Cardinal Arithmetic and the Axiom of Choice

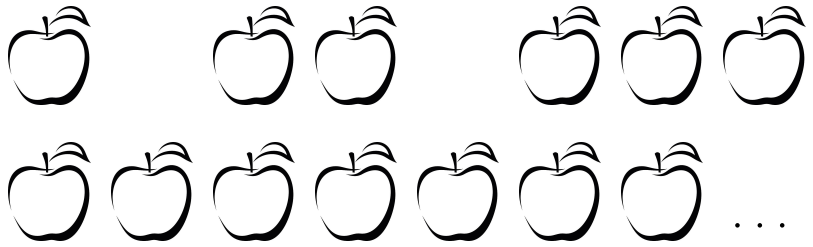
Instructor: Nikhil Sahoo

18 September 2019

Note to the reader: The exercises given below come with three sorts of punctuation. An exclamation mark means that the exercise should be attempted before moving forward. A period means that the exercise is good practice, but not strictly necessary. A question mark means that the exercise is harder than usual.

1 Introduction

You may have heard of the infamous *Axiom of Choice* (if you haven't, don't worry; we'll start from scratch). It has a history of controversy and notoriety, which has permeated into popular mathematics, philosophy and to some extent, the public eye. Some mathematicians (particularly near the turn of the 20^{th} century) saw the Axiom of Choice as a sort of cheating. Although it has some mind-boggling and paradoxical consequences, my goal is to show you how this axiom is a necessary tool in dealing with the vast eccentricity of infinities.



When we first learn about whole numbers, it is through counting concrete objects. Another way to say this is that we define whole numbers to measure the size of finite *sets* of objects. Here, we will deal with infinities in much the same way, by considering the sizes of various sets. First, we must become acquainted with sets.

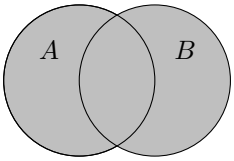
Georg Cantor, the founder of modern set theory, defined a set $\mathfrak{M}$ as "a collection into a whole, of definite, well-distinguished objects [called the *elements* of $\mathfrak{M}$] of our perception or of our thought." This definition is somewhat informal, so we must be careful to avoid such paradoxical notions as "the set of all sets." However, on first encounter, the fully axiomatic and rigorous constructions of sets can mask simple, valuable intuition.

Definition. Fix sets A and B. If x is an element of A, we write $x \in A$. If every $x \in A$ also satisfies $x \in B$, then we say that A is a *subset* of B and we write $A \subset B$. Given some condition P, which might be satisfied by elements of B, we write $\{x \in B : x \text{ satisfies } P\}$ for the subset of B consisting of elements satisfying P.

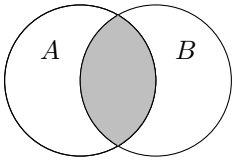
For two sets A and B, we can form their *union* $A \cup B$, which consists of all elements that are in A OR B. We can also form their *intersection* $A \cap B$, which consists of all elements that are in A AND B. In general, we can consider a large collection of sets $\{A_i : i \in I\}$, where I is some auxilliary indexing set. We define:

$$\text{The union } \bigcup_{i \in I} A_i \text{ is the set of all elements contained in AT LEAST ONE OF the } A_i;$$
$$\text{The intersection } \bigcap_{i \in I} A_i \text{ is the set of all elements contained in ALL OF the } A_i.$$

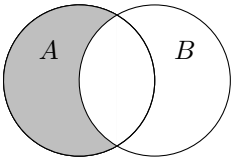
We can also define the set difference $A - B$, which consists of all elements that are in A BUT NOT in B. The union, intersection and difference of two sets A and B can be represented pictorially by Venn diagrams.



$A \cup B$



$A \cap B$



$A - B$

Exercise 1! Show that the following are equivalent: $A \subset B$, $A \cup B = B$, $A \cap B = A$.

The simplest set is the empty set $\emptyset$, which has no elements whatsoever. Other examples of sets include:

$$\text{The natural, whole or counting numbers } \mathbb{N} = \{0, 1, 2, 3, \dots\}.$$
$$\text{The integers } \mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\} \text{ (short for } \textit{zahlen}).$$
$$\text{The rational/quotient numbers } \mathbb{Q} = \{a/b : a, b \in \mathbb{Z} \text{ and } b \neq 0\}.$$
$$\text{The real numbers } \mathbb{R} \text{ and complex numbers } \mathbb{C} = \{a + ib : a, b \in \mathbb{R}\}.$$

Note that $\emptyset \subset \mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$. Sets do not need to consist of numbers. If P is a set of people, then

$$\{x \in P : x \text{ likes apples}\} \subset \{x \in P : x \text{ likes some fruit}\}.$$

We may also form sets of sets. For any set A, its power set $\mathcal{P}(A)$ is the set of all subsets of A. Notice that $\emptyset \subset A$ for any set A, so the power set is always nonempty. In §3, we will prove that $\mathcal{P}(A)$ is actually *larger* than A. In Von Neumann's approach to set theory, all sets can be built up from the power set operation.

Definition. Given two sets A and B, we can form their *Cartesian product* $A \times B$, which consists of all ordered pairs (a, b) with $a \in A$ and $b \in B$. For example, the Euclidean plane can be expressed as $\mathbb{R} \times \mathbb{R}$, because every point in the plane is given by an ordered pair of coordinates. For a simpler example, consider

$$\{1, 2\} \times \{1, 2, 3\} = \{(1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (2, 3)\}.$$

Exercise 2! Prove that the Cartesian product $A \times B$ is nonempty if and only if A and B are nonempty.

Definition. We may now define the notion of a *relation* on a set A, which is just a subset $R \subset A \times A$. This set R encodes a relationship between certain elements of A, satisfied precisely by the pairs $(a, b) \in R$. Rather than writing $(a, b) \in R$, we will choose some symbol (e.g. $\propto$) and write $a \propto b$. We say that $\propto$ is...

(a) ... reflexive if $a \propto a$ for all $a \in A$.

(b) ... symmetric if $a \propto b$ implies that $b \propto a$.

(c) ... transitive if $a \propto b$ and $b \propto c$ implies that $a \propto c$.

(d) ... anti-symmetric if $a \propto b$ and $b \propto a$ implies that $a = b$.

A relation is said to be an *equivalence relation* if it is reflexive, symmetric and transitive (as a mnemonic device, note that RST is an "alphabetical run"). Equivalence relations generalize our usual notion of equality, so we denote them by symbols like $\sim$, $\simeq$, $\cong$, $\equiv$, etc. However, the symbol $=$ is reversed for actual equality.

Exercise 3. Prove the following results about equivalence relations.

1. Let n be a positive integer. Given $a, b \in \mathbb{Z}$, we say that $a \equiv_n b$ if $a - b$ is divisible by n. Show that this defines an equivalence relation on $\mathbb{Z}$. Given $a \equiv_n b$ and $c \equiv_n d$, show that $a + c \equiv_n b + d$ and $ac \equiv_n bd$.
2. Do symmetric and transitive imply reflexive? Prove this implication or construct a counter-example.

Exercise 4? Let us first define a couple of concepts related to equivalence relations:

(a) A partition of a set A is a set $P \subset \mathcal{P}(A)$ of nonempty subsets, such that $\bigcup_{B \in P} B = A$ and for any $B, C \in P$, we have $B \cap C = \emptyset$. In other words, each element $x \in A$ is contained in a unique $P_x \in P$.

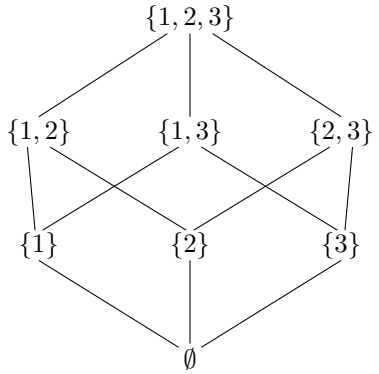
(b) If $\sim$ is an equivalence relation on A, the equivalence class of $x \in A$ is the set $[x] = \{y \in A : x \sim y\}$. We write $A/\!\sim$ for the set $\big\{[x] \in \mathcal{P}(A) : x \in A\big\}$ of equivalence classes.

If P is a partition of A, we define $x \sim_P y$ if and only if $P_x = P_y$. Show that this is an equivalence relation. If $\sim$ is an equivalence relation on A, show that $A/\!\sim$ is a partition of A. Show that these operations are inverses of each other. (See §2 for the definition of an inverse. What are the domain and range involved?)

Definition. A relation is said to be a *partial order* if it is reflexive, anti-symmetric and transitive. Partial orders are usually denoted by $\leq$ or $\preceq$. We write $a < b$ to mean that $a \leq b$ and $a \neq b$ (and similarly with $\prec$). Some common examples are sets of sets (e.g. a power set), with the ordering given by inclusion of subsets.

