MATH 320 SET THEORY

BURAK KAYA

Abstract. These are the lecture notes I used for a 14-week introductory set theory class I taught at the Department of Mathematics of Middle East Technical University during Spring 2018. In order to determine the course content and prepare the lecture notes, I mainly used the textbook by Hrbacek and Jech [1] which I also listed as a supplementary resource for the course.

Contents

0. Prelude

0.1. **Some historical remarks.** If one examines the history of mathematics, one sees that towards the end of 19^{th}century, some mathematicians started to investigate the "nature" of mathematical objects. For example, Dedekind gave a construction for the real numbers, Peano axiomatized the natural numbers, Cantor established a rigorous way to deal with the notion of infinity. These works may be considered as first steps to understand *what* mathematical objects are.

In early 20^{th}century, arose what is known as *the foundational crisis of mathematics*. Mathematicians searched for a proper foundations of mathematics which is free of contradictions and is sufficient to carry out all traditional mathematical reasoning. There were several philosophical schools having different views on how mathematics should be done and what mathematical objects are. Among these philosophical schools, the leading one was Hilbert's formalist approach, according to which mathematics is simply an activity carried out in some formal system[1]. On the one hand, mathematics had already been done "axiomatically" since Euclid. On the other hand, Hilbert wanted to provide a **rigorous** axiomatic foundation to mathematics[2]. With the work of Dedekind and Cantor, the idea that mathematics can be founded on set theory became more common. This eventually led[3] to the development of the **Z**ermelo-**F**raenkel set theory with the axiom of **C**hoice, by Ernst Zermelo, with the later contributions of Abraham Fraenkel, Thoralf Skolem and John von Neumann.

Today, some mathematicians consider ZFC as *the* foundation of mathematics, in which one can formalize virtually all known mathematical reasoning. In this course, we aim to study the axioms of ZFC and investigate their consequences. That said, we should note there are many other set theories with different strengths introduced for various purposes, such as von Neumann-Gödel-Bernays set theory, Morse-Kelley set theory, New Foundations, Kripke-Platek set theory and the Elementary Theory of the Category of Sets.

0.2. **The language of set theory and well-formed formulas.** We shall work in first-order logic with equality symbol whose language consists of a single binary relation symbol $\in$. For those who are not familiar with first-order logic, we first review how the well-formed formulas in the language of set theory are constructed. Our basic symbols consist of the symbols

$$\in \; = \; \forall \; \exists \; \neg \; \wedge \; \vee \; \rightarrow \; \leftrightarrow \; (\;)$$

together with an infinite supply of variable symbols

$$a \; b \; c \; d \; e \; \ldots$$

The *(well-formed) formulas* in the language of set theory are those strings that can be obtained in finite numbers of steps by application of the following rules.

- Strings of the form $x \in y$ and $x = y$, where x and y are variable symbols, are formulas.

[1]To illustrate this point,we should perhaps remind Hilbert's famous saying: "Mathematics is a game played according to certain simple rules with meaningless marks on paper."

[2]In fact, Hilbert wanted more than this. Those who wish to learn more should google the term "Hilbert's program".

[3]We refer reader to the web page https://plato.stanford.edu/entries/settheory-early/ for a detailed and more accurate historical description.

- If φ and ψ are formulas and x is any variable symbol, then the following strings are formulas

$$\neg\varphi \quad \exists x\varphi \quad \forall x\varphi \quad (\varphi \wedge \psi) \quad (\varphi \vee \psi) \quad (\varphi \to \psi) \quad (\varphi \leftrightarrow \psi)$$

For example, the string $\exists x \forall y \neg y \in x$ is a well-formed formula in the language of set theory, whereas, the string $\exists x \forall \neg x \to \exists \vee$ is not. A variable in a formula is said to be *bound* if it is in the scope of a quantifier; otherwise, it is said to be *free*. A formula with no free variables is called a *sentence*. For example, the string $\exists x \forall y \neg y \in x$ is a sentence, and the string $\exists z \exists t((\neg z = t \wedge z \in x) \wedge t \in y)$ is a formula with two free variables x and y and hence not a sentence.

"Officially", we work in an axiomatic system that consists of the axioms of ZFC and the standard logical axioms (in the language of set theory) together with a sound and complete proof system[4]. "Unofficially", we are going to work in natural language and carry out our mathematical arguments informally, as is the case in any other branch of mathematics. Nevertheless, if necessary, the reader should be able to convert arguments in natural language to formal proofs in first-order logic and vice versa.

0.3. **What are sets anyway?** Up to this point, we have not mentioned anything related to the *meaning* of the formulas in the language of set theory. For example, what does $x \in y$ really *mean*?

On the one hand, we note that it is perfectly possible to take a purely formalist approach and simply derive theorems in the aforementioned axiomatic system with attaching no meaning to symbols. On the other hand, we believe that this approach is pedagogically inappropriate for students who are exposed to set theory for the first time; and that it fails to acknowledge the role of mathematical intuition, which not only manipulates symbols but also understands what they *refer to*. Consequently, we shall adopt a Platonist point of view that we think is better-suited for teaching purposes[5]. Back to the question... What does $x \in y$ really *mean*?

A long time ago in a galaxy far, far away.... existed the universe of mathematical objects called *sets* which is denoted by $\mathbf{V}$. We shall not try to define what a set is. You should think of sets as primitive objects, perhaps by comparing it to points of Euclid's Elements. Sets are to us like points are to Euclid. Sets are simply the objects in the universe of sets.

Between certain sets holds the *membership* relation which we denote by $x \in y$. Our intuitive interpretation of the relation $\in$ is that $x \in y$ holds if the set y contains the set x as its element. In this sense, sets are objects that contain certain other sets as their members.

Quantifiers ranging over the universe of sets and logical connectives having their usual intended meanings, a sentence in the language of set theory is simply an assertion about the universe of sets that is either true or false, depending on how the membership relation holds between sets.

---

[4]Details of our proof system are not really relevant for this course, since most of our arguments are going to be done informally. Moreover, there are many (essentially equivalent) proof systems that are sufficient for our purposes. Those students who wish to learn how a sound and complete proof system for first-order logic may be set up should google the term *Hilbert(-style) proof system*.

[5]However, I personally do not consider myself as a follower of mathematical Platonism.

We assume that the axioms of ZFC are true sentences about the universe of sets, whose truth is self-evident and dictated by our mathematical intuition[6]. In this course, we shall study the logical consequences of the axioms of ZFC and try to understand the structure of the universe of sets $\mathbf{V}$.

0.4. **Classes vs. Sets.** A *class* is simply a collection of sets and hence is a subcollection of the universe of sets. We remark that classes are not (necessarily) objects in the universe of sets according to this definition. Consequently, we cannot directly talk about them in our axiomatic system by referring to them via variable symbols[7]. However, there is a way to get around this problem and make assertions about classes in a meaningful manner.

Let $\varphi(x)$ be a *property* of sets, i.e. a formula in the language of set theory with one free variable. The collection C of sets satisfying the formula $\varphi(x)$ is a class and is denoted by

$$\{x : \varphi(x)\}$$

In this case, the class C is said to be *defined* by the formula $\varphi(x)$. We also allow multiple free variables to appear in the defining formula, in which case the class

$$\{x : \psi(x, p, q, \ldots, t)\}$$

is said to be defined by ψ with parameters $p, q, \ldots, t$, where $p, q, \ldots, t$ are fixed sets.

For the rest of this course, we shall restrict our attention to those classes that are defined by some formula in the language of set theory possibly via some parameters. As such, we can meaningfully make assertions about classes in our axiomatic system by identifying formulas with the corresponding classes. For example, if C and D are classes that are defined by the formulas $\varphi(x)$ and $\psi(x)$ respectively, then the assertion $C = D$ can be stated by the sentence $\forall x(\varphi(x) \leftrightarrow \psi(x))$. We can also "quantify" over a class C defined by the formula $\varphi(x)$ using the formulas

$$\forall x(\varphi(x) \to \psi) \text{ and } \exists x(\varphi(x) \wedge \psi)$$

which would intuitively correspond to $\forall x \in C\ \psi$ and $\exists x \in C\ \psi$ respectively if we could have quantified over the classes in the first place. One can similarly define quantification over classes defined via parameters.

It is clear that every set, being a collection of sets, is a class. More precisely, given a set x, we can simply define it by the formula $y \in x$ using the set x itself as a parameter, i.e. $x = \{y : y \in x\}$. On the other hand, not every class is a set.

Theorem 1 (Russell's paradox)**.** *The class $R = \{x : \neg x \in x\}$ is not a set. More precisely,*

$$\neg\exists x \forall y(y \in x \leftrightarrow \neg y \in y)$$

Proof. Assume to the contrary that there exists x such that $\forall y(y \in x \leftrightarrow \neg y \in y)$. Then, letting y be the set x, we have $\neg x \in x \leftrightarrow x \in x$, which is a contradiction. □

Classes that are not sets are called *proper classes*. For example, the class R defined above is a proper class. As we shall see later, another example of a proper class is the universe of sets $\mathbf{V}$ which can be defined by the formula $x = x$.

---

[6]Those students with philosophical tendencies may read Penelope Maddy's famous articles *Believing the Axioms, I*, *Believing the Axioms, II* and her book *Defending the Axioms* after completing this course.

[7]We note that some of the set theories we mentioned earlier are capable of talking about classes directly. For example, this can be done in NBG and MK.

0.5. **Notational remarks.** In what follows, our assertions about sets should ideally be written in the language of set theory, having only $\in$ as a non-logical symbol. However, this approach is cumbersome and for convenience we will often expand our language by introducing new non-logical symbols that are abbreviations for certain formulas of set theory. For example, the formula $\neg x \in y$ is abbreviated as $x \notin y$. The reader is expected to keep track of introductions of such abbreviations.

Another notational convenience we shall adopt is to write $\forall z \in x\ \varphi$ instead of $\forall z(z \in x \to \varphi)$ and to write $\exists z \in x\ \varphi$ instead of $\exists z(z \in x \wedge \varphi)$ where φ is a formula in the language of set theory. Finally, we note that parentheses are usually omitted whenever there is no ambiguity.

1. Some axioms of ZFC and their elementary consequences

1.1. **And G said, "Let there be sets"; and there were sets.** We begin our discussion by introducing the axiom which asserts that the universe of sets is not void.

Axiom 1 (The axiom of empty set)**.** *There exists a set with no elements.*

$$\exists x \forall y \; y \notin x$$

A set with no elements will be referred to as an *empty set*. One can ask whether there may be more than one empty set. Unfortunately, we cannot answer this question without additional axioms.

Intuitively speaking, the only feature of sets is to contain certain other sets. Thus one may argue that a set should be completely determined by its elements. This suggests the following axiom.

Axiom 2 (The axiom of extensionality)**.** *Two sets are equal if and only if they have the same elements.*

$$\forall x \forall y \; (x = y \leftrightarrow \forall z (z \in x \leftrightarrow z \in y))$$

Now we are in a position to prove our first theorem in set theory.

Theorem 2. *There exists a unique set with no elements.*

Proof. Assume to the contrary that x and y are sets with no elements such that $x \neq y$. Then, by the axiom of extensionality, there exists z such that either that $z \in x$ and $z \notin y$, or that $z \notin x$ and $z \in y$. In both cases, we have a contradiction since x and y have no elements. □

From now on, the (unique) empty set with no elements will be denoted by $\emptyset$. At this point, we cannot prove the existence of sets other than the empty set without further axioms. For all we know, the universe of sets could consist of only the empty set.

1.2. **Constructing more sets.** In this section, we introduce several axioms that enable us to define some elementary operations on the universe of sets and construct sets other than the empty set.

Axiom 3 (The axiom of pairing)**.** *For any sets x and y, there exists a set z which consists of the elements x and y.*

$$\forall x \forall y \exists z \forall t (t \in z \leftrightarrow (t = x \vee t = y))$$

In other words, for any sets x and y, the collection $\{x, y\}$ is indeed a set. We shall call this set the *unordered pair* of x and y. Here are two applications of the axiom of pairing.

- By pairing $\emptyset$ with itself, we can now prove that the set $\{\emptyset\}$ exists.
- By pairing the set $\{\emptyset\}$ with $\emptyset$, we can also construct the set $\{\emptyset, \{\emptyset\}\}$.

Next follows an important application of the axiom of pairing. Let x and y be sets. Then, by the axiom of pairing, the sets $\{x\}$ and $\{x, y\}$ both exist. By pairing these sets, we obtain the set $\{\{x\}, \{x, y\}\}$.

Definition 1 (Kuratowski)**.** *The set $\{\{x\}, \{x, y\}\}$ is called the ordered pair of x and y and is denoted by (x, y).*

The reason (x, y) is called the *ordered* pair is easily seen from the next lemma.

Lemma 1. *Let x, y, x', y' be sets. $(x, y) = (x', y')$ if and only if $x = x'$ and $y = y'$.*

Proof. Left to the reader as an exercise. □

We next introduce an axiom that allows us to collect the elements of elements of a set into a single set.

Axiom 4 (The axiom of union)**.** *For any set x, there exists a set y which consists of exactly the elements of elements of x.*

$$\forall x \exists y \forall z (z \in y \leftrightarrow \exists s (s \in x \wedge z \in s))$$

We are used to thinking of union as an operation applied to a collection of sets instead of a single set. In the axiom above, you should think of the set x as the collection of sets whose union is to be taken. In this case, the set y is the union of elements of x. We shall call y simply the *union* of x and denote it by $\bigcup x$. In other words,

$$\bigcup x = \{z : \exists s \in x \ z \in s\}$$

Next follows the definition of the union of two sets. Let x and y be sets. Then, by pairing, the set $\{x, y\}$ exists.

Definition 2. *The set $\bigcup\{x, y\}$ is called the union of x and y and is denoted by $x \cup y$.*

Exercise 1. *Let x and y be sets. Prove that for all z, we have that $z \in x \cup y$ if and only if $z \in x$ or $z \in y$.*

The dual notion of the union of a set is the *intersection* of a set x, which can be defined as follows.

$$\bigcap x = \{z : \forall s \in x \ z \in s\}$$

Exercise 2. *Show that every set belongs to the class $\bigcap \emptyset$. In other words, $\bigcap \emptyset = \mathbf{V}$.*

Note that we do not know yet whether or not the class $\bigcap x$ is indeed a set for every non-empty set x. In order to show this, we shall need the following axiom.

Axiom 5 (The axiom of separation)**.** *Let $\varphi(z, p)$ be a formula in the language of set theory with two variables z and p. For any p and for any x, there exists a set y that consists of elements of x satisfying the property $\varphi(\cdot, p)$.*

$$\forall p \forall x \exists y \forall z (z \in y \leftrightarrow (z \in x \wedge \varphi(z, p)))$$

We would like to emphasize that the axiom of separation is an axiom *schema* that consists of infinitely many axioms, one for each formula in the language of set theory with two free variables. In each such axiom, you should think of the variable p as a parameter which, when fixed, defines a property $\varphi(\cdot, p)$ of sets. In some textbooks, the axiom of separation is stated for formulas that may have an arbitrary number of parameters. Together with other axioms, one can prove that our formulation of the axiom of separation implies this formulation and vice versa.

One consequence of the axiom of separation is the existence of the intersection of a non-empty set. Let A be a non-empty set, B be an element of A and $\varphi(x, y)$ be the formula $\forall s (s \in y \to x \in s)$. Then, by an instance of the axiom of separation, the class $\{x : x \in B \wedge \varphi(x, A)\}$ forms a set. But this set is precisely $\bigcap A$. Having shown that the intersection of a non-empty set exists, we now define the intersection of two sets.

Definition 3. *The set* $\bigcap\{x, y\}$ *is called the intersection of* x *and* y *and is denoted by* $x \cap y$.

Exercise 3. *Let* x *and* y *be sets. Prove that for all* z*, we have that* $z \in x \cap y$ *if and only if* $z \in x$ *and* $z \in y$.

Two sets x and y are said to be *disjoint* if $x \cap y = \emptyset$. It is trivial to observe[1] that the axiom of separation tells us that the subclass of a set consisting of elements satisfying a certain property is indeed a set, i.e. if a is a set and $\varphi(x)$ is a property of sets, then the class

$$\{x : x \in a \wedge \varphi(x)\} = \{x \in a : \varphi(x)\}$$

is a set. An important consequence of this observation is that the universe of sets $\mathbf{V}$ is a proper class.

Theorem 3. *There does not exist a set which contains all sets, i.e.* $\neg\exists x \forall y\ y \in x$.

Proof. Assume to the contrary that there exist a set U which contains all sets. Then, by separation, there exists a set R such that

$$R = \{x \in U : x \notin x\}$$

But then, since $R \in U$, we have $R \in R \leftrightarrow R \notin R$, which is a contradiction. □

We next introduce some standard operations between sets. Note that for any x and y, the set $\{z \in x : z \notin y\}$ exists by the axiom of separation.

Definition 4. *The set* $\{z \in x : z \notin y\}$ *is called the difference of* x *and* y *and is denoted by* $x - y$.

By taking the union of the sets $x - y$ and $y - x$, we obtain the operation known as the symmetric difference.

Definition 5. *The set* $(x - y) \cup (y - x)$ *is called the symmetric difference of* x *and* y *and is denoted by* $x \triangle y$.

We shall not include here the list of basic properties of the operations introduced so far and refer the reader to any elementary textbook on set theory.

Before introducing the next axiom, we will need the notion of a subset of a set. Let x and y be sets. The set x is said to be a *subset* of y if every element of x belongs to y. More precisely, x is a subset of y if we have $\forall z(z \in x \rightarrow z \in y)$. We shall write $x \subseteq y$ if x is a subset of y; and write $x \subsetneq y$ if $x \subseteq y$ and $x \neq y$. In the latter case, x is said to be a *proper subset* of y. The reader can easily verify that for all x, y, z and non-empty w, we have that

- $\emptyset \subseteq x$ and $x \subseteq x$,
- $\{t : t \in x \wedge \varphi(t)\} \subseteq x$ for any property φ,
- $(x \subseteq y \wedge y \subseteq x) \leftrightarrow x = y$,
- $(x \subseteq y \wedge y \subseteq z) \rightarrow x \subseteq z$,
- $\bigcap w \subseteq \bigcup w$
- $y \in x \rightarrow \bigcap x \subseteq y \subseteq \bigcup x$

The next axiom guarantees the existence of the set of all subsets of a set.

[1]To derive this from our formulation of the axiom of separation, given a formula $\varphi(x)$ and a set x, apply the axiom of separation to a with using formula $\psi(x, y) :\ \varphi(x) \wedge y = y$.

Axiom 6 (The axiom of power set)**.** *For any set x there exists a set y that consists of all subsets of x.*

$$\forall x \exists y \forall z (z \subseteq x \leftrightarrow z \in y)$$

The set $\{z : z \subseteq x\}$ is called the *power set* of x and is denoted by $\mathcal{P}(x)$. When we introduce infinite sets, the power set of an infinite set will be a central object to study, some fundamental properties of which cannot be decided[2] via the axioms of ZFC.

Exercise 4. *Prove that for any set x, the set $\mathcal{P}(x)$, together with the binary operation $\triangle$, forms an abelian group in which every non-identity element has order 2.*

Exercise 5. *Prove that for any non-empty set x, the set $\mathcal{P}(x)$ forms a commutative ring in which every element equals its square, where the binary operations for addition and multiplication are $\triangle$ and $\cap$ respectively.*

Axioms 1-6 are far from being complete to serve as a foundation of mathematics. For once, we cannot prove the existence of an "infinite" set without further axioms. Before introducing more axioms, in the next section, we are going to study how various mathematical concepts can be represented by sets.

[2]The proper term for this phenomenon is *independence*. A sentence φ is said to be *independent* of ZFC in the case that neither φ nor $\neg\varphi$ can be proven from ZFC.

2. From Pairs to Products

2.1. **Relations.** One of the most fundamental concepts in mathematics is the concept of a relation and we begin this section by introducing the set-theoretic definition of a relation.

Definition 6. *A set R is said to be a (binary) relation if it consists of ordered pairs, i.e. $\forall z \in R\ \exists x\ \exists y\ z = (x, y)$*

Given a relation R, using the union and separation axioms, we can form the sets

$$dom(R) = \{a : \exists b\ (a, b) \in R\} \text{ and } ran(R) = \{b : \exists a\ (a, b) \in R\}$$

These sets are called the *domain* of R and the *range* of R respectively. Intuitively speaking, one can think of the relation R as a "rule" that *relates* certain sets in $dom(R)$ to certain sets in $ran(R)$. If R is a relation and $(a, b) \in R$, then one says that "a is in relation R with b" or "a is related to b under the relation R". It is common practice to write aRb instead of $(a, b) \in R$.

Definition 7. *Let A be a set and R be a relation. The image of the set A under the relation R is the set*

$$\{y : \exists x \in A\ xRy\}$$

and is denoted by $R[A]$.

Definition 8. *Let B be a set and R be a relation. The inverse image of the set B under the relation R is the set*

$$\{x : \exists y \in B\ xRy\}$$

and is denoted by $R^{-1}[B]$.

Exercise 6. *Let a, b, c be sets. Show that the set*

$$\{(a, b), (a, a), (c, a), (b, b)\}$$

is a relation and find its domain and range. Then find the image and the inverse image of the set $\{a, c\}$ under this relation.

We will not list many exercises regarding these basic notions and refer the reader who wish to practice to any elementary textbook on set theory.

Definition 9. *Let R be a relation. The inverse relation of R is the set*

$$\{(b, a) : (a, b) \in R\}$$

and is denoted by R^{-1}.

At this point, one may object that our notation creates an ambiguity since it is not clear whether the set $R^{-1}[A]$ denotes the image of A under R^{-1} or the inverse image of A under R. This objection is resolved by the following exercise which justifies our usage of the notation $R^{-1}[A]$ to denote both sets.

Lemma 2. *Let R be a relation and A be a set. Show that the image of A under the relation R^{-1} is the same as the inverse image of A under R.*

Proof. Left to the reader as an exercise. □

Next comes the definition of the cartesian product of two sets. One can easily check that if $a \in A$ and $b \in B$, then $(a, b) \in \mathcal{P}(\mathcal{P}(A \cup B))$. Thus, given two sets A and B, using the axioms introduced so far, we can form the set of all ordered pairs whose first components belong to A and whose second components belong to B.

Definition 10. *Let A and B be sets. The cartesian product of A and B is the set*

$$\{(a,b) \in \mathcal{P}(\mathcal{P}(A \cup B)) : a \in A \wedge b \in B\}$$

and is denoted by $A \times B$.

Definition 11. *Let R be a relation and A, B be sets. The relation R is said to be*

- *a relation from A to B if $R \subseteq A \times B$;*
- *a relation on A if $R \subseteq A \times A$.*

In particular, every relation R is a relation from $dom(R)$ to $ran(R)$. However, notice that a relation R being from the set A to the set B does not necessarily mean that $A = dom(R)$ and $B = ran(R)$.

Definition 12. *Let R and S be relations. Then the composition of S and R is the relation*

$$\{(a,b) : \exists c \ (a,c) \in R \ \wedge \ (c,b) \in S\}$$

and is denoted by $S \circ R$.

The notion of composition of two relations is most frequently used when both relations are a special type of relations called *functions*. On the other hand, some useful properties of the operation $\circ$ still hold for arbitrary relations.

Exercise 7. *Let R and S be relations. Prove that $(S \circ R)^{-1} = R^{-1} \circ S^{-1}$.*

Exercise 8. *Let R, S and T be relations. Prove that $T \circ (S \circ R) = (T \circ S) \circ R$.*

Before introducing the notion of a function, we would like to mention two relations defined on an arbitrary set, which will be useful in later sections.

