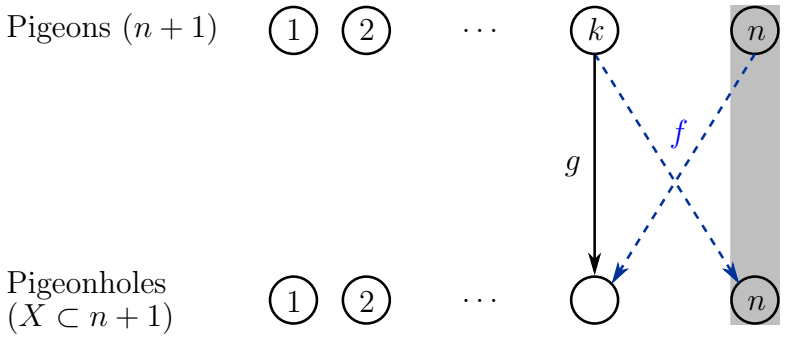


Figure 2: Illustration of the action of g.

Then g is one-to-one and maps n onto $X - \{n\} \subset n$. See fig. 2 for an illustration.

In both cases, the inductive hypothesis is contradicted. ■

Corollary 4.4.1. *The following are true:*

i. If $n \neq m$ then there is no one-to-one mapping of n onto m.

ii. If $|S| = n$ and $|S| = m$ then $n = m$.

iii. $\mathbb{N}$ is infinite.

Proof. We skip the proof for (i) and (ii). For (iii), the successor function is a one-to-one mapping of $\mathbb{N}$ onto its proper subset $\mathbb{N} - \{0\}$. ■

4.1 Countable Sets

Definition 4.3. A set S is *countable* if $|S| = |\mathbb{N}|$. A set S is called *at most countable* if $|S| \leq |\mathbb{N}|$. □

The following theorem tells us that a countable set is like the smallest kind of infinity.

Theorem 4.5. *Any subset A of a countable set is either finite or countable.*

Proof. Without loss of generality we may consider A as a subset of $\mathbb{N}$ We have to show that it is either finite or equipotent with $\mathbb{N}$.

The usual ordering $<$ on $\mathbb{N}$ is a well ordering. Its restriction to A well-orders A. For any non-empty subset of A, there is a least element. Let c be a fixed element not in A. Define $f : \mathbb{N} \to A \cup \{c\}$ recursively:

$$f(n) = \begin{cases} \min(A - \operatorname{ran}(f \restriction n)), & \text{if } A - \operatorname{ran}(f \restriction n) \neq \emptyset \\ c, & \text{otherwise} \end{cases}$$

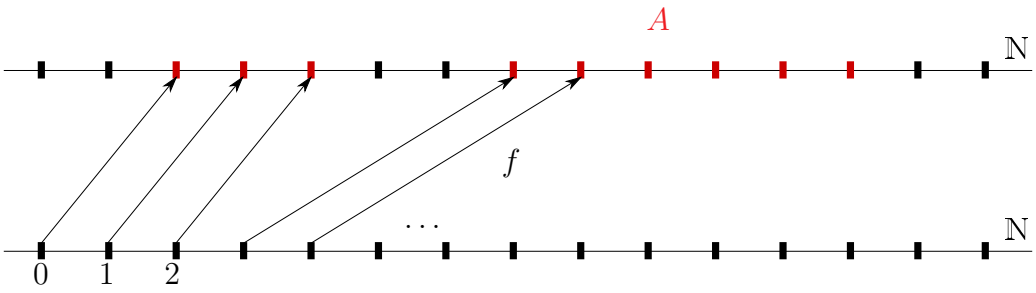


Figure 3: An illustration of the behaviour of f.

If $c \in \operatorname{ran}(f)$, let n be the least element such that $f(n) = c$. Then $f \restriction n : n \to A$ is injective since it is always increasing, and it is surjective since $A - \operatorname{ran}(f \restriction n) = \emptyset$.

If $c \notin \operatorname{ran}(f)$, then $\operatorname{ran}(f) \subseteq A$. $f : \mathbb{N} \to A$ is again injective for the same reason, and can be shown to be surjective through induction, which we skip. ∎

Corollary 4.5.1.

1. *A set is at most countable iff it is either finite or countable*
2. *If X is countable and $f : X \to Y$ is onto, then Y is at most countable.*
3. *An equivalence relation on an at most countable set has at most countably many equivalence classes.*

Theorem 4.6. *The union of two countable sets is a countable set.*

Proof. Suppose $A = \{a_n \mid n \in \mathbb{N}\}$ and $B = \{b_n \mid n \in \mathbb{N}\}$. Then consider $h : \mathbb{N} \to A \cup B$:

$$h(n) = \begin{cases} a_k, & \text{if } n = 2k \\ b_k, & \text{if } n = 2k + 1 \end{cases}$$

This is a surjection, and from corollary 4.5.1 we know that $A \cup B$ is at most countable. We can inductively generalize this to arbitrary unions if we wanted to. ∎

Theorem 4.7. *If A and B are countable, then $A \times B$ is countable.*

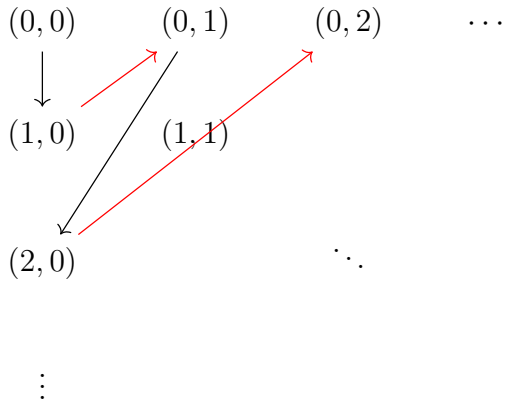


Figure 4: Behaviour of the pairing function.

Proof. It suffices to show that $|\mathbb{N} \times \mathbb{N}| = |\mathbb{N}|$. Figure 4 illustrates a possible function that enumerates all pairs.

To put the formula down, we begin by observing that the k-th line in red encounters $k+1$ elements. Furthermore, (m, n) is on line $m + n$. Thus define

$$\begin{aligned} f(m,n) &= n + \overbrace{\sum_{i=1}^{k} i}^{\substack{\text{number of} \\ \text{points prior}}} \\ &= n + \frac{1}{2}(k+1)k \\ &= n + \frac{1}{2}(m+n+1)(m+n) \end{aligned}$$

We can show that it is bijective by finding its inverse. This part is skipped. ■

Theorem 4.8. *Let $\langle A_n \mid n \in \mathbb{N} \rangle$ be a countable system of at most countable sets and $\langle f_n \rangle$ be the system of enumeration of A_n, i.e. $A_n = \{f_n(k) \mid k \in \mathbb{N}\}$. Then $\bigcup_{n=0}^{\infty} A_n$ is at most countable.*

Proof. Define $F : \mathbb{N} \times \mathbb{N} \to \bigcup_{u=0}^{\infty} A_n$ by $F(n, k) = f_n(k)$. F is surjective, so $\bigcup_{u=0}^{\infty} A_n$ is at most countable. ■

We cannot directly prove this fact without the enumeration functions. This can only be proved using the Axiom of Choice. We will delay the proof until the section where we cover the Axiom of Choice.

Definition 4.4. For all countable sets A, we define $|A| = |\mathbb{N}|$. We define the cardinality of $\mathbb{N}$ as $\aleph_0$. □

The results above can be reformulated in terms of cardinals.

1. $\forall n \in \mathbb{N}\ (\aleph_0 > n)$

- If $\kappa \leq \aleph_0$ then either $\kappa = \aleph_0$ or there is some $n \in \mathbb{N}$ such that $\kappa = n$.

2. If $|A| = \aleph_0$ and $|B| = \aleph_0$ then
   - $|A \cup B| = \aleph_0$
   - $|A \times B| = \aleph_0$.

The following theorem tells us that there is an uncountable set.

Theorem 4.9 (Cantor's Theorem)**.** $|\mathcal{P}(\mathbb{N})| > |\mathbb{N}|$.

Proof. The function $f : \mathbb{N} \to \mathcal{P}(\mathbb{N})$ defined by $f(n) = \{n\}$ is injective and hence $|\mathbb{N}| \leq |\mathcal{P}(\mathbb{N})|$.

Next we show that for every sequence $\langle S_n \mid n \in \mathbb{N} \rangle$ of subsets of $\mathbb{N}$ there is some $S \subseteq \mathbb{N}$ such that $S \neq S_n$ for all $n \in \mathbb{N}$. Define S

$$\forall n \in \mathbb{N}\ (n \in S \iff n \notin S_n)$$

$S \neq S_n$ since they always differ at element n. This is also the famous diagonal argument. ■

Theorem 4.10. $|\mathcal{P}(\mathbb{N})| = \left|2^{\mathbb{N}}\right|$.

Proof. Recall that $2^{\mathbb{N}} = \{f \mid f : \mathbb{N} \to \{0, 1\}\}$. The indicator function $\mathbb{1}_S : \mathbb{N} \to \{0, 1\}$ creates a correspondence between sets and their respective indicator functions. This correspondence is a bijection from $\mathcal{P}(\mathbb{N})$ to $2^{\mathbb{N}}$, since $\mathbb{1}_S$ is bijective for all S. ■

4.2 Countable linear orderings

From the previous sections, $\mathbb{Z}$ is countable because it is a union of two countable sets. Furthermore, $\mathbb{Q}$ is also countable because each rational number can be represented as an ordered pair of $\mathbb{Z}$, and $\mathbb{Z} \times \mathbb{Z}$ is countable. We will give precise definitions of these sets in the future. However what matters is that we cannot distinguish these sets on the basis of their cardinality. What makes them different (on the real line) is the way that they are ordered.

Definition 4.5 (Similar sets)**.** Linearly ordered sets are *similar* (have the same *order type*) if they are isomorphic. □

Example 4.1. $(\mathbb{N}, <)$ and $(\mathbb{Z}, <)$ are not similar since they are not isomorphic ($\mathbb{N}$ has a $<$-least element but $\mathbb{Z}$ does not). $(\mathbb{Z}, <)$ and $(\mathbb{Q}, <)$ are not similar since $\mathbb{Q}$ is dense (we will cover this later on). ◊

It is easy to check that "similar" behaves like an equivalence relation, i.e. it is reflexive, symmetric, and transitive. However it is not an equivalence relation since the collection of all the linear orders do not form a set.

We first begin by establishing that a finite set can only be linearly ordered in one way up to isomorphism.

Lemma 4.11. *Every linear ordering on a finite set is a well-ordering.*

Proof. Let A be a finite set and $<$ is a linear ordering on A. We show that every non-empty subset B of A has a $<$-least element.

B is finite. Perform induction on the number of elements of B. If B has a single element, then that element is the least element. Suppose the statement is true for n elements. Let $B = \{b\} \cup B'$ where B' has n elements. By comparing b with the least element in B' we can find a least element in B. ■

Theorem 4.12. *If $(A_1, <_1)$ and $(A_2, <_2)$ are finite linearly ordered sets and $|A_1| = |A_2|$, then $(A_1, <_1)$ and $(A_2, <_2)$ are similar.*

Proof. We perform induction on the number of elements n in A_1 and A_2. If $n = 0$, then the empty function is the isomorphism between them. We leave out the checking. Assume that the statement is true for some n. Let $|A_1| = |A_2| = n + 1$. Let a_1 and a_2 be the $<_1$ and $<_2$ least elements respectively. By lemma 4.11 they exist. By the inductive hypothesis there exists an isomorphism g between $(A_1 \setminus \{a_1\}, <_1 \cap (A_1 \cap \{a_1\})^2)$ and $(A_2 \setminus \{a_2\}, <_1 \cap (A_2 \cap \{a_2\})^2)$. Define

$$f(a) = \begin{cases} g(a), & \text{if } a \in A_1 \setminus \{a_1\} \\ a_2, & \text{if } a = a_1 \end{cases}$$

Then f is an isomorphism between $(A_1, <_1)$ and $(A_2, <_2)$. ■

Definition 4.6 (Dense sets)**.** An linearly ordered set $(X, <)$ is *dense* if it has at least two elements and

$$\forall a, b \in X, \exists x \in X \ [a < b \implies a < x < b]$$

□

The restriction on two elements is to avoid the situation where we can call a single point dense.

Example 4.2. $(\mathbb{Q}, <)$ is a dense linearly ordered set without endpoints (least or greatest elements). We will prove these facts later one after we define $\mathbb{Q}$. ◊

Definition 4.7. We call a function $f : P \to Q$ a *partial* isomorphism if it is only defined on a subset of P and

$$\forall p, p' \in \operatorname{dom}(h) \ [p \prec p' \iff h(p) < h(p')].$$

□

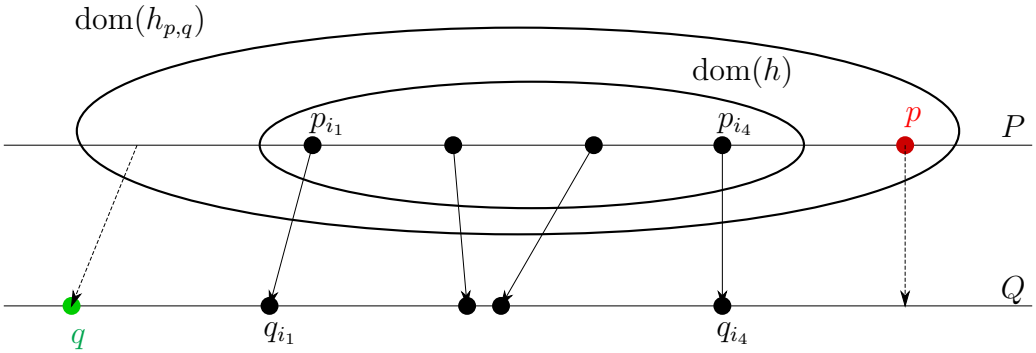
Lemma 4.13 (One step extension lemma)**.** *Let $(P, \prec)$ and $(Q, <)$ be countable dense linearly ordered sets without endpoints and $h : P \to Q$ be a partial isomorphism. If $\operatorname{dom} h$ is finite, then for any $p \in P$ and $q \in Q$, there is a partial isomorphism $h_{p,q} \supseteq h$ such that $p \in \operatorname{dom}(h_{p,q})$ and $q \in \operatorname{ran}(h_{p,q})$.*

Proof. This proof method is also called the "back and forth" method.

Let $\langle p_n \mid n \in \mathbb{N} \rangle$ be a one-to-one sequence such that $P = \{p_n \mid n \in \mathbb{N}\}$ and $\langle q_n \mid n \in \mathbb{N} \rangle$ be a one-to-one sequence such that $Q = \{q_n \mid n \in \mathbb{N}\}$.

Let $h = \{(p_{i_1}, q_{j_1}), \ldots, (p_{i_k}, q_{j_k})\}$ where $p_{i_1} \prec \cdots p_{i_j}$ and thus $q_{j_1} < \cdots q_{j_k}$.

If $p \notin \operatorname{dom}(h)$, we have either $p \prec p_{i_1}$ or $p_{i_e} \prec p \prec p_{i_{e+1}}$ or $p_{i_k} \prec p$. Take the least natural number n such that for $q_n \in Q$,



- $p \prec p_{i_1} \implies q_n < q_{j_1}$,
- $p_{i_e} \prec p \prec p_{i_{e+1}} \implies q_{j_e} \prec q_n \prec q_{j_{e+1}}$,
- $p_{i_k} \prec p \implies q_{j_k} < q_n$.

The existence of such a q_n is guaranteed by $(Q, <)$ being a dense linear ordering without endpoints. Now $h' = h \cup \{(p, q_n)\}$ is a partial isomorphism.