Exercise 5! Consider a set A and show that $\subset$ is a partial order on the power set $\mathcal{P}(A)$.

Below, we illustrate an example of this partial order on the eight subsets of the set $\{1, 2, 3\}$.



Exercise 6? Let $\leq$ be a reflexive, transitive relation on the set X. Given any $x, y \in X$, we say that $x \sim y$ if $x \leq y$ and $y \leq x$. Show that $\sim$ is an equivalence relation on X. If $x \sim x'$ and $y \sim y'$, show that $x \leq y$ implies $x' \leq y'$. Thus we define a relation on $X/\sim$ by setting $[x] \leq [y]$ if and only if $x \leq y$ (see Exercise 4). Show that this is a partial order on $X/\sim$.

Let $\leq$ be a partial order on the set A. We say that two elements $a, b \in A$ are comparable if $a \leq b$ or $b \leq a$. In the above example, note that $\{1, 2\}$ and $\{1, 3\}$ are not comparable, since neither is a subset of the other.

Definition. A *total order* on a set A is a partial order such that every two elements of A are comparable. A *well-order* is a total order on a set A, such that every nonempty subset $B \subset A$ contains a least element. More specifically, a least element of the subset B is some $m \in B$ such that $m \leq b$ for all $b \in B$.

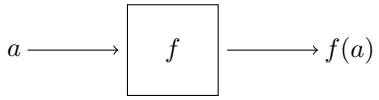
The typical example of a total order is the usual ordering on the real number line. This order also restricts to a total order on the subsets $\mathbb{N}$, $\mathbb{Z}$ and $\mathbb{Q}$ (as well as any other subset of the real line). For the time being, we do not know if many other total orders exist. But in §5, we will prove that any set possesses a well-order.

Exercise 7! Consider a total order $\leq$ on any finite set A and show that $\leq$ is a well-order. Now consider some $n \in B \subset \mathbb{N}$ and show that $B \cap \{0, \ldots, n\}$ contains a minimal element. Conclude that $\mathbb{N}$ is well-ordered.

Exercise 8. Consider the sets $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\{p \in \mathbb{N} : p \text{ is prime}\}$ and $\{x \in \mathbb{R} : 0 \leq x \leq 1\}$. With their usual ordering as real numbers, which are well-ordered? Can you construct a well-order on the set of integers $\mathbb{Z}$?

2 The Axiom of Choice

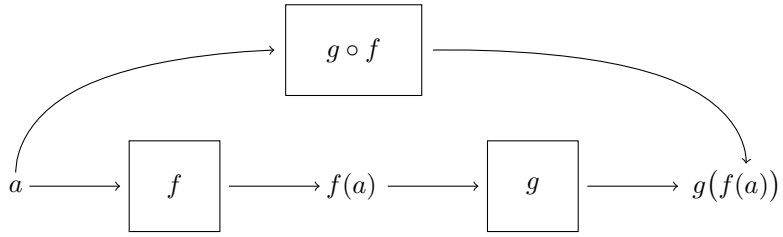
We have discussed relations on one set, but it will also be useful to have objects that relate different sets. The most common such object is a *function*. Consider sets A and B. A function f from A to B (we write $f : A \to B$) is a rule, which assigns to every element $a \in A$ a unique element of B, which we denote $f(a)$. You can think of this f as a machine or a black box, which takes inputs from A and gives outputs from B.



We call A the domain of the function and B the range. The most commonly used functions in mathematics can be described by equations. For example, we may have $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = 2x$. The same formula could describe a function $\mathbb{Z} \to \mathbb{R}$ or $\mathbb{Z} \to \mathbb{Z}$, but these are all distinct, because they have different domain and range. On the other hand, there is no reason why functions need to involve numbers. We might define a function F to take a person as an input and output their favorite fruit. Or maybe their best friend's second favorite fruit. Or basically anything else, as long as every input uniquely determines a well-defined output.

Exercise 9? For set-theorists, the above notion of a function as a "rule" is too informal. Instead, a function $f : A \to B$ is defined by its graph $\Gamma_f = \{(a, b) \in A \times B : b = f(a)\}$. Show that the set $\Gamma \subset A \times B$ is the graph of some function if and only if it satisfies: for every $a \in A$, there is exactly one $b \in B$ such that $(a, b) \in \Gamma$.

Consider functions $f : A \to B$ and $g : B \to C$. We can form a new function $g \circ f : A \to C$, called their composition, given by performing one function after the other: $(g \circ f)(a) = g\big(f(a)\big)$. We apply f to the input $a \in A$ and then apply g to the resulting element $f(a) \in B$. Writing $g \circ f$ in the correct order can be confusing at first, but when composing several functions in succession, this notation is vastly more convenient. Below, we give a schematic representation of the "machine" or "black box" viewpoint for the composition $g \circ f$.



The simplest function of all is the *identity*, which just outputs its input. For any set A, we define $\mathrm{Id}_A : A \to A$ by the formula $\mathrm{Id}_A(a) = a$. This rather boring function has some nice properties with respect to composition.

Exercise 10! Consider any function $f : A \to B$ and show that $f \circ \mathrm{Id}_A = \mathrm{Id}_B \circ f = f$.

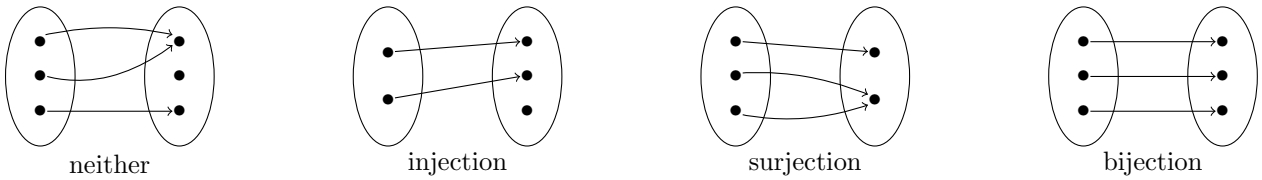
Definition. We will now define some nice properties that a function $f : A \to B$ may possess:

(a) We say the function f is an *injection* (or *injective* or *one-to-one*) if no two elements of A are mapped to the same element of B. This condition is most commonly expressed as: if $f(a) = f(b)$, then $a = b$. Such a function makes A "look like" a subset of B (so A sits inside of B), hence the name injection.

(b) We say the function f is a *surjection* (or *surjective* or *onto*) if every element of B is mapped to by an element of A. This condition is most commonly expressed as: for all $b \in B$, there exists $a \in A$ with $f(a) = b$. Such a function makes A cover all of B, hence the name onto ("sur" is French for "on").

(c) We say the function f is a *bijection* (or *bijective* or a *one-to-one correspondence*) if it is both surjective and injective. Equivalently, the function f is a bijection if every element of the range B is mapped to by a unique element of the domain A. Such a function makes the sets A and B "look the same."

The identity $\mathrm{Id}_A : A \to A$ is clearly a bijection. We illustrate some other examples of these properties below.



Given any function $f : A \to B$ and a subset $C \subset A$, we define $f(C) = \{b \in B : b = f(c) \text{ for some } c \in C\}$, which is called the *image* of C under the function f. This is the set of all elements of B that are hit by C. Similarly, given a subset $D \subset B$, we define $f^{-1}(D) = \{a \in A : f(a) \in D\}$, called the *inverse image* of D. This is the set of all elements of A that hit D. This gives us another way to characterize in/sur/bijections:

The function f is an injection if and only if the set $f^{-1}(b)$ contains ≤ 1 elements for every $b \in B$;

The function f is a surjection if and only if the set $f^{-1}(b)$ contains ≥ 1 elements every $b \in B$;

The function f is a bijection if and only if the set $f^{-1}(b)$ contains exactly 1 element for every $b \in B$.

We may also say that a function $f : A \to B$ is a surjection if and only if $f(A) = B$.

Exercise 11! Consider two functions $f : A \to B$ and $g : B \to C$.

(a) If $g \circ f$ is an injection, show that f is also an injection.

(b) If $g \circ f$ is a surjection, show that g is also a surjection.

Consider some $f : A \to B$ with $A \neq \emptyset$. If there exists some $g : B \to A$ with $g \circ f = \mathrm{Id}_A$, then Exercise 9(a) shows that f is an injection. We call such a function g a *left inverse* of f (because we get the identity map by composing with g on the left). Conversely, suppose that f is an injection. Then we can write $B = f(A) \cup B'$, where $B' = B - f(A)$. Choose some $a_1 \in A$. If $b \in f(A)$, then there exists a unique $a \in A$ with $f(a) = b$, by injectivity. We set $g(b) = a$. But if $b \in B'$, we set $g(b) = a_1$. This defines a function $g : B \to A$. For any $a \in A$, we have $f(a) \in f(A)$ and therefore $g \circ f(a) = a$. Hence $g \circ f = \mathrm{Id}_A$. Thus the function f is injective if and only if it possesses a left inverse. The exercise given below hints at why we needed to assume $A \neq \emptyset$.