Definition 13. *Let A be a set. The membership relation on A is the relation*

$$\{(a,b) \in A \times A : a \in b\}$$

and is denoted by $\in_A$.

Definition 14. *Let A be a set. The identity relation on A is the relation*

$$\{(a,b) \in A \times A : a = b\}$$

and is denoted by Δ_A.

The notion of a binary relation can be generalized to that of an n-ary relation, which is a relation that holds or not holds between n many sets. However, the most convenient way to define n-ary relations requires the construction of natural numbers and the n-fold cartesian product of sets. Consequently, we postpone the definition of an n-ary relation until Section 3.

2.2. **Functions.** Recall that one can think of a relation R as a "rule" that *relates* certain sets in $dom(R)$ to certain sets in $ran(R)$. If this "rule" happens to *uniquely assign* each set $dom(R)$ to a certain set in $ran(R)$, then the corresponding relation is said to be a function. More precisely,

Definition 15. *Let R be a relation. The relation R is said to be a function if*

$$\forall a \forall b \forall c \ (aRb \wedge aRc \rightarrow b = c)^1$$

[1]Some authors call a relation satisfying this property *well-defined*. In this terminology, functions are simply relations that are well-defined.

The simplest example of a function is the empty set $\emptyset$. Notice that the definition of a function vacuously holds for the empty set for it has not elements.

Definition 16. *Let R be a relation and A, B be sets. The relation R is said to be a function from A to B if R is a function, $dom(R) = A$ and $ran(R) \subseteq B$. In this case, R is said to have domain A and codomain B.*

An important point to realize is that, according to this definition, the *very same set* can be considered as a function from the same domain to different codomains. For this reason, whenever it is necessary, we shall always specify the codomain of a function.

Definition 17. *Let R be a function and $x \in dom(R)$. The (necessarily) unique element $y \in ran(R)$ for which $(x, y) \in R$ is called the value of R at x.*

Before we proceed, we introduce some notation regarding functions. From now on, we shall write $R : A \to B$ whenever we need to denote a set R which is a function from the set A to the set B. The value of R at a will be denoted by $R(a)$.

We would also like to emphasize that functions *are* relations and hence all notions introduced for relations so far are applicable to functions as well. We next introduce the notion of a bijective function, which will be central to our study of infinite sets.

Definition 18. *Let $f : A \to B$ be a function with domain A and codomain B. Then f is said to be*

- *one-to-one (or injective) if for all $x, y \in A$ we have $f(x) = f(y) \to x = y$.*
- *onto (or surjective) if $ran(f) = f[A] = B$.*
- *one-to-one correspondence (or bijection) if it is both one-to-one and onto.*

Observe that surjectivity and bijectivity of a function both depend on the specified codomain, unlike injectivity. Consequently, the *very same set* can be surjective for some codomain and not surjective for some other codomain. The following exercise illustrates this fact.

Exercise 9. *Prove that the empty set $\emptyset$ is a bijection as a function from $\emptyset$ to $\emptyset$ and not a surjection as a function from $\emptyset$ to $\{\emptyset\}$.*

The notion of injectivity can be generalized to arbitrary relations. More specifically, a relation R is said to be injective if and only if $\forall x \forall y \forall z\ (xRz \wedge yRz \to x = y)$. It is easily seen that a relation being injective is equivalent to its inverse relation being a function and vice versa. Consequently, we have the following fact.

Lemma 3. *Let R be a relation. Then the relation R is an injective function if and only if the inverse relation R^{-1} is an injective function.*

Proof. Let R be a relation that is an injective function. Since R is injective, $\forall x \forall y \forall z\ (xRz \ \wedge yRz \to x = y)$ and hence $\forall x \forall y \forall z\ (zR^{-1}x \ \wedge zR^{-1}y \to x = y)$, which is exactly what it means for R^{-1} to be a function. Since R is a function, $\forall x \forall y \forall z\ (xRy \ \wedge xRz \to y = z)$ and hence $\forall x \forall y \forall z\ (yR^{-1}x \ \wedge zR^{-1}x \to y = z)$, which is exactly what it means for R^{-1} to be injective. By changing the roles of R and R^{-1}, the proof of the right-to-left direction can be done similarly. □

One can easily verify that any subset of a function is itself a function. This observation suggests the following definition.

Definition 19. *Let f be a function and A be a set. The restriction of f to A is the function*

$$\{(a, b) \in f : a \in A\}$$

and is denoted by $f \restriction A$.

Definition 20. *Two functions f and g are said to be compatible if $f(x) = g(x)$ for all $x \in dom(f) \cap dom(g)$.*

In other words, two functions are compatible if the values they take agree at every element in the intersection of their domains. The following lemma shows that two functions being compatible is equivalent to their union being a function.

Lemma 4. *Let f and g be functions. Then f and g are compatible if and only if $f \cup g$ is a function.*

Proof. For the left-to-right direction, assume that f and g are compatible functions. Clearly $f \cup g$ is a relation. We want to show that for all x, y, z if $(x, y) \in f \cup g$ and $(x, z) \in f \cup g$, then $y = z$. Let $(x, y) \in f \cup g$ and $(x, z) \in f \cup g$. There are four cases.

- If $(x, y) \in f$ and $(x, z) \in f$, then $y = z$ since f is a function.
- If $(x, y) \in g$ and $(x, z) \in g$, then $y = z$ since g is a function.
- If $(x, y) \in f$ and $(x, z) \in g$, then $y = z$ since f and g are compatible.
- If $(x, y) \in g$ and $(x, z) \in f$, then $y = z$ since f and g are compatible.

Thus, $f \cup g$ is a function. For the converse direction, assume that $f \cup g$ is a function. Let $x \in dom(f) \cap dom(g)$. Since f and g are functions, there exist y and z such that $f(x) = y$ and $g(x) = z$. Then, clearly we have $(x, y), (x, z) \in f \cup g$. However, by assumption $f \cup g$ is a function and hence $f(x) = y = z = g(x)$. Thus, f and g are compatible. □

The following exercise shows that the lemma above can be generalized to arbitrary collections of compatible functions.

Exercise 10. *Let S be a set such that elements of S are functions which are pairwise compatible. Show that $\bigcup S$ is a function with domain $\bigcup\{dom(f) : f \in S\}$.*

The next lemma shows that the class of functions are closed under the operation of composition.

Lemma 5. *Let f and g be functions. Then the composition $g \circ f$ is a function.*

Proof. Let x, y, z be sets such that $(x, y) \in g \circ f$ and $(x, z) \in g \circ f$. We want to show that $y = z$. By definition of composition, there exist y' and z' such that $(x, y') \in f$ and $(y', y) \in g$; and $(x, z') \in f$ and $(z', z) \in g$. Since f is a function, $(x, y') \in f$ and $(x, z') \in f$ implies that $y' = z'$. Since g is a function and $y' = z'$, $(y', y) \in g$ and (z', z) implies that $y = z$. □

Exercise 11. *Let f and g be functions. Show that the domain of the function $g \circ f$ is $dom(f) \cap f^{-1}[dom(g)]$ and that $(g \circ f)(x) = g(f(x))$ for all x in this domain.*

Given two sets x and y, a function f from x to y is an element of $\mathcal{P}(x \times y)$ and hence we can form the set of all functions from x to y

$$\{f \in \mathcal{P}(x \times y) : \forall a \forall b \forall c \ ((a, b) \in f \wedge (a, c) \in f) \rightarrow b = c \ \wedge \ dom(f) = x\}$$

using the axioms introduced so far. From now on, the set of all functions from the set x to the set y will be denoted by ${}^{x}y$. Some authors use the notation y^x to denote this set, however, we reserve this notation for exponentiation on ordinal and cardinal numbers in order to avoid ambiguities.

2.3. **Products and sequences.** Next will be discussed how to define the product of an arbitrary collection of sets.

Recall that when we defined the cartesian product $A \times B$ of two sets, the order of the sets A and B mattered. Even though the cartesian product $B \times A$ is in a natural bijection with the cartesian product $A \times B$, these are different objects in the universe of sets. Therefore, in order to generalize the concept of cartesian product to arbitrarily many sets, we first need to *label* the sets whose product is to be taken. This labeling can be done through some function.

Let J be a set which contains the sets whose product is to be taken and possibly other sets. Let $F : I \to J$ be an arbitrary function. We will refer to the function F an *indexed system of sets* with the *index set* I. Here we think of the set $i \in I$ as the *label* of the set $F(i)$ for all $i \in I$. While talking about indexed systems of sets, it is customary to write F_i instead of $F(i)$ and write $\{F_i\}_{i\in I}$ instead of $F[I]$, which we will also refer to as an indexed system of sets.

Definition 21. *Let $\{F_i\}_{i\in I}$ be an indexed system of sets with the index set I. The product of the indexed system $\{F_i\}_{i\in I}$ is the set*

$$\{f : I \to \bigcup\{F_i\}_{i\in I} \mid \forall i \in I \ f(i) \in F_i\}$$

and is denoted by $\prod_{i\in I} F_i$.

In other words, the product $\prod_{i\in I} F_i$ is the set of all functions f with domain I such that $f(i) \in F_i$ for all $i \in I$. One usually denotes a set $f \in \prod_{i\in I} F_i$ using the sequence notation $(f(i))_{i\in I}$ since f can be considered as a *sequence* which takes values in F_i at each component i.

Indeed, this is exactly how we define sequences over arbitrary sets. Let $\{S_i\}_{i\in I}$ be an indexed family of sets for some index set I such that $S_i = S$ for all $i \in I$. An element f of the product $\prod_{i\in I} S$ is called a *sequence* over S with the index set I and is denoted by $(f(i))_{i\in I}$.

We have not constructed the natural numbers yet. For the following exercises, the reader should assume[2] that $0 = \emptyset$, $1 = \{0\}$ and $2 = \{0, 1\}$.

Exercise 12. *Let $\{A_i\}_{i\in 2}$ be an indexed system of set with the index set 2. Show that the map $f : \prod_{i\in 2} A_i \to A_0 \times A_1$ given by $f(g) = (g(0), g(1))$ is a bijection.*

Consequently, the notion of cartesian product can be considered as a special case of the product of an indexed system of sets. As the reader may guess, once we define natural numbers, the cartesian product of sets $A_1, \ldots, A_n$ will simply be defined as the product of the indexed family $\{A_i\}_{i\in n}$.

The next exercise shows that there is a natural bijection between the power set of any set X and the product of an appropriately chosen system with index set X.

Exercise 13. *Let X be any set. Show that ${}^{X}2 = \prod_{i\in X} 2$ and that the map f from $\prod_{i\in X} 2$ to $\mathcal{P}(X)$ given by $f(g) = \{x \in X : g(x) = 1\}$ is a bijection.*

[2]Since the notions introduced so far are enough to carry out the construction of natural numbers, the curious reader may read the first subsection of Section 4 for a precise construction at this point.

We next focus on a seemingly simple question. Assume that $\{A_i\}_{i\in I}$ is an indexed system of sets such that $A_i \neq \emptyset$ for all $i \in I$. Is the product $\prod_{i\in I} A_i$ necessarily non-empty?

Given a fixed finite set I such as $I = \{0, 1, 2\}$, the reader can prove as an exercise that the answer is affirmative. As can be seen from the exercise above, the answer is also affirmative when $A_i = 2$.

However, it is not clear whether or not $\prod_{i\in I} A_i \neq \emptyset$ for all indexed system of sets $\{A_i\}_{i\in I}$ with $A_i \neq \emptyset$ for all $i \in I$. It turns out that this statement cannot be proven or disproven from Axioms 1-6 plus the axioms of Infinity, Replacement and Foundation. In the next section, we shall introduce an axiom that settles this question.

Before we conclude this subsection, we would like to mention two notations regarding indexed systems of sets. From now on, given an indexed system of set $\{A_i\}_{i\in I}$, we shall denote the sets $\bigcup\{A_i\}_{i\in I}$ and $\bigcap\{A_i\}_{i\in I}$ by $\bigcup_{i\in I} A_i$ and $\bigcap_{i\in I} A_i$ respectively.

2.4. **To choose or not to choose.** In this section, we shall introduce the axiom of choice, one of the most famous axioms of ZFC. For historical reasons, the axiom of choice became so famous that the letter **C** of ZFC stands for this axiom.

There are literally dozens of equivalent formulations of the axiom of choice. Below, we introduce the formulation which states that the product of an indexed system of non-empty sets is non-empty. Some equivalent formulations of this axiom will be mentioned in later sections.

Axiom 7 (The axiom of choice)**.** *For all sets I and for all indexed systems of sets $\{A_i\}_{i\in I}$ with $A_i \neq \emptyset$ for all $i \in I$, the product $\prod_{i\in I} A_i$ is non-empty*[3]*.*

Recall that an element f of the product $\prod_{i\in I} A_i$ is a function with $f(i) \in A_i$ for all $i \in I$. Loosely speaking, the function f *chooses* one element from each A_i. In this sense, the axiom of choice allows us to "simultaneously choose" an element from each set in a set of non-empty sets. The reader who does not feel comfortable with indexed systems may find the following lemma more intuitive.

Lemma 6. *Let M be a set whose elements are non-empty sets. Then there exists a function $f : M \to \bigcup M$ such that $f(x) \in x$ for all $x \in M$.*

Proof. Notice that every set can be indexed by itself through identity function. More precisely, let $M = I$ and $\{M_i\}_{i\in I}$ be the indexed system of sets with $M_i = i$. Then, since $M_i \neq \emptyset$ for all $i \in I$, by the axiom of choice, there exists $f : I \to \bigcup M$ such that $f(i) \in i$ for all $i \in I$, which is precisely what we wanted to prove. □

It is easily seen that the axiom of choice is implied by the statement of the lemma above together with Axioms 1-6.

[3]Unlike the other axioms introduced up to now, we shall not attempt to write the axiom of choice in the language of set theory since it will require **too much** space. A curious reader may attempt to do this in his or her free time!

3. Equivalence Relations and Order Relations

In this section, we will learn several important types of relations. In order to provide nice examples, we shall assume throughout this chapter that the set of natural numbers $\mathbb{N}$ is already constructed together with its usual arithmetical operations and relations. The reader who wish to see a precise construction of these may read Section 4.

3.1. **Equivalence relations, partitions and transversals.** We begin by introducing the notion of an equivalence relation, which is frequently used when different mathematical objects are needed to be considered "the same" for various purposes.

Definition 22. *Let X be a set and E be a relation on X, i.e. $E \subseteq X \times X$. The relation E is said to be an equivalence relation if it is*

- *reflexive, i.e. for all $x \in X$ we have xEx,*
- *symmetric, i.e. for all $x, y \in X$ we have $xEy \to yEx$, and*
- *transitive, i.e. for all $x, y, z, \in X$ we have $(xEy \wedge yEz) \to xEz$.*

For example, the identity relation Δ_X is an equivalence relation on X for any set X. Indeed, we have $\Delta_X \subseteq E$ for all equivalence relations E on X. Note that the empty set is an equivalence relation on itself and, indeed, is the unique equivalence relation on the empty set, which will be referred to as the *empty relation*.

Exercise 14. *Let X be any set and define the relation $E \subseteq X \times X$ by*

$$xEy \leftrightarrow \text{ There exists a bijection between } x \text{ and } y$$

for all $x, y \in X$. Show that E is an equivalence relation on X.

Exercise 15. *Let X be a non-empty set such that the elements of X are equivalence relations on some fixed set Y. Show that if we have $E \subseteq F$ or $F \subseteq E$ for all $E, F \subseteq X$, then $\bigcup X$ is an equivalence relation on Y.*

We next introduce some terminology to talk about elements that are related to each other under some equivalence relation.

Definition 23. *Let X be a set, E be an equivalence relation on X and $x \in X$. The equivalence class of x modulo E is the set*

$$\{y \in X : yEx\}$$

and is denoted by $[x]_E$.

Given an equivalence relation E on some set X, the sets of the form $[x]_E$ for some $x \in X$ are referred to as *E-equivalence classes*. Two elements in the same E-equivalence class are said to be *E-equivalent*. Observe that, according to this definition, the empty relation on the empty set has no equivalence classes, since we require each equivalence class of E to be of the form $[x]_E$ for *some* $x \in X$.

The following lemma shows that equivalence classes of an equivalence relation on a non-empty set are either identical or disjoint.

Lemma 7. *Let X be a non-empty set and E be an equivalence relation X. Then for all $x, y \in X$ we have that either $[x]_E = [y]_E$ or $[x]_E \cap [y]_E = \emptyset$.*

Proof. Let $x, y \in X$ and assume that $[x]_E \cap [y]_E \neq \emptyset$. Then there exists $z \in X$ such that zEx and zEy and hence xEy by symmetry and transitivity of E. Now pick $w \in [x]_E$, then wEx and xEy and hence $w \in [y]_E$. This shows that $[x]_E \subseteq [y]_E$.

By a symmetric argument, one can show that $[y]_E \subseteq [x]_E$ and hence $[x]_E = [y]_E$. This shows that $[x]_E = [y]_E$ or $[x]_E \cap [y]_E = \emptyset$. Since $[x]_E$ and $[y]_E$ are both non-empty, both cases cannot occur simultaneously. Therefore, either $[x]_E = [y]_E$ or $[x]_E \cap [y]_E = \emptyset$. □

Exercise 16. *Let E be the equivalence relation on $\mathbb{N} \times \mathbb{N}$ defined by*

$$(p, q)E(r, s) \leftrightarrow p + s = q + r$$

Show that E is an equivalence relation and find the equivalence class $[(2, 0)]_E$.

Next, we define a partition of a set and the quotient set of an equivalence relation, notions which will be related through a fundamental theorem.

Definition 24. *Let X be a set and E be an equivalence relation on X. The quotient set of X with respect to E is the set*

$$\{[x]_E : x \in X\}$$

which consists of the equivalence classes of E and is denoted by X/E.

Definition 25. *Let X be a set. Then a subset $S \subseteq \mathcal{P}(X)$ is said to be a partition of X if*

- *elements of S are non-empty, i.e. for all $A \in S$ we have $A \neq \emptyset$.*
- *distinct elements of S are disjoint, i.e. for all $A, B \in S$ if $A \neq B$ then $A \cap B = \emptyset$, and*
- *the union of S is the set X, i.e. $\bigcup S = X$*

According to this definition, the empty set has a unique partition, which is the empty set itself. We remark that, in some textbooks, the notion of a partition of a set may not be defined for the empty set. The following theorem is an easy consequence of Lemma 7.

Theorem 4. *Let X be a set and E be an equivalence relation on X. Then X/E is a partition of X.*

Proof. Left to the reader as an exercise. □

In other words, every equivalence relation on a set induces a partition. It turns out that the converse is also true, i.e. every partition is induced by some equivalence relation.

Lemma 8. *Let S be a partition of a set X. Then the relation E_S defined by*

$$xE_Sy \leftrightarrow \exists D \in S \ (x \in D \wedge y \in D)$$

is an equivalence relation on X.

Proof. The claim clearly holds when X is the empty set. Now, assume that X is non-empty and let $x, y, z \in X$ be such that xE_Sy and yE_Sz. Then, by definition, there exist $C, D \in S$ such that $x \in C$ and $y \in C$; and, $y \in D$ and $z \in D$. However, since S is a partition, $C \cap D \neq \emptyset$ implies that $C = D$. But then $x \in C$ and $z \in C$, which implies that $(x, z) \in E_S$. Hence, E_S is transitive.

Showing that E_S is reflexive and symmetric is left as an exercise to the reader. □

It is not difficult to check that different partitions of the same set induce different equivalence relations, i.e. if $S \neq D$ are partitions of a set X, then $E_S \neq E_D$. Consequently, we have the following theorem, which some authors refer to as the fundamental theorem of equivalence relations.

Theorem 5. *Let X be any set. The map f from the set of partitions of X to the set of equivalence relations on X given by $f(S) = E_S$ where*

$$E_S = \{(x, y) \in X \times X : \exists D \in S \; x \in D \wedge y \in D\}$$

is a bijection.

Next will be discussed the notion of a transversal of an equivalence relation.

Definition 26. *Let X be a set and E be an equivalence relation on X. A subset $T \subseteq X$ is said to be a transversal (or, a set of representatives) for the equivalence relation E if for every $x \in X$ there exists $y \in X$ such that $[x]_E \cap T = \{y\}$.*

In other words, a transversal for an equivalence relation is a set that contains exactly one element from each equivalence class. Given a transversal $T \subseteq X$ for an equivalence relation $E \subseteq X \times X$, the unique set in $[x]_E \cap T$ is called the *representative* of the equivalence class $[x]_E$.

By the axiom of choice, since we can simultaneously pick one element from each set in a set of non-empty sets, transversals exist for non-empty equivalence relations. More precisely, we have the following theorem.

Theorem 6. *Let X be a non-empty set and E be an equivalence relation on X. Then there exists a transversal T for E.*

Proof. Since X is non-empty, the partition X/E is a set of non-empty sets. By Lemma 6, there exists a function $f : X/E \to X$ such that $f(x) \in x$ for all $x \in X/E$. Notice that f picks exactly one element from each equivalence class. Thus, the range $f[X]$ of this function is a transversal for E. □

In the next subsection, we will solve a seemingly-impossible-to-solve puzzle using the existence of transversals. For this puzzle, we will assume the familiarity of the reader with *finite* and *countably infinite* sets. The reader who does not feel comfortable using these concepts at this point may skip the next subsection and read it after completing Section 5.

3.2. **A Game of Thrones, Prisoners and Hats.** After the battle of the Blackwater, King Joffrey of Westeros captured countably infinitely many soldiers of Stannis Baratheon as his prisoners and put the set of prisoners in a bijection with the set of natural numbers. In other words, every prisoner is uniquely labeled by some natural number.

King Joffrey, who has been known for his cruel games, explained to the prisoners that they would be executed the next morning, unless they succeed in the following game that will take place before the execution:

The prisoners will be standing in a straight line in such a way that every prisoner will be able to see the infinitely many prisoners whose labels are greater than his label, i.e. the prisoners are standing on the number line facing the positive direction.

Then each prisoner will be *randomly* given a hat that is either red or blue. The prisoners can see *all* the hats in front of them but cannot see their own hats. Moreover, they are not allowed to move or communicate in any way. After all the hats are distributed, each prisoner will be asked to guess the color of his own hat and write his guess in a piece of paper.

The rules of the game are as follows: If there are only finitely many prisoners who guess wrong, then all the prisoners are set free. Otherwise, they all are executed.

Once the rules are explained to the prisoners, they immediately think that it is impossible to succeed since they are in no position to obtain information about the colors of their own hats by looking at the colors of other prisoners' hats.

Tyrion Lannister, who is not fond of King Joffrey and who has studied set theory in his youth, decides to help the prisoners. Soldiers of Stannis are so smart that they have been known to memorize infinite amount of information if necessary. Knowing this fact, Tyrion realizes that he can set the prisoners free.

Theorem 7. *There exists a survival strategy for the prisoners.*

Proof. Let E be the equivalence relation on the set ${}^{\mathbb{N}}2$ defined by

$$fEg \leftrightarrow \exists m \in \mathbb{N}\ \forall n \in \mathbb{N}\ (n \geq m \rightarrow f(n) = g(n))$$

In other words, two functions from $\mathbb{N}$ to 2 are E-equivalent if and only if they take the same values at sufficiently large natural numbers. We skip the details of checking that E is indeed an equivalence relation and leave this as an exercise to the reader.