If $q \notin \mathrm{ran}(h')$, then by the same argument as before (with roles of P and Q reversed) we find a $p_m \in P$, and let $h_{p,q} = h' \cup \{p_m, q)\}$. ■

Theorem 4.14. *Let $(P, \prec)$ and $(Q, <)$ be countable dense linearly ordered sets without endpoints. Then $(P, \prec)$ and $(Q, <)$ are similar.*

Proof. We construct a sequence of compatible partial isomorphisms recursively.

$$h_0 = \emptyset$$
$$h_{n+1} = (h_n)_{p_n, q_n}$$

where $(h_n)_{p_n,q_n}$ is the extension of h such that $p_n \in \mathrm{dom}((h_n)_{p_n,q_n})$ and $q_n \in \mathrm{ran}((h_n)_{p_n,q_n})$, as per lemma 4.13. Let $h = \bigcup_{n \in \mathbb{N}} h_n$. Then $h : P \to Q$ is an isomorphism between $(P, \prec)$ and $(Q, <)$. ■

The next theorem shows that there is an universal linear ordering of countable sets.

Theorem 4.15. *Every countable linearly ordered set can be isomorphically mapped into* [5] *any countable dense linearly ordered set without endpoints.*

Proof. This is the forward direction of the proof of lemma 4.13. ■

4.3 Cardinals

The notion that every set A has a cardinality denoted by $|A|$ requires the Axiom of choice. We will not worry about this right now, and assume that this is true.

Definition 4.8 (Cardinal addition)**.** If $|A| = \kappa$ and $|B| = \lambda$, and if $A \cap B = \emptyset$, then

$$\kappa + \lambda = |A \cup B|$$

□

[5] Here this means that we do not require the function to be onto, only that it preserves the order.

This definition depends on sets A and B but we can show that in fact this does not matter.

Lemma 4.16. *Addition of cardinals is well-defined. If* $|A| = |A'|$, $|B| = |B'|$, *and* $A \cap B = A' \cap B' = \emptyset$, *then* $|A \cup B| = |A' \cup B|$.

Proof. Take bijections $\alpha : A \to A'$ and $\beta : B \to B'$. We can check that $h : A \cup B \to A' \cup B'$ where $h = \alpha \cup \beta$ is also a bijection since h is a function, and because of the disjoint properties of the sets. ■

Definition 4.9 (Cardinal multiplication)**.** If $|A| = \kappa$ and $|B| = \lambda$,

$$\kappa \cdot \lambda = |A \times B|$$

□

Similar to addition, this also does not depend on the representatives.

Lemma 4.17. *Multiplication of cardinals is well-defined. If* $|A| = |A'|$, $|B| = |B'|$, *then* $|A \times B| = |A' \times B|$.

Theorem 4.18 (Arithmetic laws)**.**

i. $\kappa + \lambda = \lambda + \kappa$ *and* $\kappa \cdot \lambda = \lambda \cdot \kappa$

ii. $\kappa + (\lambda + \mu) = (\kappa + \lambda) + \mu$ *and* $\kappa \cdot (\lambda \cdot \mu) = (\kappa \cdot \lambda) \cdot \mu$

iii. $\kappa \cdot (\lambda + \mu) = \kappa \cdot \lambda + \kappa \cdot \mu$.

Proof. Let $\kappa = |A|$, $\lambda = |B|$, and $\mu = |C|$.

i. It is quite clear that $|A \cup B| = |B \cup A|$. Also, $|A \times B| = |B \times A|$ since we can find a bijection between them that maps (a, b) to (b, a).

ii. Addition is straightforward. For multiplication, we can also find a bijection as above.

iii. For this we assume that $B \cap C = \emptyset$. We have to show that $|A \times (B \cup C)| = |(A \times B) \cup (A \times C)|$. Note that $(A \times B) \cap (A \times C) = \emptyset$. It is not hard to find a bijection from the pairs in the left to the right. ■

However, not all the laws of addition and multiplication of finite number hold for cardinals. For example, $\aleph_0 + \aleph_0 = \aleph_0$, so there is no cancellation law.

Definition 4.10 (Cardinal exponentiation)**.** If $|A| = \kappa$ and $|B| = \lambda$, then

$$\kappa^\lambda = \left|A^B\right|$$

□

Lemma 4.19. *Cardinal exponentiation is well defined If* $|A| = |A'|$ *and* $|B| = |B'|$, *then* $\left|A^B\right| = \left|{A'}^{B'}\right|$.

Proof. Take bijections $\alpha : A \to A'$ and $\beta : B \to B'$. For any function $f : B \to A$, we can define $f' : B' \to A'$ by $f' = \alpha \circ f \circ \beta^{-1}$. Then $f \mapsto f'$ is a bijection. ■

Theorem 4.20 (Exponentiation laws)**.**

i. $\kappa^{\lambda+\mu} = \kappa^\lambda \cdot \kappa^\mu$.

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

iii. $(\kappa \cdot \lambda)^\mu = \kappa^\mu \cdot \lambda^\mu$.

Proof. Let $\kappa = |A|$, $\lambda = |B|$, and $\mu = |C|$.

i. We want to show that $\left|A^{B\cup C}\right| = \left|A^B \times A^C\right|$. Here we have to assume $B \cap C = \emptyset$. Then for each $(f_1, f_2) \in A^B \times A^C$, define $g : B \cup C \to A = f_1 \cup f_2$. Then let $F(f_1, f_2) = g$. F is a bijection from $A^B \times A^C$ to $A^{B\cup C}$.

ii. We want to show that $\left|(A^B)^C\right| = \left|A^{B\times C}\right| = \left|A^{C\times B}\right|$. Take any $f \in A^{C\times B}$. Define $F(f) = \{(c, (b, a)) \mid ((c, b), a) \in f\}$. F is a bijection from $A^{C\times B}$ to $(A^B)^C$.

iii. We want to show that $\left|(A \times B)^C\right| = \left|A^C \times B^C\right|$. For each $(f_1, f_2) \in A^C \times B^C$, define $g : C \to (A \times B)$ where $g(c) = (f_1(c), f_2(c))$. Then let $F(f_1, f_2) = g$. F is a bijection from $A^C \times B^C$ to $(A \times B)^C$.

■

Theorem 4.21 (Cantor's theorem)**.** $|X| \leq |\mathcal{P}(X)|$ *for every set* X.

Proof. Consider $f : X \to \mathcal{P}(X)$ defined by $f(x) = \{x\}$. This is one to one so $|X| \leq |\mathcal{P}(X)|$.

We have to then show that for any function $f : X \to \mathcal{P}(X)$ it cannot be surjective. Note that for some $x \in X$, $f(x) \subseteq X$. Then consider the set $S = \{x \in X \mid x \notin f(x)\}$. We claim that S is not in the range of f. Suppose that $S = f(z)$ for some $z \in X$, then

$$z \in S \iff z \in f(z) \iff z \notin S.$$

■

Theorem 4.22. $|\mathcal{P}(X)| = \left|2^X\right|$ *for every set* X.

Proof. Refer to theorem 4.10. ■

Corollary 4.22.1. *For any system of sets* S *there is a set* Y *such that* $|Y| > |X|$ *holds for all* $X \in S$.

Proof. Let $Y = \mathcal{P}(\bigcup S)$. ■

Theorem 4.23 (Miscellaneous identities)**.** *Here are some useful identities.*

i. $\kappa + \kappa = 2 \cdot \kappa$.

ii. $\kappa^2 = \kappa \cdot \kappa$.

iii. $\kappa^\kappa \leq 2^{\kappa\cdot\kappa}$.

Proof. For the following, let $|A| = |B| = \kappa$.

i. For this we have to assume $A \cap B = \emptyset$. Let $f : A \to B$ be a bijection. Also let $x \in A \cup B$. Either $x \in A$ or $x \in B$. Define $g : A \cup B \to \{0, 1\} \times A$ by

$$g(x) = \begin{cases} (0, f^{-1}(x)), & \text{if } x \in B \\ (1, x), & \text{if } x \in A \end{cases}$$

g is a bijection.

ii. define $h : A^{\{0,1\}} \to A \times A$ by

$$h(f) = (f(0), f(1))$$

h is a bijection.

iii. Take any $f : B \to A$. Define $h : A^B \to \{0, 1\}^{A \times B}$ by

$$(b, a) \in f \iff ((a, b), 0) \in h(f) \qquad (b, a) \notin f \iff ((a, b), 1) \in h(f)$$

h is surjective, because given any $g : A \times B \to \{0, 1\}$, create $f' : B \to A$ such that

$$g(a, b) = 0 \implies f(b) = a \qquad g(a, b) = 1 \implies f(b) = a$$

■

5 Other sets

5.1 Integers

We now give a set-theoretic definition of integers. Many proofs will be skipped since they are just definition checking. The purpose of this is to show that set theory can construct these concepts, and not to replace them.

Let $\mathbb{Z}'$ be the set $\mathbb{N} \times \mathbb{N}$, and define a relation $\sim$ on $\mathbb{Z}'$ by

$$(a, b) \sim (c, d) \iff a + d = b + c.$$

Then $\sim$ is an equivalence relation on $\mathbb{Z}'$. The purpose of this is to mimic subtraction which we do not have right now. Thus (a, b) looks something like $a - b$.

Let $\mathbb{Z}$ be the set $\mathbb{Z}'/\sim$. We call $\mathbb{Z}$ the set of all integers.

The ordering $<$ on $\mathbb{Z}$ is defined by

$$[(a, b)] < [(c, d)] \iff a + d <_{\mathbb{N}} b + c.$$

Lemma 5.1. *$<$ is well-defined and it is a linear ordering.*

Theorem 5.2. *$\mathbb{Z}$ is countable.*

Proof. Since $\mathbb{N} \times \mathbb{N}$ is countable, and any equivalence classes over $\mathbb{Z}'$ is also countable. See corollary 4.5.1. ■

We can embed $\mathbb{N}$ into $\mathbb{Z}$ using the function $i : \mathbb{N} \to \mathbb{Z}$ defined by $i(n) = [(n, 0)]$ for any $n \in \mathbb{N}$. It is order preserving and one-to-one.

Next we define addition $+$ and multiplication $\cdot$ in terms of the addition and multiplication operations in $\mathbb{N}$:

$$[(a, b)] + [(c, d)] = [(a + c, b + d)]$$
$$[(a, b)] \cdot [(c, d)] = [(ac + bd, ad + bc)]$$

We can also define the additive inverse $-[(a, b)] = [(b, a)]$. Then subtraction can be defined in terms of addition.

Theorem 5.3. *The operations of addition and multiplication are well-defined and satisfy the usual laws of algebra.*

5.2 Rational Numbers

Let $\mathbb{Q}'$ be the set $\mathbb{Z} \times (\mathbb{Z} - \{0\})$, and define a relation $\sim$ on $\mathbb{Q}'$ by

$$(a, b) \sim (c, d) \iff ad = bc.$$

Then $\sim$ is an equivalence relation on $\mathbb{Q}'$. Again here we are mimicking division which we do not have. (a, b) looks like $\frac{a}{b}$.

Let $\mathbb{Q}$ be the set $\mathbb{Q}'/\sim$. We call $\mathbb{Q}$ the set of all rational numbers.

Theorem 5.4. $\mathbb{Q}$ *is countable.*

Proof. $\mathbb{Q}'$ is countable and again by corollary 4.5.1 the set of equivalence classes are countable. ■

The ordering $<$ on $\mathbb{Q}$ is defined by

$$[(a, b)] < [(c, d)] \iff ad <_{\mathbb{N}} bc$$

when both $b, d >_{\mathbb{N}} 0$.

$\mathbb{Z}$ can be embedded in $\mathbb{Q}$ using the function $i : \mathbb{Z} \to \mathbb{Q}$ defined by $i(z) = [(z, 1)]$ for any $z \in \mathbb{Z}$. It is one-to-one and order preserving.

Addition $+$ and multiplication $\cdot$ can be defined in terms of the same operations in $\mathbb{N}$:

$$[(a, b)] + [(c, d)] = [(ad + bc, bd)]$$
$$[(a, b)] \cdot [(c, d)] = [(ac, bd)]$$

For all $[(a, b)] \neq [(0, 1)]$, we can define its multiplicative inverse $[(a, b)]^{-1} = [(b, a)]$. Then we can define division $/$ in terms of multiplication.

Theorem 5.5. *The operations of addition and multiplication are well-defined and satisfy the usual laws of algebra.*

Theorem 5.6. $(Q, <)$ *is a dense linearly ordered set without endpoints.*

Proof. Let us now write a/b instead of $[(a, b)]$.

Since $a/b - 1 < a/b < a/b + 1$, $\mathbb{Q}$ has no endpoints.

For any $r, s \in \mathbb{Q}$ such that $r < s$, assume that $r = a/b$ and $s = c/d$ where $b, d > 0$. Then let $x = (ad + bd)/2bd = (r + s)/2$. Then $r < x < s$. ■

5.3 Real Numbers

The rational numbers has gaps in the form of irrational numbers. We will define the real numbers to fill in these gaps. One method is the Dedekind cut, and another is using Cauchy sequences. We will explore the Dedekind cut in the following section. We will also explore some interesting properties of the real numbers.

Definition 5.1. Let $(P, <)$ be a dense linearly ordered set. We say P is *complete* if every non-empty $S \subseteq P$ bounded from above has a supremum. □

Definition 5.2 (Dedekind cuts)**.** A Dedekind *cut* in $\mathbb{Q}$ is a subset $A \subseteq \mathbb{Q}$ such that

- $A \neq \emptyset$ and $A \neq \mathbb{Q}$.
- $\forall p \in \mathbb{Q}, \forall q \in A \ (p < q \implies p \in A)$.
- A does not have a greatest element. □

The set of all real numbers $\mathbb{R}$ is the set of all cuts in $\mathbb{Q}$.

Since a cut is never $\mathbb{Q}$, we can always find some element outside the cut and so $\mathbb{R}$ has no endpoints.

Define the relation $<$ on $\mathbb{R}$ to be

$$A < B \iff A \subset B.$$

We can embed $\mathbb{Q}$ in $\mathbb{R}$. The function $i : \mathbb{Q} \to \mathbb{R}$ defined by $i(p) = \{q \in \mathbb{Q} \mid q < p\}$.

Theorem 5.7. $\mathbb{Q}$ *is dense in* $\mathbb{R}$.

Proof. For any $s, t \in \mathbb{R}$ where $s < t$, take any $r \in t \setminus s$. Then $s < i(r) < t$. We skip the checking. ■

Theorem 5.8. $\mathbb{R}$ *is complete.*

Proof. Let S be a non-empty subset of $\mathbb{R}$ bounded from above. Then there exists some $A_0 \in \mathbb{R}$ such that for all $A \in S$, $A \subseteq A_0$. Let

$$A_S = \bigcup \{A \mid A \in S\}.$$

We can check that A_S is a cut.

- $A_S \neq \emptyset$ obviously.
- Furthermore since S is bounded from above so $A_S \neq \mathbb{Q}$.
- If $q \in A_S$ and $p < q$, then there is some $A \subset A_S$ such that $q \in A$. Then $p \in A$ as well and so $p \in A_S$.
- Suppose there was a greatest element. This element would come from one of the cuts. But since cuts do not have greatest elements, we can simply find another member larger than the greatest element.

We check that A_S is the supremum of S. Since for all $A \in S$, $A \subseteq A_S$, so A_S is an upper bound for S. If A' is another upper bound for S, then for all $A \in S$, $A \subseteq A'$. Then the union of all A's is still a subset of or equal to A'. Therefore A_S is the supremum of S. ■

We can define addition $+$ by

$$A + B = \{p +_{\mathbb{Q}} q \mid p \in A \wedge q \in B\}$$

Using this we can define the additive inverse, subtraction, and absolute value.

For multiplication $\cdot$ it is more troublesome. If A and B are non-negative,

$$A \cdot B = 0_{\mathbb{R}} \cup \{rs \mid 0 \leq r \in x \wedge 0 \leq s \in y\}.$$

If they are both negative, then

$$A \cdot B = |A| \cdot |B|$$

and if only one of them is negative,

$$A \cdot B = -(|A| \cdot |B|)$$

Theorem 5.9 (Cardinality of the continuum)**.** *The cardinality of* $\mathbb{R}$ *is* $2^{\aleph_0}$.