Exercise 12. Consider some set A. How many functions $\emptyset \to A$ exist? How many functions $A \to \emptyset$ exist?

Exercise 13! Show that $f : A \to B$ is a bijection if and only there exists a function $g : B \to A$ such that $g \circ f = \mathrm{Id}_A$ and $f \circ g = \mathrm{Id}_B$. We call such a function g an *inverse* (or a *two-sided inverse*) of the function f.

We proceed along similar lines for the notion of surjection. Consider a function $f : A \to B$. If there exists some $g : B \to A$ with $f \circ g = \mathrm{Id}_B$, then Exercise 9(b) shows that f is a surjection. We call such a function g a *right inverse* of f. Conversely, suppose that f is a surjection and $g : B \to A$ as follows. For any $b \in B$, there exists an $a \in A$ with $f(a) = b$, by surjectivity. We choose such an a and let $g(b) = a$. Then we have $f \circ g(b) = f(a) = b$, so $f \circ g = \mathrm{Id}_B$. Thus the function f is surjective if and only if it has a right inverse. However, something differs significantly from the case of injections. There, we arbitratily chose some $a_1 \in A$. But to construct a right inverse, we defined each $g(b)$ as an arbitrary element of $f^{-1}(b)$, so we made as many choices as there are elements of B. The ability to make an infinite number of arbitrary choices is given by:

Axiom of Choice. Consider a set of sets $\{A_i : i \in I\}$ indexed by a set I. If each A_i is nonempty, then it is possible to choose elements $a_i \in A_i$ for every $i \in I$. (This choice is made instantaneously, not one at a time.)

This axiom seems perfectly reasonable. In fact, it almost seems to restate the definition of a nonempty set. If a set A contains at least one element, then surely we can choose an element of A! But the crux of this axiom is in making an arbitrary number of choices at once. This has some bizarre consequences and thus stirred up controversy in the early part of the 20^{th} century. In fact, a major achievement of 20^{th} century logic was the discovery that this axiom could not be proved or disproved from the other axioms of set theory. We will go into more detail about some of the mind-boggling consequences of the Axiom of Choice in §5.

The notion of "choosing" some $a_i \in A_i$ for every $i \in I$ still needs to be formalized in set-theoretic language. Towards this end, we first define $A = \bigcup_{i \in I} A_i$. A choice function is any $f : I \to A$ with the property that $f(i) \in A_i$ for every $i \in I$. In other words, a choice function "chooses" an element from each A_i. We define:

The set of all choice functions for $\{A_i : i \in I\}$ is the *Cartesian product* of these sets, denoted $\prod_{i \in I} A_i$.

This is a broad generalization of the Cartesian product defined in §1. Given two sets A and B, the ordered pairs $(a, b) \in A \times B$ are exactly equivalent to the functions $f : \{1, 2\} \to A \cup B$ with $f(1) \in A$ and $f(2) \in B$. In each case, we are accounting for all ways to choose one element of A and one element of B. When $A = \emptyset$, there are no ordered pairs (a, b) with $a \in A$, so $A \times B = \emptyset$. More generally, consider the following exercise.

Exercise 14! Consider a set of sets $\{A_i : i \in I\}$. If there is some $i \in I$ with $A_i = \emptyset$, show that $\prod_{i \in I} A_i = \emptyset$.

On the other hand, the Axiom of Choice can be formalized by saying that if each of the sets A_i is nonempty, then $\prod_{i \in I} A_i$ is nonempty, i.e. there exists a choice function $f : I \to A$ corresponding to this set of sets.

Excercise 15? Show that the Axiom of Choice (in terms of Cartesian products) is equivalent to the fact that every surjection possesses a right inverse, i.e. show that either statement can be proved from the other.

3 Comparizing Sizes

At last, we are in a position to talk about the sizes of infinite sets, the *cardinal numbers*. Again, our definition of this notion will be somewhat informal. For a set A, we write $\#(A)$ to represent the "number of elements" in A. We say that $\#(A) = \#(B)$ if there is a bijection from A to B. We also refer to a cardinal number α, with the implicit understanding that α denotes some $\#(A)$, where the set A may not be explicitly mentioned (just as we abbreviate "the size of the set $\{a, b, c\}$" by the symbol 3). The finite cardinals are the natural numbers $n \in \mathbb{N}$, which can be recursively defined by $0 = \#(\emptyset)$ and $n = \#(\{0, 1, 2, \ldots, n-1\})$. The smallest infinite cardinal is $\aleph_0 = \#(\mathbb{N})$ (this is pronounced "aleph-naught" and comes from the Hebrew alphabet).

As well as knowing when two sets have the same size, we want a notion of one set being bigger than another. We say that $\#(A) \leq \#(B)$ if there exists an injection from A to B (this injection may just be the inclusion of a subset $A \subset B$). For example, every set A satisfies $\emptyset \subset A$, so every cardinal number α satisfies $0 \leq \alpha$.

Exercise 16! Suppose $A \neq \emptyset$. Show that $\#(A) \leq \#(B)$ if and only if there is a surjection from B to A.

We are now able to compare the sizes of various sets, but there is still one huge problem. We may have $\#(A) \leq \#(B)$ and $\#(B) \leq \#(A)$, but not know of any bijection showing that $\#(A) = \#(B)$. Our intuition tells us that having both $\#(A) < \#(B)$ and $\#(B) < \#(A)$ should be impossible. Indeed, this is the case. Remarkably, this is one of the few facts of cardinal arthimetic that does not require the Axiom of Choice.

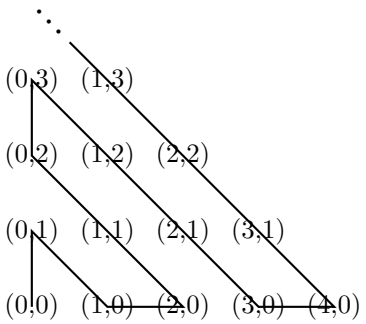
Theorem. Suppose that there exist injections $f : A \to B$ and $g : B \to A$. Then we have $\#(A) = \#(B)$. (This result was conjectured by Cantor and proved independently by Bernstein, Dedekind and Schröder.)

Proof. Since g is an injection, it restricts to a bijection $g : B \to f(B)$. By identifying B with $f(B)$, we may assume that $B \subset A$ and g is simply the inclusion of this subset. We now inductively define subsets $C_n \subset A$ for all $n \in \mathbb{N}$. Let $C_0 = A - B$ and $C_n = f(C_{n-1})$ for any $n > 0$. Next, consider their union $D = \bigcup_{n\in\mathbb{N}} C_n$.

We now define a function $h : A \to B$ as follows. If $a \in D$, then $h(a) = f(a)$. If $a \in A - D$, then $h(a) = a$. Suppose that $a \in D$ and $b \in A - D$. Then $a \in C_n$ for some $n \in \mathbb{N}$ and thus $h(a) = f(a) \in C_{n+1}$. But $b \notin D$ and $C_{n+1} \subset D$, so $h(b) = b \notin C_{n+1}$. Therefore $h(a) \neq h(b)$. Now consider some $x, y \in A$ with $h(x) = h(y)$. Then we must have $x, y \in D$ or $x, y \in A - D$. But if $x, y \in A - D$, then $x = h(x) = h(y) = y$; if $x, y \in D$, then $f(x) = h(x) = h(y) = f(y)$, so $x = y$ by the injectivity of f. In either case, we see that $h(x) = h(y)$ implies $x = y$. Therefore, the function h is an injection. Next, consider $b \in B$. If $b \in A - D$, then $h(b) = b$. If $b \in D$, then $b \in C_n$ for some $n \in \mathbb{N}$. But we can't have $b \in B$ and $b \in C_0 = A - B$, so we must have $n > 0$. Thus $b \in C_n = f(C_{n-1})$, so there exists $a \in C_{n-1} \subset D$ with $h(a) = f(a) = b$. In either case, we have shown that $b \in h(A)$. Therefore, the function $h : A \to B$ is a surjection and thus a bijection. □

Exercise 17! Prove that $\leq$ is a partial order on cardinal numbers. (Recall that there are too many sets to form a "set of all sets." Similarly, there are too many cardinals to form a "set of all cardinal numbers." Thus, we technically cannot speak of a relation on cardinal numbers without the use of further machinery. But we can still confirm that $\leq$ satisfies the defining properties of a partial order.)