By Theorem 6, there exists a transversal $T \subseteq {}^{\mathbb{N}}2$ for the equivalence relation E. Let $F : {}^{\mathbb{N}}2 \rightarrow {}^{\mathbb{N}}2$ be the function defined by

$$F(f) = \text{ the unique element of } [f]_E \cap T$$

for all $f \in {}^{\mathbb{N}}2$, that is, the function F sends each f to the unique element of T which is E-equivalent to f. The survival strategy of the prisoners is as follows. When asked the color of his hat, Prisoner n first constructs the function $f : \mathbb{N} \rightarrow 2$ defined by

- $f(i) = 0$ for all $i \leq n$,
- $f(i) = 0$ if prisoner i has red hat and $i > n$, and
- $f(i) = 1$ if prisoner i has blue hat and $i > n$

In other words, Prisoner n first "encodes" the colors of the hats into a function from $\mathbb{N}$ to 2, assuming that the colors of the hats he does not see are all red. Then he guesses red if $(F(f))(n) = 0$ and guesses blue if $(F(f))(n) = 1$.

We claim that the prisoners survive if they use this strategy. To see this, let $g : \mathbb{N} \rightarrow 2$ be the function that encodes the actual state of the hats after the game starts, i.e. $g(i) = 0$ if and only if the hat of Prisoner i is red.

Let $h : \mathbb{N} \rightarrow 2$ be the unique function such that $h \in T$ and hEg. By construction, Prisoner i will guess the color of his hat based on the value $h(i)$, i.e. he guesses red if $h(i) = 0$ and blue if $h(i) = 1$. However, by definition, there exists $m \in \mathbb{N}$ such that for all $n \geq m$ we have $h(n) = g(n)$. This means that Prisoner n will guess the color of his hat correctly for all $n \geq m$. Therefore, the prisoners survive. □

3.3. **Order relations.** In this subsection, we will learn about order relations, which are frequently used in mathematics when different mathematical objects are needed to be "compared" for various purposes.

Definition 27. *Let X be a set and E be a relation on X, i.e. $E \subseteq X \times X$. The relation E is said to be a (partial) order relation if it is*

- *reflexive, i.e. for all $x \in X$ we have xEx,*
- *anti-symmetric, i.e. for all $x, y \in X$ we have $(xEy \wedge yEx) \rightarrow x = y$, and*
- *transitive, i.e. for all $x, y, z, \in X$ we have $(xEy \wedge yEz) \rightarrow xEz$.*

We shall often use the symbols $\leq$ or $\preccurlyeq$ to denote various partial order relations and read $x \leq y$ as "x is less than or equal to y". The reader should keep in mind that, depending on the context, relations denoted by these symbols may have nothing to do with their usual intended meaning on various number systems.

Exercise 17. *Let $\preccurlyeq$ be the relation on $\mathbb{N}^+ = \{k \in \mathbb{N} : k \neq 0\}$ defined by*

$$x \preccurlyeq y \leftrightarrow \exists k \in \mathbb{N}^+ \ y = k \cdot x$$

for all $x, y \in \mathbb{N}$. Show that $\preccurlyeq$ is a partial order relation.

Given a partial order relation $\leq$ on some set X, it will sometimes be more convenient to work with the relation $<$ defined by $x < y \rightarrow x \leq y \ \wedge x \neq y$ for all $x, y \in X$. It turns out that those relations that are obtained from partial order relations in this way are exactly those that are transitive and asymmetric.

Definition 28. *Let X be a set and E be a relation on X, i.e. $E \subseteq X \times X$. The relation E is said to be a strict (partial) order relation if it is*

- *asymmetric, i.e. for all $x, y \in X$ we have $xEy \rightarrow \neg yEx$, and*
- *transitive, i.e. for all $x, y, z, \in X$ we have $(xEy \wedge yEz) \rightarrow xEz$.*

Lemma 9. *Let F be a strict partial order relation on X and E be the relation on X defined by $xEy \leftrightarrow xFy \ \vee \ x = y$. Then E is a partial order relation on X.*

Proof. E is clearly reflexive since $x = x$ for all $x \in X$. Let $x, y \in X$ such that xEy and yEx. Then, by definition, we have $xFy \ \vee \ x = y$ and $yFx \ \vee \ y = x$. Since F is asymmetric, it cannot be that xFy and yFx. Thus, $x = y$ and hence E is anti-symmetric. To see that E is transitive, let $x, y, z \in X$ such that xEy and yEz. Then, by definition, we have $xFy \ \vee \ x = y$ and $yFz \ \vee \ y = z$. If $x = y$, then xEz and we are done. So suppose we have xFy. Now we have $yFz \ \vee \ y = z$. If yFz , then by the transitivity of F we have xFz which implies xEz. If $y = z$, then $x = y = z$ and so xEz by definition. In all cases, we obtained xEz. Therefore E is transitive. □

Lemma 10. *Let E be a partial order relation on X and F be the relation on X defined by $xFy \leftrightarrow xEy \ \wedge \ x \neq y$. Then F is a strict partial order relation on X.*

Proof. Let $x, y \in X$ such that xFy. Then, by definition, $xEy \ \wedge \ x \neq y$. If it were the case that yFx, then we would have yEx which would imply $x = y$ by anti-symmetry of E, which gives a contradiction. Thus, $\neg yFx$ and hence F is asymmetric. To see that F is transitive, let $x, y, z \in X$ such that xFy and yFz. Then, by definition, $xEy \wedge x \neq y$ and $yEz \wedge y \neq z$. By transitivity of E, we have xEz. If it were the case that $x = z$, then xEy and yEz would imply $x = y$, which is a contradiction. Thus, $x \neq z$ and hence xFz, which completes the proof that F is a strict partial order. □

From now on, whenever we mention the *induced* strict partial order relation $<$ of a partial order relation $\leq$ or the *induced* partial order relation $\leq$ of a strict partial order relation $<$, the reader should understand that $x \leq y \leftrightarrow x < y \ \vee x = y$ and $x < y \leftrightarrow x \leq y \ \wedge \ x \neq y$.

Given a partial order relation $\leq$ on some set, we say that two elements a and b are said to be *comparable* (with respect to $\leq$) if $a \leq b$ or $b \leq a$. Similarly, given a strict partial order relation $<$, two elements a and b are said to be *comparable* (with respect $<$) if $a = b$ or $a < b$ or $b < a$. If two elements are not comparable,

then they are called *incomparable*. Partial orders in which any two elements are comparable will be of special importance to us.

Definition 29. *Let $\leq$ be a partial order relation on a set X. The relation $\leq$ is said to be a linear order relation if for all $a, b \in X$, the elements a and b are comparable (with respect to $\leq$).*

Definition 30. *Let $<$ be a strict partial order relation on a set X. The relation $<$ is said to be a strict linear order relation if for all $a, b \in X$, the elements a and b are comparable (with respect to $<$).*

Exercise 18. *Let X be a set that contains at least two elements. Show that the relation E on $\mathcal{P}(X)$ given by*

$$xEy \leftrightarrow x \subseteq y$$

for all $x, y \in \mathcal{P}(X)$ is a partial order relation which is not a linear order.

Next will be introduced several notions of "bigness" and "smallness" for a partial order relation.

Definition 31. *Let $\leq$ be a partial order relation on a set X and $Y \subseteq X$. Then an element $y \in Y$ is said to be*

- *a least element of Y with respect to $\leq$ if $\forall x \in Y \ y \leq x$.*
- *a minimal element of Y with respect to $\leq$ if $\forall x \in Y \ (x \leq y \rightarrow x = y)$.*
- *a greatest element of Y with respect to $\leq$ if $\forall x \in Y \ x \leq y$.*
- *a maximal element of Y with respect to $\leq$ if $\forall x \in Y \ (y \leq x \rightarrow x = y)$.*

It follows from the transitivity of $\leq$ that least and greatest elements, if they exist, are unique. Hence, we can talk about *the* least element of Y and *the* greatest element of Y with respect to $\leq$. Being least is clearly stronger than being minimal, i.e. least elements are also minimal elements. However, the converse is not true as shown by the following exercise

Exercise 19. *Let $\preccurlyeq$ be the relation on $A = \mathbb{N} - \{0, 1\}$ defined by*

$$x \preccurlyeq y \leftrightarrow \exists k \in \mathbb{N}^+ \ y = k \cdot x$$

for all $x, y \in \mathbb{N}$. Show that there is no least element of A with respect to $\preccurlyeq$ and that the set of minimal elements of A with respect to $\preccurlyeq$ is exactly the set of prime numbers.

On the other hand, being minimal implies being least whenever any two elements are comparable.

Exercise 20. *Let $\leq$ be a partial order relation on a set X and $Y \subseteq X$ be a subset such that x and y are comparable with respect to $\leq$ for all $x, y \in Y$. Show that if $y \in Y$ is a minimal element of Y with respect to $\leq$, then it is also the least element of Y with respect to Y.*

As can be seen from the previous exercise, subsets in which any two elements are comparable with respect to an order relation are of importance and deserve a special name.

Definition 32. *Let $\leq$ be a partial order relation on a set X. A subset $Y \subseteq X$ is said to be a chain (with respect to $\leq$) if x and y are comparable (with respect to $\leq$) for all $x, y \in Y$.*

From now on, while referring to properties of elements with respect to some order relation $\leq$, we may sometimes omit the phrase "*with respect to* $\leq$" if the order relation is understood from the context.

3.4. **Well-orders.** In this subsection, we shall learn the notion of a well-order relation. Since the theory of ordinal numbers will be based on well-orders, the reader is expected to get a solid grasp of this notion.

Definition 33. *Let $\leq$ be a partial order relation on a set X. The relation $\leq$ is said to be a well-order relation if $\leq$ is a linear order relation on X and every non-empty subset of X has a least element.*

Definition 34. *Let $<$ be a strict partial order relation on a set X. The relation $<$ is said to be a strict well-order relation if the induced partial order $\leq$ is a well-order relation.*

Given a partial (respectively, linear, well-order) order relation $\leq$ on a set X, the pair $(X, \leq)$ is called a *partially* (respectively, *linearly*, *well-*) *ordered set*. We can similarly define *strictly partially* (respectively, *linearly*, *well-*) *ordered sets*.

An example of a well-ordered set is $(\mathbb{N}, \leq)$ where the relation $\leq$ is the usual linear order relation defined on natural numbers. Once we introduce the class of ordinal numbers, we will be able to prove that every-well ordered set is "represented" by some ordinal number. For this reason, we shall not provide many examples of well-ordered sets at this point and proceed to prove several important results. We first introduce some terminology.

Definition 35. *Let $(S, \leq)$ be a partially ordered set, $<$ be the induced strict partial order and $s \in S$. The set $pred(s) = \{i \in S : i < s\}$ is called the predecessors of s.*

Definition 36. *Let $(S, \leq)$ be a well-ordered set, $<$ be the induced strict well-order and $s \in S$ be a non-maximal element. The least element of the set $\{t \in S : s < t\}$ is called the successor of s in $(S, \leq)$ and is denoted by s^+.*

Definition 37. *Let $(S, \leq)$ be a linearly ordered set and $I \subseteq S$. The set I is called an initial segment of S if $pred(i) \subseteq I$ for every $i \in I$. An initial segment I of S is called proper if $I \neq S$.*

In other words, initial segments in strictly linearly ordered set are those sets that are "closed downwards". The next lemma shows that every proper initial segment of a strictly well-ordered set is indeed the predecessors of some element.

Lemma 11. *Let $(S, \leq)$ be a well-ordered set and $I \subseteq S$ be a proper initial segment. Then there exists $s \in S$ such that $I = pred(s)$.*

Proof. Since I is a proper initial segments, $I \neq S$ and hence $S - I$ is non-empty. By definition, there exists a minimal element of $S - I$, say $s \in S - I$. Let $x \in pred(s)$. Then $x \notin S - I$ by the minimality of s and hence $x \in I$. Conversely, let $x \in I$. Since I is a proper initial segment and $s \notin I$, we cannot have $s \leq x$. Thus, $x < s$ and hence $x \in pred(s)$, which completes the proof that $I = pred(s)$. □

Our aim in this subsection is to prove that given two well-ordered sets, either one of them is a proper initial segment of the other or they are "the same". Of course, we first have to define what it means for two partially ordered sets to be "the same".

Definition 38. *Let $(P, \leq)$ and $(Q, \preccurlyeq)$ be two partially ordered sets. A function $f : P \to Q$ is said to be order-preserving if $x \leq y \leftrightarrow f(x) \preccurlyeq f(y)$ for all $x, y \in P$.*

It is easily seen that any order-preserving map is necessarily one-to-one[1]. We can similarly define order-preserving functions for strict partial order relations[2]. Moreover, we have the following.

Exercise 21. *Let $\leq$ and $\preccurlyeq$ be linear order relations on P and Q respectively; and let $<$ and $\prec$ be the induced strict linear order relations. Show that for any function $f : P \to Q$ we have that*

$$(\forall x, y \in P\ (x < y \leftrightarrow f(x) \prec f(y))) \leftrightarrow (\forall x, y \in P\ (x \leq y \leftrightarrow f(x) \preccurlyeq f(y)))$$

Consequently, whenever we need to show that a function is order-preserving between linear order relations, we may work with the order relations themselves or with their induced strict order relations. Two partially ordered sets are considered to be "the same" if there exists a bijective order-preserving map between them.

Definition 39. *Let $(P, \leq)$ and $(Q, \preccurlyeq)$ be two partially ordered sets. A function $f : P \to Q$ is said to be an order-isomorphism if it is bijective and order-preserving.*

Definition 40. *Two partially ordered sets $(P, \leq)$ and $(Q, \preccurlyeq)$ are said to be order-isomorphic if there exists an order-isomorphism between them, in which case we shall write $(P, \leq) \cong (Q, \preccurlyeq)$.*

The notion of order-isomorphism for strictly partially ordered sets can be similarly defined. It is not difficult to check that two partially ordered sets are isomorphic if and only if the corresponding strictly partially ordered sets are isomorphic.

The following exercise shows that the requirements for a function to be order-preserving can be slightly weakened if we are working with strictly linearly ordered sets.

Exercise 22. *Let $(P, <)$, $(Q, \prec)$ be two strictly linearly ordered sets and $f : P \to Q$ be a function such that $x < y \to f(x) \prec f(y)$ for all $x, y \in P$. Show that we indeed have $x < y \leftrightarrow f(x) \prec f(y)$ for all $x, y \in P$.*

Next will be shown that an order-preserving map from a well-order to itself is non-decreasing.

Lemma 12. *Let $(P, \leq)$ be a well-ordered set and $f : P \to P$ be order-preserving. Then $x \leq f(x)$ for all $x \in P$.*

Proof. Let $<$ denote the induced strict well-order relation. Consider the set

$$Q = \{x \in P : f(x) < x\}$$

If Q is non-empty, then it has a minimal element, say $q \in Q$. Then, $f(q) < q$ and hence $f(f(q)) < f(q)$ since f is an order-embedding. This implies that $f(q) \in Q$, which contradicts the minimality of q. Thus, Q is empty, which completes the proof. □

[1]We would like to point out that some authors define order-preserving functions to be those functions that only satisfy $x \leq y \to f(x) \preccurlyeq f(y)$, in which case constant functions between partially ordered sets are automatically order-preserving.

[2]An important point to realize is that only requiring $x < y \leftrightarrow f(x) \prec f(y)$ does not guarantee injectivity unless we are working with linear orders.

An immediate consequence of Lemma 12 is that any order-isomorphism from a well-ordered set to itself is the identity function. The next exercise shows that subsets of ordered sets are themselves ordered sets with respect to the same relation.

Exercise 23. *Let (X, E) be a partially (respectively, linearly, well-) ordered set and $Y \subseteq E$. Define the relation F on Y as the restriction of the relation E onto Y, i.e. $F = E \cap (Y \times Y)$. Show that (Y, F) is a partially (respectively, linearly, well-) ordered set.*

Another corollary of Lemma 12 is that no well-order is isomorphic to a proper initial segment of itself.

Corollary 8. *Let $(P, \leq)$ be a well-ordered set and $p \in P$. Let $\leq_p$ be the restriction of $\leq$ to the set $pred(p)$. Then $(P, \leq)$ and $(pred(p), \leq_p)$ are not isomorphic.*

Proof. Assume to the contrary that there is an order-isomorphism $f : P \to pred(p)$. By Lemma 12, we have $x \leq f(x)$ for all $x \in P$. In particular, $p \leq f(p)$ which is a contradiction since $f(p) \in pred(p)$. □

We are now ready to prove the main theorem of this subsection.

Theorem 9. *Let $(P, \leq)$ and $(Q, \preccurlyeq)$ be well-ordered sets. Then either*

a. *$(P, \leq)$ and $(Q, \preccurlyeq)$ are isomorphic,*
b. *$(P, \leq)$ is isomorphic to some proper initial segment of Q, or*
c. *$(Q, \preccurlyeq)$ is isomorphic to some proper initial segment of P.*

Proof. We begin by showing that these cases are mutually exclusive. If any two of a, b and c held simultaneously, then we would have a well-order which is isomorphic to a proper initial segment of itself, which is impossible. Thus these cases are mutually exclusive. In order to show that at least one of these cases holds, consider the relation

$$f = \{(a, b) \in P \times Q : pred(a) \text{ and } pred(b) \text{ are isomorphic}\}$$

We first show that f is a function. If f were not a function, then there would be $a \in P$ and $b, c \in Q$ such that $pred(a)$, $pred(b)$ and $pred(c)$ are isomorphic to each other. But this contradicts Corollary 8, since either $b \in pred(c)$ or $c \in pred(b)$. By a similar argument, one can easily show that f is one-to-one. The proof that f is order-preserving is left to the reader as an exercise.

Finally, we need to show that at least one of three cases holds. If $dom(f) = P$, then either a or b holds. So it is sufficient to prove that if $dom(f) \neq P$, then c holds. Assume that $dom(f) \neq P$.

We first argue that $dom(f)$ and $ran(f)$ are initial segments of P and Q respectively. Let $a \in dom(f)$ and $a' \in pred(a)$. Since $a \in dom(f)$, there exists $b \in ran(f) \subseteq Q$ such that $pred(a)$ and $pred(b)$ are isomorphic via some order-preserving function $h : pred(a) \to pred(b)$. Let $b' = h(a')$. Then $pred(a')$ and $pred(b')$ are isomorphic via the restriction of h onto $pred(a')$. Thus, $(a', b') \in f$ and hence $dom(f)$ is an initial segment. By a symmetric argument, one can easily show that $ran(f)$ is also an initial segment.

Assume that $ran(f)$ is a proper initial segment. Then there exists $b \in Q$ such that $ran(f) = pred(b)$. But then, since $f^{-1}[pred(b)]$ is an initial segment of P, it is of the form $pred(a)$ for some $a \in P$. Consequently, $(a, b) \in f$ which implies that $b \in ran(f) = pred(b)$, which is a contradiction. Therefore, $ran(f) = Q$ and hence

Q is order-isomorphic to the initial segment $dom(f)$ via the function f^{-1}. This completes the proof. □

3.5. **Well-founded relations and the Axiom of Foundation.** In this subsection, we will learn an important generalization of the notion of a strict well-order relation, namely, the notion of a well-founded relation.

Definition 41. *Let E be a relation on a set X. The relation E is said to be well-founded if $\forall M \subseteq X\ (M \neq \emptyset \rightarrow (\exists m \in M\ \forall s \in M\ (s, m) \notin E))$.*

In other words, a relation E is well-founded if every non-empty subset M has an *E-minimal* element, i.e. an element to which there are no elements in M that are related under the relation E. For example, any strict well-order relation is well-founded.

For technical reasons we shall not explain at this moment, we want the membership relation $\in_X$ on a set X to be well-founded. In order to be able to prove this, we assert the following axiom.

Axiom 8 (The Axiom of Foundation)**.** *If S is a non-empty set, then there exists an element of S which is disjoint from S, i.e.*

$$\forall S(S \neq \emptyset \rightarrow (\exists s \in S\ s \cap S = \emptyset))$$

Theorem 10. *Let X be a set and $\in_X$ be the membership relation on X. Then the relation $\in_X$ is well-founded.*

Proof. If X is empty, then $\in_X$ is the empty relation which is well-founded. If X is non-empty, let $M \subseteq X$ be any non-empty set. We wish to show that there exists an $\in_X$-minimal element of M. Since M is non-empty, then, by the axiom of foundation, there exists $m \in M$ with $m \cap M = \emptyset$. If it were the case that $\exists x \in M\ (x, m) \in \in_X$, then $x \in (m \cap M)$ which is a contradiction. Thus, $\forall x \in M\ (x, m) \notin X$, which shows that $m \in M$ is $\in_X$-minimal and hence $\in_X$ is well-founded. □

Another consequence of the axiom of foundation is that there are no sets that are members of themselves.

Theorem 11. *For any set x, we have that $x \notin x$.*

Proof. Assume towards a contradiction that there exists a set x such that $x \in x$. Let y be the set $\{x\}$. Since y is non-empty, by foundation, there exists $z \in y$ such that $z \cap y = \emptyset$. However, the only element of y is the set x and $y \cap x = \{x\}$, which is a contradiction. Thus there cannot exist such a set x. □

One can similarly prove with a similar argument that there cannot be sets which form a "loop" under the membership relation.

Exercise 24. *There are no sets x an y such that we have both $x \in y$ and $y \in x$.*

Next will be shown that the *successor* operation, which will be central to the following sections, is one-to-one.

Definition 42. *Let x be a set. The successor of x is defined to be the set $x \cup \{x\}$ and is denoted by $S(x)$.*

Lemma 13. *Let x and y be sets. If $S(x) = S(y)$, then $x = y$.*

Proof. Assume that $S(x) = S(y)$. Then, by definition, $x \cup \{x\} = y \cup \{y\}$. It follows that $x \in y \ \vee \ x = y$ and $y \in x \ \vee \ y = x$. If $x \neq y$, then we have both $x \in y$ and $y \in x$, which contradicts the previous exercise. Thus, $x = y$. $\square$

4. Natural Numbers

As we have discussed before, ZFC is supposed to be a foundation of mathematics. Consequently, we should be able to "represent" numbers as sets. In this section, we shall construct the set of natural numbers together with its usual arithmetic operations.

4.1. **The construction of the set of natural numbers.** The idea is to define the natural number n to be a set with n elements. However, since there are infinitely many such sets, we should choose a "canonical" set with n elements to be the natural number n.

We define the natural number 0 to be the empty set $\emptyset$. The natural number n is the set obtained by applying the successor operation to the empty set n times. In other words,

$$\begin{aligned} 0 &= \emptyset \\ 1 &= S(0) = \{0\} \\ 2 &= S(S(0)) = S(1) = \{0, 1\} \\ 3 &= S(S(S(0))) = S(2) = S(S(1)) = \{0, 1, 2\} \\ &\dots \end{aligned}$$

Notice that each *specific* natural number can be constructed from the axioms introduced so far. However, we cannot prove without additional axioms that there exists a set which contains $0, S(0), S(S(0)), \dots$. Before we introduce an axiom that asserts the existence of such a set, we shall define the notion of an inductive set.

Definition 43. *A set x is said to be inductive if*

- $\emptyset \in x$ *and*
- $\forall y \ (y \in x \to S(y) \in x)$.

Observe that an inductive set is "infinite" and necessarily contains $0, S(0), \dots$. However, there may be other elements contained in an inductive set. We would like the set of natural numbers to be the smallest inductive set. Of course, in order to construct this set, we first have to assert that an inductive set exists.