Proof. Since $\mathbb{R} \subseteq \mathcal{P}(\mathbb{Q})$, so $|\mathbb{R}| \leq 2^{|\mathbb{Q}|} = 2^{\aleph_0}$. Therefore $|\mathbb{R}| \leq 2^{\aleph_0}$.

On the other hand, let S be the set of $x \in [0, 1)$ whose decimal expansion consists of only 0's and 1's. Then S is in a one-to-one correspondence with the est of all infinite sequences of 0's and 1's. So $|S| = 2^{\aleph_0}$ and $|\mathbb{R}| \geq 2^{\aleph_0}$.

By theorem 4.3, $|\mathbb{R}| = 2^{\aleph_0}$. ■

Theorem 5.10. $2^{\aleph_0} \cdot 2^{\aleph_0} = 2^{\aleph_0}$.

Proof. It suffices to show that $|(0, 1] \times (0, 1]| = |(0, 1]|$.

Take the decimal expansions of every $x \in (0, 1]$. To ensure we always get an unique expansion, let us take the expansion that does not end with infinite 9's, i.e. $0.10\ldots$ instead of $0.09\ldots$. Now divide the sequence of digits in the decimal expansion into blocks from left to right, where we find the first digit that is not 9 and form a block ending with that digit. For example:

$$0.98765\ldots \qquad 0.36912\ldots$$

Take any $(x, y) \in (0, 1] \times (0, 1]$. Form a read number z by interlacing the blocks of decimal expansions of x and y. For example with the two numbers above, we can form

$$z = 0.98376691\ldots$$

Furthermore, given a z we can uniquely reconstruct x and y. Hence this forms a bijection as required.

The rationale for the blocks is to prevent cases such as $z = 0.090909\ldots$ that has no pre-image. In our case every block will have a digit different from 9, so we do not encounter this problem. ■

Corollary 5.10.1.

i. $|\mathbb{R}^n| = 2^{\aleph_0}$.

ii. $|\mathbb{C}| = 2^{\aleph_0}$.

Theorem 5.11 (Arithmetic properties of $2^{\aleph_0}$)**.**

i. $\forall n \in \mathbb{N}\ \left[n + 2^{\aleph_0} = \aleph_0 + 2^{\aleph_0} = 2^{\aleph_0} + 2^{\aleph_0} = 2^{\aleph_0}\right]$

ii. $\forall n \in \mathbb{N}, n > 0\ \left[n \cdot 2^{\aleph_0} = \aleph_0 \cdot 2^{\aleph_0} = 2^{\aleph_0} \cdot 2^{\aleph_0} = 2^{\aleph_0}\right]$

iii. $\forall n \in \mathbb{N}, n > 0\ \left[\left(2^{\aleph_0}\right)^n = \left(2^{\aleph_0}\right)^{\aleph_0} = 2^{\aleph_0}\right]$

iv. $\forall n \in \mathbb{N}, n > 1\ \left[n^{\aleph_0} = {\aleph_0}^{\aleph_0} = 2^{\aleph_0}\right]$

Proof. For the proof on why $\aleph_0 \cdot \aleph_0 = \aleph_0$, see theorem 4.7. For $\kappa + \kappa = 2\kappa$, see theorem 4.23.

i. $2^{\aleph_0} \leq n + 2^{\aleph_0} \leq \aleph_0 + 2^{\aleph_0} \leq 2^{\aleph_0} + 2^{\aleph_0} = 2 \cdot 2^{\aleph_0} = 2^{1+\aleph_0} = 2^{\aleph_0}$.

ii. $2^{\aleph_0} \leq n \cdot 2^{\aleph_0} \leq \aleph_0 \cdot 2^{\aleph_0} \leq 2^{\aleph_0} \cdot 2^{\aleph_0} = 2^{\aleph_0+\aleph_0} = 2^{\aleph_0}$.

iii. $2^{\aleph_0} \leq 2^{n\cdot\aleph_0} = \left(2^{\aleph_0}\right)^n \leq \left(2^{\aleph_0}\right)^{\aleph_0} = 2^{\aleph_0\cdot\aleph_0} = 2^{\aleph_0}$.

iv. $2^{\aleph_0} \leq n^{\aleph_0} \leq {\aleph_0}^{\aleph_0} \leq \left(2^{\aleph_0}\right)^{\aleph_0} = 2^{\aleph_0\cdot\aleph_0} = 2^{\aleph_0}$.

■

Corollary 5.11.1.

i. The set of all infinite sequences of natural numbers has cardinality $2^{\aleph_0}$.

ii. The set of all finite sequences of real numbers has cardinality $2^{\aleph_0}$.

Proof.

i. First of all we can treat an infinite sequence on $\mathbb{N}$ as a function from $\mathbb{N}$ to $\mathbb{N}$ and so the set in question is $\mathbb{N}^{\mathbb{N}}$. Furthermore, each function forms a subset of $\mathbb{N} \times \mathbb{N}$. Then $2^{\aleph_0} = \left|2^{\mathbb{N}}\right| \leq \left|\mathbb{N}^{\mathbb{N}}\right| \leq |\mathcal{P}(\mathbb{N} \times \mathbb{N})| = 2^{\aleph_0}$.

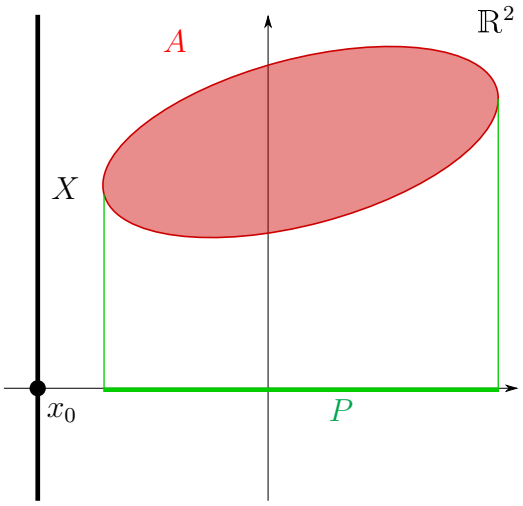
ii. Call the set in question S. Certainly $|\mathbb{R}| < |S|$ since we can form a injection $x \mapsto \{x\}$. Next, observe that the set of all sequences of length n is a subset of $\mathbb{R}^n$. Then S is the countably infinite union of all such sets, and so $|S| \leq \aleph_0 \cdot 2^{\aleph_0} = 2^{\aleph_0}$.

■

Theorem 5.12. *If A is a countable subset of $\mathbb{R}$, then $|\mathbb{R} - A| = 2^{\aleph_0}$.*

Proof. We use the fact that $|\mathbb{R} \times \mathbb{R}| = 2^{\aleph_0}$ and show the statement for $|\mathbb{R} \times \mathbb{R}|$ instead.

So suppose $A \subseteq \mathbb{R} \times \mathbb{R}$ is countable. Let $P = \text{dom}(A)$.



Since $|A| = \aleph_0$, we have $|P| \leq \aleph_0$. Thus there exists $x_0 \in \mathbb{R}$ such that $x_0 \notin P$. Consequently, $X = \{x_0\} \times \mathbb{R}$ is disjoint from A, so $X \subseteq (\mathbb{R} \times \mathbb{R}) - A$. It is clear that $|X| = 2^{\aleph_0}$, and we have $|(\mathbb{R} \times \mathbb{R}) - A| \geq 2^{\aleph_0}$. ■

Corollary 5.12.1.

i. The set of all irrational numbers has cardinality $2^{\aleph_0}$.

ii. The set of all infinite subsets of $\mathbb{N}$ has cardinality $2^{\aleph_0}$.

iii. The set P of all bijections from $\mathbb{N}$ to $\mathbb{N}$ has cardinality $2^{\aleph_0}$.

Proof.

i. The rationals are countable and so the complement of the rationals in the reals has cardinality $2^{\aleph_0}$.

ii. The set of all finite subsets of $\mathbb{N}$ is countable. This is the set $\{\mathbb{N} \cup \mathbb{N}^2 \cup \cdots\}$. We also know that $|\mathcal{P}(\mathbb{N})| = 2^{\aleph_0}$. Then taking complements we get our result.

iii. Since $\left|\mathbb{N}^{\mathbb{N}}\right| = 2^{\aleph_0}$, it suffices to show that $|P| \geq 2^{\aleph_0}$. Let E and O be the set of all even and odd numbers respectively. If $X \subseteq E$ is infinite, define a function $f_X : \mathbb{N} \to \mathbb{N}$ with $n \in \mathbb{N}$ as

$$f_X(2n) = n\text{-th element of } X$$
$$f_X(2n+1) = n\text{-th element of } \mathbb{N} - X.$$

f_X is surjective since we cover every element in X and also in $\mathbb{N} - X$. f_X is injective because the images obviously do not coincide. Thus f_X is a bijection.

Furthermore, $X_1 \neq X_2 \implies f_{X_1} \neq f_{X_2}$. Therefore we have a one-to-one map from the infinite subsets of E to P defined by $X \mapsto f_X$. Therefore $|P| \geq 2^{\aleph_0}$.

■

Theorem 5.13. *The set of all real-valued functions on real numbers has cardinality* $2^{2^{\aleph_0}} > 2^{\aleph_0}$.

Proof. The cardinal number of $\mathbb{R}^{\mathbb{R}}$ is $\left(2^{\aleph_0}\right)^{2^{\aleph_0}} = 2^{\aleph_0 \cdot 2^{\aleph_0}} = 2^{2^{\aleph_0}}$. See theorem 5.11 for a justification of the last step. ■

Theorem 5.14. *The set C of all continuous functions from $\mathbb{R}$ to $\mathbb{R}$ has cardinality* $2^{\aleph_0}$.

Proof. A function $f \in C$ is determined by its values on $\mathbb{Q}$. This means if $f_1, f_2 \in C$, and $f_1 \restriction \mathbb{Q} = f_2 \restriction \mathbb{Q}$, then $f_1 = f_2$. This is because we can approximate any $a \in \mathbb{R}$ by a sequence of rationals $\{q_n\} \to a$. Then $f_1(a) = \lim f_1(q_n) = \lim f_2(q_n) = f_2(a)$. Therefore $|C| \leq \left|\mathbb{R}^{\mathbb{Q}}\right| = 2^{\aleph_0}$.

On the other hand, considering constant values functions alone we see that $|C| \geq 2^{\aleph_0}$. ■

5.4 Topics related to $\mathbb{R}$

Since we have achieved our goal of demonstrating how set theory can be used to construct $\mathbb{Z}$, $\mathbb{Q}$, and $\mathbb{R}$, we will now use them in a more common-sense way, i.e. we will not be using the set-theoretic definitions unless stated explicitly.

Definition 5.3 (Open sets)**.** A set $A \subseteq \mathbb{R}$ is open if

$$\forall a \in A, \exists \delta > 0 \ (|x - a| < \delta \implies x \in A).$$

In other words, there is an open interval (neighbourhood) $(a - \delta, a + \delta) \subseteq A$.

A set B is *closed* if $\mathbb{R} \setminus B$ is open. □

Theorem 5.15. *Every system of mutually disjoint open intervals in $\mathbb{R}$ is at most countable.*

Proof. Call the system S. Let $P = \mathbb{Q} \cap S$. Since $\mathbb{Q}$ is dense in $\mathbb{R}$, each open interval in S contains at least one element of P. Now define a function $f : P \to S$ by mapping a point p to the interval in S containing it. As elements of S are mutually disjoint, f is surjective, so S is at most countable. ■

Corollary 5.15.1. *The set of all closed subsets of reals has cardinality* $2^{\aleph_0}$.

Proof. We can make a one-to-one map from each closed subset to its complement, which is open. ■

Lemma 5.16.

i. The intersection of a finite system of open sets is open.

ii. The union of a finite system of closed sets is closed.

Proof.

i. Let $a \in A \cap B$. Then

$$\exists \delta_A > 0, \forall x \ (|x - a| < \delta_A \implies x \in A)$$
$$\exists \delta_B > 0, \forall x \ (|x - a| < \delta_B \implies x \in A)$$

Let $\delta = \min(\delta_A, \delta_B)$, then $\forall x \ (|x - a| < \delta \implies x \in A \cap B)$.

ii. This follows from the above and De Morgan's laws.

■

Lemma 5.17. *Every open set is an union of a system of open intervals with rational endpoints.*

Proof. Let A be open and let S be the system of all open intervals with rational endpoints included in A. Then $\bigcup S \subseteq A$.

If $a \in A$, then $(a - \delta, a + \delta) \subseteq A$ for some $\delta > 0$. Since $\mathbb{Q}$ is dense in $\mathbb{R}$, we can find $r_1, r_2 \in \mathbb{Q}$ such that

$$a - \delta < r_1 < a < r_2 < a + \delta$$

Then $a \in (r_1, r_2) \subseteq A$ so $a \in \bigcup S$. ■

Theorem 5.18. *The set O of all open subsets of the reals has cardinality $2^{\aleph_0}$.*

Proof. $|O| \geq 2^{\aleph_0}$, since for all $x \in \mathbb{R}$, (x, ∞) is open.

On the other hand, each open set is a union of a system of open intervals with rational endpoints. Therefore there are at most $2^{\aleph_0}$ such systems, and so $|O| \leq 2^{\aleph_0}$. ■

Definition 5.4 (Accumulation points)**.** $a \in \mathbb{R}$ is an *accumulation (limit) point* of $A \subseteq \mathbb{R}$ if

$$\forall \delta > 0, \exists x \in A \ (x \neq a \ \wedge |x - a| < \delta).$$

□

Definition 5.5 (Isolated points)**.** $a \in \mathbb{R}$ is an *isolated point* of $A \subseteq \mathbb{R}$ if

$$\exists \delta > 0, \forall x \neq a \ (|x - a| < \delta \implies x \notin A).$$

□

Lemma 5.19. *A set $A \subseteq \mathbb{R}$ is closed iff all accumulation points of A belong to A.*

Proof.

($\implies$): Assume that A is closed. Then $\mathbb{R} \setminus A$ is open. If $a \in \mathbb{R} \setminus A$, then

$$\exists \delta > 0, \forall x \ (|x - a| < \delta \implies x \in \mathbb{R} \setminus A \implies x \notin A)$$

so a cannot be an accumulation point of A.

($\impliedby$): Take any $a \in \mathbb{R} \setminus A$. By supposition a is not an accumulation point. Therefore, $\exists \delta > 0, \forall a \in A \ (x = a \vee |x - a| > \delta)$ which means that $\forall a \notin A, \exists \delta > 0 \ (|x - a| < \delta \implies x \notin A)$. ■

Although a closed set contains all of their accumulation points, a closed set may or may not contain isolated points.

Definition 5.6 (Perfect sets)**.** A non-empty set A is called a *perfect set* if A is closed without isolated points. □

Example 5.1. Closed intervals are perfect sets. ◊

Let $S = \bigcup_{n\in\mathbb{N}} \{0,1\}^n$ be the set of all finite sequences of 0's and 1's. We construct a system of closed intervals $D = \langle D_s \mid s \in S\rangle$ as follows [6]:

$$D = [0,1]$$
$$D_0 = \left[0, \frac{1}{3}\right] \qquad D_1 = \left[\frac{2}{3}, 1\right]$$
$$D_{0,0} = \left[0, \frac{1}{9}\right] \qquad D_{0,1} = \left[\frac{2}{9}, \frac{1}{3}\right] \qquad D_{1,0} = \left[\frac{2}{3}, \frac{7}{9}\right] \qquad D_{1,1} = \left[\frac{8}{9}, 1\right]$$
$$\vdots$$

In general if $D_{s_0,\ldots,s_n} = [a,b]$, then $D_{s_0,\ldots,s_n,0}$ is the left one third of $[a,b]$ and $D_{s_0,\ldots,s_n,1}$ is the right one third of $[a,b]$.