Example. We will show that $\#(\mathbb{Q}) = \#(\mathbb{N})$. First, we construct a bijection $\mathbb{N} \to \mathbb{N} \times \mathbb{N}$, which just amounts to giving an exhaustive list of the elements of $\mathbb{N} \times \mathbb{N}$. But if we just start listing off $(0, 0)$, $(0, 1)$, $(0, 2)$, ..., we'll end up going to infinity before ever hitting $(1, 0)$. Instead, we must take a strategic zig-zag approach, illustrated below. We arrange the pairs $(a, b) \in \mathbb{N} \times \mathbb{N}$ in an infinite rectangular array and follow a snake-like pattern. This allows us to write every element of $\mathbb{N}\times\mathbb{N}$ in an infinite list, which defines a bijection $\mathbb{N} \to \mathbb{N}\times\mathbb{N}$. Notice that each diagonal line runs through all pairs $(a, b) \in \mathbb{N} \times \mathbb{N}$ with $a + b = n$ for some fixed $n \in \mathbb{N}$.



Proving that $\#(\mathbb{N}) = \#(\mathbb{Z})$ is straightforward, since the elements of $\mathbb{Z}$ can be listed as $0, 1, -1, 2, -2, \ldots$. Thus $\#(\mathbb{Z} \times \mathbb{Z}) = \#(\mathbb{N} \times \mathbb{N}) = \#(\mathbb{N})$. Let $Q = \{(m, n) \in \mathbb{Z} \times \mathbb{Z} : n \neq 0\}$. Then we can define a map $f : Q \to \mathbb{Q}$ by $f(m, n) = m/n$. This map is surjective by the definition of $\mathbb{Q}$. Since $Q \subset \mathbb{Z} \times \mathbb{Z}$ and $\mathbb{N} \subset \mathbb{Q}$, we then have $\#(\mathbb{N}) \leq \#(\mathbb{Q}) \leq \#(Q) \leq \#(\mathbb{Z} \times \mathbb{Z}) = \#(\mathbb{N})$. By the above theorem, we have $\#(\mathbb{Q}) = \#(\mathbb{N})$.

Exercise 18. To fill in a detail of the above argument, prove the following result. If A, B, C, D are sets with $\#(A) = \#(C)$ and $\#(B) = \#(D)$, then $\#(A \times B) = \#(C \times D)$ (= can also be replaced here with $\leq$).

Example. What does it mean for a set to be finite or infinite? As a preliminary definition, we will say that A is finite if $\#(A)$ is a natural number. Otherwise, we will say that A is infinite. The empty set is finite, since $\#(\emptyset) = 0$. If A is nonempty, then we can recursively choose distinct elements of A. First, we choose some $a_0 \in A$. If $a_0, \ldots, a_{n-1}$ have been chosen, we choose some $a_n \in A - \{a_0, \ldots, a_{n-1}\}$. By construction, these elements never repeat. The only way that this process will fail is if it is impossible to choose a_n because $A - \{a_0, \ldots, a_{n-1}\}$ is empty. In that case, we have $A = \{a_0, \ldots, a_{n-1}\}$ and $\#(A) = n$, so the set A is finite. Otherwise, we have $\#(A) > n$ for all $n \in \mathbb{N}$ and thus the set A is infinite. Using this sequence $a_0, a_1, a_2, \ldots$, we define an injective function $f : \mathbb{N} \to A$ by setting $f(n) = a_n$. This proves that $\#(A) \geq \#(\mathbb{N}) = \aleph_0$.

Exercise 19. Using the above example and theorem, prove the following statements.

(a) A set A is infinite if and only if $\#(A) \geq \aleph_0$.

(b) A set A is finite if and only if $\#(A) < \aleph_0$.

We now have a lot more tools to handle infinite cardinals, but we have still only seen one: $\aleph_0$. If this were the only infinite cardinal, then the whole theory of cardinal numbers would be extremely boring. Luckily, there is a standard construction, which takes any cardinal number and constructs a strictly larger one.

Theorem. For any set A, the power set $\mathcal{P}(A)$ is strictly larger, i.e. $\#(A) < \#\big(\mathcal{P}(A)\big)$.

Proof. We can form an injection $f : A \to \mathcal{P}(A)$ by setting $f(a) = \{a\}$. Therefore $\#(A) \leq \#\big(\mathcal{P}(A)\big)$. Suppose, for the sake of contradiction, that there is a bijection $\Phi : A \to \mathcal{P}(A)$. Then we can form the subset $U = \{a \in A : a \notin \Phi(a)\}$. Since Φ is a bijection, there is a unique $u \in A$ with $\Phi(u) = U$. We treat two cases:

If $u \in U$, then $u \notin \Phi(u)$. But then $u \in U$ and $u \notin U$. This is a contradiction.
If $u \notin U$, then $u \in \Phi(u)$. But then $u \notin U$ and $u \in U$. This is a contradiction.

We must have $u \in U$ or $u \notin U$, but both cases lead to a contradiction, so the bijection Φ cannot exist. $\square$

This theorem shows that there is an infinite hierarchy of infinite cardinals. A concrete example of a set A with $\#(A) > \aleph_0$ is the power set $A = \mathcal{P}(\mathbb{N})$. But there exists a more familiar object of the same cardinality, namely the real number line $\mathbb{R}$. We will prove this in the next section.

4 Arithmetic Operations

We will now demonstrate how to add and multiply cardinal numbers. Again, this is based on the intuition given by finite sets. Consider two cardinal numbers α and β. Choose some sets A and B such that $\alpha = \#(A)$, $\beta = \#(B)$ and $A \cap B = \emptyset$. Then we define the sum as $\alpha + \beta = \#(A \cup B)$ and the product as $\alpha\beta = \#(A \times B)$. If we don't have $A \cap B = \emptyset$, we can replace them by $A \times \{1\}$ and $B \times \{2\}$. Technically, we must confirm that this definition does not depend on the choice of A and B, but this verification is not particularly interesting.

Exercise 20. Show that the sum $\alpha + \beta$ and the product $\alpha\beta$ do not depend on the choice of sets A and B.

Next, we wish to prove some properties of addition and multiplication that are analogous to the finite case. Consider $\alpha = \#(A)$, $\beta = \#(B)$ and $\gamma = \#(C)$, with $A \cap B = \emptyset$. Then $(C \times A) \cap (C \times B) = \emptyset$ and we have

$$\begin{aligned} C \times (A \cup B) &= \big\{(c, x) : c \in C \text{ and } (x \in A \text{ or } x \in B)\big\} \\ &= \big\{(c, x) : (c \in C \text{ and } x \in A) \text{ or } (c \in C \text{ and } x \in B)\big\} \\ &= \big\{(c, x) : c \in C \text{ and } x \in A\big\} \cup \big\{(c, x) : c \in C \text{ and } x \in B\big\} \\ &= (C \times A) \cup (C \times B). \end{aligned}$$

This proves the distributive property $\gamma(\alpha + \beta) = \gamma\alpha + \gamma\beta$. Also consider some $\alpha' = \#(A')$ and $\beta' = \#(B')$, with $A' \cap B' = \emptyset$. Suppose that $\alpha \leq \alpha'$ and $\beta \leq \beta'$. Then there exist injections $f : A \to A'$ and $g : B \to B'$. Define a new function $h : A \cup B \to A' \cup B'$ as follows. If $a \in A$, then $h(a) = f(a)$; if $b \in B$, then $h(b) = g(b)$. Since $A \cap B = \emptyset$, this function is well-defined. Since $A' \cap B' = \emptyset$, the function is also injective. Therefore, $\alpha + \beta \leq \alpha' + \beta'$. We give a longer list of results relating to the addition and multiplication of cardinals.

$$\alpha + 0 = \alpha. \qquad 1 \cdot \alpha = \alpha. \qquad \gamma(\alpha + \beta) = \gamma\alpha + \gamma\beta.$$

$$\text{If } \alpha \leq \alpha' \text{ and } \beta \leq \beta', \text{ then } \alpha + \beta \leq \alpha' + \beta' \text{ and } \alpha\beta \leq \alpha'\beta'.$$

$$\alpha\beta = 0 \text{ if and only if } \alpha = 0 \text{ or } \beta = 0.$$

Exercise 21. Prove all of the above assertions that we have not proven already.