Axiom 9 (The Axiom of Infinity)**.** *An inductive set exists, i.e.*

$$\exists x \ (\emptyset \in x \ \wedge (\forall y (y \in x \to S(y) \in x)))$$

By the axiom of infinity, we know that there exists an inductive set I. It follows from the axiom of separation that the collection

$$\{x \in I : \forall J \ (\text{“}J \text{ is inductive”} \to x \in J)\}$$

is a set. This set is called *the set of natural numbers* and is denoted by $\mathbb{N}$. Any element of $\mathbb{N}$ is said to be a *natural number*.

A trivial but important observation is that $\mathbb{N}$ is inductive and that $\mathbb{N} \subseteq I$ for every inductive set I. The following principle immediately follows from this observation.

Theorem 12 (The Principle of Induction)**.** *Let $\varphi(x)$ be a property of sets. If $I = \{n \in \mathbb{N} : \varphi(n)\}$ is inductive, then $I = \mathbb{N}$.*

Proof. Assume that $I = \{n \in \mathbb{N} : \varphi(n)\}$ is inductive. Then $\mathbb{N} \subseteq I$ by our previous observation. On the other hand, by definition, $I \subseteq \mathbb{N}$. Thus, $I = \mathbb{N}$. □

The principle of induction is a fundamental tool to prove statements about the set of natural numbers. It is easily checked that the principle of induction holds even if we allow parameters $\varphi(x, y, z, \ldots, w)$ in the property defining the set I in the statement. We should perhaps summarize the induction principle as follows: Any inductive subset of the set of natural numbers equals the set of natural numbers.

We next define the order relation on $\mathbb{N}$. Define the relation $<$ on $\mathbb{N}$ as follows

$$m < n \leftrightarrow m \in n$$

for all $m, n \in \mathbb{N}$. At this point, we do not know that $<$ is a strict order relation. In order to prove this, we shall some basic properties of $<$.

Lemma 14. *For all $n \in \mathbb{N}$, we have $0 \leq n$, where $\leq$ is the relation defined by $x \leq y \leftrightarrow x < y \vee x = y$.*

Proof. We shall prove this by the principle of induction. Let I be the set of natural numbers for which the claim holds. We clearly have $0 \in I$. Now, let $n \in I$. Then, $0 \in n$ or $0 = n$. In both cases, $0 \in S(n)$ and hence $0 \leq S(n)$. Thus, I is inductive and hence, by induction $I = \mathbb{N}$. □

Exercise 25. *Prove that for all $k, n \in \mathbb{N}$, we have $k < S(n) \leftrightarrow k \leq n$.*

Exercise 26. *Using the principle of induction, prove that for all $n \in \mathbb{N}$, we have that $n = 0$ or $n = S(k)$ for some $k \in n$.*

We shall next establish that the set of natural numbers together with the relation defined above is a strictly well-ordered set.

Theorem 13. *$(\mathbb{N}, <)$ is a strictly well-ordered set.*

Proof. The proof of this theorem is a long induction argument.

- $<$ is transitive.
  Let $I = \{z \in \mathbb{N} : \forall x, y \in \mathbb{N}(x < y \ \wedge \ y < z \rightarrow x < z)\}$. We shall prove by induction that $I = \mathbb{N}$. It is easily seen that $0 \in I$ since the property defining elements of I vacuously holds for 0. To complete the proof that I is inductive, we need to prove that $n \in I \rightarrow S(n) \in I$. Let $n \in I$ and $x, y \in \mathbb{N}$ such that $x < y$ and $y < S(n)$. Since $y < S(n)$, by previous exercise, we know that $y \leq n$ and hence $y < n$ or $y = n$. If $y < n$, then $x < n$ since $n \in I$. If $y = n$, then $x < n$ by assumption. Since I is inductive, it equals $\mathbb{N}$ and hence $<$ is transitive.
- $<$ is asymmetric.
  Let $x, y \in \mathbb{N}$ such that $x < y$. Then, by definition, $x \in y$. It follows from an exercise in the previous section that $y \notin x$. Thus $\neg y < x$ and hence $<$ is asymmetric[1].
- Any two natural numbers are comparable with respect to $<$.
  Let $I = \{n \in \mathbb{N} : \forall m \in \mathbb{N} \ (m < n \ \vee m = n \ \vee n < m)\}$. By Lemma 14, we know that $0 \in I$. Let $n \in I$. We wish to show that $S(n) \in I$. Let $m \in \mathbb{N}$. We know that $m < n \ \vee m = n \ \vee n < m$. If either one of the first two cases holds, then $m \in S(n)$ and so $m < S(n)$. Thus, in order to complete the proof that $S(n) \in I$, it is sufficient to prove that if $n < m$, then $S(n) \leq m$.

[1]Notice that we used the axiom of foundation here. In truth, we do not need the axiom of foundation to prove this statement. Proving this statement via the principle of induction is left the reader as an exercise

We shall prove this via another induction argument.

Let
$$J = \{j \in \mathbb{N} : \forall k \in \mathbb{N}\ (k < j \to S(k) \leq j)\}$$
Clearly, $0 \in J$. Let $j \in J$. We wish to prove that $S(j) \in J$. To prove this, pick $k \in \mathbb{N}$ such that $k < S(j)$. Then, either $k \in j$ or $k = j$. If $k \in j$, then we have $S(k) \leq j$ since $j \in J$. But this means that $S(k) < S(j)$. If $k = j$, then we have $S(k) = S(j)$ and hence $S(k) \leq S(j)$. Thus, $S(j) \in J$ and hence J is inductive. Consequently, $J = \mathbb{N}$.

Going back to the main proof. If $n < m$, then $S(n) \leq m$ by the proof above. Thus, $S(n) \in I$, which completes the proof I is inductive and hence equals $\mathbb{N}$.

- Every non-empty subset of $\mathbb{N}$ has a least element with respect to $<$.
  This proof is left to the reader as an exercise.

□

Exercise 27. *Using the principle of induction, prove that $(n, \in_n)$ is a strictly well-ordered set for all $n \in \mathbb{N}$, where $\in_n$ is the membership relation on n.*

Now that we have defined the set of natural numbers, we provide some definitions that were promised to the reader earlier.

Definition 44. *Let X be a set and $n \in \mathbb{N}$. A sequence of length n over the set X is a function $f : n \to X$. A sequence (indexed by $\mathbb{N}$) over the set X is a function $f : \mathbb{N} \to X$.*

As noted earlier, while working with sequences, it is convenient to use the notation $(f_i)_{i\in n}$ or $(f_0, f_1, \ldots, f_{n-1})$ to denote the sequence $f : n \to X$, where $f_i = f(i)$. Similarly, we shall write $(f_i)_{i\in\mathbb{N}}$ to denote a sequence $f : \mathbb{N} \to X$ indexed by $\mathbb{N}$.

Next will be given the definition of the cartesian product of finitely many sets, which is simply a special case of the product of indexed systems.

Definition 45. *Given an indexed system of sets $\{A_i\}_{i\in n}$ indexed by some natural number $n \in \mathbb{N}$, we define the (n-fold) cartesian product $A_0 \times A_1 \times \ldots A_{n-1}$ to be the product $\prod\{A_i\}_{i\in n}$.*

We note that all properties of products of arbitrary indexed systems holds for n-fold cartesian products. If $\{A_i\}_{i\in n}$ is an indexed system of sets such that $A_i = X$ for all $i \in n$ for some fixed set X, then we use the notation X^n to denote the n-fold cartesian product $A_0 \times A_1 \times \cdots \times A_{n-1}$ since this set is identically the set of functions from n to X, for which we use the notations X^n or ${}^nX$. Before we conclude this subsection, we finally define n-ary operations on a set.

Definition 46. *Let X be a set and $n \in \mathbb{N}$. An n-ary operation on the set X is a function from X^n to the set X.*

It is common practice to call 1-ary, 2-ary and 3-ary functions *unary*, *binary* and *ternary* functions respectively. Notice that there exist natural bijections between X and X^1, and $X \times X$ and X^2, and $(X \times X) \times X$ and X^3, and so on[2]. For this reason, we shall often not make a distinction between these sets and use them interchangeably.

[2]For the case $n = 2$, see Exercise 13. This idea can easily be generalized to other values of n.

4.2. **Arithmetic on the set of natural numbers.** In this subsection, we shall define the usual arithmetic operations on the set of natural numbers. It is convenient to define these operations recursively. In order to be able to justify the usage of recursive definitions, we shall next prove the following fundamental theorem.

Theorem 14 (The Recursion Theorem)**.** *Let X be a non-empty set, $x \in X$ and $f : X \to X$ be a function. Then there exists a unique function $g : \mathbb{N} \to X$ such that*

- $g(0) = x$ *and*
- $g(S(n)) = f(g(n))$ *for all* $n \in \mathbb{N}$.

The function g, considered as a sequence over the set X with the index set $\mathbb{N}$, is simply the sequence $(x, f(x), f(f(x)), \dots)$. Intuitively speaking, the function f can be considered as instructions to compute the value of g at the successor of a natural number using the value of g at that natural number.

Proof. For a natural number $n \in \mathbb{N}$, call a function $t : S(n) \to X$ an n-step computation if $t(0) = x$ and $t(S(k)) = f(t(k))$ for all $k \in S(n)$. Define

$$G = \{t \subseteq \mathbb{N} \times X : \exists n \in \mathbb{N} \text{ “}t \text{ is an n-step computation”}\}$$

Let $g = \bigcup G$. We shall prove several claims.

- g is a function.
  By Exercise 10, it is sufficient to prove that element of G are compatible functions. Let $t, u \in G$ and set $n = dom(t)$ and $m = dom(u)$. Without loss of generality, assume that $n \in m$. Then it follows from an easy induction argument that $n \subseteq m$. If t and u were not compatible, then there would be a least $k' \in n$ such that $t(k') \neq u(k')$. Since $t(0) = u(0) = x$, we know that $k' \neq 0$. Then, there exists $k \in n$ such that $k' = S(k)$. But then, since $t(k) = u(k)$, $t(S(k)) = f(t(k)) = f(u(k)) = u(S(k))$, which is a contradiction. Thus, t and u are compatible.
- $dom(g) = \mathbb{N}$
  Since we have $dom(g) \subseteq \mathbb{N}$, it suffices to show that $dom(g)$ is inductive. Observe that $\{(0, x)\}$ is a 0-step computation and hence $0 \in dom(g)$. Let $n \in dom(g)$. Then there exists $t \in G$ such that $n \in dom(t)$. If it is the case that $S(n) \in dom(t)$, then we are done. If not, then $S(n) = dom(t)$. Let
  $$t' = t \cup \{(S(n), f(t(n)))\}$$
  It is easily checked that t' is an $S(n)$-step computation, so $S(n) \in dom(g)$. It follows that $dom(g)$ is inductive and hence equals $\mathbb{N}$.
- g satisfies conditions in the statement.
  This part is left to the reader as an exercise.
- g is unique.
  Let $h : \mathbb{N} \to X$ be a function that satisfies the conditions in the statement. Let $I = \{n \in \mathbb{N} : g(n) = h(n)\}$. Clearly we have $0 \in I$. Assume that $n \in I$. Then $g(S(n)) = f(g(n)) = f(h(n)) = h(S(n))$ and hence $S(n) \in I$. By the principle of induction, $I = \mathbb{N}$ and hence $g = h$.

This completes the proof of the theorem. □

The recursion theorem has many variants that allow “parameters” in the recursive definition. Here we shall not present these variants and simply assume that our usage of recursive definitions in the rest of these notes is justified. We refer the curious reader to the textbook for proofs of these variants.

We shall next define the binary operation of addition $+$ on the set $\mathbb{N}$ of natural numbers by recursion. For any $m, n \in \mathbb{N}$, define

$$+(m, 0) = m$$

$$+(m, S(n)) = S(+(m, n))$$

Notice that we are defining the value of $+$ at the pair $(m, S(n))$ using its value at the pair (m, n) recursively. Why does there exist such a binary operation?

It follows from the recursion theorem (and also from an easy induction argument) that for each $m \in \mathbb{N}$, there exists a (unique) function $f_m : \mathbb{N} \to \mathbb{N}$ such that $f_m(0) = m$ and $f_m(S(n)) = S(f_m(n))$ for all $n \in \mathbb{N}$. Using the axioms we have introduced so far, one can easily prove that the collection

$$\{((m, n), p) \in \mathbb{N}^2 \times \mathbb{N} : \ f_m(n) = p\}$$

is a set and indeed is a binary operation[3] from $\mathbb{N}^2$ to $\mathbb{N}$.

Lemma 15. *For all $n \in \mathbb{N}$, $+(n, 0) = n$.*

Proof. This follows from the fact that $f_n(0)$ for all $n \in \mathbb{N}$. □

Lemma 16. *For all $m, n \in \mathbb{N}$, $+(m, S(n)) = S(+(m, n))$.*

Proof. Let $m, n \in \mathbb{N}$. Then, $+(m, S(n)) = f_m(S(n)) = S(f_m(n)) = S(+(m, n))$ by the properties of each f_m. □

From now on, we shall use the usual notation $m+n$ to denote the value $+(m, n)$. Next are proven some basic properties of the addition operation.

Lemma 17. *For all $m, n \in \mathbb{N}$, $m + S(n) = S(m) + n$.*

Proof. We prove this by induction on n. Let

$$I = \{n \in \mathbb{N} : \forall m \in \mathbb{N}\ m + S(n) = S(m) + n\}$$

We wish to prove that I is inductive.

- $m + S(0) = S(m + 0) = S(m) = S(m) + 0$ by the previous lemmas and hence $0 \in I$.
- Assume that $n \in I$. It follows from the previous lemmas and $n \in I$ that $m + S(S(n)) = S(m + S(n)) = S(S(m) + n)) = S(m) + S(n)$ and hence $S(n) \in I$.

This show that I is inductive and hence $I = \mathbb{N}$ by the principle of induction. □

Exercise 28. *Using the principle of induction, show that $0 + n = n$ for all $n \in \mathbb{N}$.*

Lemma 18. *For all $m, n \in \mathbb{N}$, $m + n = n + m$.*

Proof. We prove this by induction on n. Let $I = \{n \in \mathbb{N} : \forall m \in \mathbb{N}\ m+n = n+m\}$. We wish to prove that I is inductive.

- By the previous exercise, $0 + m = m = m + 0$ for all $m \in \mathbb{N}$ and hence $0 \in I$.

[3]That this relation is indeed a function follows from the uniqueness of f_m for each $m \in \mathbb{N}$.

- Assume that $n \in I$. Let $m \in \mathbb{N}$. Then, since $n \in I$,
$$m + S(n) = S(m + n) = S(n + m) = n + S(m)$$
On the other hand, we know that $n + S(m) = S(n) + m$ by the previous lemma. Thus
$$m + S(n) = n + S(m) = S(n) + m$$
This shows that $S(n) \in I$.

Thus I is inductive and hence $I = \mathbb{N}$, which completes the proof. □

Using the recursion theorem, one can similarly prove that there exists a binary operation $\cdot$ on $\mathbb{N}$ such that
$$m \cdot 0 = 0$$
$$m \cdot S(n) = (m \cdot n) + m$$
for all $m, n \in \mathbb{N}$, which is the multiplication operation on $\mathbb{N}$. The exponentiation operation on $\mathbb{N}$ can be similarly defined as the unique binary operation on $\mathbb{N}$ such that for all $m, n \in \mathbb{N}$,
$$m^0 = 1$$
$$m^{S(n)} = m^n \cdot m.$$
We shall not prove here all properties of these operations such as commutativity, associativity and distributivity. The reader is expected to imitate the proofs in this subsection to prove the usual identities involving these operations.

Having defined recursion on the set of natural numbers, we give a characterization of a linearly ordered set being well-ordered using infinite descending sequences.

Lemma 19. *Let $(W, \prec)$ be a strictly linearly ordered set. Then $(W, \prec)$ is strictly well-ordered if and only if there exists no sequence $(w_i)_{i \in \mathbb{N}}$ of elements of W such that $w_{i+1} \prec w_i$ for all $i \in \mathbb{N}$.*

Proof. Assume that $(W, \prec)$ is strictly well-ordered. Then, for any sequence $(w_i)_{i \in \mathbb{N}}$ of elements of W, the set $\{w_i : i \in \mathbb{N}\}$ is non-empty and hence has a least element with respect to $\preceq$ which is w_k for some $k \in \mathbb{N}$. Then $w_{k+1} \not\prec w_k$.

Now assume that $(W, \prec)$ is not strictly well-ordered. Then there exists a non-empty subset $S \subseteq W$ which does not have a least element with respect to $\preceq$. Since S is non-empty, we can select an element $w_0 \in S$. By recursion, define w_{i+1} to be some element of $S_i = \{w \in S : w \prec w_i\}$ if this set is non-empty[4]; and w_0 otherwise.

Since S does not have a least element with respect to $\preceq$, it can easily be proven via induction that S_i is non-empty for all $i \in \mathbb{N}$. Consequently, the sequence $(w_i)_{i \in \mathbb{N}}$ of elements of W satisfy the property that $w_{i+1} \prec w_i$ for all $i \in \mathbb{N}$. □

[4]Even though it is a minor point, we note that we are using the axiom of choice in order to choose an element from S_i for every $i \in \mathbb{N}$ whenever it is possible.

5. Equinumerosity

In this section, we shall learn the notion of equinumerosity which allows us to compare the "size" of sets. Historically speaking, this concept is what led to the development of set theory.

Definition 47. *Two sets A and B are said to be equinumerous if there exists a bijection from A to B, in which case we write $|A| = |B|$.*

If two sets are equinumerous, then we can use the elements of one of them as labels to "count" the elements of the other set. For this reason, we consider equinumerous sets as sets that have the same "size".

5.1. Finite sets.

Having constructed natural numbers, we can now give a precise meaning to the notion of "finite".

Definition 48. *A set A is said to be finite if there exists $n \in \mathbb{N}$ such that A is equinumerous with the natural number n, in which case we say A has n elements and write $|A| = n$.*

By definition, each natural number is a finite set since it is equinumerous with itself via the identity function. The following lemma shows that a natural number cannot be equinumerous with a proper subset of itself.

Lemma 20. *For each $n \in \mathbb{N}$, any injective function from n to n is surjective.*

Proof. We prove this by induction on n. The claim vacuously holds for $n = 0$. Assume that it holds for $n \in \mathbb{N}$. We wish to show that it also holds for $S(n)$. Assume to the contrary that there exists a function $f : S(n) \to S(n)$ which is injective but not surjective.

- If $n \notin ran(f)$, then $f \restriction_n$ is an injective function from n to $n - \{f(n)\}$ which contradicts the induction assumption.
- If $n \in ran(f)$, then $n = f(k)$ for some (unique) $k \in S(n)$. If $k = n$, then the function $f \restriction_n$ is an injective function from n to n which is not surjective, contradicting the induction assumption. If $k \neq n$, then the relation

  $$g = (f - \{(k, n), (n, f(n))\}) \cup \{(k, f(n))\}$$

  is an injective function from n to n such that $ran(g) = ran(f) - \{n\}$. Since $ran(f) \neq S(n)$ and $n \in ran(f)$, we have that $ran(g) \neq n$. Thus, g is an injective function from n to n which is not surjective, contradicting the induction assumption.

Therefore, the claim holds for $S(n)$ and hence holds for all $n \in \mathbb{N}$ by induction. □

This lemma has some important corollaries.

Corollary 15. *The set $\mathbb{N}$ is not finite.*

Proof. Since there exists an injective function on $\mathbb{N}$ which is not surjective, if $\mathbb{N}$ were equinumerous with some natural number $n \in \mathbb{N}$, then there would exist an injective function on n that is not surjective, which contradicts the previous lemma. □

Corollary 16 (The Pigeonhole Principle)**.** *Let $m, n \in \mathbb{N}$ such that $m < n$. Then there does not exist an injective function from n to m.*

Proof. Assume to the contrary that $f : n \to m$ is an injective function. Since $m \subseteq n$, the function $f \restriction_m$ is an injective function from m to $m - \{f(m)\}$, contradicting the previous lemma. □

Exercise 29. *Using the previous lemma, show that if S is finite and equinumerous with both $n \in \mathbb{N}$ and $m \in \mathbb{N}$, then $m = n$.*

We will next prove some basic properties of finite sets.

Theorem 17. *Any subset of a finite set is finite.*

Proof. We shall prove this by induction. Let

$$I = \{n \in \mathbb{N} : \forall X(|X| = n \to \text{Every subset of X is finite})\}$$

Clearly $0 \in I$ since the unique subset of the empty set is the empty set, which is finite. Assume that $n \in I$, that is, the claim holds for all finite sets with n elements. We wish to prove that $S(n) \in I$, that is, the claim holds for all finite sets with $S(n)$ elements. Let A be a finite set such that $|A| = S(n)$, say $f : S(n) \to A$ is a bijection. Let $B \subseteq A$. To complete the inductive step, we need to prove that B is finite.

If $B \subseteq f[n]$, then, since $f[n]$ is a finite set with n elements, B is finite by the induction assumption. If $B \nsubseteq f[n]$, then $f(n) \in B$ and $B - \{f(n)\} \subseteq f[n]$. In this case, $B - \{f(n)\}$ is finite by the induction assumption and hence there exists a bijection $h : m \to B - f(n)$. Then $h \cup \{(m, f(n))\}$ is a bijection from $S(m)$ to B, which shows that B is finite.

Thus the claim holds for finite sets with $S(n)$ elements and hence $S(n) \in I$. By the principle of induction, $I = \mathbb{N}$ and hence the claim is true for all finite sets. □

Lemma 21. *If X and Y are finite sets, then $X \cup Y$ is finite.*

Proof. Let X and Y be finite sets and $f : n \to X$ and $g : m \to Y$ be bijections. Then the function $h : n + m \to X \cup Y$ given by

$$h(i) = \begin{cases} f(i) & \text{if } i < n \\ g(i - n) & \text{if } n \leq i < n + m \end{cases}$$

is a surjection. For each $z \in X \cup Y$, let w_z be the least element of $h^{-1}[\{z\}]$. Then the map $F : X \cup Y \to n + m$ given by $F(z) = w_z$ for all $z \in X \cup Y$ is an injection. By the previous lemma, $F[X \cup Y]$ is a subset of a finite set and hence is finite. On the other hand, $X \cup Y$ is equinumerous to the set $F[X \cup Y]$ and hence is finite. □

Lemma 22. *The power set of a finite set is finite.*

Proof. We shall prove this by induction. Let

$$I = \{n \in \mathbb{N} : \forall X(|X| = n \to \mathcal{P}(X) \text{ is finite})\}$$

Clearly $0 \in I$ since $|\mathcal{P}(\emptyset)| = 1$. Assume that $n \in I$. We wish to show that $S(n) \in I$. Let X be a set with $S(n)$ elements, say $f : S(n) \to X$ is a bijection. Notice that $\mathcal{P}(X) = \mathcal{P}(f[n]) \cup \{Y \cup \{f(n)\} : \ Y \in \mathcal{P}(f[n])\}$ and that both sets on the right hand side are finite by the induction assumption since they are equinumerous with power sets of sets of n elements. By the previous lemma, the union of two finite sets is finite and hence $\mathcal{P}(X)$ is finite. Therefore, $S(n) \in I$ and hence $I = \mathbb{N}$ by the principle of induction. □

5.2. **To infinity and beyond.** In this subsection, we shall investigate some basic properties of infinite sets. The reader is expected to develop a good understanding of infinite sets since modern set theory, at its core, is simply the study of the behaviour of infinite sets.