Now let $C_n = \bigcup\{D_s \mid s \in \{0,1\}^n\}$, so that $C_0 = [0,1], C_1 = \left[0, \frac{1}{3}\right] \cup \left[\frac{2}{3}, 1\right]$, etc. Since the C_n's are all unions of a finite system of closed intervals, they are closed.

The set $C = \bigcap_{n\in\mathbb{N}} C_n$ is called *Cantor's set*. This is closed, since it is the intersection of closed intervals.

Theorem 5.20. *C has cardinality of the continuum $2^{\aleph_0}$,*

Proof. We wish to find a bijection from $\{0,1\}^{\mathbb{N}}$ onto C.

For any $f \in \{0,1\}^{\mathbb{N}}$ we let $D_f = \bigcap_{n\in\mathbb{N}} D_{f\restriction n}$ [7]. Note that $D_f \subseteq C$ and $D_f \neq \emptyset$. Furthermore the length of the interval $D_{f\restriction n}$ is $\frac{1}{3^n}$. Thus D_f contains a single unique element d_f.

Conversely, for any $a \in C$, there is an unique $f \in \{0,1\}^{\mathbb{N}}$ such that $a = d_f$. In fact $f = \bigcup\{s \mid a \in D_s\}$. Therefore this mapping from f to D_f is a bijection. ■

Theorem 5.21. *C is a perfect set.*

Proof. We know that C is closed and non-empty, so it remains to show that it has no isolated points.

Let $a \in C$ and $\delta > 0$. We need to find $x \in C, x \neq a$ where $|x - a| < \delta$. Take $n \in \mathbb{N}$ such that $\frac{1}{3^n} < \delta$. From the proof of the previous theorem, we know that $a = d_f$ for some $f \in \{0,1\}^{\mathbb{N}}$. Let $x = d_g$ where $g \in \{0,1\}^{\mathbb{N}}$ is another sequence such that $g \restriction n = f \restriction n$ and $f \neq g$. Then $a \neq x$. Furthermore $a, x \in D_{f\restriction n} = D_{g\restriction n}$, which has length $\frac{1}{3^n} < \delta$ so $|x - a| < \delta$. ■

[6] We write the sequences as though they are just lists of numbers for convenience

[7] This is because our definition is only for finite sequences. However this lets us generalise it to infinite sequences f.

Theorem 5.22. *The relative complement of C in $[0,1]$ is dense in $[0,1]$.*

Proof. Let $0 \leq a < b \leq 1$. We will show that the interval (a,b) contains elements not in C. Take a $n \in \mathbb{N}$ such that $\frac{1}{3^n} < \frac{b-a}{2}$, and let $k \in \mathbb{N}$ be the smallest natural number such athat $\frac{k}{3^n} \geq a$. This means that $a \leq \frac{k}{3^n} < \frac{k+1}{3^n} < b$. However, the middle third of the interval $\left[\frac{k}{3^n}, \frac{k+1}{3^n}\right]$ is removed at some point and so is disjoint from C. ■

6 The continuum hypothesis

Definition 6.1 (Continuum hypothesis)**.** The continuum hypothesis (CH) states that there is no uncountable cardinal number κ such that $\kappa < 2^{\aleph_0}$. □

This problem was originally raised by Cantor. There is a partial result, that CH holds for open and closed subsets of $\mathbb{R}$. The question is much more difficult for arbitrary sets, even for arbitrary subsets of natural numbers.

Theorem 6.1. *Every non-empty open set of reals has cardinality $2^{\aleph_0}$.*

Proof. Every open set contains an open interval, and every open interval has cardinality $2^{\aleph_0}$. ■

Recall that a *perfect* set is a closed set without isolated points. For example, a closed interval and the Cantor set are perfect sets. The proof that the Cantor set has cardinality of the continuum serves as a template for our next theorem. Before that, we will need two lemmas. Let P denote the system of all perfect subsets of $\mathbb{R}$.

Lemma 6.2. *There are functions $G_0, G_1 : P \to P$, such that for each $F \in P$, $G_0(F) \subseteq F$ and $G_1(F) \subseteq F$ and $G_0(F) \cap G_1(F) = \emptyset$. In other words, every perfect set contains two disjoint perfect subsets.*

Proof. We show that for every perfect set F there are rational numbers $r < s$, such that $F \cap (-\infty, r]$ and $F \cap [s, \infty)$ are both perfect.

Let $\alpha = \inf F$, and if F is not bounded below then let $\alpha = -\infty$. Similarly let $\beta = \sup F$, and if F is not bounded above then let $\beta = \infty$. There are two cases.

- $(\alpha, \beta) \subseteq F$. Then any rational numbers r,s such that $\alpha < r < s < \beta$ have the desired property.

  First of all $(-\infty, r] \cap F$ is non-empty. It is closed because it is the intersection of two closed sets. There are no isolated points, since there are no isolated points in intervals, namely, the interval (α, β). The argument is the same for the other side.

- $(\alpha, \beta) \not\subseteq F$. Then there exists $a \in (a, \beta)$ such that $a \notin F$. Therefore there exists $\delta > 0$ such that $(a - \delta, a + \delta) \cap F \neq \emptyset$. This is because F is perfect. If for all neighbourhoods of a, there is a point in that neighbourhood that is also in F, then a is a limit point. But perfect sets contain all of its limit points. Furthermore, $\alpha < a - \delta$ and $a + \delta < \beta$, since they are the infimum and supremum respectively. Then any rational numbers r,s such that $a - \delta < r < s < a + \delta$ have the desired property.

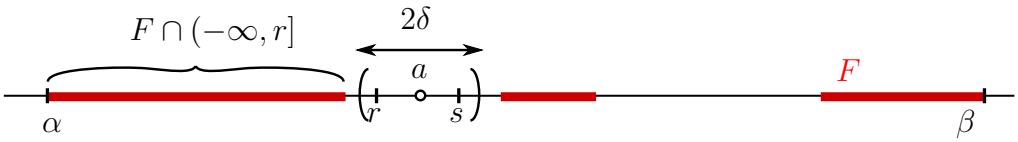


Figure 5: Illustration of finding a witness for non-isolation.

Again, $(-\infty, r] \cap F$ is non-empty. It is closed since the intersection of two closed sets is closed. There are no isolated points because F originally has no isolated points. For any point in the intersection, we can always choose the neighbourhood small enough such that it never meets $a - \delta$ and so there exists a witness for the non-isolation of the point. An illustration is given in fig. 5. The argument is the same for the other side.

Now to complete the proof, we fix an enumeration $\langle (r_n, s_n) \mid n \in \mathbb{N} \rangle$ of the countable set of all ordered pairs of $\mathbb{Q}$, and define $G_0(F) = F \cap (-\infty, r_n]$, and $G_1(F) = [s_n, \infty)$, where (r_n, s_n) is the first pair in our enumeration for which $r_n < s_n$ and both $F \cap (-\infty, r_n]$ and $F \cap [s_n, \infty)$ are perfect. ■

Lemma 6.3. *There is a function $H : P \times \mathbb{N}_1 \to P$ such that for each $F \in P$ and each $n \in \mathbb{N}_1$[8], $A = H(f, n) \subseteq F$ and $\sup(A) - \inf(A) \leq \frac{1}{n}$. In other words, each perfect set has a perfect subset of an arbitrarily small diameter.*

Proof. Let $F \in P$ and $n \in N_1$. There exists integer m such that $F \cap (\frac{m}{n}, \frac{m+1}{n}) \neq \emptyset$. If not, then $F \subseteq \{\frac{m}{b} \mid m \in \mathbb{Z}\}$, which means it is not perfect. Then take the least $m \geq 0$, or the greatest $m < 0$ with this property.

Let $a = \inf[F \cap (\frac{m}{n}, \frac{m+1}{n})]$ and $b = \sup[F \cap (\frac{m}{n}, \frac{m+1}{n})]$. Then $\frac{m}{n} \leq a \leq b \leq \frac{m+1}{n}$, therefore $b - a \leq \frac{1}{n}$. Now define $H(F, n) = F \cap [a, b]$. We can check that this is a perfect set. ■

Theorem 6.4. *Every perfect set has cardinality $2^{\aleph_0}$.*

Proof. The proof is very similar to the proof of the cardinality of the Cantor set.

Let $F \in P$. We construct a system of its perfect subsets $\langle F_s \mid f \in \{0,1\}^{\mathbb{N}} \rangle$

$$F_{s_0,\ldots,s_{n-1},0} = H(G_0(F_{s_0,\ldots,s_{n-1}}), n)$$
$$F_{s_0,\ldots,s_{n-1},1} = H(G_1(F_{s_0,\ldots,s_{n-1}}), n)$$

where H and G_0, G_1 are the functions we saw in the previous lemmas.

For any $f \in \{0,1\}^{\mathbb{N}}$ we let $F_f = \bigcap_{n \in \mathbb{N}} F_{f \restriction n}$. Then $F_f = \{d_f\}$. Thus $\langle d_f \mid f \in \{0,1\}^{\mathbb{N}} \rangle$ is a one-to-one mapping of $\{0,1\}^{\mathbb{N}}$ into F. Hence $|F| \geq 2^{\aleph_0}$. ■

We now move on to show that the closed sets also behave in accordance with the continuum hypothesis.

[8] Here $\mathbb{N}_1 = \mathbb{N} - \{0\}$.

Theorem 6.5. *Every uncountable closed set contains a perfect subset.*

Proof. For any $A \subseteq \mathbb{R}$, we call $a \in \mathbb{R}$ a *condensation point* if for every $\delta > 0$, the set $\{x \in A \mid |x - a| < \delta\}$ is uncountable. Let A_c denote the set of all condensation points of A.

The proof comes in three claims.

- A_c is closed.

  We show that every accumulation point of A_c belongs to A_c. Let a be an accumulation point of A_c and $\delta > 0$. Then there is $x \in A_c$ such that $|x - a| < \delta$. Since $x \in A_c$, there are uncountably many $y \in A$ such that for any $\epsilon > 0$, $|y - x| < \epsilon$. We choose $\epsilon = \delta - |x - a|$. The ϵ-neighbourhood is entirely contained within the δ-neighbourhood. This means that there are now uncountably many $y \in A$ contained in the δ-neighbourhood.

- If F is a closed set, $C = F \setminus F_c$ is at most countable. Note that if F is closed, then $F_c \subseteq F$ since a closed set contains all of its accumulation points.

  If $a \in C$, then a is not a condensation point of F, so there is a $\delta > 0$ such that $F \cap (a - \delta, a + \delta)$ is at most countable. Since $\mathbb{Q}$ is dense in $\mathbb{R}$, there is an open interval (r, s) with rational endpoints such that $a \in F \cap (r, s)$, and $F \cap (r, s)$ is at most countable. Therefore, $C \subseteq \bigcup\{F \cap (r, s) \mid r, s \in \mathbb{Q} \wedge F \cap (r, s) \text{ is countable}\}$. The countable union of countable sets is at most countable, so C is also at most countable[9].

- If F is an uncountable closed set, then F_c is perfect.

  By the two claims above, F_c is closed and non-empty. Now we have to show that F_c has no isolated points. Assume $a \in F_c$ is an isolated point of F_c. Then $\exists \delta > 0, \forall x \neq a\ (|x - a| < \delta \implies x \in F \setminus F_c)$. By the second claim, $F \setminus F_c$ is at most countable. Therefore there are at most countably many $x \in F$ for which $|x - a| < \delta$, contradicting the assumptions that $a \in F_c$.

The last claim proves the main theorem. ■

Corollary 6.5.1. *Every closed set is either at most countable or of size* $2^{\aleph_0}$.

There is an alternative proof without using the axiom of choice.

Definition 6.2 (Derivative sets)**.** Let $A \subseteq \mathbb{R}$. The *derivative* of A, written A', is the set of all accumulation points of A. □

Lemma 6.6. *Every closed set has at most countably many isolated points.*

Proof. Let $F \neq \emptyset$ be a closed set. For every isolated point a, by definition there exists an interval (r, s) with rational endpoints such that a is the only element of F in (r, s). Now the function mapping (r, s) to a witnesses the countability of F, since the rationals are countable. ■

Theorem 6.7. *Every uncountable closed set contains a perfect subset.*

[9] The last statement requires the Axiom of Choice.

Proof. Let $F \neq \emptyset$ be a closed set. If $F = F'$, then F is perfect and $|F| = 2^{\aleph_0}$. Otherwise, $F = (F - F') \cup F'$ and $|F - F'| \leq \aleph_0$ by lemma 6.6.

In general we can define an infinite sequence $\langle F_n \mid n \in \mathbb{N} \rangle$ through recursion:

$$F_0 = F \qquad F_{n+1} = F'_n$$

Then all F_n are closed sets. If for some $n \in \mathbb{N}$, $F_n = \emptyset$ then $|F| \leq \aleph_0$. If F_n is perfect, then $|F| = 2^{\aleph_0}$. In the case that all F_n's have isolated points, we can define

$$F_\omega = \bigcap_{n \in \mathbb{N}} F_n$$

which is smaller than all F_n. If F_ω is perfect, then $|F| \geq |F_\omega| = 2^{\aleph_0}$. If $F_\omega = \emptyset$, then $F = \bigcup_{n \in \mathbb{N}} (F_n - F_{n+1})$ which is a countable union of countable sets, and so is countable.

Now another issue is that F_ω may also have isolated points. So we have to consider $F_{\omega+1}$, $F_{\omega+2}$, and perhaps even up to $F_{\omega+\omega}$. Since the sets are getting smaller, hopefully eventually we either reach a perfet set of $\emptyset$. This was the original motivation that led Cantor to his theory of transfinite or ordinal numbers. We will finish the proof after studying ordinal numbers. The rest of the proof is at section 8. ■

Example 6.1. Consider

$$F = \{m - \frac{1}{n} \mid m, n \in \mathbb{N}_1\}$$

What is F'? We can see that $1, 2, \ldots$ are the limit points. So $F' = \{1, 2, \ldots\}$. However, notice that $F'' = \emptyset$.

We can consider another set

$$G = \{m - \frac{1}{n} - \frac{1}{k} \mid m, n, k \in \mathbb{N}_1\}$$

Now $G' = F$, and $G'' = F'$. Then $G''' = \emptyset$. ◊

7 Ordinals

Recall that a set A is well-ordered by $<$ if A is linearly ordered by $<$ and any non-empty subset X of A has a $<$-least element.

Definition 7.1 (Initial segments)**.** Let $(L, <)$ be a linearly ordered set. A set $S \subsetneq L$ is called an *initial segment* if it is closed downwards, i.e. $\forall a \in S, x < a \implies x \in S$. □

Lemma 7.1. *If $(W, <)$ is a well-ordered set and if S is an initial segment of $(W, <)$, then there exists $a \in W$ such that $S = \{x \in W \mid x < a\}$. In other words, every initial segment of a well-order is determined by a single element.*

Proof. Consider the set $X = W \setminus S$. Since S is a proper subset, X is non-empty. X has a least-element a. We claim that $S = \{x \mid x < a\}$.

If there is some x where $x < a$ then $x \in S$ since a is the least element in X. For all elements $y \in S$, then $y < a$, since otherwise $a \in S$. ■

Therefore we can call the set $W[a] = \{x \in W \mid x < a\}$ the initial segment given by a.

Lemma 7.2. *If $(W, <)$ is a well-ordered set and if $f : W \to W$ is an increasing (i.e. order preserving) function, then $f(x) \geq x$ for all $x \in W$.*

Proof. We will show that $X = \{x \mid f(x) < x\}$ is empty.

Suppose not. Then let a be the least element in X. Thus $f(a) < a$, which implies $f(f(a)) < f(a)$, contradicting the fact that a is the least element in X. ■

Theorem 7.3.

i. No well-ordered set is isomorphic to an initial segment of itself.

ii. Each well-ordered set has only one automorphism[10], the identity.

iii. If W_1 and W_2 are isomorphic well-ordered sets, then the isomorphism between W_1 and W_2 is unique.