We can also define the sum and product of arbitrarily many cardinal numbers $\alpha_i = \#(A_i)$, indexed by $i \in I$:

$$\sum_{i \in I} \alpha_i = \#\left(\bigcup_{i \in I} A_i \times \{i\}\right) \quad \text{and} \quad \prod_{i \in I} \alpha_i = \#\left(\prod_{i \in I} A_i\right).$$

Recall how the Axiom of Choice was formulated at the end of §2. We can now reformulate this as follows: if $\alpha_i \neq 0$ for all $i \in I$, then $\prod_{i \in I} \alpha_i \neq 0$. Intuitively, this seems like a reasonable property of multiplication.

Exercise 22? Consider two sets of cardinals $\{\alpha_i : i \in I\}$ and $\{\beta_j : j \in J\}$. Suppose the function $f : I \to J$ is an injection, with the property that $\alpha_i \leq \beta_{f(i)}$ for all $i \in I$. Let γ be a cardinal and prove the following

$$\sum_{i \in I} \alpha_i \leq \sum_{j \in J} \beta_j, \qquad \prod_{i \in I} \alpha_i \leq \prod_{j \in J} \beta_j, \qquad \gamma \cdot \sum_{i \in I} \alpha_i = \sum_{i \in I} \gamma\alpha_i.$$

Next, we define the operation of exponentiation. Given two sets A and B, we write B^A for the set of all functions $f : A \to B$. If we write $\alpha = \#(A)$ and $\beta = \#(B)$, then the exponent is defined to be $\beta^\alpha = \#(B^A)$. At first glance, this definition may seem rather strange. The rationale is again motivated by the finite case. For some positive integers m and n, we write $m^n = m \times \cdots \times m$ multiplied n times. For infinite cardinals, we similarly can see that multiplication is iterated addition and exponentiation is iterated multiplication.

Exercise 23! Consider cardinal numbers α and β. Suppose that $\alpha = \#(A)$. Show that

$$\sum_{a \in A} \beta = \alpha \times \beta \qquad \text{and} \qquad \prod_{a \in A} \beta = \beta^\alpha.$$

Recall that exponents are also called powers. We can now explain why $\mathcal{P}(A)$ is called the power set of A. For any set A, we can biject the sets $\mathcal{P}(A)$ and 2^A as follows. Given $f : A \to \{0, 1\}$, we take $f^{-1}(1) \in \mathcal{P}(A)$. Given $B \in \mathcal{P}(A)$, we define a function $f : A \to \{0, 1\}$ by setting $f(b) = 1$ if $b \in B$ and $f(b) = 0$ if $b \notin B$.

Exercise 24. Show that these operations are inverses $\mathcal{P}(A) \longleftrightarrow 2^A$. For any cardinal α, show that $\alpha < 2^\alpha$.

Again, we have not confirmed that exponentiation and aribtrary addition and multiplication are well-defined, but we leave this as an exercise for the curious reader. Instead, we move on to properties of exponentiation:

$$\alpha^0 = 1^\alpha = 1. \qquad \text{If } \alpha \geq 1, \text{ then } 0^\alpha = 0. \qquad \alpha^1 = \alpha.$$

$$\gamma^{\alpha+\beta} = \gamma^\alpha \gamma^\beta. \qquad (\alpha\beta)^\gamma = \alpha^\gamma \beta^\gamma. \qquad \gamma^{\beta\alpha} = (\gamma^\beta)^\alpha.$$

$$\text{If } \alpha \leq \beta \text{ and } \gamma \geq 1, \text{ then } \gamma^\alpha \leq \gamma^\beta. \qquad \text{If } \alpha \leq \beta, \text{ then } \alpha^\gamma \leq \beta^\gamma.$$

We will show that $\gamma^{\beta\alpha} = (\gamma^\beta)^\alpha$. Consider sets A, B, C. We define functions $(C^B)^A \leftrightarrow C^{A\times B}$ as follows:

$$\text{For any } f : A \to C^B, \text{ we define } \Phi(f) : A \times B \to C \text{ by } \Phi(f)(a, b) = f(a)(b).$$
$$\text{For any } g : A \times B \to C, \text{ we define } \Psi(g) : A \to C^B \text{ by } \Psi(g)(a)(b) = g(a, b).$$

This defines functions $\Phi : (C^B)^A \to C^{A\times B}$ and $\Psi : C^{A\times B} \to (C^B)^A$. At first, the mess of parentheses looks quite cryptic, so let's go through it step-by-step. The function Φ takes inputs from $(C^B)^A$, which consists of functions $f : A \to C^B$. The range of Φ is $C^{A\times B}$, so we should have $\Phi(f) : A \times B \to C$. Given an input $(a, b) \in A \times B$, what is the output $\Phi(f)(a, b)$? We get a function $f(a) : B \to C$, since C^B is the range of f. Applying this function to b, we get an output $f(a)(b) \in C$. This is what is meant by $\Phi(f)(a, b) = f(a)(b)$. Going through the definitions, we can similarly parse what is meant by $\Psi(g)(a)(b) = g(a, b)$. Next, we wish to show that Φ is a bijective function. By Exercise 13, it suffices to show that $\Psi \circ \Phi(f) = f$ and $\Phi \circ \Psi(g) = g$.

First consider $g : A \times B \to C$ and let $f = \Psi(g)$. Then we have $\Phi(f)(a, b) = f(a)(b) = \Psi(g)(a)(b) = g(a, b)$. Since g and $\Phi(f) : A \times B \to C$ have the same output for every $(a, b) \in A \times B$, we have $g = \Phi(f) = \Phi \circ \Psi(g)$.

Now consider $f : A \to C^B$ and let $g = \Phi(f)$. Then we have $\Psi(g)(a)(b) = g(a, b) = \Phi(f)(a, b) = f(a)(b)$. Since $f(a) : B \to C$ and $\Psi(g)(a) : B \to C$ have the same output for every $b \in B$, we have $f(a) = \Psi(g)(a)$. Since $f : A \to B^C$ and $\Psi(g) : A \to C^B$ have the same output for every $a \in A$, we have $f = \Psi(g) = \Psi \circ \Phi(f)$.

Exercise 25? Prove the rest of the above properties of exponentiation.

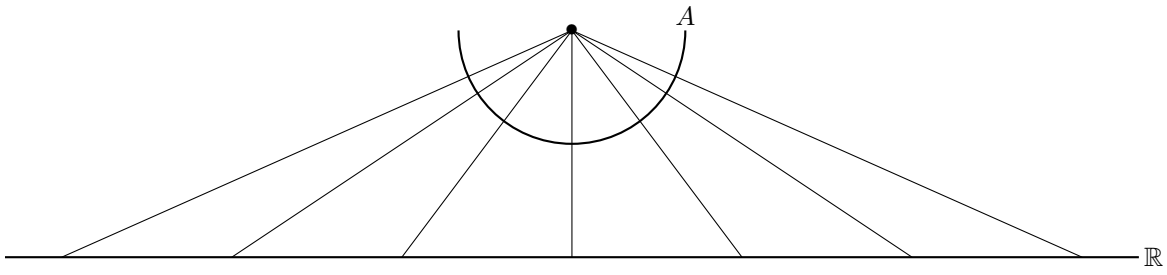
Exercise 26. Prove that cardinal exponentiation and arbitrary addition and multiplication are well-defined.

We are now in a position to prove that $\#(\mathbb{R}) = 2^{\aleph_0} > \aleph_0 = \#(\mathbb{N})$. First, consider some finite number $k \geq 2$. In §3, we proved that $\aleph_0 \cdot \aleph_0 = \aleph_0$. For any finite $n \geq 1$, we then have $\aleph_0 \leq n\aleph_0 \leq \aleph_0 \cdot \aleph_0 = \aleph_0$, so $n\aleph_0 = \aleph_0$. Choose $n \geq 1$ large enough that $k \leq 2^n$. Then $2^{\aleph_0} \leq k^{\aleph_0} \leq \left(2^n\right)^{\aleph_0} = 2^{n\aleph_0} = 2^{\aleph_0}$ and therefore $2^{\aleph_0} = k^{\aleph_0}$. Define $A = \{s \in \mathbb{R} : 0 \leq s \leq 1\}$. We will first prove that $\#(A) = 10^{\aleph_0}$ and then prove that $\#(\mathbb{R}) = \#(A)$. Together, these assertions will prove that $\#(\mathbb{R}) = 10^{\aleph_0} = 2^{\aleph_0}$ (we could also circumvent 10 by using binary).