Definition 49. *A set X is said to be infinite if it is not finite, i.e. there does not exist a natural number $n \in \mathbb{N}$ such that X and n are equinumerous.*

Even though this is the "official" definition of being infinite, we will often use alternative characterizations of being infinite in order to show that certain sets are infinite.

Definition 50. *A set X is said to be Dedekind-infinite if there exists an injection from X onto a proper subset of X.*

Next will be proven that being infinite is equivalent to being Dedekind-infinite.

Lemma 23. *Let X be an infinite set. Then there exists an injection from $\mathbb{N}$ to X.*

Proof. By Lemma 6, there exists a function $g : \mathcal{P}(X) - \{\emptyset\} \to X$ such that $g(U) \in U$ for all $U \in \mathcal{P}(X) - \{\emptyset\}$. Let $x \in X$ be a fixed element. Let $G : \mathcal{P}(X) \to X$ be the function given by $G(U) = g(U)$ for non-empty $U \subseteq X$ and $G(\emptyset) = x$. By recursion, define a function $f : \mathbb{N} \to X$ as follows.

- $f(0) = x$
- $f(S(i)) = G(X - \{f(0), f(1), \ldots, f(i)\})$ for all $i \in \mathbb{N}$.

An easy induction argument shows that $X - \{f(0), \ldots, f(i)\} \neq \emptyset$ for all $i \in \mathbb{N}$ since X is infinite. It follows that for all $i \in \mathbb{N}$ we have that $f(S(i)) \neq f(j)$ for each $0 \leq j \leq i$. This implies that f is an injective function. □

Theorem 18. *A set is infinite if and only if it is Dedekind-infinite.*

Proof. Let X be a set. Assume that X is infinite. Then, by the previous lemma, there exists an injection $f : \mathbb{N} \to X$. Consider the function $g : X \to X$ given by

$$g(x) = \begin{cases} x & \text{if } x \notin ran(f) \\ f(k+1) & \text{if } x = f(k) \text{ for some } k \in \mathbb{N} \end{cases}$$

It is easily checked that g is an injection and $ran(g) = X - \{f(0)\}$ and hence X is Dedekind-infinite.

For the converse direction, assume that X is Dedekind-infinite, say $g : X \to X$ is an injection which is not a surjection. If X were equinumerous with some natural number $n \in \mathbb{N}$ via some bijection $h : X \to n$, then $h \circ g \circ h^{-1} : n \to n$ would be an injection which is not a surjection, contradicting Lemma 20. Thus X is infinite. □

Having established another characterization of being infinite, we next define an important class of infinite sets.

Definition 51. *Let A be a set. Then A is said to be countably infinite if A and $\mathbb{N}$ are equinumerous, and said to be at most countable (or simply, countable) if it is either finite or countably infinite.*

We next prove a series of lemmas that will eventually be used to prove that the set of sequences of finite length over an at most countable set is at most countable.

Lemma 24. *If A is at most countable, then there is a surjection from $\mathbb{N}$ onto A.*

Proof. Trivial. □

Lemma 25. *If there exists a surjection from A to B, then there exists an injection from B to A.*

Proof. Let $f : A \to B$ be a surjection. Let U_b denote the set $f^{-1}[\{b\}]$ for each $b \in B$. Since U_b is non-empty for each $b \in B$, Lemma 6 implies that there exists $G : \{U_b : b \in B\} \to A$ such that $G(U_b) \in U_b$. It is easily seen that the function $g : B \to A$ defined by $g(b) = G(U_b)$ is an injective function. □

Lemma 26. *The set $\mathbb{N} \times \mathbb{N}$ is countably infinite.*

Proof. The function $f : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ given by $f(m, n) = 2^m(2n+1)-1$ is a bijection since every positive natural number can be uniquely expressed as a product of a power of two and an odd number. □

An immediate corollary of this lemma is the following fact.

Corollary 19. *If A and B are countably infinite sets, then $A \times B$ is countably infinite.*

Proof. Exercise. □

Lemma 27. *Any subset of a countably infinite set is at most countable.*

Proof. Let A be a countably infinite set and $B \subseteq A$. Let $f : \mathbb{N} \to A$ be a bijection. If B is finite, then it is at most countable by definition. Assume that B is infinite. We need to prove that B and $\mathbb{N}$ are equinumerous. Let $b = f(k)$ where k is the least natural number such that $f(k) \in B$, which exists since B is infinite. By recursion, define the function $g : \mathbb{N} \to B$ as follows.

- $g(0) = b$
- $g(S(i)) = f(k)$ where k is the least natural number such that $f(k) \in B$ and $f(k) \notin \{g(0), g(1), \ldots, g(i)\}$. (Notice that such k always exists since B is infinite.)

In this case, g is a bijection from $\mathbb{N}$ to B, which completes the proof that B is at most countable. We leave the details of checking that g is a bijection to the reader. □

Lemma 28. *Let $\{A_i\}_{i\in\mathbb{N}}$ be an indexed system of sets such that A_i is at most countable for all $i \in \mathbb{N}$. Then $\bigcup_{i\in\mathbb{N}} A_i$ is at most countable.*

Proof. Since each A_i is at most countable, the set B_i consisting of surjections from $\mathbb{N}$ onto A_i is non-empty for all $i \in \mathbb{N}$. Consequently, we can choose a surjection $g_i : \mathbb{N} \to A_i$ from each B_i for each $i \in \mathbb{N}$. Let $f : \mathbb{N} \times \mathbb{N} \to \bigcup_{i\in\mathbb{N}} A_i$ be the function given by $f(m, n) = g_m(n)$. Since each g_m is a surjection, we have that f is a surjection. Thus, by Lemma 25, there exists an injection from $\bigcup_{i\in\mathbb{N}} A_i$ to $\mathbb{N} \times \mathbb{N}$. However, the latter set is countably infinite and it follows from the previous lemma that $\bigcup_{i\in\mathbb{N}} A_i$ is at most countable. □

Exercise 30. *Using induction, show that if A is an at most countable set, then A^n is at most countable for any $n \in \mathbb{N}$.*

Recall that, given a natural number $n \in \mathbb{N}$, we defined a sequence of length n over a set A to be a function from n to A. Thus the set of sequences of finite length over a set A is simply $\bigcup_{n\in\mathbb{N}} A^n$.

Lemma 29. *Let A be an at most countable set. Then the set $\bigcup_{n\in\mathbb{N}} A^n$ of sequences of finite length over A is at most countable.*

Proof. By the previous exercise, A^n is countable for all $n \in \mathbb{N}$. By Lemma 28, we know that countable union of at most countable sets is at most countable and hence $\bigcup_{n\in\mathbb{N}} A^n$ is at most countable. □

Are all sets countable?.. The answer to this question turns out to be negative, as was proven by Georg Cantor in 1874 in his groundbreaking paper "Über eine Eigenschaft des Inbegriffes aller reellen algebraischen Zahlen". Historically speaking, this was the birth of set theory. Here we shall present not the original proof of this observation but another proof which is also due to Cantor. Before we present this proof, we introduce some terminology.

Definition 52. *Let A and B be sets. Then we say that the cardinality of A is*

- *less than equal to the cardinality of B, denoted by $|A| \leq |B|$, if there exists an injection from A to B.*
- *strictly less than the cardinality of B, denoted by $|A| < |B|$, if there exists an injection from A to B but there is no bijection from A to B.*

We shall next prove a remarkable theorem which is usually referred to as Cantor's theorem. Cantor's theorem has an elegant and simple proof that involves the technique called *diagonalization*[1].

Theorem 20 (Cantor's Theorem)**.** *For any set X, we have $|X| < |\mathcal{P}(X)|$, i.e. the cardinality of X is strictly less than the cardinality of its power set $\mathcal{P}(X)$.*

Proof. It is easily seen that the function $g : X \to \mathcal{P}(X)$ given by $g(x) = \{x\}$ for all $x \in X$ is an injection. Thus it is sufficient to prove that there cannot be any bijection from X to $\mathcal{P}(X)$.

Let $f : X \to \mathcal{P}(X)$ be any function. Consider the set $W = \{x \in X : x \notin f(x)\}$. Clearly we have $W \in \mathcal{P}(X)$. We claim that $W \notin ran(f)$. Assume to the contrary that $W \in ran(f)$. Then there exists $w \in X$ such that $f(w) = W$. It follows from the definition of W that

$$w \in W \leftrightarrow w \notin W$$

which is a contradiction. Therefore $W \notin ran(f)$. This shows that *no* function from X to $\mathcal{P}(X)$ is surjective, which completes the proof. □

A set is said to be *uncountable* if it is not at most countable. Cantor's theorem shows that, for example, the set $\mathcal{P}(\mathbb{N})$ is uncountable.

Exercise 31. *Prove that the set ${}^{\mathbb{N}}\mathbb{N}$ is uncountable. (**Hint**: Use Exercise 13.)*

Our next goal is to prove that $\mathcal{P}(\mathbb{N})$ and the set of real numbers $\mathbb{R}$ are equinumerous. However, we do not know at this point how to construct the set $\mathbb{R}$ and represent real numbers as sets. In the next section, we shall provide the construction of some standard number systems such as integers, rational numbers and real numbers.

Before that, we would like to conclude this section with another important theorem that is extremely useful for proving that various sets are equinumerous.

[1]The author personally considers the proof of Cantor's theorem as one of the most beautiful arguments in mathematics.

Theorem 21 (Cantor-Schröder-Bernstein)**.** *Let A and B be sets. If there exist injections $f : A \to B$ and $g : B \to A$, then there exists a bijection $h : A \to B$.*

Proof. Let $(A_n)_{n\in\mathbb{N}}$ and $(B_n)_{n\in\mathbb{N}}$ be sequences of subsets of A and B respectively, defined as follows by recursion.

$$A_0 = A \quad A_1 = g[B_0] \ \dots \ A_{n+1} = g[B_n] \dots$$
$$B_0 = B \quad B_1 = f[A_0] \ \dots \ B_{n+1} = f[A_n] \dots$$

It is easily seen that $A_{n+1} \subseteq A_n$ and $B_{n+1} \subseteq B_n$ for all $n \in \mathbb{N}$. Let $D = \bigcap_{n\in\mathbb{N}} A_n$ and $E = \bigcap_{n\in\mathbb{N}} B_n$. It follows from the definitions that

$$A = D \cup \bigcup_{n\in\mathbb{N}} A_n - A_{n+1} \text{ and } B = E \cup \bigcup_{n\in\mathbb{N}} B_n - B_{n+1}$$

For each even natural number $n \in \mathbb{N}$, define the function h_n from $A_n - A_{n+1}$ to $B_{n+1} - B_{n+2}$ by $h_n(a) = f(a)$. Similarly, for each odd natural number $n \in \mathbb{N}$, define the function h_n from $A_n - A_{n+1}$ to $B_{n-1} - B_n$ by $h_n(a) = g^{-1}(a)$. It is straightforward to check that the relation $\bigcup_{n\in\mathbb{N}} h_n$ is a bijective function from the set $\bigcup_{n\in\mathbb{N}} A_n - A_{n+1}$ to the set $\bigcup_{n\in\mathbb{N}} B_n - B_{n+1}$. Moreover, since f is injective, we have that

$$f[D] = f[\bigcap_{n\in\mathbb{N}} A_n] = \bigcap_{n\in\mathbb{N}} f[A_n] = \bigcap_{n\in\mathbb{N}} B_{n+1} = E$$

Thus, $\bigcup_{n\in\mathbb{N}} h_n \cup f \restriction_D$ is a bijection from A to B. □

Written using the notation we introduced, Cantor-Schröder-Bernstein theorem simply states that if $|A| \leq |B|$ and $|B| \leq |A|$, then $|A| = |B|$. Indeed, this is exactly the reason we chose the notation $|A| \leq |B|$ to denote the existence of an injection from A to B.

6. Construction of various number systems

Recall that ZFC is supposed to provide a foundation of mathematics and hence we should be able to treat all mathematical objects as sets. In this section, we shall construct the set $\mathbb{R}$ of real numbers, which is unarguably one of the most important mathematical objects. In order to construct the set of real numbers, we will need the set of integers and the set of rational numbers constructed first.

6.1. **Integers.** In this subsection, we will construct the set $\mathbb{Z}$ of integers together with its arithmetical operations and linear order relation, which turn $\mathbb{Z}$ into an ordered ring. Let $\sim$ be the relation on $\mathbb{N} \times \mathbb{N}$ defined by

$$(p, q) \sim (r, s) \longleftrightarrow p + s = q + r$$

for all $p, q, r, s \in \mathbb{N}$. It follows from Exercise 16 that $\sim$ is an equivalence relation. We define the set of integers to be the quotient set

$$\mathbb{Z} = \mathbb{N} \times \mathbb{N} \,/\, \sim$$

Intuitively speaking, the equivalence class $[(p, q)]_\sim$ is supposed to represent the integer "$p - q$". In other words, the integers

$$\ldots \quad -2 \quad -1 \quad 0 \quad 1 \quad 2 \quad \ldots$$

are represented by the sets

$$\ldots \quad [(0, 2)]_\sim \quad [(0, 1)]_\sim \quad [(0, 0)]_\sim \quad [(1, 0)]_\sim \quad [(2, 0)]_\sim \quad \ldots$$

For notational convenience, the equivalence class $[(p, q)]_\sim$ will be denoted by $[p, q]$ throughout this subsection. We will next define the arithmetical operations and the linear order structure on the set $\mathbb{Z}$ of integers.

Having the interpretation that $[p, q]$ is supposed to represent the integer "$p - q$" in mind, we wish to define $[p, q] +_\mathbb{Z} [r, s]$ to be the integer $[p + r, q + s]$ where $+$ denotes the addition defined on the set $\mathbb{N}$ of natural numbers. More precisely, we define $+_\mathbb{Z}$ to be the relation from $\mathbb{Z} \times \mathbb{Z}$ to $\mathbb{Z}$ given by

$$+_\mathbb{Z} = \{(([p, q], [r, s]), [u, v]) : p, q, r, s, u, v \in \mathbb{N} \wedge p + r = u \ \wedge q + s = v\}$$

At this point, it not clear that this relation is well-defined and hence defines a binary operation on $\mathbb{Z}$. The next lemma shows that $+_\mathbb{Z}$ is indeed well-defined.

Lemma 30. *$+_\mathbb{Z}$ is well-defined.*

Proof. Let $p, q, r, s, u, v, p', q', r', s', u', v'$ be in $\mathbb{N}$. Assume that

- $[p, q] = [p', q']$
- $[r, s] = [r', s']$ and
- $(([p, q], [r, s]), [u, v])$ and $(([p', q'], [r', s']), [u', v'])$ are both in $+_\mathbb{Z}$.

We wish to prove that $[u, v] = [u', v']$. Since $(([p, q], [r, s]), [u, v])$ is in $+_\mathbb{Z}$, it follows from the definition that $p + r = u$ and $q + s = v$. Similarly, $(([p', q'], [r', s']), [u', v'])$ being in $+_\mathbb{Z}$ implies that $p' + r' = u'$ and $q' + s' = v'$. On the other hand, since $[p, q] = [p', q']$ and $[r, s] = [r', s']$, we have $p+q' = p'+q$ and $r+s' = r'+s$. Combining these equations together and using the properties of addition on natural numbers, we obtain that

$$\begin{aligned}(p + q') + (r + s') &= (p' + q) + (r' + s)\\ (p + r) + (q' + s') &= (q + s) + (p' + r')\\ u + v' &= v + u'\end{aligned}$$

and hence we have $[u, v] = [u', v']$ by the definition of $\sim$. $\square$

The binary operation $+_{\mathbb{Z}}$ is called addition on $\mathbb{Z}$. Similarly, we can define a binary operation $\cdot_{\mathbb{Z}}$ called multiplication on $\mathbb{Z}$ as follows.

$$[p, q] \cdot_{\mathbb{Z}} [r, s] = [p \cdot r + q \cdot s, p \cdot s + q \cdot r]$$

where $\cdot$ refers to multiplication on $\mathbb{N}$. That the relation $\cdot_{\mathbb{Z}}$ is well-defined $\cdot_{\mathbb{Z}}$ can be proven with a proof similar to that of Lemma 30. Finally, consider the relation $<_{\mathbb{Z}}$ defined on $\mathbb{Z}$ as follows.

$$[p, q] \leq_{\mathbb{Z}} [r, s] \longleftrightarrow p + s \leq q + r$$

where $\leq$ refers to the usual linear order relation on $\mathbb{N}$. Together with $+_{\mathbb{Z}}$, $\cdot_{\mathbb{Z}}$ and $\leq_{\mathbb{Z}}$, the set of integers satisfies its usual properties. More precisely, we have that

Theorem 22. *The structure $(\mathbb{Z}, +_{\mathbb{Z}}, \cdot_{\mathbb{Z}}, \leq_{\mathbb{Z}})$ is an ordered ring.*

Since it is not essential for the objectives of this course, we shall not carry out the tedious task of proving this fact here and refer the reader to various textbooks and lecture notes that cover the construction of basic number systems. Having assumed Theorem 22 without proof, we will freely use various notations reserved for integers. For example, the inverse of an integer p with respect to $+_{\mathbb{Z}}$ will be denoted by $-p$.

We would like to mention that the well-known secondary school fact $\mathbb{N} \subseteq \mathbb{Z}$ is *not* true with our constructions of $\mathbb{N}$ and $\mathbb{Z}$. The reader can easily verify, for example, that $\emptyset = 0 \notin \mathbb{Z}$ and but $\emptyset = 0 \in \mathbb{N}$. However, there exists a canonical *copy* of $\mathbb{N}$ inside $\mathbb{Z}$, namely, the range of the function $f : \mathbb{N} \to \mathbb{Z}$ given by

$$f(n) = [n, 0]$$

for all $n \in \mathbb{N}$. Identifying $\mathbb{N}$ and $f[\mathbb{N}]$, we may consider $\mathbb{N}$ as a subset of $\mathbb{Z}$. Consequently, we may use the Arabic numerals $0, 1, 2, \ldots$ to denote both the actual natural numbers

$$\emptyset, \quad S(\emptyset), \quad S(S(\emptyset)), \ \ldots$$

and the corresponding integers

$$[\emptyset, \emptyset], \quad [S(\emptyset), \emptyset], \quad [S(S(\emptyset)), \emptyset], \ \ldots$$

The reader is expected to not worry about such minor issues for they will create no confusion for the purposes of this course. We conclude this subsection by proving that the set of integers is countable

Theorem 23. *$\mathbb{Z}$ is countable.*

Proof. Consider the map $g : \mathbb{N} \to \mathbb{Z}$ given by

$$g(i) = \begin{cases} [j, 0] & \text{if } i \text{ is even and } i = 2 \cdot j \\ [0, j] & \text{if } i \text{ is odd and } i + 1 = 2 \cdot j \end{cases}$$

for all $i \in \mathbb{N}$. In other words, the map g sends the natural numbers $0, 1, 2, 3, 4, \ldots$ to $0, -1, 1, -2, 2, \ldots$ respectively. It is an exercise to the reader to check that the map g is a bijection. $\square$

6.2. **Rational numbers.** In this subsection, we will construct the set $\mathbb{Q}$ of rational numbers together with its arithmetical operations and linear order relation, which turn $\mathbb{Q}$ into an ordered field. Let $\mathbb{Z}^*$ denote the set $\mathbb{Z}-\{[0,0]\}$ and $\sim$ be the relation on $\mathbb{Z}\times\mathbb{Z}^*$ defined by

$$(p,q)\sim(r,s) \longleftarrow p\cdot_{\mathbb{Z}} s = q\cdot_{\mathbb{Z}} r$$

It is easily verified that $\sim$ is an equivalence relation. We define the set of rational numbers to be the quotient set

$$\mathbb{Q} = \mathbb{Z}\times\mathbb{Z}^* \,/\sim$$

Intuitively speaking, the equivalence class $[(p,q)]_\sim$ is supposed to represent the fraction $\frac{p}{q}$. For this reason, the equivalence class $[(p,q)]_\sim$ will be denoted by $\frac{p}{q}$ through this section. It easily follows from Theorem 22 that $\frac{p}{q} = \frac{-p}{-q}$ for all $p\in\mathbb{Z}$ and $q\in\mathbb{Z}^*$. Consequently, while referring to the rational number $\frac{p}{q}$, we may assume without loss of generality that $0<_{\mathbb{Z}} q$. We next define two binary operations $+_{\mathbb{Q}}$ and $\cdot_{\mathbb{Q}}$ on $\mathbb{Q}$ as follows.

$$\frac{p}{q}+_{\mathbb{Q}}\frac{r}{s} = \frac{(p\cdot_{\mathbb{Z}} s)+_{\mathbb{Z}}(q\cdot_{\mathbb{Z}} r)}{q\cdot_{\mathbb{Z}} s}$$
$$\frac{p}{q}\cdot_{\mathbb{Q}}\frac{r}{s} = \frac{p\cdot_{\mathbb{Z}} r}{q\cdot_{\mathbb{Z}} s}$$

for all $p,r\in\mathbb{Z}$ and $q,s\in\mathbb{Z}^*$. That the relations $+_{\mathbb{Q}}$ and $\cdot_{\mathbb{Q}}$ are well-defined can be checked using Theorem 22, with a proof similar to that of Lemma 30. We now define a relation $\leq_{\mathbb{Q}}$ on $\mathbb{Q}$ as follows.

$$\frac{p}{q}\leq_{\mathbb{Q}}\frac{r}{s} \longleftarrow p\cdot_{\mathbb{Z}} s \leq_{\mathbb{Z}} q\cdot_{\mathbb{Z}} r$$

for all $p,r\in\mathbb{Z}$ and $q,s\in\mathbb{Z}$ with $q,s>_{\mathbb{Z}}0$. Together with $+_{\mathbb{Q}}$, $\cdot_{\mathbb{Q}}$ and $\leq_{\mathbb{Q}}$, the set of rational numbers satisfies its usual properties. More precisely, we have that

Theorem 24. *The structure* $(\mathbb{Q},+_{\mathbb{Q}},\cdot_{\mathbb{Q}},\leq_{\mathbb{Q}})$ *is an ordered field.*

As was the case before, we shall not prove Theorem 24 since its proof is long, tedious and does not contain any important ideas. However, the author thinks that every mathematician should see such proofs at least once in his or her lifetime to be completely convinced that such constructions indeed work the way they are supposed to work.

Once again, the well-known secondary school fact $\mathbb{Z}\subseteq\mathbb{Q}$ is *not* true with our constructions of $\mathbb{Z}$ and $\mathbb{Q}$. However, there exists a canonical *copy* of $\mathbb{Z}$ inside $\mathbb{Q}$, namely, the range of the function $f:\mathbb{Z}\to\mathbb{Q}$ given by

$$f(i) = \frac{i}{1}$$

for all $i\in\mathbb{Z}$. Identifying $\mathbb{Z}$ and $f[\mathbb{Z}]$, we may consider $\mathbb{Z}$ as a subset of $\mathbb{Q}$. Consequently, we may use a numeral p to denote both the integer p and the rational number $\frac{p}{1}$. It turns out that the set of rational numbers is also countable.