Proof. These follow from lemma 7.2. We give a brief proof.

i. Such an isomorphism f will require $x < f(x)$ at some point.

ii. Shuffling the elements with an isomorphism will cause either the function, or the inverse of the function, to go against the lemma.

iii. Suppose there are two isomorphisms $f_1, f_2 : W_1 \to W_2$, then $f_1 \circ f_2^{-1}$ is the identity.

■

Theorem 7.4. *If W_1 and W_2 are well-ordered sets, then one and exactly one of the following holds:*

i. W_1 and W_2 are isomorphic, or

ii. W_2 is isomorphic to an initial segment of W_1, or

iii. W_1 is isomorphic to an initial segment of W_2.

In all cases the isomorphism is unique.

Proof. Theorem 7.3 shows that the three cases are mutually exclusive. Furthermore it also shows that the isomorphism in each case is unique. It remains to show that one of the three cases always holds. The idea is to match the least elements of the two sets, then the second least elements, and so on.

Let

$$f = \{(x, y) \in W_1 \times W_2 \mid W_1[x] \text{ is isomorphic to } W_2[y]\}$$

Firstly, f is a function. If $(x, u_1) \in f$ and $(x, u_2) \in f$, then $W_1[x] \cong W_2[y_1]$ and $W_2[y_2]$, so $W_2[y_1]$ is isomorphic to $W_2[y_2]$, so $y_1 = y_2$ by theorem 7.3.

[10] An automorphism is an isomorphism to itself.

Next, f is also injective. Suppose $f(x_1) = f(x_2) = y$. Then $W_1[x_1] \cong W_2[y] \cong W_1[x_2]$ which means $x_1 = x_2$ again by theorem 7.3. Therefore f is an isomorphism between its domain and its range.

Furthermore, f is order preserving, i.e. $x < x' \implies f(x) < f(x')$. Let h is the isomorphism between $W_1[x']$ and $W_2[f(x')]$. $h \restriction W_1[x]$ is an isomorphism between $W_1[x]$ and $W_2[h(x)]$, so $f(x) = h(x) < f(x')$. This also means that $f(x)$ is defined.

Now, for each of the respective cases:

i. If $\mathrm{dom}(f) = W_1$ and $\mathrm{ran}(f) = W_2$ then the statement holds.

ii. If $\mathrm{dom}(f) \neq W_1$, then $\mathrm{dom}(f)$ is an initial segment of W_1 by the argument above. Therefore we can write $\mathrm{dom}(f) = W_1[a]$ for some $a \in W_1$. We claim that $\mathrm{ran}(f) = W_2$. Suppose this does not hold. Then $\mathrm{ran}(f)$ is an initial segment of W_2 (the argument for this is similar to the argument for the domain). We can write $\mathrm{ran}(f) = W_2[b]$ for some $b \in W_2$. Then f is an isomorphism between $W_1[a]$ and $W_2[b]$, which contradicts the fact that $a \notin \mathrm{dom}(f)$.

iii. If $\mathrm{dom}(f) = W_1$, but $\mathrm{ran}(f) \neq W_2$, this is similar to the second case above.

■

This serves as a motivation for the creation of ordinals. The order isomorphism acts like an equivalence relation on well-ordered sets, and we may want to consider the equivalence class induced by this relation[11]. However such a class is too big to be a set. What we can do is to instead find a member from each class, and these will represent all possible isomorphic well-orders. In a certain sense, ordinals are a generalization of natural numbers.

Definition 7.2 (Transitive sets)**.** A set T is *transitive* if every element of T is also a subset of T. In other words, $u \in v \in T \implies u \in T$. □

Example 7.1. The natural numbers are transitive. For example, take $m \in n = \{0, \ldots, n-1\} \in \mathbb{N}$, then $m \in \mathbb{N}$. ◊

Theorem 7.5. *A set X is transitive iff $X \subset \mathcal{P}(X)$.*

Proof. If X is transitive, then it contains subsets of itself. Hence $X \subset \mathcal{P}(X)$. Equality cannot hold since $\mathcal{P}(X) \ni X \notin X$.

On the other hand, if $X \subset \mathcal{P}(X)$, then an element of X is also a subset of X. Hence it is transitive. ■

Theorem 7.6.

i. If X is a transitive set then $\bigcup X$ is also a transitive set.

ii. Let S be a system of transitive sets. $\bigcup S$ is transitive.

Proof. The two proofs are very similar and just involve element chasing.

[11] The purpose for doing this is similar to Russell's definition of a natural number n as the class of all sets with n items, which is also an equivalence class.

i. By definition $x \in \bigcup X$ iff there is some $a \in A$ such that $x \in a$. However, since A is transitive, $x \in A$, which means that for any $y \in x$, $y \in \bigcup X$.

ii. By definition $u \in \bigcup S$ iff there is some $s \in S$ such that $u \in s$. However, since S is transitive, $u \in A$, which means that for any $v \in u$, $v \in \bigcup S$.

■

Definition 7.3 (Ordinals)**.** A set α is an *ordinal* or *ordinal number* if α if transitive and α is well-ordered by $\in_\alpha$, which is the membership relation restricted to α. □

Every natural number is an ordinal. We have already shown that it is transitive, and $<_{\mathbb{N}}$ is defined using $\in_{\mathbb{N}}$. The set of all natural numbers $\mathbb{N}$ is also an ordinal, and we denote it by ω.

Theorem 7.7 (Successor ordinals)**.** *If α is an ordinal, then $S(\alpha) = \alpha \cup \{\alpha\}$ is also an ordinal. $S(\alpha)$ is called the* successor *of α, and we can also write it as $\alpha + 1$.*

Proof. Take any $x \in S(\alpha)$. Either $x \in \alpha$ or $x \in \{\alpha\}$. If $x \in \alpha$ then for any $y \in x$, $\alpha \ni y \in S(\alpha)$. If $x \in \{\alpha\}$, then $x = \alpha$ which means for any $y \in x$, $\alpha \ni y \in S(\alpha)$. Thus $S(\alpha)$ is transitive.

It is easy to see that $\in_{S(\alpha)}$ is a total ordering. What we need to show is that it is also a well-ordering. Take a non-empty subset A of $S(\alpha)$. If $A = \{\alpha\}$ then α is the least element. Otherwise, since $\in_\alpha$ is a well-order, the least element of A is the $\in_\alpha$-least element of $A \setminus \{\alpha\}$. ■

Definition 7.4 (Successor ordinals)**.** An ordinal α is called a *successor ordinal* if $\alpha = \beta + 1$ for some ordinal β. □

Definition 7.5 (Limit ordinals)**.** If an ordinal is not a successor ordinal then it is called a *limit ordinal*. □

Example 7.2. 2020 is a successor ordinal, and ω is a limit ordinal. ◊

Definition 7.6. For all ordinals, we define $\alpha < \beta$ iff $\alpha \in \beta$. □

7.1 Order properties of ordinals

We want to show that $\in$ is in fact a well-order over the ordinals.

Lemma 7.8. *If α is an ordinal, then $\alpha \notin \alpha$.*

Proof. Otherwise $\alpha \in \alpha$. So the strict linear ordering $(\alpha, \in_\alpha)$ has an element x such that $x \in x$. This contradicts the asymmetry of $\in_\alpha$. ■

Lemma 7.9. *Every element of an ordinal is an ordinal.*

Proof. Let there be an ordinal α and take $x \in \alpha$. If $u \in v \in x \in \alpha$, then $v \in \alpha$ since α is transitive. Then this will mean $u \in \alpha$. Since $\in_\alpha$ is a linear ordering, $u \in x$. So x is transitive.

Since α is transitive, $x \in \alpha$ implies $x \subseteq \alpha$. If $y \subseteq x$, then $y \subseteq \alpha$ as well. From here it can be seen that $\in_x = \in_\alpha \restriction x$ will be a well-ordering. ■

Lemma 7.10. *If α and β are ordinals such that $\alpha \subset \beta$, then $\alpha \in \beta$.*

Proof. Suppose $\alpha \subset \beta$. Then $\beta \setminus \alpha \neq \emptyset$, and so it contains a least element γ. We claim that $\gamma = \alpha$.

$\gamma \subseteq \alpha$ because otherwise $\delta \in (\gamma \setminus \alpha) \in (\beta \setminus \alpha)$ will mean that γ is not the least element.

For any $\delta \in \alpha$, $\delta \in \beta$ since $\alpha \subset \beta$. So either $\delta \in \gamma$ or $\delta = \gamma$ or $\gamma \in \delta$. The last two cases imply that $\gamma \in \alpha$. This contradicts the choice of γ. Therefore, $\delta \in \gamma$. So $\alpha \subseteq \gamma$. ■

Corollary 7.10.1. *Let α and β be two ordinals. $\alpha \subset \beta \iff \alpha \in \beta$.*

Proof. This follows from the previous lemma as well as the fact that if $\alpha \in \beta$ then $\alpha \subseteq \beta$ as β is transitive. ■

Theorem 7.11. *Let α, β, γ be ordinal numbers.*

i. If $\alpha < \beta$ and $\beta < \gamma$, then $\alpha < \gamma$.

ii. $\alpha < \beta$ and $\beta < \alpha$ cannot both be true.

iii. $\alpha < \beta$ or $\alpha = \beta$ or $\beta < \alpha$.

iv. Every set of ordinals is well-ordered by $<$.

v. For every set of ordinal numbers X, there is an ordinal number $\alpha \notin X$.

Proof.

i. This follows from the transitivity of γ.

ii. If both are true, then it contradicts lemma 7.8.

iii. Consider $\delta = \alpha \cap \beta$. δ is an ordinal. It is easy to show that it is transitive and well-ordered by $\in_\alpha \restriction \delta$.

So now $\delta \subseteq \alpha$ and $\delta \subseteq \beta$. If $\delta \subset \alpha$ and $\delta \subset \beta$, then $\delta \in \alpha$ and $\delta \in \beta$ by lemma 7.10, which means $\delta \in \delta$, which contradicts lemma 7.8. If $\delta = \alpha$ then $\alpha \subseteq \beta$ and lemma 7.10 says that either $\alpha = \beta$ or $\alpha \in \beta$. Similarly so for the case $\delta = \beta$.

iv. Let $X \neq \emptyset$ be a set of ordinals. Choose $\alpha \in X$. If $\alpha \cap X = \emptyset$, then $\forall y \in X,\ (y \notin \alpha)$. By (iii) this means either $y = \alpha$ or $\alpha < y$, i.e. α is the least element. If instead $\alpha \cap X \neq \emptyset$, then there exists a least element in $\alpha \cap X$ by the ordering $\in_\alpha$.

v. Consider $\beta = \bigcup X$. β is an ordinal. Since all its elements are transitive, with some extra steps we can show that β itself is transitive. Furthermore (iv) shows that β is well-ordered by ϵ.

Let $\alpha = S(\beta)$, then α is an ordinal but $\alpha \notin X$. If $\alpha \in X$ this would mean that $\alpha \subseteq \beta$. By lemma 7.10 this would mean either $\alpha = \beta$ or $\alpha \in \beta$. In both cases it would mean $\alpha \in S(\beta) = \alpha$, contradicting lemma 7.8.

■

The last item also shows that we cannot have a set of all ordinals. Another way to show this is through the Burali-Forti paradox. If there was a set of ordinals, call it O, points (i) — (iv) from theorem 7.11 tells us that O itself is transitive and well-ordered by ϵ. Then $O \in O$, which is a contradiction.

Lemma 7.12. *Let S be a set of ordinals. Then $\bigcup S$ is an ordinal.*

Proof. First of all theorem 7.6 tells us that $\bigcup S$ is transitive. Next, $\bigcup S$ collects elements of S, which are ordinals by lemma 7.9 and so by theorem 7.11 $\bigcup S$ is well-ordered by $\in$. ■

Theorem 7.13 (Supremum)**.** *Let S be a set of ordinals. Then $\bigcup S$ is the least ordinal greater than or equal to all elements in S (in other words, a least upper bound). We call it the* supremum *of S, and denote it by* $\sup S$.

Proof. For any $s \in S$, for all $u \in s$, $u \in \bigcup S$ so $s \subseteq \bigcup S$. Then corollary 7.10.1 tells us that $s \leq \bigcup S$.

Next, consider any ordinal X that is an upper bound of S. For any $u \in \bigcup S$, there is some $s \in S$ such that $u \in s$. However $s \leq X$, which means that $y < X$, i.e. $y \in X$. Therefore $\bigcup S \subseteq X$. ■

Theorem 7.14. *Let S be a set of ordinals. If S does not have a greatest element, then* $\sup S$ *is a limit ordinal.*

Proof. If $\sup S$ was a successor, suppose $\sup S = \alpha + 1$, then α will be the greatest element in $\sup S$. So we proceed by showing $\sup S$ has no greatest element.

Suppose there is a greatest element $u \in \bigcup S$. Then there exists a $s \in S$ such that $u \in s$. Since S has no greatest element, there exists some $x \in S$ such that $s < x$, or equivalently $s \in x$. Therefore $s \in \bigcup S$. However, $u < s$, contradiction. ■

7.2 Order types of well-orderings

Our next goal is to link ordinals with well-orderings and show that ordinals act as some kind of representative for well-orderings.

Definition 7.7 (Order types)**.** Let W_1 and W_2 be well-ordered sets. If $W_1 \cong W_2[a]$ for some $a \in W_2$, then we say that W_1 has *smaller order type* than W_2. □

Theorem 7.15 (Induction principle for well-ordered sets)**.** *Let P be a property. If for every well-ordered set W, every well-ordered set with smaller order type than W has property P implies W has property P, then every well-ordered set has property P.*

Proof. Suppose some well-ordered set W fails to have property P. By the assumption, there is some well-ordering of a smaller order type that also fails to satisfy P. Let $a \in W$ by the least element of W such that some well-ordered set $W' \cong W[a]$ does not satisfy P. Then every well-ordering of smaller order type than W' has property P by the choice of a. Then every well-ordering of smaller order type than W' will have property P, and by the assumption W' will have property P, which is a contradiction. ■

To show that ordinals represent well-orderings, surprisingly we will need a new axiom. Since we are not going in depth into axiomatic set theory, we will just briefly talk about this. Let ω be the set of natural numbers, which exists from the axiom of infinity. From ω we can get sets $\omega + 1$, $\omega + 2$, etc. from the axiom of union and the axiom of pairing. Intuitively we can come up with $\omega + \omega$ as the union of all $\omega + n$ for all $n \in \omega$. However the existence of $\omega + \omega$ cannot be proved from the axioms that we have so far.

The axiom schema of replacement Let $P(x, y)$ be a property such that $\forall x, \exists! y\ P(x, y)$. Then for every set A, there is a set B, such that for every $x \in A$, there is a $y \in B$ for which $P(x, y)$ holds.

If we think of $P(x, y)$ as "x nominates y", then if each member of A nominates at most one object, the collection of all nominees is a set. The name "replacement" comes from the idea of replacing each $x \in A$ by its nominees to obtain a new set B.

In Zermelo's original axioms for set theory this axiom was not included. It was only included later on by Fraenkel.

Theorem 7.16. *Every well-ordered set is isomorphic to an unique ordinal.*

Proof. Apply induction on well-ordered sets to the following property P: there is an ordinal α such that $(W, <)$ is isomorphic to $(\alpha, \in)$. The uniqueness comes from the fact that two distinct ordinals cannot be isomorphic.

Let $(W, <)$ be a well-ordered set such that every well-ordered set with smaller order type than W has property P. We will then show that W has property P.

For each $a \in W$, $W[a]$ is isomorphic to some ordinal α_a. By the axiom of replacement, $S = \{\alpha_a \mid a \in W\}$ is a set. We claim that S is a ordinal.

S is well-ordered by $\in$ by theorem 7.11 since it is a set of ordinals. S is transitive because if $\gamma \in \alpha_a \in S$, let f be the isomorphism between $W[a]$ and α_a, and let $c = f^{-1}(\gamma)$. Then $f \restriction c$ is an isomorphism between $W[c]$ and γ. This can be easily checked, but intuitively this is because the initial segment $W[c]$ is "covered by" the initial segment $W[a]$. In the end this means that $\gamma \in S$.