For any natural number $n > 0$, we write n to denote the set $\{0, 1, \ldots, n-1\}$ (as well as the size of this set). Every number $s \in A$ is represented by a decimal expansion $0.a_0a_1a_2\ldots$ with $a_k \in \{0, 1, \ldots, 9\}$ for all $k \in \mathbb{N}$. This sequence is equivalent to a function $\mathbb{N} \to \{0, 1, \ldots, 9\}$. Therefore, we have a surjection $F : 10^{\mathbb{N}} \to A$, which takes the sequence $(a_0, a_1, \ldots) \in 10^{\mathbb{N}}$ to the number $0.a_0a_1\ldots \in A$. But there is a problem: not every real number has a *unique* decimal representation, e.g. $0.999\ldots = 1$. Given some $a_0, \ldots, a_n \in \{0, 1, \ldots, 9\}$ with $a_n > 0$, we have $F(a_0, \ldots, a_n, 0, 0, 0, \ldots) = F(a_0, \ldots, a_n - 1, 9, 9, 9, \ldots)$. Hence, if the number $0 < s < 1$ has a terminating decimal expansion, then $F^{-1}(s)$ contains two elements. Otherwise, the set $F^{-1}(s)$ contains one element. For each integer $n > 0$, let $B_n = \{s \in A : s = 0.a_1a_2\ldots a_n \text{ with } a_n \neq 0\}$. Let $B = \bigcup_{n>0} B_n$. Note that each B_n is finite and thus $\#(B) = \sum_{n>0} \#(B_n) \leq \sum_{n>0} \aleph_0 = \aleph_0 \cdot \aleph_0 = \aleph_0$. But B is infinite, since $10^{-n} \in B$ for all $n > 0$. Thus $\aleph_0 \leq \#(B) \leq \aleph_0$ and so $\#(B) = \aleph_0$. If we write $\alpha = \#(A - B)$, then

$$10^{\aleph_0} = \#\big(10^{\mathbb{N}}\big) = \#(2 \times B) + \#(A - B) = 2\aleph_0 + \alpha = \aleph_0 + \alpha = \#(B) + \#(A - B) = \#(A).$$

It remains to show that $\#(A) = \#(\mathbb{R})$. Since $A \subset \mathbb{R}$, we clearly have $\#(A) \leq \#(\mathbb{R})$. To prove $\#(\mathbb{R}) \leq \#(A)$, we will geometrically establish an injection $f : \mathbb{R} \to A$. Draw the segment A as a semi-circle with $\mathbb{R}$ below.



For any $s \in \mathbb{R}$, we draw a line segment from s to the center of the semicircle. This intersects the semicircle in a unique point, which we call $f(s)$. To see that this is an injection, note that we can recover s from $f(s)$: if we draw the line going through $f(s)$ and the center of the circle, then s is where this line intersects $\mathbb{R}$.

Exercise 27? Why does this construction give an injection, but not a bijection?

5 Choice Revisited

We are now ready to revisit the Axiom of Choice (AC) and use it prove some interesting facts about cardinals. Exercise 15 showed that AC is equivalent to the fact that every surjection has a right inverse. There are many other equivalent statements, some of which we will prove. First, we will revisit the notions of partial, total and well-orders. Let $\leq$ be a partial order on A and consider a subset $B \subset A$. We define the following:

(a) A *maximal element* of B is an $m \in B$, such that if $m \geq b$ and $b \in B$, then $m = b$. In other words, there does not exist $b \in B$ with $m < b$. A minimal (or least) element of B is defined analogously.

(b) An *upper bound* for B is some $u \in A$ (not necessarily in B), such that $u \geq b$ for all $b \in B$. As the name suggests, a *least upper bound* for B is a minimal element among the set of upper bounds of B.

Notice that maximal elements of B need not be upper bounds, nor conversely. For example, consider the set $A = \mathcal{P}\big(\{1,2,3,4\}\big) - \big\{\{1,2\}\big\}$ with the partial order of subsets. Define $B = \big\{\{1\},\{2\}\big\} \subset A$. Both elements of B are maximal. The least upper bounds of B are $\{1,2,3\}$ and $\{1,2,4\}$ (since we ensured that $\{1,2\} \notin A$).

For any partial order $(A, \leq)$, we will say that a subset $B \subset A$ is a *chain* if $\leq$ restricts to a well-order on B. If $B \subset A$ is a chain with upper bound $u \notin B$, notice that $B \cup \{u\}$ is also a chain. For two chains $B \subset C \subset A$, we say that B is an *initial segment* of C if $B = C$ or if there exists some $c_0 \in C$ with $B = \{c \in C : c < c_0\}$.

Our first result, *Zorn's Lemma*, gives a sufficient condition for maximal elements to exist. The intuitive idea is to keep choosing larger and larger elements, until we get to a maximal element. This requires the ability to "cap off" infinite chains and to "keep choosing" elements to add to a chain of arbitrarily large cardinality. However, doing this in practice requires a good deal of formal tools, which we do not have. This formalism can be skipped, but only with great care. The proof that follows is definitely hard, but the payoff is huge.

Lemma 1. Suppose that $(A, \leq)$ is well-ordered. If $B \subset A$ is an initial segment and $m \in A - B$ is minimal, then $B = \{a \in A : a < m\}$ and $B \cup \{m\}$ is also an initial segment. Initial segments are preserved by unions.

Proof. Since $B \subset A$ is an initial segment, there is some $b_0 \in A$ with $B = \{a \in A : a < b_0\}$. Then $b_0 \in A - B$, so $m \leq b_0$ by minimality. But $m \notin B$, so we must have $m \geq b_0$. Hence $m = b_0$ and $B = \{a \in A : a < m\}$. Thus we have $B \cup \{m\} = \{a \in A : a \leq m\}$. If $B \cup \{m\} = A$, then $B \cup \{m\}$ is trivially an initial segment. Otherwise, we have $A - \big(B \cup \{m\}\big) \neq \emptyset$, so we can choose a minimal $a_0 \in A - \big(B \cup \{m\}\big)$. Next, we define the initial segment $M = \{a \in A : a < a_0\}$. Since $m \in A - B$ is minimal and $a_0 \in A - \big(B \cup \{m\}\big) \subset A - B$, we have $a_0 > m$. Thus we have $B \cup \{m\} = \{a \in A : a \leq m\} \subset \{a \in A : a < a_0\} = M$. But if $a \notin B \cup \{m\}$, then $a \geq a_0$ by the minimality of $a_0 \in A - \big(B \cup \{m\}\big)$, so $a \notin M$. Thus $B \cup \{m\} = M$ is an initial segment.

Now consider a set $\{B_i \subset A : i \in I\}$ of initial segments and define $B = \bigcup_{i \in I} B_i$. If $B = A$, then B is trivially an initial segment. Otherwise, let $m \in A - B$ be minimal and define $M = \{a \in A : a < m\}$. For each $i \in I$, we define $b_i \in A - B_i$ to be minimal, so that $B_i = \{a \in A : a < b_i\}$. Since $B_i \subset B$, we have $A - B \subset A - B_i$ and thus $m \in A - B_i$. By the minimality of $b_i \in A - B_i$, it follows that $b_i \leq m$ for all $i \in I$. When $a \in B$, we have some $i \in I$ with $a \in B_i$ and it follows that $a < b_i \leq m$. Thus $B \subset M$. When $a \notin B$, we have $a \geq m$ by the minimality of $m \in A - B$ and it follows that $a \notin M$. Thus $B = M$ is an initial segment. □

Lemma 2. Let $(A, \leq)$ be a partially ordered set with chains $X, Y \subset A$. Then there exists a maximal subset $M \subset A$, such that M is an initial segment of both X and Y (this set M is unique, but we do not need this).

Proof. Let $\mathcal{C}$ be the set of $Z \subset A$ that are initial segments of both X and Y. Notice that $\emptyset \in \mathcal{C}$, so $\mathcal{C} \neq \emptyset$. Let $M = \bigcup_{Z \in \mathcal{C}} Z$. By Lemma 1, the set M is an initial segment of both X and Y. Note that M is maximal among the sets in $\mathcal{C}$, because $Z_0 \subset \bigcup_{Z \in \mathcal{C}} Z = M$ for any $Z_0 \in \mathcal{C}$. □

Lemma 3. Suppose that $(A, \leq)$ is a partially ordered set and $\mathcal{C} \subset \mathcal{P}(A)$ is a set of chains in A. For any elements $X, Y \in \mathcal{C}$, we suppose that either X is an initial segment of Y or Y is an initial segment of X. Then $B = \bigcup_{X \in \mathcal{C}} X$ is also chain. Moreover, every $X \in \mathcal{C}$ is an initial segment of B.