Theorem 25. $\mathbb{Q}$ *is countable.*

Proof. We shall use Cantor-Schröder-Bernstein theorem. It is straightforward to check that the map $f:\mathbb{N}\to\mathbb{Q}$ given by $f(n)=\frac{n}{1}$ for all $n\in\mathbb{N}$ is an injection. Note that the map $g:\mathbb{Z}\times\mathbb{Z}^*\to\mathbb{Q}$ given by $g(m,n)=\frac{m}{n}$ for all $m\in\mathbb{Z}$ and $n\in\mathbb{Z}^*$ is a surjection and hence, by Lemma 25, there exists an injection $h:\mathbb{Q}\to\mathbb{Z}\times\mathbb{Z}^*$.

Since $\mathbb{Z}$ is countable by Theorem 23, we have that the set $\mathbb{Z} \times \mathbb{Z}^*$ is countable by Corollary 19. It follows that there exists an injection $k : \mathbb{Q} \to \mathbb{N}$ and hence there exists a bijecton between $\mathbb{Q}$ and $\mathbb{N}$ by Cantor-Schröder-Bernstein theorem. □

Finally, we would like to note that our construction of $\mathbb{Q}$ is a special case of a construction called *localization of a ring*, which produces field when it is done with integral domains such as $\mathbb{Z}$. The curious reader is referred to any graduate abstract algebra book for the general construction.

6.3. **Real numbers.** In this subsection, we will construct the set $\mathbb{R}$ of real numbers together with its arithmetical operations and linear order relation, which turn $\mathbb{R}$ into an ordered field.

There are many different constructions of $\mathbb{R}$, although, two of these have been popular among mathematicians, namely, the Cauchy-completion construction and the Dedekind-completion construction. Since the Cauchy-completion construction is usually covered in real analysis courses, we shall cover the Dedekind-completion construction hoping that the reader will learn about the Cauchy-completion construction in other courses.

Before we proceed, we would like to mention a few words regarding the motivation behind this construction. When plotted on a number line with respect to their usual linear ordering, rational numbers will form a scattered set of points. There will be "holes" between rational numbers. For example, consider the sets $A = \{x \in \mathbb{Q} : x \leq_{\mathbb{Q}} 0 \wedge x^2 <_{\mathbb{Q}} 2\}$ and $B = \mathbb{Q} - A$. Since there is no rational number whose square is 2, there cannot be an element of $\mathbb{Q}$ "right between" the elements of A and the elements of B, that is, there is a hole between A and B in the number line. We would like to "complete" these holes by inserting real numbers there. How on earth are we going to do that?..

If there is a hole in the number line between rational numbers, then this hole determines two different subsets of $\mathbb{Q}$, namely, the set of rational numbers which are on the left of this hole and the complement of this set. For example, the hole that is supposed to correspond to $\sqrt{2}$, a real number which we have not yet constructed, determines the set A and its complement B defined above. Our idea is to "fill" the hole determining the pair (A, B) with the pair (A, B) itself. We would like the real number $\sqrt{2}$ to *be* the pair (A, B). On the other hand, the set B is automatically determined by the set A, since we defined B to be $\mathbb{Q} - A$. Therefore, instead of using the pair (A, B), we can simply use the set A. This brings us to the main definition of this section.

Definition 53. *A Dedekind cut of $\mathbb{Q}$ is a subset $A \subseteq \mathbb{Q}$ such that*

- *$A \neq \emptyset$ and $A \neq \mathbb{Q}$,*
- *For all $x, y \in \mathbb{Q}$, if $y \in A$ and $x <_{\mathbb{Q}} y$, then $x \in A$.*
- *A has no greatest element with respect to $<_{\mathbb{Q}}$.*

In other words, a Dedekind cut is a subset of rational numbers which is nonempty, proper, closed downwards with respect to $<_{\mathbb{Q}}$ and which has no greatest element with respect to $<_{\mathbb{Q}}$[1]. For example, the set $A = \{x \in \mathbb{Q} : x \leq_{\mathbb{Q}} 0 \wedge x^2 <_{\mathbb{Q}} 2\}$ is a Dedekind cut.

[1]We would like to remark that some textbooks may define Dedekind cuts as pairs (A, B) of subsets of rational numbers where the set A satisfies our requirements and the set B is its complement. As we have noted before, since the set A determines the set B, we find it unnecessary to involve B.

Our idea is to collect all Dedekind cuts of $\mathbb{Q}$ together so that they are going to fill all the holes in the number line. But what about the points in the number line that do not correspond to holes, namely, the rational numbers? Should we construct them separately as in the previous subsection?

A moment's thought shows that not every Dedekind cut is determined by a hole. For example, for every rational number $q \in \mathbb{Q}$, the set $C = \{x \in \mathbb{Q} : x <_{\mathbb{Q}} q\}$ is a Dedekind cut, however, there is no hole between C and $\mathbb{Q} - C$ in the number line. The reason is that the set $\mathbb{Q} - C$ has a least element, namely, the rational number q. We wish to consider a Dedekind cut D as the real number which is supposed to be "right after" all the elements of D. With this interpretation, such Dedekind cuts as C whose complement has a least element will represent rational numbers and other Dedekind cuts such as A will represent real numbers that fill the holes between rational numbers. Having this intuitive picture in mind, we define the set of real numbers to be

$$\mathbb{R} = \{A \subseteq \mathbb{Q} : A \text{ is a Dedekind cut}\}$$

In other words, a real number is simply a Dedekind cut. We next define the linear order relation $\leq_{\mathbb{R}}$ on the set $\mathbb{R}$ as follows.

$$r \leq_{\mathbb{R}} s \longleftarrow r \subseteq s$$

It should be clear to the reader that the relation $\leq_{\mathbb{R}}$ overlaps with our intuitive understanding of the real number line.

Proposition 1. *The relation $\leq_{\mathbb{R}}$ is a linear order relation.*

Proof. That $\leq_{\mathbb{R}}$ is a partial order relation easily follows from the properties of $\subseteq$ and is left to the reader to be proven as an exercise. We shall only show that any two real numbers are comparable with respect to $\leq_{\mathbb{R}}$.

Let $r, s \in \mathbb{R}$. If $r = s$, then we clearly have $r \leq_{\mathbb{R}} s$. Assume now that $r \neq s$. Then we have $r \nsubseteq s$ or $s \nsubseteq r$. Without loss of generality, we may assume that $r \nsubseteq s$. By definition, there exists $x \in r$ such that $x \notin s$. We claim that $s \subseteq r$.

Let $y \in s$. Since $\leq_{\mathbb{Q}}$ is a linear order relation $x \notin s$, we have that $y <_{\mathbb{Q}} x$ or $x <_{\mathbb{Q}} y$. If it were the case that $x <_{\mathbb{Q}} y$, then, since s is a Dedekind cut, we would have $x \in s$, which is a contradiction. Therefore, we have $y <_{\mathbb{Q}} x$. But then, since r is a Dedekind cut and $x \in r$, we have that $y \in r$. Therefore $s \subseteq r$ and hence $s \leq_{\mathbb{R}} r$ by definition. □

Next will be proven a marvelous property of $\mathbb{R}$, which is fundamental to mathematical analysis and with which the reader is probably familiar from his or her first year calculus courses.

Theorem 26. *Any non-empty subset of $\mathbb{R}$ which is bounded above has a least upper bound.*

Proof. Let $S \subseteq \mathbb{R}$ be a non-empty bounded subset of $\mathbb{R}$. Consider the set

$$s = \bigcup S$$

Since each element of S is a Dedekind cut, the set s is a subset of $\mathbb{Q}$. We claim that s is the least upper bound of S. We first show that $s \in \mathbb{R}$, that is, s is indeed a Dedekind cut.

- s is non-empty and proper.
  Since S is bounded above, there exists $r \in \mathbb{R}$ such that $x \subseteq r$ for all $x \in S$. Since r is a Dedekind cut, it is a proper subset of $\mathbb{Q}$ and hence there exists $y \in \mathbb{Q}$ such that $y \notin r$. Since r is closed downwards with respect to $<_{\mathbb{Q}}$, that $y \notin r$ implies that $z <_{\mathbb{Q}} y$ for all $z \in r$. It then follows that, for all $x \in S$, we have that $z <_{\mathbb{Q}} y$ whenever $z \in x$. In other words, $y \notin \bigcup S = s$. Thus s is a proper subset of $\mathbb{Q}$. Since the elements of S are non-empty, so is s.
- s has no greatest element.
  Assume to the contrary that there exists $q \in s$ such that $x \leq_{\mathbb{Q}} q$ for all $x \in s$. Then, by definition, there would exist $r \in S$ such that $q \in r$, in which case q would be the greatest element of r, contradicting that r is a Dedekind cut. Therefore, s has no greatest element.
- s is closed downwards with respect to $<_{\mathbb{Q}}$.
  Let $p, q \in \mathbb{Q}$ be such that $p <_{\mathbb{Q}} q$ and $q \in s$. Since $q \in s$, there exists $r \in S$ such that $q \in r$. As r is a Dedekind cut, we have that $p \in r$ and hence $p \in s$ by definition. Therefore, s is closed downwards.

We now show that s is the least upper bound of S. Let $r \in \mathbb{R}$ be an upper bound of S. Then, for every $x \in S$, by definition, we have that $x \subseteq r$ and hence $s = \bigcup S \subseteq r$ which means $s \leq_{\mathbb{R}} r$. Therefore, s is the least upper bound of S. □

The property of having least upper bounds for non-empty subsets which are bounded above in a linearly ordered set is known as *completeness*. We have proven above that the linearly ordered set $(\mathbb{R}, \leq_{\mathbb{R}})$ is complete. Before we proceed to define arithmetical operations on $\mathbb{R}$, we would like to remark that our construction of $\mathbb{R}$, where we put a complete linear order structure on the set of Dedekind cuts of the linearly ordered set $(\mathbb{Q}, \leq_{\mathbb{Q}})$, may be carried in a general setting with arbitrary linearly ordered sets. We refer the reader to any book on order theory for the general construction.

We will next define the addition operation $+_{\mathbb{R}}$ on $\mathbb{R}$. Given two real numbers $r, s \in \mathbb{R}$, we define $r +_{\mathbb{R}} s$ to be the real number

$$r +_{\mathbb{R}} s = \{x +_{\mathbb{Q}} y : x \in r \ \wedge \ y \in s\}$$

The reader is expected to check that the set $r +_{\mathbb{R}} s$ is indeed a Dedekind cut of $\mathbb{Q}$ and hence is an element of $\mathbb{R}$.

Contrary to the definition of $+_{\mathbb{R}}$, the multiplication operation $\cdot_{\mathbb{R}}$ is going to have, aesthetically speaking, an "ugly" definition which will be handled case-by-case. This may be considered as a downside of the Dedekind-completion construction. Let $0_{\mathbb{R}}$ be the Dedekind cut $\{q \in \mathbb{Q} : q <_{\mathbb{Q}} 0\}$ consisting of negative rational numbers, that is, the real number zero. Given two real numbers $r, s \in \mathbb{R}$, we define $r \cdot_{\mathbb{R}} s$ to be the real number

$$r \cdot_{\mathbb{R}} s = \begin{cases} \{x \cdot_{\mathbb{Q}} y : 0 \leq_{\mathbb{Q}} x, y \ \wedge \ x \in r \ \wedge y \in s\} \cup 0_{\mathbb{R}} & \text{if } 0 \leq_{\mathbb{R}} r, s \\ -(-r \cdot_{\mathbb{R}} s) & \text{if } 0 \leq_{\mathbb{R}} s \text{ and } r <_{\mathbb{R}} 0 \\ -(r \cdot_{\mathbb{R}} -s) & \text{if } 0 \leq_{\mathbb{R}} r \text{ and } s <_{\mathbb{R}} 0 \\ -r \cdot_{\mathbb{R}} -s & \text{if } r <_{\mathbb{R}} 0 \text{ and } s <_{\mathbb{R}} 0 \end{cases}$$

where $-p$ denotes the real number $\{x +_{\mathbb{Q}} (-y) : x <_{\mathbb{Q}} 0 \ \wedge y \in \mathbb{Q} - p\}$ for a real number $p \in \mathbb{R}$. Together with $+_{\mathbb{R}}$, $\cdot_{\mathbb{R}}$ and $\leq_{\mathbb{R}}$, the set of real numbers behaves the way it is "supposed" to behave. More precisely, we have that

Theorem 27. *The structure* $(\mathbb{R}, +_{\mathbb{R}}, \cdot_{\mathbb{R}}, \leq_{\mathbb{R}})$ *is a complete ordered field.*

Indeed, up to isomorphism, $(\mathbb{R}, +_{\mathbb{R}}, \cdot_{\mathbb{R}}, \leq_{\mathbb{R}})$ is the *unique* complete ordered field. As was the case before, we shall not attempt to prove this important fact which has a long and tedious proof. From now on, we shall assume all the facts regarding real numbers that one learns in a first-year calculus, whenever such facts are necessary in our proofs.

We conclude this section by proving that the set of real numbers has the same cardinality as $\mathcal{P}(\mathbb{N})$.

Theorem 28. *There exists a bijection between* $\mathbb{R}$ *and* $\mathcal{P}(\mathbb{N})$.

Proof. We plan to use Cantor-Schröder-Bernstein theorem. Let $h : \mathbb{Q} \to \mathbb{N}$ be a fixed bijection. Consider the map $g : \mathbb{R} \to \mathcal{P}(\mathbb{N})$ given by $g(r) = h[r]$ for all $r \in \mathbb{R}$. It is easily checked that g is injective since, if two Dedekind cuts r and s are distinct, then one of them has to contain a rational number which is not in the other.

In order to construct an injection in the other direction, we shall assume the familiarity of the reader with n-ary expansion of real numbers. Consider the map $f : {}^{\mathbb{N}}2 \to \mathbb{R}$ given by

$$f((a_i)_{i\in\mathbb{N}}) = \sum_{i=0}^{\infty} \frac{2 \cdot a_i}{3^{i+1}}$$

for all $(a_i)_{i\in\mathbb{N}} \in {}^{\mathbb{N}}2$. In other words, the map f sends the sequence $(a_i)_{i\in\mathbb{N}}$ to the real number whose $(i+1)$-th digit in its *ternary* expansion is $2 \cdot a_i$. A first-year calculus knowledge on the sum of geometric series immediately implies that f is injective[2]. By Exercise 13, there exists a bijection between $\mathcal{P}(\mathbb{N})$ and ${}^{\mathbb{N}}2$. Consequently, we can find an injection from $\mathcal{P}(\mathbb{N})$ to $\mathbb{R}$ by composing these maps. As there are injections in both directions, it follows from Cantor-Schröder-Bernstein theorem that there is a bijection between $\mathcal{P}(\mathbb{N})$ to $\mathbb{R}$.

□

[2]The curious reader may wish to google the term "Cantor set", which is the range of f.

7. Ordinal numbers

In this section, we shall learn the concept of an ordinal number, which arguably is the most important concept in modern set theory. Historically, ordinal numbers were first defined as isomorphism classes of strictly well-ordered sets. Later on, John von Neumann presented a simpler definition, which has been the standard definition since then.

Definition 54. *A set x is said to be transitive if every element of x is also a subset of x, i.e. $\forall y(y \in x \to y \subseteq x)$.*

In other words, transitive sets are those sets that contain elements of elements of themselves. Equivalently, a set x is transitive if and only if $\bigcup x \subseteq x$.

Exercise 32. *Using induction, prove that n is transitive for all $n \in \mathbb{N}$.*

Ordinal numbers are supposed to "represent" strictly well-ordered sets. This is why ordinal numbers were first defined as isomorphism classes of strictly well-ordered sets. However, this definition brings some technical difficulties since an isomorphism class of a strictly well-ordered sets is not a set but a proper class.

John von Neumann's idea was to find "canonical" representatives in isomorphism classes of strictly well-ordered sets. It turns out that every strictly well-ordered set is isomorphic to some transitive set which is strictly well-ordered by the membership relation $\in$. This leads us to the following simple but extra-ordinarily useful definition.

Definition 55. *A set α is an ordinal number if*

- *α is transitive, and*
- *$(\alpha, \in_\alpha)$ is a strictly well-ordered set.*

What are some examples of ordinal numbers? It follows from Exercise 27 and 32 that each natural number is an ordinal number. Moreover, the set of natural numbers $\mathbb{N}$ itself is also an ordinal number. From now on, we shall use ω to denote the set $\mathbb{N}$. The next lemma shows that the successor of an ordinal number is also an ordinal number.

Lemma 31. *If α is an ordinal number, then so is $S(\alpha)$.*

Proof. Let α is an ordinal number. Let $\gamma \in S(\alpha)$. Then either $\gamma \in \alpha$ or $\gamma = \alpha$. In both cases, we have $\gamma \subseteq S(\alpha)$ since α is transitive. Checking that $\in_{S(\alpha)}$ is a strict well-order relation is left to the reader as an exercise. **Hint**. Notice that the pair $(S(\alpha), \in_{S(\alpha)})$ is obtained by joining α to the strictly well-ordered set $(\alpha, \in_\alpha)$ as the greatest element. □

Ordinal numbers have various equivalent characterization. The reader is expected to be able to prove equivalence of some of these characterizations.

Exercise 33. *Prove that α is an ordinal if and only if α is transitive and for all $\gamma, \delta \in \alpha$ either $\gamma \in \delta$, $\gamma = \delta$ or $\delta \in \gamma$.*

Next will be proven that the class of ordinal numbers together with the membership relation $\in$ is strictly well-ordered[1].

[1]We have not officially defined what it means for an arbitrary class to be strictly well-ordered. The reader need not learn this and should simply know that the membership relation $\in$, considered on the class of ordinal numbers, has all the properties of a strict well-order relation.

Theorem 29. *Given ordinal numbers α, β, γ. Then*

i. *If $\alpha \in \beta$ and $\beta \in \gamma$, then $\alpha \in \gamma$.*
ii. *If $\alpha \in \beta$, then $\beta \notin \alpha$.*
iii. *Either $\alpha \in \beta$, $\alpha = \beta$ or $\beta \in \alpha$.*
iv. *Every non-empty set of ordinals has a least element with respect to $\in$.*

We note that some bullet points in the statement of this theorem easily follows from the axiom of foundation. However, as will be seen from the proof, the axiom of foundation is not really necessary. In order to prove this theorem, we shall need the following lemmas.

Lemma 32. *If α is an ordinal, then $\alpha \notin \alpha$.*

Proof. Let α be an ordinal. Assume to the contrary that $\alpha \in \alpha$. Then $\in_\alpha$ cannot be asymmetric on α since $\alpha \in \alpha$. □

Lemma 33. *If α is an ordinal and $\gamma \in \alpha$, then γ is an ordinal.*

Proof. Let α be an ordinal and $\gamma \in \alpha$. We first show that γ is transitive. Pick $y \in \gamma$ and $x \in y$. Since α is transitive, $\gamma \subseteq \alpha$ hence $y \in \alpha$, which again implies $y \subseteq \alpha$ by the transitivity of α. Thus $x \in \alpha$. Since α is strictly well-ordered by $\in$ and $x, y, \gamma \in \alpha$, it follows from $x \in y$ and $y \in \gamma$ that $x \in \gamma$. Therefore γ is transitive.

Since $\gamma \subseteq \alpha$ and $(\alpha, \in_\alpha)$ is a strictly well-ordered set, the pair $(\gamma, \in_\alpha \cap(\gamma \times \gamma))$ is a strictly well-ordered set. However, the transitivity of α implies that

$$\in_\alpha \cap(\gamma \times \gamma) = \in_\gamma$$

Thus, $(\gamma, \in_\gamma)$ is a strictly well-ordered set and hence γ is an ordinal. □

Lemma 34. *If α and β are ordinals and $\alpha \subsetneq \beta$, then $\alpha \in \beta$.*

Proof. Let α and β be ordinals such that $\alpha \subsetneq \beta$. Since $\beta - \alpha \neq \emptyset$, by definition, there exists $\gamma \in \beta - \alpha$ which is minimal with respect to $\in$. We claim that $\gamma = \alpha$.

To see that $\gamma \subseteq \alpha$, let $\delta \in \gamma$. By the transitivity of β, we have $\delta \in \beta$. Since γ is a minimal element of $\beta - \alpha$ with respect to $\in$, we have $\delta \notin \beta - \alpha$ and hence $\delta \in \alpha$.

To see that $\alpha \subseteq \gamma$, let $\delta \in \alpha$. If it were not the case that $\delta \in \gamma$, since $\gamma \subseteq \beta$ and β is linearly ordered by $\in$, we would have $\gamma = \delta$ or $\gamma \in \delta$, both of which implies $\gamma \in \alpha$ by the transitivity of α. This contradicts the choice of γ. Therefore, it has to be the case that $\delta \in \gamma$. Therefore $\alpha \subseteq \gamma$ and hence $\alpha = \gamma$, which completes the proof. □

Exercise 34. *If α and β are ordinals, then so is $\alpha \cap \beta$.*

Proof of Theorem 29.

i. Let $\alpha \in \beta$ and $\beta \in \gamma$. We have $\alpha \in \gamma$ since γ is transitive.
ii. Let $\alpha \in \beta$. If it were the case that $\beta \in \alpha$, then $\alpha \in \alpha$ by transitivity of α, which contradicts Lemma 32.
iii. Let α, β be ordinals. Then $\gamma = \alpha \cap \beta$ is an ordinal and, $\gamma \subseteq \alpha$ and $\gamma \subseteq \beta$. We wish to show that $\gamma = \alpha$ or $\gamma = \beta$. Assume to the contrary that $\gamma \neq \alpha$ and $\gamma \neq \beta$. Then it follows from the previous lemma that $\gamma \in \alpha \cap \beta = \gamma$, a contradiction. Thus $\gamma = \alpha$ or $\gamma = \beta$.

    If $\gamma = \alpha$, then $\alpha = \gamma = \beta \cap \alpha$ and so $\alpha \subseteq \beta$. In this case, the previous lemma implies that $\alpha = \beta$ or $\alpha \in \beta$. If $\gamma \neq \alpha$, then $\beta = \gamma = \beta \cap \alpha$ and so $\beta \subseteq \alpha$. Similarly, the previous lemma implies that $\beta = \alpha$ or $\beta \in \alpha$. Thus, we have $\alpha = \beta$, $\alpha \in \beta$ or $\beta \in \alpha$.

iv. Let X be a non-empty set whose elements are ordinal numbers. Pick $\alpha \in X$. If $\alpha \cap X = \emptyset$, then α is the least element of X with respect to $\in$. If $\alpha \cap X \neq \emptyset$, then $\alpha \cap X$ has a least element γ with respect to $\in$ since $\alpha \cap X \subseteq \alpha$ and α is an ordinal. If there were $\delta \in X$ such that $\delta \in \gamma$, then $\delta \in \alpha$ by transitivity and hence $\delta \in \alpha \in X$, which contradicts the minimality of γ in $\alpha \cap X$ with respect to $\in$. Thus, γ is the least element of X with respect to $\in$.

□

Lemma 35. *For any set X of ordinals, there exists an ordinal γ such that $\gamma \notin X$.*

Proof. Let X be a set of ordinals and set $\gamma = \bigcup X$. Since the elements of X are transitive sets, so is γ. Moreover, it follows from Lemma 33 and 29 that γ is strictly well-ordered by $\in$. Therefore, γ is an ordinal.

Consider the ordinal $\alpha = S(\gamma)$. If it were the case that $\alpha \in X$, then $\alpha \subseteq \bigcup X = \gamma$ and hence $\alpha = \gamma$ or $\alpha \in \gamma$ by Lemma 34. In both cases, we have that $\alpha \in S(\gamma) = \alpha$, which is a contradiction. Therefore, $\alpha \notin X$. □

Therefore, the collection of ordinal numbers is a proper class. As was pointed out before, Theorem 29 shows that the membership relation $\in$, considered on the class of ordinal numbers, has all the properties of a strict well-order relation. In the light of this observation, from now on, given two ordinal numbers α and β, we shall write $\alpha < \beta$ if $\alpha \in \beta$. As before, we shall write $\alpha \leq \beta$ if $\alpha < \beta$ or $\alpha = \beta$.