Furthermore, $g : W \to S$ defined by $g(a) = \alpha_a$ is an isomorphism. From the definition of W it is quite clear that g is a bijection. Furthermore, $a < b \implies \alpha_a < \alpha_b$. If we take f to be the isomorphism between $W[b]$ and α_b, then $f \restriction a$ is an isomorphism between $W[a]$ and some initial segment of α_b. Then it follows that $\alpha_a < \alpha_b$. So g preserves the order and therefore is an isomorphism. Therefore $(W, <)$ has property P. ∎

7.3 Transfinite induction and recursion

Theorem 7.17 (Transfinite induction)**.** *Let $P(x)$ be a property. Assume that for all ordinals α, $P(\beta)$ is true for all $\beta < \alpha$ then $P(\alpha)$ holds. Then $P(\alpha)$ holds for all ordinals α.*

Proof. Look at the smallest counterexample, call it α'. Every for all $\beta < \alpha'$, $P(\beta)$ holds since α' is the least. Then by the assumption $P(\alpha')$ holds. ∎

It is often more convenient to use the following equivalence version.

Theorem 7.18 (Transfinite induction)**.** *Let $P(x)$ be a property. Assume that*

- *$P(0)$ holds.*
- *$\forall \alpha\ [P(\alpha) \implies P(\alpha + 1)]$.*

- *For all limit ordinals $\alpha \neq 0$, if $P(\beta)$ holds for all $\beta < \alpha$, then $P(\alpha)$ holds.*

Then $P(\alpha)$ holds for all ordinals α.

Proof. It suffices to show that the assumptions made here are equivalent to the assumption in the original version of the transfinite inductive principle above.

Let α be an ordinal such that $P(\beta)$ holds for all $\beta < \alpha$. If $\alpha = 0$, then $P(\alpha)$ holds. If α is a successor, then $P(\alpha)$ holds. If $\alpha \neq 0$ is a limit ordinal, then $P(\alpha)$ again holds. ■

We now generalize the recursion theorem of natural numbers to ordinals.

Theorem 7.19 (Transfinite recursion)**.** *Let G be an operation*[12] *defined on the class of all sets. Then there is a unique operation F defined on the class of all ordinals such that for all ordinals α:*

$$F(\alpha) = G(F \restriction \alpha).$$

Proof. Similar as the case for $\mathbb{N}$. ■

Theorem 7.20 (Parametric transfinite recursion)**.** *Let G_1, G_2, and G_3 be operations defined on the class of all sets. Then there is an unique operation $F(z, \alpha)$ defined on the class of all sets z and all ordinals α such that*

- $F(z, 0) = G_1(z)$,
- $F(z, \alpha + 1) = G_2(z, \alpha, F(z, \alpha))$,
- $F(z, \alpha) = G_3(z, \{(\beta, F(z, \beta)) \mid \beta < \alpha\})$ *if $\alpha \neq 0$ is a limit ordinal.*

Proof. Similar as the case for $\mathbb{N}$. ■

Definition 7.8 (Addition of ordinals)**.** For all ordinals α:

- $\alpha + 0 = \beta$,
- $\alpha + (\alpha + 1) = (\alpha + \beta) + 1$,
- $\alpha + \beta = \sup\{\alpha + \gamma \mid \gamma < \beta\}$ for all limit ordinals $\alpha \neq 0$

□

Some identities from the natural numbers might hold, but not all. We can give an example to show that right cancellation does not hold. Let m be a natural number. Then $m + \omega = \sup\{m + n \mid n < \omega\} = \omega$. Hence $m + \omega \neq \omega + m$. Another example, $1 \neq 2$ but $1 + \omega = \omega = 2 + \omega$. However, left cancellation holds.

Theorem 7.21 (Left cancellation)**.** *If α_1, α_2, β are ordinals,*

i. $\alpha_1 < \alpha_2 \iff \beta + \alpha_1 < \beta + \alpha_2$.

ii. $\beta + \alpha_1 = \beta + \alpha_2 \iff \alpha_1 = \alpha_2$.

[12] Here an operation is a formula that gives a unique output. We do not have to dwell too much on this.

Proof. (ii) follows easily from (i) by considering cases $\alpha_1 < \alpha_2$, $\alpha_2 < \alpha_1$ which contradict the LHS. So we only show (i).

($\implies$): We will perform transfinite induction on α_2. Let $P(x)$ be the property that either $x \leq \alpha_1$ or $\beta + \alpha_1 < \beta + x$. That this is the logically equivalent to the claim. We will show that for all ordinals α_2, $P(\alpha_2)$ holds.

Firstly $P(0)$ holds since $0 \leq \alpha_1$.

Assume that $P(\alpha_2)$ holds for some ordinal α_2. If $\alpha_2 < \alpha_1$, then $\alpha_2 + 1 \leq \alpha_1$[13] so Then $P(\alpha_2 + 1)$ holds. If $\alpha_2 = \alpha_1$, then $\beta + \alpha_1 = \beta + \alpha_2 < \beta + \alpha_2 + 1$, so $P(\alpha_2 + 1)$ holds. Finally, if $\alpha_1 < \alpha_2$, then by the inductive hypothesis, $\beta + \alpha_1 < \beta + \alpha_2 < \beta + \alpha_2 + 1$ so again $P(\alpha_2 + 1)$ holds.

Now assume $\alpha_2 \neq 0$ is a limit ordinal. Without loss of generality, assume $\alpha_1 < \alpha_2$. Then $\alpha_1 + 1 < \alpha_2$. This is because otherwise $\alpha_1 + 1 \leq \alpha_2$ but $\alpha_2 \neq \alpha_1 + 1$. Therefore

$$\begin{aligned} \beta + \alpha_1 &< \beta + \alpha_1 + 1 \\ &\leq \sup\{\beta + \gamma \mid \gamma < \alpha_2\} \\ &= \beta + \alpha_2 \end{aligned}$$

so $P(\alpha_2)$ holds.

($\impliedby$): Assume that $\beta + \alpha_1 < \beta + \alpha_2$. Now either $\alpha_1 < \alpha_2$ or $\alpha_1 = \alpha_2$ or $\alpha_1 > \alpha_2$. The last two cases will imply that $\beta + \alpha_1 \geq \beta + \alpha_2$. So only the first case can occur. ∎

Theorem 7.22 (Associative law for addition)**.** *For all ordinals α, β, γ, $(\alpha+\beta)+\gamma = \alpha+(\beta+\gamma)$.*

Proof. Perform transfinite induction on γ. Let $P(x)$ be the property that $(\alpha+\beta)+x = \alpha+(\beta+x)$.

$P(0)$ holds quite obviously. For ordinals the proof is identical to the case for natural numbers, theorem 3.14.

So assume that γ is a limit ordinal and for all $\delta < \gamma$, $P(\delta)$ holds. Then

$$\begin{aligned} (\alpha + \beta) + \gamma &= \sup\{(\alpha + \beta) + \delta \mid \delta < \gamma\} && \text{(Definition of addition)} \\ &= \sup\{\alpha + (\beta + \delta) \mid \delta < \gamma\} && \text{(Inductive hypothesis)} \\ &\leq \alpha + (\beta + \gamma) \end{aligned}$$

The last step is justified by theorem 7.21. $\delta < \gamma \implies \beta + \delta < \beta + \gamma$. Applied once more, we get $\alpha + (\beta + \delta) < \alpha + (\beta + \gamma)$.

Now we want to show the other direction. First we show that $\beta + \gamma$ is a limit ordinal. $\xi < \beta + \gamma$ implies $\xi \leq \beta + \delta$ for some $\delta < \gamma$. Therefore $\xi + 1 \leq (\beta + \delta) + 1 = \beta + (\delta + 1) < \beta + \gamma$ by the inductive hypothesis.

So now $\alpha + (\beta + \gamma) = \sup\{\alpha + \xi \mid \xi < \beta + \gamma\}$. Again $\xi < \beta + \gamma$ implies $\xi \leq \beta + \delta$ for some $\delta < \gamma$. Then

$$\begin{aligned} \alpha + \xi &\leq \alpha + (\beta + \delta) && \text{(Theorem 7.21)} \\ &= (\alpha + \beta) + \delta && \text{(Inductive hypothesis)} \\ &< (\alpha + \beta) + \gamma. && \text{(Theorem 7.21)} \end{aligned}$$

Therefore $\sup\{\alpha + \xi \mid \xi < \beta + \gamma\} \leq (\alpha + \beta) + \gamma$. ∎

[13] This needs a small argument.

Theorem 7.23. *Let α and β be ordinals and $\alpha \leq \beta$. Then the equation $\alpha + \xi = \beta$ has a unique solution.*

Proof. Perform transfinite induction. Let $P(\beta)$ be the statement that for all $\alpha \leq \beta$, there is an unique solution to the equation $\alpha + \xi = \beta$. First we show existence.

$P(0)$ is clearly true. Suppose $P(\beta_0)$ holds for some ordinal β_0. Consider $\beta_0 + 1$. Now $\alpha \leq \beta_0 + 1$ implies either $\alpha = \beta_0 + 1$ or $\alpha \leq \beta_0$. For the first case, $\xi = 0$, and for the second case, by the inductive hypothesis there is an unique ξ such that $\alpha + \xi = \beta_0$, so $\alpha + (\xi + 1) = \beta_0 + 1$.

Suppose $P(\beta)$ holds for all $\beta < \lambda$ where λ is a non-zero limit ordinal. We want to show that $P(\lambda)$ holds. Take any $\alpha \leq \lambda$. If $\alpha = \lambda$ then $\xi = 0$. If $\alpha < \lambda$, then for any $\beta < \lambda$ such that $\alpha < \beta$, there is some ξ_β such that $\alpha + \xi_\beta = \beta$. We claim that the solution we are seeking here is given by $\xi = \sup\{\xi_\beta \mid \beta < \lambda\}$.

First, we check that $\xi = \sup\{\xi_\beta \mid \beta < \lambda\}$ is a limit ordinal. Suppose $\nu < \xi$. Then for some β with $\alpha < \beta < \lambda$, $\nu < \xi_\beta$. Then $\nu + 1 < \xi_\beta + 1 = \xi_{\beta+1} < \xi$.

Next, we check that ξ is indeed a solution.

$$\begin{aligned} \alpha + \xi &= \sup\{\alpha + \xi_\beta \mid \beta < \lambda\} \\ &= \sup\{\beta \mid \beta < \lambda\} \\ &= \lambda. \end{aligned}$$

Uniqueness follows directly from left cancellation. ■

Definition 7.9 (Ordinal multiplication)**.** For all ordinals α,

- $\alpha \cdot 0 = 0$.
- $\alpha \cdot (\beta + 1) = (\alpha \cdot \beta) + \alpha$.
- $\alpha \cdot \beta = \sup\{\alpha \cdot \gamma \mid \gamma < \alpha\}$ for all limit ordinals $\beta \neq 0$.

□

Similar to addition, multiplication is not commutative and there is no right cancellation.

Theorem 7.24 (Left distributive law for multiplication)**.** *For all ordinals α, β, γ, $\alpha \cdot (\beta + \gamma) = \alpha \cdot \beta + \alpha \cdot \gamma$.*

Theorem 7.25 (Associative law for multiplication)**.** *For all ordinals α, β, γ, $(\alpha \cdot \beta) \cdot \gamma = \alpha \cdot (\beta \cdot \gamma)$.*

Example 7.3. $2 \cdot \omega = \sup\{2 \cdot n \mid n < \omega\} = \omega \neq \omega \cdot 2 = \omega + \omega$. ◊

Definition 7.10 (Ordinal exponentiation)**.** For all ordinals α,

- $\alpha^0 = 1$.
- $\alpha^{\beta+1} = \alpha^\beta \cdot \alpha$.
- $\alpha^\beta = \sup\{\alpha^\gamma \mid \gamma < \beta\}$ for all limit ordinals $\beta \neq 0$.

□

Ordinal exponentiation is very different from cardinal exponentiation.

Example 7.4. $2^{\omega} = \sup\{2^n \mid n < \omega\} = \omega$. However we have previously shown that $2^{\aleph_0} > \aleph_0$. ◊

Example 7.5. We will evaluate a few statements.

- $(\omega + 1) + \omega$. By associativity, this is equivalent to $\omega + (1 + \omega)$. Keeping in mind ω is a limit ordinal, we have

$$\begin{aligned} 1 + \omega &= \sup\{1 + n \mid n \in \omega\} \\ &= \omega \end{aligned}$$

  Thus the sum evaluates to $\omega + \omega$.

- $\omega + \omega^2$.

$$\begin{aligned} \omega + \omega^2 &= \sup\{\omega + \omega \cdot n \mid n < \omega\} \\ &= \sup\{\omega \cdot (n + 1) \mid n < \omega\} \\ &= \omega^2 \end{aligned}$$

◊

Definition 7.11. Functions whose domain is an ordinal α are called *transfinite sequences* of length α. □

8 Alephs

We have talked about cardinalities but have not really defined what is meant by $|X|$. However intuitively we believe that every set has a size or cardinality. So what we wish to accomplish is to assign an ordinal α to X and defined $|X| = \alpha$. To do this we will need to introduce the axiom of choice.

Definition 8.1 (Initial ordinal)**.** An ordinal α is called an *initial ordinal* if it is not equipotent (equinumerous) to any $\beta < \alpha$. □

Example 8.1. ω is an initial ordinal whereas $\omega + 1$ is not. We cannot find a bijection from a natural number to ω. However we can find a bijection from $\omega + 1$ to ω by shifting everything by 1. ◊

Theorem 8.1 (Cardinal numbers)**.** *Each well-ordered set X is equipotent to an unique initial ordinal α_0. We will call α_0 the* cardinal number *of X.*

Proof. Since X is well-ordered, X is equipotent to some ordinal α, by theorem 7.16. Let α_0 by the least such α, then α_0 is an initial ordinal, since if $\beta < \alpha$ then $|\beta| \neq |X|$ (α_0 is the least).

If $|X| = |\alpha_1|$, then $\alpha_0 \leq \alpha_1$ by the choice of α_0. If $\alpha_0 < \alpha_1$ then α_1 is not initial. Therefore $\alpha_0 = \alpha_1$. ■

Are there always larger and larger initial ordinals? Perhaps using Cantor's theorem we may conclude yes, since given some $|X| = \alpha$ there is always set $\mathcal{P}(X)$ with a larger cardinality. However the problem here is that we are not sure that $\mathcal{P}(X)$ can be well-ordered.

Definition 8.2 (Hartogs number)**.** For any set A, let $h(A)$ be the least ordinal which is not equipotent to any subset of A. We call $h(A)$ the *Hartogs number* of A. In other words, $h(A)$ is the least ordinal such that $|h(A)| \not\leq |A|$. □

In other words, even though we might not be able to well-order A, some subsets of it might be well-ordered. Then $h(A)$ forms the upper bound.

Lemma 8.2. *For any set A,*

i. $h(A)$ is an initial ordinal.

ii. $h(A)$ exists.

Proof.

i. If $h(A)$ is not an initial ordinal, then there is some ordinal $\alpha < h(A)$ equipotent to $h(A)$. Then α is equipotent to a subset of A, which means $h(A)$ is also equipotent to a subset of A, which is a contradiction.

ii. For every well-ordered set $W \subseteq A$, there is an unique ordinal α that is isomorphic to W by theorem 7.16. The axiom schema of replacement (section 7.2) implies that there exists a set H such that for every well-ordering $<\in \mathcal{P}(A \times A)$, the ordinal α isomorphic to $<$ is in H.

   Next, we claim that H is an ordinal. Since H is a set of ordinals, H is well-ordered by $\in$. Now take any $\beta \in \alpha \in H$. α is isomorphic to some well-ordered set $W \subseteq A$. Let $f : \alpha \to W$ be the isomorphism. Then $f \restriction \beta$ is an isomorphism between β and some $C \subseteq W \subseteq A$. Hence $\beta \in H$ and H is transitive.