Proof. Let $B = \bigcup_{X \in \mathcal{C}} X$. For any $x, y \in B$, there exist $X, Y \in \mathcal{C}$ with $x \in X$ and $y \in Y$. By assumption, either X is an initial segment of Y or vice versa. In particular, we either have $X \subset Y$ or $Y \subset X$. If $X \subset Y$, then $x, y \in Y$ and Y is a chain, so x and y are comparable. If $Y \subset X$, then x and y are again comparable. Thus B is totally ordered by $\leq$. Now consider a nonempty subset $D \subset B$. If we choose some $x_0 \in D \subset B$, then there exists some $X_0 \in \mathcal{C}$ with $x_0 \in X_0$. Since X_0 is a chain, it is well-ordered, so there is a minimal element $m \in X_0 \cap D$. We contend that m is actually minimal in D. Fix $y \in D$ and choose $Y \in \mathcal{C}$ with $y \in Y$. If $y \in X_0$, then $y \in X_0 \cap D$, so we have $m \leq y$. When Y is an initial segment of X_0, we have $y \in Y \subset X_0$ and thus $m \leq y$. Otherwise, we know that X_0 is an initial segment of Y. When $X_0 = Y$, we have $y \in X_0$ and therefore $m \leq y$. If we have $Y - X_0 \neq \emptyset$, then Lemma 1 gives some $c_0 \in Y$ with $X_0 = \{x \in Y : x < c_0\}$. If $y < c_0$, then $y \in X_0$ and thus $m \leq y$. But if $y \geq c_0$, then we have $m < c_0 \leq y$ (because $m \in X_0$).

We have proven that T is a chain. Consider some $X_0 \in \mathcal{C}$. Let $t \in T - X_0$ be minimal and define the initial segment $M = \{a \in T : a < t\}$. Since $t \in T$, there exists some $X_1 \in \mathcal{C}$ with $t \in X_1$. Since $t \in X_1$ but $t \notin X_0$, we can see that X_1 is not an initial segment of X_0. Thus X_0 is an initial segment of X_1, so there is some $x_0 \in X_1$ such that $X_0 = \{a \in X_1 : a < x_0\}$. Since $t \notin X_0$, we have $t \geq x_0$. Therefore, we have the inclusion

$$X_0 = \{a \in X_1 : a < x_0\} \subset \{a \in T : a < x_0\} \subset \{a \in T : a < t\} = M.$$

If $a \in T - X_0$, then $a \geq t$ by the minimality of t, so $a \notin M$. Thus $X_0 = M$ is an initial segment of T. Thus there is some $x_0 \in X_1$ such that $X_0 = \{a \in X_1 : a < x_0\}$. □

Zorn's Lemma. (First proved by Kazimierz Kuratowski.) Let $\leq$ be a partial order on $A \neq \emptyset$, such that every chain $B \subset A$ has an upper bound. Then for any $p \in A$, there is a maximal element $m \in A$ with $m \geq p$.

Proof. Fix $p \in A$. Assume for the sake of contradiction that there is no maximal element $m \in A$ with $m \geq p$. Let $\mathcal{D}$ be the set of chains in A with least element p. Then $\mathcal{D}$ is nonempty, since $\{p\} \in \mathcal{D}$. For each $B \in \mathcal{D}$, there exists an upper bound b for the chain B. Since $b \geq p$, we have assumed that b cannot be maximal in A. Thus we can find some $u > b$. Then u is clearly an upper bound for B and $u \notin B$. Hence, for each $B \in \mathcal{D}$, we can define the nonempty set $U(B) = \{u \in A : u \notin B \text{ and } u \text{ is an upper bound for } B\}$. By applying AC, we can pick a choice function $\Phi : \mathcal{D} \to A$ with $\Phi(B) \in U(B)$ for all $B \in \mathcal{D}$. In picking this choice function, we can choose to have $\Phi(\emptyset) = \{p\}$. We define $\mathcal{C} \subset \mathcal{D}$ by setting $X \in \mathcal{C}$ if and only if:

$$\text{for every initial segment } S \subset X \text{ with } S \neq X, \text{ the least element of } X - S \text{ is } \Phi(S).$$

The least element of $X - S$ is well-defined because any $X \in \mathcal{D}$ is well-ordered. Now consider some $X, Y \in \mathcal{C}$. We contend that either X is an initial segment of Y or vice versa. By Lemma 2, there is a maximal $M \subset A$, such that M is an initial segment of both X and Y. First, we suppose that $X - M$ and $Y - M$ are non-empty. Since $X, Y \in \mathcal{C}$, it follows that $\Phi(M)$ is the least element of both $X - M$ and $Y - M$. But then $M \cup \{\Phi(M)\}$ is a bigger initial segment of both X and Y, contradicting the definition of M. Therefore, either $X - M = \emptyset$ or $Y - M = \emptyset$. Equivalently, we have $M = X$ or $M = Y$. But if $X = M$, then X is an initial segment of Y, while if $Y = M$, then Y is an initial segment of X. Thus $\mathcal{C}$ satisfies the hypothesis of Lemma 3, so the set

$T = \bigcup_{X \in \mathcal{C}} X$ is a chain in A. Notice that T is non-empty, because $\{p\} \in \mathcal{C}$ and thus $p \in T$. For any $x \in T$, there is some $X \in \mathcal{C} \subset \mathcal{D}$ with $x \in X$, so $x \geq p$ by definition of $\mathcal{D}$. Thus T is a chain with least element p.

Now consider an initial segment $S \subset T$ with $S \neq T$. Choose $m \in T - S$ to be minimal. Then by Lemma 1, we have $S = \{a \in T : a < m\}$. Because $m \in T$, there exists some $X_0 \in \mathcal{C}$ with $m \in X_0$. Then by Lemma 3, the set X_0 is an initial segment of T. If we have $X_0 = T$, then S is an initial segment of X_0. If not, we have $X_0 = \{a \in T : a < x_0\}$ for some $x_0 \in T$. Since $m \in X_0$, we have $a < m < x_0$ for all $a \in S$. It follows that

$$S = \{a \in T : a < m < x_0\} = \{a \in T : a < m \text{ and } a < x_0\} = \{a \in X_0 : a < m\}.$$

Thus S is an initial segment of X_0. If $S = X_0$, then $\Phi(X_0) = \Phi(S)$ is the least element of $T - X_0 = T - S$. If $T = X_0$, then $\Phi(S)$ is the least element of $X_0 - S = T - S$. Now suppose that $S \neq X_0 \neq T$. Since $X_0 \in \mathcal{C}$, we know that $\Phi(S)$ is the least element of $X_0 - S$. Suppose that $a \in T - S$. If $a \in X_0 - S$, then $a \geq \Phi(S)$. But $\Phi(S) \in X_0 = \{a \in T : a < x_0\}$, so if $a \notin X_0$, then $a \geq x_0 > \Phi(S)$. Thus $\Phi(S)$ is also minimal in $T - S$. This proves that $T \in \mathcal{C}$. But then $\Phi(T)$ is an upper bound for T and $\Phi(T) \notin T$, so $T \cup \{\Phi(T)\}$ is a larger set in $\mathcal{C}$. This is a contradiction, since $\Phi(T) \notin T = \bigcup_{X \in \mathcal{C}} X$, so a maximal element $m \geq p$ does exist! □

Now that we have finished the grueling proof of Zorn's lemma, the remaining theorems will go much smoother.

Well-Ordering Theorem. Any set A admits a well-order.

Proof. Consider the set $\mathcal{W}$ whose elements are pairs $(B, \leq)$, such that $B \subset A$ and $\leq$ is a well-order on B. We give $\mathcal{W}$ a partial ordering $\subset$ by saying that $(B_1, \leq_1) \subset (B_2, \leq_2)$ if $B_1 \subset B_2$, the order $\leq_2$ restricts to $\leq_1$ on the set B_1, and B_1 is an initial segment of B_2. Fix a chain $\mathcal{C} = \{(B_i, \leq_i) : i \in I\} \subset \mathcal{W}$. Let $B = \bigcup_{i \in I} B_i$. We define a partial order $\leq$ on B as follows. Suppose that $x \in B_i$ and $y \in B_j$ with $(B_i, \leq_i) \subset (B_j, \leq_j)$. Then $x \in B_i \subset B_j$, so we will say that $x \leq y$ if and only if $x \leq_j y$ (and vice versa). The reader may check that this is indeed a partial order and that the definition is independent of i and j. Thus B is a partially ordered set and $\mathcal{C}$ is a set of chains $B_i \subset B$, satisfying the hypothesis of Lemma 3 (because $\leq$ restricts to $\leq_i$ on B_i for each $i \in I$). Therefore $B = \bigcup_{i \in I} B_i$ is a chain in itself, so we have constructed an upper bound B for the chain $\mathcal{C}$. By Zorn's Lemma, there is a maximal element $(M, \leq)$ of $\mathcal{W}$. If $A - M \neq \emptyset$, then we choose some $a \in A - M$ and define a well-order on $M \cup \{a\}$ by setting $a \geq m$ for all $m \in M$. However, this is in $\mathcal{W}$ and $M \subset M \cup \{a\}$, so this contradicts the maximality of M. Hence, we must have $A = M$. □

Exercise 28! Prove that the relation $\leq$ defined on B is a well-defined partial order.