It follows from Lemma 35 that for any set X of ordinals, there exists a least ordinal γ such that $\beta \leq \gamma$ for all $\beta \in X$, namely the ordinal number $\bigcup X$. We shall call this ordinal the *supremum* of X and denote it by $sup(X)$. For example, $sup(S(n)) = n$ for all $n \in \mathbb{N}$ and $sup(\omega) = \omega$.

Before we conclude this subsection, we introduce two more important definitions.

Definition 56. *An ordinal α is said to be a successor ordinal if $\alpha = S(\beta)$ for some ordinal β. An ordinal α is said to be a limit ordinal if $\alpha \neq 0$ and α is not a successor ordinal.*

7.1. **How do the ordinals look like?** We have established in the previous subsection that any two ordinals are comparable with respect to $<$. It follows that 0 is the least of all ordinals. Then follows natural numbers in their usual ordering[2]. Thus we have the following picture.

$$0\ 1\ 2\ 3\ \ldots$$

It is straightforward to check that natural numbers are precisely the finite ordinal numbers. We already know that each natural number is a finite ordinal. The converse follows from the following lemma.

Lemma 36. *If α is a finite ordinal number, then $\alpha \in \omega$.*

Proof. Let α be a finite ordinal number. Assume towards a contradiction that $\alpha \notin \omega$. Then, since any two ordinals are comparable with respect to $\in$, we have either $\omega = \alpha$ or $\omega \in \alpha$. In both cases, we have $\omega \subseteq \alpha$. If α were finite, say $|\alpha| = |n|$ for some $n \in \mathbb{N}$, then there would be an injection from $S(n)$ to n contradicting the pigeonhole principle. □

[2]It is an easy exercise to the reader to check that there can be no ordinals between α and $S(\alpha)$.

Given any infinite ordinal α, we have $n \in \alpha$ for all $n \in \omega$. It follows that $\omega \subseteq \alpha$ and hence $\omega = \alpha$ or $\omega \in \alpha$. This shows that ω is the first infinite ordinal number. Thus we have the following picture.

$$0\ 1\ 2\ 3\ \ldots \omega$$

By iteratively applying the successor operation, we can construct the following ordinals.

$$0\ 1\ 2\ 3\ \ldots\ \omega\ S(\omega)\ S(S(\omega))\ \ldots$$

It is easily seen that the collection $\{0, 1, 2, 3, \ldots, \omega, S(\omega), S(S(\omega)), \ldots\}$, if it is a set, has to be an ordinal. It turns out that, without further axioms, we cannot prove that this collection is a set. We next introduce an axiom that allows us to construct this ordinal.

Axiom 10 (The axiom of replacement)**.** *Let $\varphi(x, y)$ be a formula in the language of set theory with two free variables. If for all x there exists a unique y such that $\varphi(x, y)$, then for any A, there exists B such that for all y we have $y \in B$ if and only if there exists $x \in A$ such that $\varphi(x, y)$, i.e.*

$$(\forall x \exists y \forall y'(\varphi(x, y') \leftrightarrow y = y') \rightarrow (\forall A \exists B \forall y(y \in B \leftrightarrow \exists x(x \in A \wedge \varphi(x, y)))))$$

As was the case with the axiom of separation, the axiom of replacement is not a single axiom but rather an axiom *schema*. We have an axiom for each formula $\varphi(x, y)$ in the language of set theory.

Axioms 1-10 together form the axiomatic system known as ZFC. Thus, ZFC consists of eight axioms and two axiom schemas. It can be shown that some of these axioms are redundant[3]. However, ZFC is traditionally presented in this form.

Of all the axioms of ZFC, the axiom of replacement is the most difficult to understand. In order for the reader to have a good understanding of this axiom, we next introduce the concept of a class function.

Let $\varphi(x, y)$ be a formula in the language of set theory such that for all x there exists a unique y such that $\varphi(x, y)$. In this case, we can think of the formula $\varphi(x, y)$ as "defining a function" on the class of all sets $\mathbf{V}$ since it "maps" each set x to a unique set y. For each set x, we shall write $F_\varphi(x)$ to denote the unique set y such that $\varphi(x, y)$ and say that φ defines the *class function* F_φ. We remark that F_φ is not a function in the technical sense of the word "function". Even though F_φ is not technically a function, we shall use the notations introduced for functions. More specifically, we will write $F_\varphi[A]$ to denote the set $\{y : \exists x \in A\ F_\varphi(x) = y\}$ and write $F \restriction_A$ to denote the set $\{(x, F(x)) : x \in A\}$.

The axiom of replacement simply asserts that the image of a set under a class function is a set, i.e. if F_φ is a class function and A is a set, then the collection $F_\varphi[A]$ is a set.

It turns out that the *parametric* version of the axiom of replacement can be proven from the axioms of ZFC[4]. In other words, if $\varphi(p_1, p_2, \ldots, p_n, x, y)$ is a formula in the language of set theory such that for all $p_1, p_2, \ldots, p_n$ the formula $\varphi(p_1, p_2, \ldots, p_n, x, y)$ defines a class function F_φ, then for any $p_1, p_2, \ldots, p_n$ and

[3]For example, the axiom of empty set can be proven from the axiom of infinity and the axiom of separation. Similarly, the axiom of separation can be proven from the remaining axioms.

[4]The curios reader may read the note "ZFC without parameters" by Ralf Schindler and Philipp Schlicht via the link http://www.math.uni-bonn.de/people/schlicht/publications.html as of 30 April 2018.

for any set A the collection $F_\varphi[A]$ is a set. Indeed, some textbooks introduce the axiom of replacement in this parametric form.

Back to ordinals... How does the axiom of replacement help us construct the next limit ordinal after the first infinite ordinal?

Consider the class function F_φ which maps each natural number n to the infinite ordinal which contains exactly n ordinals that contain limit ordinals and maps other sets to the empty set. Then $F_\varphi(0) = \omega$, $F_\varphi(1) = S(\omega)$, $F_\varphi(2) = S(S(\omega))$ and so on. By the axiom of replacement, the set $F_\varphi[\omega]$ exists. Taking the union of this set, we obtain

$$\bigcup F_\varphi[\omega] = \{0, 1, 2, \dots, \omega, S(\omega), S(S(\omega)), \dots\}$$

For the reasons we shall not explain at this moment, let us name this ordinal $\omega + \omega$. At the moment, we have the following picture of the ordinals.

$$0\ 1\ 2\ 3\ \dots\ \omega\ S(\omega)\ S(S(\omega))\ \dots\ \omega + \omega$$

Having constructed the second limit ordinal $\omega + \omega$, we can keep applying the successor operation and taking the unions of the ordinals constructed via the axiom of replacement.

$$0\ 1\ 2\ 3\ \dots\ \omega\ S(\omega)\ S(S(\omega))\ \dots\ \omega + \omega\ S(\omega + \omega)\ S(S(\omega + \omega))\ \dots$$

Where does this process end? The reader may have realized that all the ordinals that can be constructed after finitely many stages via this "bottom-up" process are countable sets.

Are there uncountable ordinals? Before we answer this question, we need to prove the following fact which shows that ordinal numbers indeed "represent" strictly well-ordered sets.

Theorem 30. *Let $(W, \prec)$ be a strictly well-ordered set. Then there exists a unique ordinal α such that $(W, \prec)$ and $(\alpha, \in_\alpha)$ are isomorphic.*

Proof. Recall that for any $w \in W$, the set $pred(w)$ of predecessors of w together with the relation $\prec_w = \prec \cap (pred(w) \times pred(w))$ is a strictly well-ordered set. Let

$$A = \{w \in W : \exists \alpha \text{ "}\alpha \text{ is an ordinal" } \wedge (pred(w), \prec_w) \cong (\alpha, \in_\alpha)\}$$

As strictly well-ordered sets, given two distinct ordinals, one of them is a proper initial segment of the other one. Therefore, for any $w \in W$, if $(pred(w), \prec_w)$ is isomorphic to some ordinal, then this ordinal has to be unique and we shall denote it by α_w.

It follows from the axiom of replacement that the collection $\Omega = \{\alpha_w : \exists w\ w \in A\}$ is a set. We claim that Ω is an ordinal. Since the elements of Ω are ordinals, it is strictly well-ordered by $\in$. Therefore, it suffices to prove that Ω is transitive. Let $\alpha_w \in \Omega$ and $\gamma \in \alpha_w$. By definition, there exists an isomorphism $g : pred(w) \to \alpha_w$. Then, since γ is a proper initial segment of the ordinal α_w, the set $g^{-1}[\gamma]$ is a proper initial segment of $pred(w)$ and hence is of the form $pred(w')$ for some $w' \in pred(w)$. This means that $\gamma \in \Omega$ since $g \restriction_{pred(w')} : pred(w') \to \gamma$ is an order isomorphism. This completes the proof that Ω is transitive.

Consider the function $f : A \to \Omega$ given by $f(w) = \alpha_w$ for all $w \in A$. We claim that f is an order isomorphism. That f is surjective follows from the definition. Checking that $a \prec b \to f(a) \in f(b)$ is left to the reader as an exercise.

Finally, we show that $A = W$. If it were the case that $W - A \neq \emptyset$, then there would be a least element $w \in W - A$ with respect to $\prec$ and $A = pred(w)$. This

would imply $w \in A$, contradicting the choice of w. Thus $(W, \prec)$ is isomorphic to $(\Omega, \in_\Omega)$. As before, notice that, given two distinct ordinals, one of them is a proper initial segment of the other one. Therefore, $(W, \prec)$ cannot be isomorphic to two distinct ordinals. □

In the light of this theorem, we can associate an *order type* to each strictly well-ordered set.

Definition 57. *Let $(W, \prec)$ be a strictly well-ordered set. The order type of $(W, \prec)$ is the unique ordinal α such that $(W, \prec) \cong (\alpha, \in_\alpha)$ and is denoted by $ot(W, \prec)$.*

Exercise 35. *Let $2\mathbb{N}$ and $2\mathbb{N}+1$ denote the sets of even natural numbers and odd natural numbers respectively. Let $\prec$ be the relation defined on $\mathbb{N}$ as follows.*

$$m \prec n \leftrightarrow (m + n \in 2\mathbb{N} \ \wedge \ m < n) \vee (m \in 2\mathbb{N} \ \wedge n \in 2\mathbb{N} + 1)$$

where $<$ is the usual order relation on $\mathbb{N}$. You are given the fact that $\prec$ is a strict well-order relation on $\mathbb{N}$. Prove that the order type of $(\mathbb{N}, \prec)$ is $\omega + \omega$ by explicitly constructing an order isomorphism from $\mathbb{N}$ to $\omega + \omega$.

7.2. **Hartogs numbers.** We are now ready to prove that uncountable ordinals exist. This result trivially follows from the fact that every set can be well-ordered, which we shall see later is a consequence of the axiom of choice. However, it is possible to prove that existence of uncountable ordinals using only the axioms of ZF[5].

Theorem 31. *Let X be a set. Then there exists an ordinal λ such that there is no injection from λ to X.*

Proof. Let

$$W = \{R \in \mathcal{P}(X \times X) : \exists Y(Y \subseteq X \ \wedge (Y, R) \text{ is a well-ordered set})\}$$

By the previous theorem, for each $R \in W$, there exists a unique ordinal α_R such that $(dom(R), R) = (Y_R, R) \cong (\alpha_R, \leq)$. By the axiom of replacement, the set $\lambda = \{\alpha_R : \exists R(R \in W)\}$ exists. We claim that λ is an ordinal.

Since the elements of λ are ordinals, it is sufficient to prove that λ is transitive. Let $\gamma \in \lambda$ and $\beta \in \gamma$. Then, by definition, there exists an order isomorphism g from some well-ordered set (Y, R) to $(\gamma, \leq)$ where $Y \subseteq X$. Clearly $g^{-1}[\beta]$ and β, as well-ordered sets, are order isomorphic via the restriction of g. Thus, $\beta \in \lambda$ and hence λ is transitive. This completes the proof that λ is an ordinal.

If there were an injection from λ to X, then some subset of X could be well-ordered so that the corresponding well-ordered set is isomorphic to λ. But then, by definition, we would have $\lambda \in \lambda$, which is a contradiction. Thus there exists no injection from λ to X. □

Given any set X, the least ordinal which does not inject into X is called the *Hartogs number* of X and is denoted by $\aleph(X)$. By definition, there can be no bijection between a set X and its Hartogs number $\aleph(X)$.

In particular, the Hartogs number of ω, which is the least ordinal that does not inject into ω, is uncountable. This ordinal is known as the first uncountable ordinal and is denoted by ω_1. Since ω_1 is the least ordinal that does not inject into ω, the

[5]The axiomatic system obtained by removing the axiom of choice from the axioms of ZFC is called ZF, the **Z**ermelo-**F**raenkel set theory.

elements of ω_1 have to be countable ordinals. Conversely, any countable ordinal injects into ω and hence is an element of ω_1. Therefore, ω_1 is precisely the set of countable ordinals.

One can similarly argue that the Hartogs number $\aleph(\lambda)$ of an ordinal λ is the ordinal which consists of precisely the ordinals that are equinumerous with some subset of λ.

The existence of Hartogs numbers shows that our "bottom-up" approach of constructing ordinal numbers is useless. There are ordinal numbers that cannot be obtained from the empty set in finitely (even, countably) many stages by applying the successor operation and taking unions. Ordinal numbers exist simply because they do so.

7.3. **Transfinite induction and transfinite recursion.** In this section, we shall prove the principle of transfinite induction on the class of ordinal numbers and show that a generalization of the recursion theorem, known as the transfinite recursion theorem, holds for ordinal numbers.

Theorem 32 (The principle of transfinite induction)**.** *Let $\varphi(x)$ be a formula in the language of set theory with one free variable. Assume that for all ordinals γ, we have that if $\varphi(\beta)$ for all $\beta < \gamma$, then $\varphi(\gamma)$. Then $\varphi(\gamma)$ holds for all ordinals γ.*

Proof. Assume towards a contradiction that there exists an ordinal γ such that $\neg\varphi(\gamma)$ holds. Then, the set $\{\delta \in S(\gamma) : \neg\varphi(\delta)\}$ is non-empty and hence has a least element θ with respect to $<$. By the choice of θ, we have that $\varphi(\delta)$ holds for all $\delta < \theta$. By assumption, this implies that $\varphi(\theta)$, which is a contradiction. Thus, for all ordinals γ, we have that $\varphi(\gamma)$. □

It can easily be seen that the principle of transfinite recursion holds even if we allow fixed parameters $\varphi(x, p_1, \ldots, p_n)$ in the statement of the theorem. Note that, given a fixed ordinal δ, the principle of transfinite induction can be used to prove that a property $\varphi(x)$ holds for all ordinals $\gamma \leq \delta$ by showing that the inductive hypothesis holds for all ordinals up to δ, i.e. for all $\gamma \leq \delta$ we have if $\varphi(\beta)$ for all $\beta < \gamma$, then $\varphi(\gamma)$.

The principle of transfinite induction is the generalization of the principle of induction on $\mathbb{N}$ to the class of ordinal numbers. In practice, this principle is most commonly used in the following form.

Theorem 33 (The principle of transfinite induction, alternative formulation)**.** *Let $\varphi(x)$ be a formula in the language of set theory with one free variable. Assume that*

- *$\varphi(0)$ holds.*
- *For all ordinals γ, if $\varphi(\gamma)$ holds, then so does $\varphi(S(\gamma))$.*
- *For all limit ordinals θ, if $\varphi(\gamma)$ holds for all $\gamma < \theta$, then $\varphi(\theta)$ holds.*

Then $\varphi(\gamma)$ holds for all ordinals γ.

Proof. Exercise. **Hint.** Imitate the proof of the original formulation of the principle of transfinite induction. □

As before, this form of the principle of transfinite induction can also be used to show that a property $\varphi(x)$ holds for all ordinals up to an ordinal δ by showing that the inductive hypothesis hold for all ordinals up to δ.

We shall next prove the transfinite recursion theorem which allows us to define class functions recursively on the class of ordinal numbers.

Theorem 34 (The transfinite recursion theorem)**.** *Let F_φ be a class function. Then there exists a class function F_ψ such that $F_\psi(\alpha) = F_\varphi(F_\psi \restriction_\alpha)$ for all ordinal numbers α. Moreover, this class function is unique in the sense that if there exists another class function F_ϕ satisfying the same property, then $F_\phi(\alpha) = F_\psi(\alpha)$ for all ordinal numbers α.*

Before we proceed to the proof of this theorem, we would like to take a pause and understand certain subtleties regarding the statement of this theorem. In the statement, we are given a class function F_φ and we assert the existence of another class function F_ψ. As we emphasized before, a class function F_ψ is technically not a function, but rather, is a formula $\psi(x, y)$ that assigns a unique set y to each set x. Naturally, one should ask the following question: How on earth can we quantify over class functions in the statement of this theorem?!

What the transfinite recursion theorem really says is that, given a formula $\varphi(x, y)$ that assigns a unique set y to each set x, one can produce a formula $\psi(x, y)$ that assigns a unique set y to each set x such that $F_\psi(\alpha) = F_\varphi(F_\psi \restriction_\alpha)$ for all ordinals α; and that the values of F_ψ on the class of ordinal numbers are uniquely determined.

Therefore, one should think of the transfinite recursion theorem not as a single theorem but as a *theorem schema*. For each formula $\varphi(x, y)$ defining a class function, the corresponding instance of this schema is a theorem of ZFC. Having realized this subtle point, we are now ready to prove the transfinite recursion theorem.

Proof of Theorem 34. Let F_φ be a class function. Throughout this proof, we will refer to a function f such that $dom(f) = S(\alpha)$ and $f(\gamma) = F_\varphi(f \restriction_\gamma)$ for all $\gamma \leq \alpha$ as a *computation of length* α.

First, we show that computations of length α, if they exist, are unique for all ordinals α. Let α be an ordinal and let f and f' be computations of length α. We will prove that $f(\gamma) = f'(\gamma)$ for all $\gamma \leq \alpha$ via transfinite induction. Let $\gamma \leq \alpha$ and assume that the claim holds for all ordinals less than γ. Then

$$f(\gamma) = F_\varphi(f \restriction_\gamma)) = F_\varphi(f' \restriction_\gamma)) = f'(\gamma)$$

Thus, $f(\gamma) = f'(\gamma)$ for all $\gamma \leq \alpha$ and hence computations of a fixed length are unique if they exist.

Let $\psi(x, y)$ be the following formula with two free variables[6].

$$(\text{“}x \text{ is not an ordinal”} \ \wedge \ y = 0)$$

$$\vee$$

$$(\text{“}x \text{ is an ordinal”} \ \wedge \ \exists f \ (y = f(x) \ \wedge \text{“}f \text{ is a computation of length } x\text{”}))$$

In other words, $\psi(x, y)$ is the formula which maps each non-ordinal x to the empty set and each ordinal x to the value of some computation of length x at x. We claim that this formula defines a class function and that F_ψ satisfies the requirements in the statement of the theorem.

At this point, it is not clear that $\psi(x, y)$ indeed defines a class function, i.e. for each x there exists a unique y such that $\psi(x, y)$. Clearly, if x is not an ordinal, then $y = \emptyset$. Thus it is sufficient to show that for each ordinal x there exists a unique y such that $\psi(x, y)$ holds. Since y is given as the value of some computation of

[6]Obviously, this formula needs to be written in the language of set theory. However, we shall not carry out this tedious task since the reader should be able to convert the following informal description to a formula in the language of set theory.

length x at the set x, it suffices to prove that for all ordinals x there exists a unique computation of length x. We already know that computations of a fixed length are unique if they exist. Thus we only need to show the existence of computations of arbitrary length.

We proceed by transfinite induction. Let α be an ordinal and assume that there exists a (necessarily unique) computation of length β for all $\beta < \alpha$. We need to prove that there exists a (necessarily unique) computation of length α. Let H be the set

$$\{g : \exists \beta \in \alpha \text{ “}g \text{ is a computation of length } \beta\text{”}\}$$

which exists by the assumption and the axiom of replacement. Define

$$f = \bigcup H \cup \left\{\left(\alpha, F_\varphi\left(\bigcup H\right)\right)\right\}$$

We now prove that f is a computation of length α. Observe that, by construction, we have that $dom(f) = \bigcup_{g\in H} dom(g) \cup \{\alpha\} = S(\alpha)$. To prove that f is indeed a function, it is sufficient to prove that the collection H consists of compatible functions. Let $g, g' \in H$ be functions and assume without loss of generality that $dom(g) = \beta_1 \leq \beta_2 = dom(g')$. We prove by transfinite induction that $g(\theta) = g'(\theta)$ for all $\theta < \beta_1$. Let $\theta < \beta_1$ and assume that the claim holds for all ordinals less than θ. Then

$$g(\theta) = F_\varphi(g \restriction_\theta) = F_\varphi(g' \restriction_\theta) = g'(\theta)$$

Thus, by transfinite induction, we have $g(\theta) = g'(\theta)$ for all $\theta < \beta_1$ and hence H is a collection of compatible functions. Consequently, the relation f is indeed a function. Next, let $\beta < \alpha$ and pick $g \in H$ such that $dom(g) = S(\beta)$. Then $f(\beta) = g(\beta) = F_\varphi(g \restriction_\beta) = F_\varphi(f \restriction_\beta)$ since $g \subseteq f$ and g is a computation. Finally, $f(\alpha) = F_\varphi(\bigcup H)) = F_\varphi(f \restriction_\alpha))$ and hence, f is a computation of length α, which completes the induction.

We have proven that the formula $\psi(x, y)$ defines a class function F_ψ. We next show that this class function satisfies the conditions in the statement of the theorem. Let α be any ordinal. Then, by definition, $F_\psi(\alpha) = f(\alpha)$ where f is the unique computation length α. The crucial observation is that the restriction of a computation of length α to any ordinal $S(\beta) < \alpha$ is the unique computation of length β. Therefore, $f(\beta) = F_\psi(\beta)$ for any $\beta < \alpha$. It follows that $F_\psi(\alpha) = f(\alpha) = F_\varphi(f \restriction_\alpha) = F_\varphi(F_\psi \restriction_\alpha)$. To show that such a class function F_ψ is unique, assume that there exists another formula $\phi(x, y)$ defining a class function F_ϕ such that $F_\phi(\alpha) = F_\varphi(F_\phi \restriction_\alpha)$ for all ordinals α. By transfinite induction, we will prove that $F_\psi(\alpha) = F_\phi(\alpha)$ for all ordinals α.

Let α be an ordinal and assume that the claim holds for all ordinals less than α. Then $F_\psi(\alpha) = F_\varphi(F_\psi \restriction_\alpha) = F_\varphi(F_\phi \restriction_\alpha) = F_\phi(\alpha)$. Thus, by transfinite induction, we have that $F_\psi(\alpha) = F_\phi(\alpha)$ for all ordinals α, which completes the proof. □

In practice, the following variant of the transfinite recursion theorem is frequently used.