   Lastly we show that H is not equipotent to any subset of A. Suppose on the contrary that H is equipotent to some subset $D \subseteq A$. Let $g : D \to H$ be a bijection. Then we can define a well-ordering $<_D$ by $d <_D d' \iff g(d) \in g(d')$. Then D would be isomorphic to an unique ordinal, which is H. However $D \in H$ which would mean $H \in H$, which is a contradiction.

■

Now we know that there is an endless supply of larger and larger initial ordinals, we can define a few of them using transfinite induction.

Definition 8.3. Define

- $\omega_0 = \omega$.
- $\omega_{\alpha+1} = h(\omega_\alpha)$.
- $\omega_\alpha = \sup\{\omega_\beta \mid \beta < \alpha\}$. If $\alpha \neq 0$ is a limit ordinal.

□

A few easy facts are $|\omega_\alpha| < |\omega_{\alpha+1}|$ and that $\alpha < \beta \implies |\omega_\alpha| < |\omega_\beta|$.

Theorem 8.3.

i. ω_α *is an infinite initial ordinal for all ordinals* α*.*

ii. $\alpha < \omega_\alpha$ *for all ordinals* α*.*

iii. If Ω *is an infinite initial ordinal, then* $\Omega = \omega_\alpha$ *for some* α*.*

Proof.

i. We can perform induction on α. For the non-limit cases it is fairly obvious. When α is an limit ordinal, suppose $|\omega_\alpha| = |\gamma|$ but $\gamma < \omega_\alpha$. There exists a $\beta < \alpha$ such that $\gamma \leq \omega_\beta$ since ω_α is the supremum. However this implies $|\omega_\alpha| = |\gamma| \leq |\omega_\beta| < |\omega_\alpha|$ which is a contradiction.

ii. We perform induction on α. The non-limit cases are fairly easy to see. If α is a limit ordinal but $\omega_\alpha < \alpha$, then $\omega_{\omega_\alpha} < \omega_\alpha$ since ω_α is the supremum. However this is contradicts the inductive hypothesis.

iii. By ii., for every infinite initial ordinal Ω, there exists an ordinal α such that $\Omega < \omega_\alpha$. So we will perform induction on α and show that for any infinite initial ordinal $\Omega < \omega_\alpha$, there exists $\gamma < \alpha$ such that $\Omega = \omega_\gamma$.

If $\alpha = 0$ then this is vacuously true. If $\alpha = \beta + 1$, then $\Omega < \omega_\alpha = h(\omega_\beta)$. By definition Ω is equipotent to some subset of ω_β, so $|\Omega| \leq |\omega_\beta|$. If $\beta < \Omega$, we would have $|\Omega| \geq |\omega_\beta|$, which would mean $|\Omega| = |\omega_\beta|$ contradicting the fact that Ω is an initial ordinal. So either $\Omega = \omega_\beta$ or $\Omega < \omega_\beta$. For the latter case the induction hypothesis tells us that there is a $\gamma \leq \beta < \alpha$ such that $\Omega = \omega_\gamma$.

For the limit case, then $\Omega < \omega_\alpha = \sup\{\omega_\beta \mid \beta < \alpha\}$. Then $\Omega < \omega_\beta$ for some $\beta < \alpha$. By the inductive hypothesis there exist some $\gamma < \beta$ such that $\Omega = \omega_\gamma$.

■

We rewrite ω_α as $\aleph_\alpha$. The cardinality of a well-orderable set is either a natural number or an aleph.

Theorem 8.4. $\aleph_\alpha \cdot \aleph_\alpha = \aleph_\alpha$ *for every* α*.*

Proof. Construct the relation $\prec$ on $\omega_\alpha \times \omega_\alpha$ as $(\alpha_1, \alpha_2) \prec (\beta_1, \beta_2)$ iff

- $\max\{\alpha_1, \alpha_2\} < \max\{\beta_1, \beta_2\}$, or
- $\max\{\alpha_1, \alpha_2\} = \max\{\beta_1, \beta_2\} \wedge \alpha_1 < \beta_1$, or
- $\max\{\alpha_1, \alpha_2\} = \max\{\beta_1, \beta_2\} \wedge \alpha_1 = \beta_1 \wedge \alpha_2 < \beta_2$.

We claim that $\prec$ is a well-ordering.

First we show that it is transitive. $(\alpha_1, \alpha_2) \prec (\beta_1, \beta_2)$ and $(\beta_1, \beta_2) \prec (\gamma_1, \gamma_2)$. Then by definition $\max\{\alpha_1, \alpha_2\} \leq \max\{\beta_1, \beta_2\} \leq \max\{\gamma_1, \gamma_2\}$. If $\max\{\alpha_1, \alpha_2\} < \{\gamma_1, \gamma_2\}$ then $(\alpha_1, \alpha_2) \prec (\gamma_1, \gamma_2)$ and we are done. Otherwise, we have $\alpha_1 \leq \beta_1 \leq \gamma_1$. If $\alpha_1 < \gamma_1$ then $(\alpha_1, \alpha_2) \prec (\gamma_1, \gamma_2)$. Otherwise, we have $\alpha_1 = \beta_1 = \gamma_1$ and so necessarily $\alpha_2 < \beta_2 < \gamma_2$, which gives $(\alpha_1, \alpha_2) \prec (\gamma_1, \gamma_2)$.

Next, $\prec$ also satisfies the trichotomy condition. This is fairly easy to see from the definition.

Finally, we show that $\prec$ is a well-ordering. For any non-empty set of ordinal pairs X, let

$$A = \{\max\{\alpha, \beta\} \mid (\alpha, \beta) \in X\}.$$

A is non-empty and has a least element, call it δ. Now define

$$B = \{\alpha \mid \max\{\alpha, \beta\} = \delta \wedge (\alpha, \beta) \in X\}.$$

B is also non-empty and has a least element, call it α'. Then define

$$C = \{\beta \mid \max\{\alpha', \beta\} = \delta \wedge (\alpha', \beta) \in X\}.$$

Again C is non-empty and has a least element β'. By the sequence of choices, we can see that (α', β') is the least element of X.

Now we perform transfinite induction on α to show that $|\omega_\alpha \times \omega_\alpha| \leq \aleph_\alpha$. If $\alpha = 0$ then $\aleph_0 \cdot \aleph_0 = \aleph_0$ (see theorem 4.7).

Now assume $\aleph_\beta \cdot \aleph_\beta = \aleph_\beta$ for all $\beta < \alpha$. It suffices to show that the order type of the well-ordered set $(\omega_\alpha \times \omega_\alpha, \prec)$ is at most ω_α. Suppose the opposite is true. Then there would exist $(\gamma_1, \gamma_2) \in \omega_\alpha \times \omega_\alpha$ such that the cardinality of

$$X = \{(\xi_1, \xi_2) \in \omega_\alpha \times \omega_\alpha \mid (\xi_1, \xi_2) \prec (\gamma_1, \gamma_2)\}$$

is at least $\aleph_\alpha$. Therefore we want to show that for any $(\gamma_1, \gamma_2) \in \omega_\alpha \times \omega_\alpha$, we have $|X| < \aleph_\alpha$.

Let $\beta = \max\{\gamma_1, \gamma_2\} + 1$. Then $\beta \in \omega_\alpha$ and for all $(\xi_1, \xi_2) \in X$, we have $\max\{\xi_1, \xi_2\} \leq \max\{\gamma_1, \gamma_2\} < \beta$, so $\xi_1 \in \beta \ni \xi_2$. In other words, $X \subseteq \beta \times \beta$.

Let $\lambda < \alpha$ be such that $|\beta| \leq \aleph_\lambda$. Then $|X| \leq |\beta \times \beta| = |\beta| \cdot |\beta| \leq \aleph_\lambda \cdot \aleph_\lambda$ and by the inductive hypothesis, $\aleph_\lambda \cdot \aleph_\lambda \leq \aleph_\lambda$, hence $|X| \leq \aleph_\lambda < \aleph_\alpha$. ■

Corollary 8.4.1. *For every α and β such that $\alpha \leq \beta$, $\aleph_\alpha \cdot \aleph_\beta = \aleph_\beta$. Furthermore for every natural number n, we have $n \cdot \aleph_\alpha = \aleph_\alpha$.*

Proof. We have $\aleph_\beta = 1 \cdot \aleph_\beta \leq \aleph_\alpha \cdot \aleph_\beta \leq \aleph_\beta \cdot \aleph_\beta = \aleph_\beta$. The equality is similar for the second claim. ■

Corollary 8.4.2. *For every α and β such that $\alpha \leq \beta$, $\aleph_\alpha + \aleph_\beta = \aleph_\beta$. Furthermore for every natural number n, we have $n + \aleph_\alpha = \aleph_\alpha$.*

Proof. If $\alpha \leq \beta$ then $\aleph_\beta \leq \aleph_\alpha + \aleph_\beta \leq \aleph_\beta + \aleph_\beta = 2 \cdot \aleph_\beta = \aleph_\beta$ by the previous corollary. The second claim is proved similarly. ■

We can now resume the proof of theorem theorem 6.7. First, we will extend the definition of derivatives using transfinite recursion.

Definition 8.4 (Derivatives)**.** Let A be a set of reals. For every ordinal α, define

- $A^{(0)} = A$.
- $A^{(\alpha+1)} = (A^{(\alpha)})'$.
- $A^{(\alpha)} = \bigcap_{\xi<\alpha} A^{(\xi)}$ for limit ordinal $\alpha \neq 0$.

□

Theorem 8.5. *Let F be a closed set of reals. There exists an at most countable ordinal θ such that*

- *For all $\alpha < \theta$, the set $F^{(\alpha)} \setminus F^{(\alpha+1)}$ is non-empty and at most countable.*
- *$F^{(\theta+1)} = F^{(\theta)}$.*
- *$F^{(\theta)}$ is either empty or perfect.*
- *$F \setminus F^{(\theta)}$ is at most countable.*

Proof. Using transfinite recursion we can show that for every α, $F^{(\alpha)}$ is closed. The successor case is evident since taking the derivative gives a closed set. The limit case follows from the fact that the intersection of closed sets is still closed.

The set $F^{(\alpha)} \setminus F^{(\alpha+1)}$ is the set of all isolated points of $F^{(\alpha)}$, which is countable by lemma 6.6.

As long as $F^{(\alpha+1)} \neq F^{(\alpha)}$, the transfinite sequence $\langle F^{(\alpha)} \rangle$ is decreasing, i.e. $F^{(\alpha)} \supset F^{(\beta)}$ when $\alpha < \beta$. By the axiom of replacement, there is an ordinal θ such that $F^{(\theta+1)} = F^{(\theta)}$. Otherwise, suppose $F^{(\alpha)} \setminus F^{(\alpha+1)} \neq \emptyset$ for all α. Let $\gamma = h(\mathcal{P}(F))$. The function $g(\alpha) = F^{(\alpha)} \setminus F^{(\alpha+1)}$ is a one-to-one function from $h(\mathcal{P}(F))$ to $\mathcal{P}(F)$, which is a contradiction. So suppose θ is the least such ordinal.

Let $P = F^{(\theta)}$. We now have $P' = P$. So P is either empty or perfect.

Let $C = F \setminus P$. We want to show that C is at most countable. Let $\langle J_0, J_1, \ldots, J_n, \ldots \rangle$ be a sequence of all open intervals with rational endpoints. For each $a \in C$, let α_a be the unique ordinal $\alpha_a < \theta$ such that $a \in F^{(\alpha)} \setminus F^{(\alpha+1)}$, and let $f(a)$ be the least index n such that $J_n \cap F^{(\alpha_a)} = \{a\}$. Intuitively, α_a is the stage at which a was "thrown away" by taking the derivative.

Observe that f is a function with domain C, since it is well-defined for each $a \in C$, and it has codomain ω, since we always choose the least index. It remains to show that f is one-to-one. Let $a, b \in C$ such that $f(a) = f(b) = n$. Without loss of generality, assume that $\alpha_a \leq \alpha_b$. Hence $F^{(\alpha_a)} \supseteq F^{(\alpha_b)}$. Then $\{a\} = J_n \cap F^{(\alpha_a)} \subseteq J_n \cap F^{(\alpha_b)} = \{b\}$. So $a = b$. ■

Theorem. *Every countable closed set contains a perfect subset.*

Proof. From the previous theorem, every uncountable closed set F is decomposed into a perfect set $F^{(\theta)}$ and an at most countable set $F \setminus F^{(\theta)}$. ■

9 Axiom of choice

Is every set well-orderable? Cantor thought that the answer is yes and provided a proof.

To well-order a set A it suffices to construct a one-to-one mapping from some ordinal α onto A. We proceed by transfinite recursion. Let a be a set not in A. Define

$$f(0) = \begin{cases} \text{some element of } A, & \text{if } A \neq \emptyset \\ a, & \text{otherwise} \end{cases}$$

$$f(1) = \begin{cases} \text{some element of } A - \{f(0)\}, & \text{if } A - \{f(0)\} \neq \emptyset \\ a, & \text{otherwise} \end{cases}$$

Generally we have

$$f(\alpha) = \begin{cases} \text{some element of } A - \operatorname{ran}(f \restriction \alpha), & \text{if } A - \operatorname{ran}(f \restriction \alpha) \neq \emptyset \\ a, & \text{otherwise} \end{cases}$$

Intuitively f lists the elements of A one by one, and when A is exhausted, then f has value a. Indeed, there is an ordinal $\lambda < h(A)$, such that $f(\lambda) = a$.

For $\alpha < \beta < h(A)$, if $f(\beta) \neq a$, then $f(\beta) \in A - \operatorname{ran}(f \restriction \beta)$, $f(\alpha) \in \operatorname{ran}(f \restriction \beta)$, and so $f(\alpha) \neq f(\beta)$. If $f(\alpha) \neq a$ for all $\alpha < h(A)$, this would mean that f is an one-to-one mapping of $h(A)$ into A, contradicting the definition of $h(A)$.

The proof will be complete if $\operatorname{ran}(f \restriction \lambda) = A$. By the definition of f, $\operatorname{ran}(f \restriction \lambda) \subseteq A$. On the other hand, if $\operatorname{ran}(f \restriction \lambda) \subset A$, then $A - \operatorname{ran}(f \restriction \lambda) \neq \emptyset$. Therefore $f(\lambda) \neq a$ which contradicts the choice of λ.

This proof seems convincing yet there is a flaw. Using transfinite recursion, we will need an operation G such that f could be defined by $f(\alpha) = G(f \restriction \alpha)$. Then G will have the following properties:

$$\begin{cases} G(f \restriction \alpha) \in A - \operatorname{ran}(f \restriction \alpha), & \text{if } A - \operatorname{ran}(f \restriction \alpha) \neq \emptyset \\ G(f \restriction \alpha) = a, & \text{otherwise} \end{cases}.$$

If A is well-orderable, say with $\prec$, we can define such a G by

$$G(x) = \begin{cases} \prec\text{-least element of } A - \operatorname{ran} x, & \text{if } A - \operatorname{ran} x \neq \emptyset \\ a, & \text{otherwise} \end{cases}.$$

Definition 9.1 (Choice functions)**.** Let S be a system of sets. A function g defined on S is called a *choice function* for S if $g(X) \in X$ for all non-empty $X \in S$. □

Theorem 9.1. *A set A can be well-ordered iff the set $\mathcal{P}(A)$ has a choice function.*

Proof. If there is a choice function g for $\mathcal{P}(A)$, we can define

$$G(x) = \begin{cases} g(A - \operatorname{ran} x), & \text{if } A - \operatorname{ran}(x) \neq \emptyset \\ a, & \text{otherwise} \end{cases}.$$

For the other direction, if A is well-ordered, then the choice function g can be defined as

$$g(x) = \begin{cases} \prec\text{-least element of } x, & \text{if } x \neq \emptyset \\ \emptyset, & \text{otherwise} \end{cases}.$$

■

Theorem 9.2. *Every finite system of sets has a choice function.*

Proof. Perform induction on the number of sets n. Let us assume that every system with k elements has a choice function, and consider the case where $|S| = k+1$. Take any $X \in S$. Then the set $S - \{X\}$ has k elements. By the inductive hypothesis it has a choice function g. If $X = \emptyset$ then $g' = g \cup \{(X, \emptyset)\}$ is a choice function for S. Otherwise, for any $x \in X$, $g^x = g \cup \{(X, x)\}$ is a choice function for S. ■

This theorem cannot be generalized to the infinite case. In the finite case the axiom of comprehension was used to define g^x with x as a parameter. However, if the system is infinite, the formula will be of infinite length. Also, while it is easy to find a choice function for $\mathcal{P}(\mathbb{N})$ or $\mathcal{P}(\mathbb{Q})$ (since they are well-ordered), there is no obvious candidate for $\mathcal{P}(\mathbb{R})$.