In fact, Zorn's Lemma and the Well-Ordering theorem are not only consequences of AC, but equivalents. We have now shown that AC $\Longrightarrow$ Zorn $\Longrightarrow$ Well-Ordering. To show that all three assertions are equivalent, it remains to show that Well-Ordering $\Longrightarrow$ AC. Consider a set $\{A_i : i \in I\}$ of non-empty sets. If we assume the Well-Ordering Theorem, then we can find a well-order on $A = \bigcup_{i \in I} A_i$. For each $i \in I$, we can define the minimal element $\min(i) \in A_i \subset A$, which exists because A is well-ordered set. Then min is a choice function.

Now, we can apply these AC-equivalents to prove results about cardinal arithmetic. In fact, some of these results are also equivalent to AC, although we will not prove this. Our first result generalizes Exercise 17. (Note that there is no "set of all cardinal numbers," so the concept of a relation on cardinals is imprecise.)

Proposition. The cardinal numbers are totally ordered.

Proof. We know that the cardinal numbers are partially ordered, so we just need to prove comparability. Consider two sets A and B. If $B = \emptyset$, then we clearly have $\#(B) \leq \#(A)$. Thus we can assume that $B \neq \emptyset$. Let $\mathcal{F}$ be the set of all pairs (C, f), where $C \subset A$ and $f : C \to B$ is an injection. We let $(C_1, f_1) \subset (C_2, f_2)$ if $C_1 \subset C_2$ and $f_1(c) = f_2(c)$ for all $c \in C_1$. The reader may check that this is a partial order. Now suppose that $\mathcal{C} \subset \mathcal{F}$ is a chain. Then we may define $D = \bigcup_{X \in \mathcal{C}} X$. Define $f : D \to B$ as follows. If $d \in D$, then there is some $(C_1, f_1) \in \mathcal{C}$ with $d \in C_1$, so we set $f(d) = f_1(d)$. Since $\mathcal{C}$ is a chain, this definition is independent of the choice of f_1. Now suppose $a, b \in D$ and $f(a) = f(b)$. Then there are some $(C_1, f_1), (C_2, f_2) \in \mathcal{C}$ with $a \in C_1$ and $b \in C_2$. Since $\mathcal{C}$ is a chain, we have $(C_1, f_1) \subset (C_2, f_2)$ or $(C_2, f_2) \subset (C_1, f_1)$. Assume the former. Then $a \in C_1 \subset C_2$ and $f_2(a) = f(a) = f(b) = f_2(b)$. But f_2 is injective, so we then have $a = b$. This proves that $(D, f) \in \mathcal{F}$, so we have found an upper bound for $\mathcal{C}$. Now we apply Zorn's lemma to find a maximal element $(M, f_0) \in \mathcal{F}$. If $M = A$, then $f_0 : A \to B$ is an injection, so $\#(A) \leq \#(B)$. If $M \neq A$, we claim that $f_0 : M \to B$ is surjective. For otherwise, we could map some $a \in A - M$ to some $b \in B - f(A)$, contradicting the maximality of M. But then $\#(B) \leq \#(M) \leq \#(A)$. Therefore $\#(A)$ and $\#(B)$ are comparable. □

Exercise 29! Prove that the function $f : D \to B$ is well-defined.

Proposition. The cardinal numbers are well-ordered.

Proof. This proof follows [4]. Consider a set $\{\alpha_i : i \in I\}$ of cardinal numbers. For each $i \in I$, we may choose a set A_i with $\alpha_i = \#(A_i)$. Let $A = \prod_{i \in I} A_i$. Let $\mathcal{B}$ be the set of all $B \subset A$ satisfying the following property:

$$\text{For any } f, g \in B, \text{ if } f(j) = g(j) \text{ for some } j \in I, \text{ then } f = g.$$

We give $\mathcal{B}$ the partial order $\subset$ of subset inclusion. Let $\mathcal{C} \subset \mathcal{B}$ be a chain and define $M = \bigcup_{C \in \mathcal{C}} C$. The reader may check that $M \in \mathcal{B}$, by similar means to those used above. Hence, every chain in $\mathcal{B}$ has an upper bound, so there exists a maximal $B \in \mathcal{B}$. For each $i \in I$, we define $\pi_i : A \to A_i$ by $\pi_i(f) = f(i)$. Assume for the sake of contradiction that $\pi_i(B) \neq A_i$ for all $i \in I$. Then $A_i - \pi_i(B) \neq \emptyset$ for all $i \in I$, so there exists some choice function $f : I \to \bigcup_{i \in I} A_i$ with $f(i) \notin \pi_i(B)$ for all $i \in I$. Now suppose that $g \in B$. For each $i \in I$, we have $g(i) = \pi_i(g) \in \pi_i(B)$, so $g(i) \neq f(i)$. Thus $B \cup \{m\}$ satisfies the defining property of $\mathcal{B}$, which contradicts the maximality of $B \in \mathcal{B}$. Hence, there must be some $i_0 \in I$ with $\pi_{i_0}(B) = A_{i_0}$. Since $B \in \mathcal{B}$, we can see that if $j \in I$ and $f, g \in B$ are such that $\pi_j(f) = \pi_j(g)$, then $f(j) = g(j)$ and thus $f = g$. Therefore $\pi_j : B \to A_j$ is injective for each $j \in I$. In particular, the function $\pi_{i_0} : B \to A_{i_0}$ is a bijection, so it has an inverse $\pi_{i_0}^{-1}$. For any $j \in I$, it follows that $\pi_j \circ \pi_{i_0}^{-1} : A_{i_0} \to A_j$ is an injection. Thus $\alpha_{i_0} \leq \alpha_j$ for all $j \in I$. □

Exercise 30! Prove that we indeed have $M \in \mathcal{B}$.

Proposition. For any infinite cardinal α, we have $\alpha^2 = \alpha$.

Proof. Unfortunately, I did not have time to prove this result, but the references contain proofs. □

Corollary. If $\alpha, \beta > 0$ are cardinals, at least one of which is infinite, then $\alpha + \beta = \alpha\beta = \max\{\alpha, \beta\}$.

Proof. Suppose $\alpha \leq \beta$ and β is infinite. Then $\beta \leq \alpha + \beta \leq \beta + \beta = 2\beta \leq \beta^2 = \beta$ and $\beta \leq \alpha\beta \leq \beta^2 = \beta$. □

Alfred Tarski actually showed that the Axiom of Choice is equivalent to the assertion that $\alpha^2 = \alpha$ for all infinite cardinals α. But he claims that his proof we rejected by famous mathematicians on the editing board of the French journal *Comptes Rendus* (recounted in [6] be Jan Mycielski). The editors responded as follows:

> "Fréchet wrote that an implication between two well known propositions is not a new result."
>
> "Lebesgue wrote that an implication between two false propositions is of no interest."

This strong opposition of opinions was emblematic of the time, when some mathematicians embraced Choice, while others found it repugnant. In that same time period, L. E. J. Brouwer went so far as to retract his own proofs usimg the Axiom of Choice. The philosophical question seemed to be "which axioms are correct?" But in 1931, the Austrian mathematician Kurt Gödel revolutionized foundational mathematics. He showed that any system of logic (strong enough to do basic arithmetic) would contains statements in a sort of limbo: not false, but not provable. He also proved that a system of logic could not prove its own consistency. Thus, if the universe did favor some "correct version" of set theory, this theory would not be able to prove itself not to be contradictory. The works of Gödel and Cohen later showed that AC is independent of classical set theory, meaning that there exist models of set theory in which AC holds and other models in which it fails. In the present day, most mathematicians accept classical set theory and the Axiom of Choice. But for many in the various fields of foundational mathematics, their work centers on comparing and contrasting different frameworks of set theory and logic. The modern "right answer" is not centered on finding the omnipotent lists of "correct" axioms, but rather on maintaining the omnispective viewpoint of an open mind.

References

[1] Enderton, H. B. (1977). *Elements of set theory.* San Diego, CA: Academic Press.

[2] Gödel, K. (1931). *On formally undecidable propositions of Principia Mathematica and related systems.*

[3] Halmos, P. R. (1960). *Naive set theory.* Princeton, NJ: van Nostrand Company.

[4] Hönig, C. S. (1954). Proof of the well-ordering of cardinal numbers. *Proceedings of the AMS.*

[5] Kamke, E. (1950). *Theory of sets* (F. Bagemihl, Trans.). New York, NY: Dover.

[6] Mycielski, J. (2006). A System of axioms of set theory for the rationalists. *Notices of the AMS.*

[7] Stoll, R. R. (1963). *Set theory and logic.* New York, NY: Dover.