Axiom of choice There exists a choice function for every system of sets.

Theorem 9.3. *AC is equivalent to the well-ordering principle: every set can be well-ordered.*

Proof. This follows from theorem 9.1. ■

The axiom of choice is also equivalent to an algebraic statement called Zorn's lemma. This is useful in other fields since it is easier to apply, as we do not need to consider ordinals or transfinite induction, etc.

Definition 9.2 (Chains)**.** Let $(X, <)$ be a partially ordered set. A subset $C \subseteq X$ is called a *chain* in X if $(C, <)$ is a linear ordered subset of X. □

Theorem 9.4 (Zorn's lemma)**.** *AC is equivalent to the following:*

- *(Zorn's Lemma) If every chain in a partially ordered set has an upper bound, then the partially ordered set has a maximal element.*

Proof. We will first go from the well-ordering principle to Zorn's lemma.

Let $(A, \preceq)$ be a partial order in which every chain has an upper bound. By the well-ordering principle, there exists a transfinite sequence $\langle x_\alpha \mid \alpha < \lambda \rangle$ whose range is A. Define $F : \lambda \to \lambda + 1$ by transfinite recursion:

$$F(\nu) = \begin{cases} \text{least } \alpha < \lambda \text{ where } \forall \xi < \nu \ (x_{F(\xi)} \prec x_\alpha), & \text{if such } \alpha \text{ exists} \\ \lambda, & \text{otherwise} \end{cases}.$$

Intuitively this is a search for the smallest maximal element that is above ν. For example, $F(1)$ returns the first index α where $x_0 = x_{F(0)} \prec x_\alpha$. $F(2)$ returns the first index β such that $x_0 \prec x_\beta \wedge x_\alpha \prec x_\beta$. Notice that F is an increasing function. Let

$$\nu_0 = \begin{cases} \text{least } \nu \text{ such that } F(\nu) = \lambda, & \text{if such } \nu \text{ exists} \\ \lambda, & \text{otherwise} \end{cases}.$$

Now observe that $B = \{x_{F(\eta)} \mid \eta < \nu_0\}$ is a chain in $(A, \preceq)$. By assumption B has an upper bound b. We claim that b is a maximal element of A.

Suppose this is not the case. Then there exists some $\gamma < \lambda$ such that $b \prec x_\gamma$, which implies that for all $\eta < \nu_0$, we have $x_{F(\nu)} \prec x_\gamma$.

Take the least such γ. If $\nu_0 \leq \gamma$ then by definition $F(\nu_0) = \gamma < \lambda$ which contradicts $F(\nu_0) = \lambda$.

If on the other hand $\gamma < \nu_0$, then for all $\eta < \gamma$, we have $x_{F(\eta)} \prec x_\gamma$, so $F(\gamma) \leq \gamma$. However since it is an increasing sequence, $F(\gamma) \geq \gamma$. Therefore $F(\gamma) = \gamma$. Hence $x_\gamma = x_{F(\gamma)}$ which contradicts the fact that $\forall \eta < \gamma \ [x_{F(\eta)} \prec x_\gamma]$.

Now we will go from Zorn's lemma to the axiom of choice. Without loss of generality, let S be a system of non-empty sets. We will show that S has a choice function.

Let F be the system of all functions f for which $\operatorname{dom} f \subseteq S$ and for all $X \in \operatorname{dom} f$, $f(X) \in X$. Intuitively here f is a partial choice function. F is ordered by $\subseteq$.

If F_0 is a chain in $(F, \subseteq)$, then $f_0 = \bigcup F_0$ is a function. It is easy to check that $f_0 \in F$ and f_0 is an upper bound on F_0.

By Zorn's lemma, $(F, \subseteq)$ has a maximal element g. If $\operatorname{dom} g \neq S$, say $X \in S - \operatorname{dom} g$, and $x \in X$, then $h = g \cup \{(X, x)\} \in F$ and so $g \subset h$, a contradiction. Therefore, g is a choice function of S. ■

9.1 Consequences of the axiom of choice

Theorem 9.5. *Every infinite set A has a countable subset C.*

Proof. By the axiom of choice, $A = \{a_\alpha \mid \alpha < \gamma\}$ for some infinite ordinal γ. Let $C = \{a_\alpha \mid \alpha < \omega\}$ ■

Theorem 9.6. *For every infinite set S there exists a unique aleph $\aleph_\alpha$ such that $|S| = \aleph_\alpha$.*

Proof. As S can be well-ordered it is equipotent to some infinite ordinal and hence to an unique initial ordinal ω_α. ■

In fact the following theorem is equivalent to the axiom of choice.

Corollary 9.6.1. *For any sets A and B either $|A| \leq |B|$ or $|B| \leq |A|$.*

Proof. By the previous theorem there are some alephs $|A| = \aleph_\alpha$ and $|B| = \aleph_\beta$ and we can compare α and β. ■

Theorem 9.7. *If f is a function and A is a set, then $|f[A]| \leq |A|$.*

Proof. For each $b \in F[A]$, let $X_b = f^{-1}(\{b\})$. Note that $X_b \neq \emptyset$ and $X_{b_1} \cap X_{b_2} = \emptyset$ if $b_1 \neq b_2$. Now take

$$g \in \prod_{b \in f[A]} X_b.$$

Then g is an one-to-one mapping of $f[A]$ into A. ■

Theorem 9.8. *Every vector space has a basis.*

Proof. If C is a $\subset$-chain of independent subsets of V, then the union of C is also an independent set. By Zorn's lemma, a maximal independent set exists. ■

10 Other topics

In this section we will talk about some other topics related to ordinals.

First we will discuss Cantor's normal form. Observe that ordinal addition $\alpha + \beta$ is continuous in the second variable. If γ is a limit ordinal and $\beta = \sup_{\delta<\gamma} \beta_\delta$, then

$$\alpha + \beta = \sup_{\delta<\gamma}(\alpha + \beta_\delta)$$

The same goes for $\alpha \cdot \beta$ and α^β.

Lemma 10.1.

i. If $0 < \alpha \leq \gamma$ then there is a greatest ordinal β such that $\alpha \cdot \beta \leq \gamma$.

ii. If $1 < \alpha \leq \gamma$ then there is a greatest ordinal β such that $\alpha^\beta \leq \gamma$.

Proof. Since $\gamma < \gamma + 1 \leq \alpha \cdot (\gamma + 1)$, there is a least δ such that $\alpha \cdot \gamma > \gamma$. The least such δ must be a successor ordinal by the continuity observation, say $\delta = \beta + 1$.

The other statement follows quite similarly. ■

Lemma 10.2. *If γ is an arbitrary ordinal and if $\alpha \neq 0$, then there exists an unique ordinal β and an unique $\rho < \alpha$ such that $\gamma = \alpha \cdot \beta + \rho$.*

Proof. Let β be the greatest ordinal such that $\alpha \cdot \beta \leq \gamma$, since such an ordinal exists by the previous lemma. Then let ρ be the unique ρ such that $\alpha \cdot \beta + \rho = \gamma$.

$\rho < \alpha$, since otherwise we get $\alpha \cdot (\beta + 1) = \alpha \cdot \beta + \alpha \leq \alpha \cdot \beta + \rho = \gamma$ which contradicts the maximality of β. The uniqueness part is quite obvious. ■

Theorem 10.3 (Cantor's normal form theorem)**.** *Every ordinal $\alpha > 0$ can be expressed uniquely as*

$$\alpha = \omega^{\beta_1} \cdot k_1 + \omega^{\beta_2} \cdot k_2 + \cdots \omega^{\beta_n} \cdot k_n$$

where $\beta_1 > \beta_2 > \cdots > \beta_n$ and $k_i > 0$ are finite.

Proof. We perform transfinite induction on α. The ordinal $\alpha = 1$ can be expressed as $1 = \omega^0 \cdot 1$.

Now suppose the proposition is true for some $\alpha > 0$. By the earlier lemma, there exists a greatest β such that $\omega^\beta \leq \alpha$, and there exists an unique δ and ρ such that $\rho < \omega^\beta$ and $\alpha = \omega^\beta \cdot \delta + \rho$.

Since $\omega^\beta \leq \alpha$, we have $\delta > 0$ and $\rho < \alpha$, since if $\delta = 0$ then $\alpha = \rho < \omega^\beta$, which is a contradiction. Furthermore, δ is finite since otherwise $\alpha \geq \omega^\beta \cdot \delta \geq \omega^\beta \cdot \omega = \omega^{\beta+1}$, contradiction the maximality of β. Thus, let $\beta_1 = \beta$ and $k_1 = \delta$.

If $\rho = 0$, then $\alpha = \omega^{\beta_1} \cdot k_1$ and we are done. If $\rho > 0$, then by the inductive hypothesis,

$$\rho = \omega^{\beta_2} \cdot k_1 + \cdots \omega^{\beta_n} \cdot k_n$$

for some $\beta_2 > \cdots > \beta_n$ and finite $k_i > 0$. Since $\rho < \omega^{\beta_1}$, we have $\omega^{\beta_2} \leq \rho < \omega^{\beta_1}$, and $\beta_1 > \beta_2$. Therefore we get the normal form

$$\alpha = \omega^{\beta_1} \cdot k_1 + \omega^{\beta_2} \cdot k_2 + \cdots \omega^{\beta_n} \cdot k_n$$

Next we will show uniqueness. First observe that $\beta < \gamma$ then $\omega^\beta \cdot k < \omega^\gamma$ for every finite k. Consequently, if $\alpha = \omega^{\beta_1} \cdot k_1 + \cdots + \omega^{\beta_n} \cdot k_n$ is in normal form and $\gamma > \beta_1$, then $\alpha < \omega^\gamma$.

Next perform induction on α. For $\alpha = 1$, this is trivial. Next, let

$$\begin{aligned}\alpha &= \omega^{\beta_1} \cdot k_1 + \omega^{\beta_2} \cdot k_2 + \cdots + \omega^{\beta_n} \cdot k_n \\ &= \omega^{\gamma_1} \cdot j_1 + \omega^{\gamma_2} \cdot j_2 + \cdots + \omega^{\gamma_n} \cdot j_n\end{aligned}$$

Then by our previous observation this means $\beta_1 = \gamma_1$. Let $\delta = \omega^{\beta_1} = \omega^{\gamma_1}$ and $\rho = \omega^{\beta_2} \cdot k_2 + \cdots + \omega^{\beta_n} \cdot k_n$ and $\sigma = \omega^{\gamma_2} \cdot j_2 + \ldots + \omega^{\gamma_m} \cdot k_m$. Then $\alpha = \delta \cdot k_1 + \rho = \delta \cdot j_1 + \sigma$. However since $\rho < \delta$ and $\sigma < \delta$, by the lemma above they have to be unique, and so $k_1 = j_1$ and $\rho = \sigma$. By the inductive hypothesis, the result follows. ■

Theorem 10.4. *The ordinal*

$$\varepsilon_0 = \sup\{\omega, \omega^\omega, \omega^{\omega^\omega}, \dots\}$$

is the smallest ordinal α such that $\omega^\alpha = \alpha$.

Proof. Let β_n be a shorthand for the n-th element in the sup, i.e. $\beta_0 = \omega$, $\beta_1 = \omega^\omega$, and so on.

$$\omega^{\varepsilon_0} = \sup\{\omega^{\beta_n} \mid n \in \omega\} = \sup\{\beta_{n+1} \mid n \in \omega\} = \varepsilon_0$$

■

ε_0 is a countable ordinal and is important in Proof Theory. We will soon introduce the Goodstein theorem which is an application of ordinals and is related to Proof Theory.

We have seen the Cantor normal form for $\alpha < \epsilon_0$:

$$\alpha = \omega^{\alpha_1} \cdot k_1 + \cdots + \omega^{\alpha_n} \cdot k_n.$$

The ordinals α_i with $\alpha_i \neq 0$ can also be written in Cantor normal form with yet smaller exponents. This process terminates with finitely many steps, since $\alpha < \epsilon_0$ so there exist some n such that α is smaller than the n-th tower in ϵ_0. Therefore α can be expressed in a unique way which we call its *complete Cantor normal form* where every exponent is in Cantor normal form.

For any positive integer m and $b \geq 2$, we can have its b-representation and complete b-representation. For example:

$$\begin{aligned}2020 &= 3^6 \cdot 2 + 3^5 \cdot 2 + 3^3 \cdot 2 + 3^3 \cdot 2 + 3 + 1 \\ &= 3^{3^{3^0 \cdot 1} \cdot 2} \cdot 2 + 3^{3^{3^0 \cdot 1} \cdot 1 + 3^0 \cdot 2} \cdot 2 + \cdots\end{aligned}$$

For natural numbers $m > 0$ and $c \geq b \geq 2$ let $S_c^b(m)$ be the integer resulting from m by replacing the base b in the complete b-representation of m everywhere by c. For example

$$34 = 3^3 + 3 \cdot 2 + 1 \qquad S_4^3(34) = 4^4 + 4 \cdot 2 + 1 = 265.$$

Definition 10.1 (Goodstein sequence)**.** Define $m_0 = m$ written in complete 2-representation. Suppose that m_i is obtained when written in complete $(i+2)$-representation, then let $m_{i+1} = S_{i+3}^{i+2}(m_i) \dot{-} 1$. We call $(m_i)_{i \in \mathbb{N}}$ a *Goodstein* sequence. □

Example 10.1. Start with $m = 21$.

$$\begin{aligned} m_0 &= 2^{2^2} + 2^2 + 1 \\ m_1 &= 3^{3^3} + 3^3 \approx 10^{12} \\ m_2 &= 4^{4^4} + 4^3 \cdot 3 + 4^3 \cdot 3 + 4 \cdot 3 + 3 \approx 10^{154} \\ m_3 &= 5^{5^5} + 5^3 \cdot 3 + 5^2 \cdot 3 + 5 \cdot 3 + 2 \approx 10^{2184} \end{aligned}$$

$\Diamond$

Theorem 10.5 (Goodstein's theorem)**.** *Every Goodstein sequence terminates. For any m there exists k such that $m_i = 0$ for all $i \geq k$.*

Proof. For each m_i with terms written complete $(i+2)$-representation, we replace each $i+2$ by ω and call it α_i. For instance, using the above example:

$$\begin{aligned} \alpha_0 &= \omega^{\omega^\omega} + \omega^\omega + 1 \\ \alpha_1 &= \omega^{\omega^\omega} + \omega^\omega \\ \alpha_2 &= \omega^{\omega^\omega} + \omega^3 \cdot 3 + \omega^3 \cdot 3 + \omega \cdot 3 + 3 \end{aligned}$$

and so on. These ordinals are in Cantor normal form. It can be seen that the sequence (α_i) is a decreasing sequence of ordinals. Since there cannot be an infinite descending sequence of ordinals (ordinals are well-ordered), there exists some n such that $\alpha_n = 0$. Since $m_i < \alpha_i$ for all i, $m_n = 0$. ■

Goodstein's theorem has been shown to be unprovable in PA. One form of Gödel's incompleteness theorem states that there are true statements about the natural numbers that cannot be proved using PA. However, his construction was a pathological example using diagonalization among other tools. However, Goodstein's theorem is a more natural, number theoretic example. The only way we are able to prove it is by using stronger logical systems, such as ZFC in our case.