Theorem 35. *Let $F_{\varphi_1}, F_{\varphi_2}, F_{\varphi_3}$ be class functions. Then there exists a unique class function F_ψ such that*

- $F_\psi(0) = F_{\varphi_1}(0)$,
- $F_\psi(S(\alpha)) = F_{\varphi_2}(F_\psi(\alpha))$ *for all ordinals α, and*
- $F_\psi(\alpha) = F_{\varphi_3}(F_\psi \restriction_\alpha)$ *for all limit ordinals α*

The proof of this theorem easily follows from the proof of the transfinite recursion theorem. Finally, we note that, as was the case with the recursion theorem, there are variants of the transfinite recursion theorem that allow "parameters" in the recursive definition of F_ψ. We refer the reader to [1] for a precise statement and the proof of such a variant.

7.4. **Ordinal arithmetic.** Having proven the transfinite recursion theorem, we are ready to define some standard arithmetical operations on the class of ordinal numbers.

Given an ordinal β, we define the ordinal $\beta+\gamma$ by transfinite recursion as follows.

- $\beta + 0 = \beta$,
- $\beta + S(\gamma) = S(\beta + \gamma)$ for all ordinals γ, and
- $\beta + \gamma = sup\{\beta + \theta : \theta < \gamma\}$ for all limit ordinals γ.

Similarly, given an ordinal β, we define the ordinal $\beta \cdot \gamma$ by transfinite recursion as follows.

- $\beta \cdot 0 = 0$,
- $\beta \cdot S(\gamma) = (\beta \cdot \gamma) + \beta$ for all ordinals γ, and
- $\beta \cdot \gamma = sup\{\beta \cdot \theta : \theta < \gamma\}$ for all limit ordinals γ.

Finally, for all ordinals $\beta > 0$, we define the exponentiation on the class of ordinal numbers by transfinite recursion as follows.

- $\beta^0 = 1$,
- $\beta^{S(\gamma)} = \beta^\gamma \cdot \beta$ for all ordinals γ, and
- $\beta^\gamma = sup\{\beta^\theta : \theta < \gamma\}$ for all limit ordinals γ.

Notice that we have excluded the case $\beta = 0$ in the recursive definition of ordinal exponentiation. The reason is that, if one used this definition for $\beta = 0$, then one would obtain $0^\omega = sup\{0^n : n \in \omega\} = 1$, which we do not want to be the case. Consequently, we define $0^0 = 1$ and $0^\gamma = 0$ for all ordinals $\gamma > 0$.

We would like to emphasize that it can be proven using (the variant of) the transfinite recursion theorem that there exist formulas with two free variables in the language of set theory defining these arithmetic operations on the class of ordinal numbers. However, we shall skip the proof of this fact.

Even though ordinal numbers generalize natural numbers, arithmetic on ordinal numbers is substantially different than arithmetic on natural numbers. For example, let us calculate the ordinals $1 + \omega$ and $\omega + 1$. By definition,

$$1 + \omega = sup\{1 + n : n \in \omega\} = sup\{n : n \in \omega\} = \omega$$

On the other hand, $\omega + 1 = \omega + S(0) = S(\omega + 0) = S(\omega)$. Thus, $1 + \omega \neq \omega + 1$ and hence ordinal addition is not commutative. Similarly, one can prove that $2 \cdot \omega = \omega$ but $\omega \cdot 2 = \omega + \omega$ and hence ordinal multiplication is not commutative.

Exercise 36. *Show that there exist non-zero countable ordinals γ, δ, ϵ such that*

- $\omega + \gamma = \gamma$
- $\omega \cdot \delta = \delta$
- $\omega^\epsilon = \epsilon$

As we have noted, ordinal arithmetic is different than arithmetic on natural numbers. That said, many properties of arithmetic operations on natural numbers extend to arithmetic operations on ordinal numbers. We shall list some of these

properties and prove two of them using the principle of transfinite induction. The reader is expected to prove the rest of these properties as an exercise[7].

Lemma 37. *For all ordinals α, β and γ, we have*

a. $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$.
b. $\alpha < \beta \longleftrightarrow \gamma + \alpha < \gamma + \beta$.
c. $\alpha \leq \beta \longrightarrow \alpha + \gamma \leq \beta + \gamma$.
d. $\gamma + \alpha = \gamma + \beta \longleftrightarrow \alpha = \beta$.
e. $(\alpha \cdot \beta) \cdot \gamma = \alpha \cdot (\beta \cdot \gamma)$.
f. $\alpha \cdot (\beta + \gamma) = (\alpha \cdot \beta) + (\alpha \cdot \gamma)$.
g. $\alpha < \beta \longrightarrow \alpha \cdot \gamma \leq \beta \cdot \gamma$.
h. *If* $\gamma > 0$, *then* $\alpha < \beta \longrightarrow \gamma \cdot \alpha < \gamma \cdot \beta$.
i. $\alpha^{\beta+\gamma} = \alpha^\beta \cdot \alpha^\gamma$.
j. $(\alpha^\beta)^\gamma = \alpha^{\beta \cdot \gamma}$.
k. *If* $\gamma > 1$, *then* $\alpha < \beta \to \gamma^\alpha < \gamma^\beta$.
l. $\alpha < \beta \longrightarrow \alpha^\gamma \leq \beta^\gamma$.

Proof. [a.] We shall prove this by transfinite induction on γ. Let α and β be ordinals. Clearly $(\alpha + \beta) + 0 = \alpha + \beta = \alpha + (\beta + 0)$ and hence the claim holds for $\gamma = 0$. Let γ be an ordinal and assume that $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$. Then $(\alpha + \beta) + S(\gamma) = S((\alpha + \beta) + \gamma) = S(\alpha + (\beta + \gamma)) = \alpha + S(\beta + \gamma) = \alpha + (\beta + S(\gamma))$ and hence the claim holds for the ordinal $S(\gamma)$. Finally, let γ be a limit ordinal and assume that $(\alpha + \beta) + \theta = \alpha + (\beta + \theta)$ for all $\theta < \gamma$. First, note that by the inductive assumption

$$(\alpha + \beta) + \gamma = sup\{(\alpha + \beta) + \theta : \theta < \gamma\} = sup\{\alpha + (\beta + \theta) : \theta < \gamma\}$$

On the other hand, it follows from the next lemma[8] that

$$sup\{\alpha + (\beta + \theta) : \theta < \gamma\} = \alpha + sup\{\beta + \theta : \theta < \gamma\} = \alpha + (\beta + \gamma)$$

It follows from the principle of transfinite induction that the claim holds for all ordinals γ.

[b.] We shall prove this by transfinite induction on β. Let $P(\beta)$ be the property that β is an ordinal and $\alpha < \beta \longrightarrow \gamma + \alpha < \gamma + \beta$ for all ordinals α, γ. Note that $P(0)$ trivially holds. Assume that $P(\beta)$ holds for some ordinal β. Let γ and α be ordinals such that $\alpha < S(\beta)$. Then either $\alpha < \beta$ or $\alpha = \beta$. If $\alpha < \beta$, then we have

$$\gamma + \alpha < \gamma + \beta < S(\gamma + \beta) = \gamma + S(\beta)$$

by the inductive assumption. If $\alpha = \beta$, then $\gamma + \alpha = \gamma + \beta < S(\gamma + \beta) = \gamma + S(\beta)$. Hence $P(S(\beta))$ holds. Finally, let β be a limit ordinal and assume that $P(\theta)$ holds for all ordinals $\theta < \beta$. Then, for any $\alpha < \beta$, we have $\alpha < S(\alpha) < \beta$ and hence

$$\gamma + \alpha < S(\gamma + \alpha) = \gamma + S(\alpha) \leq sup\{\gamma + \theta : \theta < \beta\} = \gamma + \beta$$

which shows that $P(\beta)$ holds. By the principle of transfinite induction, $P(\beta)$ holds for all ordinals β. □

[7]Most of these identities can be proven via transfinite induction. However, you should know that, in order to prove some of these identities, you may need some other identities proven first.

[8]The proof of the next lemma only uses Lemma 37.b whose proof does not use Lemma 37.a. Consequently, we are justified to use this lemma.

Before concluding this subsection, we would like to mention that addition, multiplication and exponentiation on ordinal numbers are "continuous" in the second variable in the following sense.

Lemma 38. *Let X be a non-empty set of ordinals and α be an ordinal. Then*

$$sup\{\alpha + \beta : \beta \in X\} = \alpha + sup(X)$$

Proof. Let $\gamma = sup(X)$. The proof splits into two cases.

- Assume that $\gamma \notin X$. In this case, γ cannot be a successor ordinal since $\gamma = \theta + 1$ and $\gamma \notin X$ together imply that $\beta \leq \theta$ for all $\beta \in X$ and hence $\gamma = sup(X) \leq \theta$. Therefore, either $\gamma = 0$ or γ is a limit ordinal. Since X is non-empty and $\gamma \notin X$, if it were the case that $\gamma = 0$, then X would contain ordinals other than 0 contradicting that $sup(X) = \gamma$. Therefore, γ is a limit ordinal. By definition of ordinal addition,

  $$\alpha + sup(X) = \alpha + \gamma = sup\{\alpha + \beta : \beta \in \gamma\}$$

  On the one hand, we have $\{\alpha + \beta : \beta \in X\} \subseteq \{\alpha + \beta : \beta \in \gamma\}$ and hence

  $$sup\{\alpha + \beta : \beta \in X\} \leq sup\{\alpha + \beta : \beta \in \gamma\}$$

  On the other hand, since $\gamma = sup(X)$, for every ordinal $\theta \in \gamma$ there exists $\beta \in X$ such that $\theta < \beta$ and hence

  $$sup\{\alpha + \beta : \beta \in \gamma\} \leq sup\{\alpha + \beta : \beta \in X\}$$

  It follows that $sup\{\alpha + \beta : \beta \in X\} = \alpha + \gamma = \alpha + sup(X)$.
- Assume that $\gamma \in X$. Then we have that $\alpha + \gamma \leq sup\{\alpha + \beta : \beta \in X\}$. Moreover, it follows from Lemma 37.b and the definition of the supremum that $\alpha + \gamma \geq sup\{\alpha + \beta : \beta \in X\}$. Therefore

  $$sup\{\alpha + \beta : \beta \in X\} = \alpha + \gamma = \alpha + sup(X)$$

□

Using similar arguments, one can also prove the following results.

Exercise 37. *Let X be a non-empty set of ordinals and α be an ordinal. Prove that*

$$sup\{\alpha \cdot \beta : \beta \in X\} = \alpha \cdot sup(X)$$

Exercise 38. *Let X be a non-empty set of ordinals and $\alpha > 1$ be an ordinal. Prove that*

$$sup\{\alpha^\beta : \beta \in X\} = \alpha^{sup(X)}$$

7.5. **Cantor normal form of ordinal numbers.** In this subsection, we shall prove that each ordinal number γ can be expressed uniquely in a canonical form called the *Cantor normal form* of γ. The idea is to generalize the base k expansion of natural numbers to ordinal numbers and express each ordinal number in a unique "base ω" expansion. In order to prove the existence and uniqueness of Cantor normal form, we shall need several lemmas.

Lemma 39. *Let α, γ be ordinals such that $\alpha \leq \gamma$. Then there exists a unique ordinal β such that $\gamma = \alpha + \beta$.*

Proof. It follows from Lemma 37 that $\alpha+(\gamma+1) \geq (\gamma+1) > \gamma$ and hence there exists a least ordinal δ such that $\alpha+\delta > \gamma$. We claim that δ is a successor ordinal. If δ were limit, then $\alpha+\delta = sup\{\alpha+\theta : \theta < \delta\} > \gamma$ would imply that there exists an ordinal $\theta < \delta$ such that $\alpha+\theta > \gamma$, which contradicts the choice of δ. Thus δ is a successor ordinal, say $\delta = S(\beta)$. Then, by the choice of δ, we have $\alpha+\beta \leq \gamma$. If it were the case that $\alpha+\beta < \gamma$, then we would have $\gamma \geq S(\alpha+\beta) = \alpha+S(\beta) = \alpha+\delta > \gamma$, which is a contradiction. Therefore, $\alpha+\beta = \gamma$. □

Lemma 40. *Let α, γ be ordinals such that $1 \leq \alpha \leq \gamma$. Then there exists a greatest ordinal β such that $\alpha \cdot \beta \leq \gamma$.*

Proof. It follows from Lemma 37 that $\alpha \cdot (\gamma+1) \geq (\gamma+1) > \gamma$ and hence there exists a least ordinal δ such that $\alpha \cdot \delta > \gamma$. If δ were limit, then $\alpha \cdot \delta = sup\{\alpha \cdot \theta : \theta < \delta\} > \gamma$ would imply that there exists an ordinal $\theta < \delta$ such that $\alpha \cdot \theta > \gamma$, which contradicts the choice of δ. Thus δ is a successor ordinal, say $\delta = S(\beta)$. Then, by the choice of δ, we have $\alpha \cdot \beta \leq \gamma$ and $\alpha \cdot S(\beta) > \gamma$. Hence, β is the greatest ordinal such that $\alpha \cdot \beta \leq \gamma$. □

Using the exact same proof strategy in the previous lemmas, one can also prove the following.

Lemma 41. *Let α, γ be ordinals such that $2 \leq \alpha \leq \gamma$. Then there exists a greatest ordinal β such that $\alpha^\beta \leq \gamma$.*

Proof. Exercise. □

We are now ready to prove the analogue of Euclidean division for ordinals numbers.

Lemma 42. *Let α, γ be ordinals such that $\gamma \neq 0$. Then there exist unique ordinals β and ρ with $\rho < \gamma$ such that $\alpha = \gamma \cdot \beta + \rho$.*

Proof. If $\alpha < \gamma$, then clearly we can choose $\beta = 0$ and $\rho = \alpha$ since $\alpha = \gamma \cdot 0 + \alpha$. Assume that $\gamma \leq \alpha$. Then it follows from Lemma 40 that there exists β such that β is the greatest ordinal for which we have $\gamma \cdot \beta \leq \alpha$. By Lemma 39, there exists an ordinal ρ such that $\alpha = \gamma \cdot \beta + \rho$. If it were the case that $\rho \geq \gamma$, then we would have $\alpha = \gamma \cdot \beta + \rho \geq \gamma \cdot \beta + \gamma = \gamma \cdot (\beta+1)$, contradicting the choice of β. Hence we have $\rho < \gamma$.

To prove the uniqueness of β and ρ, assume that $\alpha = \gamma \cdot \beta + \rho = \gamma \cdot \beta' + \rho'$ for some ordinals $\beta, \beta', \rho, \rho'$ with $\rho, \rho' < \gamma$. If it were the case that $\beta < \beta'$, then we would have $\beta+1 \leq \beta'$ and hence

$$\alpha = \gamma \cdot \beta + \rho < \gamma \cdot \beta + \gamma = \gamma \cdot (\beta+1) \leq (\gamma \cdot \beta') \leq \gamma \cdot \beta' + \rho' = \alpha$$

which is a contradiction. Similarly, we cannot have $\beta' < \beta$. Hence $\beta = \beta'$. But then Lemma 37.d implies that $\rho = \rho'$, which completes the proof of the theorem. □

Lemma 43. *Let α, β be ordinals such that $\alpha < \beta$ and $k \in \omega$ be a natural number. Then $\omega^\alpha \cdot k < \omega^\beta$.*

Proof. By Lemma 37, we have $\omega^\alpha \cdot k < \omega^\alpha \cdot \omega = \omega^{\alpha+1} \leq \omega^\beta$. □

Next will be proven the main theorem of this subsection.

Theorem 36. *Let $\alpha > 0$ be an ordinal number. Then there exist unique ordinals $\beta_1 > \beta_2 > \cdots > \beta_n$ and positive natural numbers $k_1, k_2, \ldots, k_n$ such that*

$$\alpha = \omega^{\beta_1} \cdot k_1 + \omega^{\beta_2} \cdot k_2 + \cdots + \omega^{\beta_n} \cdot k_n$$

The expression $\omega^{\beta_1} \cdot k_1 + \omega^{\beta_2} \cdot k_2 + \cdots + \omega^{\beta_n}$ in the statement of the theorem is said to be the *Cantor normal form* of the ordinal α. We shall see later that ordinal arithmetic is substantially easier when one uses Cantor normal form of ordinals.

Proof of Theorem 36. First, we prove the existence of Cantor normal forms of ordinals by transfinite induction on $\alpha > 0$. Let $\alpha > 0$ be an ordinal and assume that all ordinals $1 \leq \gamma < \alpha$ have Cantor normal forms. If $\alpha < \omega$, then clearly $\omega^0 \cdot \alpha$ is a Cantor normal form of α. Assume that $\omega \leq \alpha$. Then, by Lemma 41, there exists a greatest ordinal β such that $\omega^\beta \leq \alpha$. It follows from Lemma 42 that there exist unique δ and ρ with $\rho < \omega^\beta$ such that $\alpha = \omega^\beta \cdot \delta + \rho$. If it were the case that $\delta \geq \omega$, then we would have

$$\alpha = \omega^\beta \cdot \delta + \rho \geq \omega^\beta \cdot \delta \geq \omega^\beta \cdot \omega = \omega^{\beta+1}$$

which contradicts the choice of β. Hence $\delta < \omega$. If $\rho = 0$, then α has a Cantor normal form $\alpha = \omega^\beta \cdot \delta$. Assume that $\rho \geq 1$. Then, by the inductive hypothesis, since $\rho < \omega^\beta \leq \alpha$, the ordinal ρ has a Cantor normal form $\omega^{\beta_2} \cdot k_2 + \cdots + \omega^{\beta_n} \cdot k_n$ for some ordinals $\beta_2 > \cdots > \beta_n$ and positive natural numbers $k_2, \ldots, k_n$. Since $\omega^{\beta_2} \leq \rho < \omega^\beta$, we have $\beta_2 < \beta$ by Lemma 37.k. It follows that

$$\omega^\beta \cdot \delta + \omega^{\beta_2} \cdot k_2 + \cdots + \omega^{\beta_n} \cdot k_n$$

is a Cantor normal form for the ordinal α. Thus, by transfinite induction, each ordinal $\alpha > 0$ has a Cantor normal form.

Next, we prove the uniqueness of the Cantor normal form of ordinals by transfinite induction on $\alpha > 0$. Let $\alpha > 0$ be an ordinal and assume that all ordinals $1 \leq \gamma < \alpha$ have unique Cantor normal forms. If $\alpha < \omega$, then clearly $\omega^0 \cdot \alpha$ is the unique Cantor normal form of α. Assume that $\omega \leq \alpha$. Let

$$\alpha = \omega^{\beta_1} \cdot k_1 + \omega^{\beta_2} \cdot k_2 + \cdots + \omega^{\beta_n} \cdot k_n = \omega^{\gamma_1} \cdot l_1 + \omega^{\gamma_2} \cdot l_2 + \cdots + \omega^{\gamma_m} \cdot l_m$$

be two Cantor normal forms of α. If $\beta_1 < \gamma_1$, then, by the previous lemma, we would have $\alpha \geq \omega^{\gamma_1} > \omega^{\beta_1} \cdot (k_1 + k_2 + \cdots + k_n) \geq \alpha$, which is a contradiction. Similarly, we cannot have $\gamma_1 < \beta_1$. Therefore we have $\gamma_1 = \beta_1$. Let

$$\begin{aligned} \delta &= \omega^{\gamma_1} \\ \rho_1 &= \omega^{\beta_2} \cdot k_2 + \cdots + \omega^{\beta_n} \cdot k_n \\ \rho_2 &= \omega^{\gamma_2} \cdot l_2 + \cdots + \omega^{\gamma_m} \cdot l_m \end{aligned}$$

Observe that $\alpha = \delta \cdot k_1 + \rho_1 = \delta \cdot l_1 + \rho_2$ that $\rho_1 < \delta$ and $\rho_2 < \delta$. Consequently, the uniqueness of ordinals in the statement of Lemma 42 implies that $k_1 = l_1$ and $\rho_1 = \rho_2$. If $\rho_1 = 0$, then $\omega^{\beta_1} \cdot k_1$ is the unique Cantor normal form of α. If $\rho_1 \neq 0$, then $\rho_2 \neq 0$ and ρ_1 has a unique Cantor normal form by the inductive assumption. It follows that $m = n$ and $\beta_i = \gamma_i$ for each $2 \leq i \leq m$. This shows that α has a unique Cantor normal form, which completes the proof. □

Having established that all ordinals have unique Cantor normal forms, we shall next prove several lemmas that allow us to easily compute sums and products of ordinals using their Cantor normal forms.

Lemma 44. *For all ordinals α, γ, if $\alpha < \gamma$, then $\omega^\alpha + \omega^\gamma = \omega^\gamma$.*

Proof. We shall prove this by transfinite induction[9] on γ.

- For $\gamma = 0$, the claim is vacuously true since there are no ordinals less than zero.
- Let γ be an ordinal and assume that the claim holds for γ. Then for any ordinal $\alpha < \gamma$ we have

$$\begin{aligned}\omega^\alpha + \omega^{S(\gamma)} = \omega^\alpha + \omega^\gamma \cdot \omega &= \omega^\alpha + \omega^\gamma \cdot (1 + \omega) \\ &= \omega^\alpha + \omega^\gamma + \omega^\gamma \cdot \omega \\ &= \omega^\gamma + \omega^\gamma \cdot \omega = \omega^\gamma \cdot (1 + \omega) = \omega^{S(\gamma)}\end{aligned}$$

  and for $\alpha = \gamma$ we have

$$\omega^\gamma + \omega^{S(\gamma)} = \omega^\gamma + \omega^\gamma \cdot \omega = \omega^\gamma \cdot (1 + \omega) = \omega^{S(\gamma)}$$

  Hence the claim holds for $S(\gamma)$.
- Let γ be a limit ordinal and assume that the claim holds for all ordinals strictly less than γ. Then, by Exercise 38, for any ordinal $\alpha < \gamma$, we have

$$\omega^\alpha + \omega^\gamma = \omega^\alpha + sup\{\omega^\theta : \theta < \gamma\} = sup\{\omega^\alpha + \omega^\theta : \theta < \gamma\}$$

  Since $\alpha < \gamma$ and γ is a limit, there exist ordinals θ such that $\alpha < \theta < \gamma$. It then follows from the inductive assumption that

$$\omega^\alpha + \omega^\gamma = sup\{\omega^\alpha + \omega^\theta : \theta < \gamma\} = sup\{\omega^\theta : \theta < \gamma\} = \omega^\gamma$$

  Therefore the claim holds for γ. By transfinite induction, the claim holds for all ordinals γ.

□

Next corollary immediately follows from the lemma above.

Corollary 37. *Let $\alpha < \gamma$ be ordinals and $m, n \in \omega$ be such that $n > 0$. Then $\omega^\alpha \cdot m + \omega^\gamma \cdot n = \omega^\gamma \cdot n$*

Proof. Exercise. (**Hint.** Use induction on m. For the base case, use the exercise[10].) □

Lemma 45. *Let $\omega^{\beta_1} \cdot k_1 + \omega^{\beta_2} \cdot k_2 + \cdots + \omega^{\beta_n} \cdot k_n$ be the Cantor normal form of a non-zero ordinal α. Then, for any $k \in \omega$ with $k > 0$, we have*

$$\alpha \cdot k = \omega^{\beta_1} \cdot (k_1 \cdot k) + \omega^{\beta_2} \cdot k_2 + \cdots + \omega^{\beta_n} \cdot k_n$$

Proof. Exercise. (**Hint.** Use induction on k and apply Corollary 37.) □

Lemma 46. *Let $\omega^{\beta_1} \cdot k_1 + \omega^{\beta_2} \cdot k_2 + \cdots + \omega^{\beta_n} \cdot k_n$ be the Cantor normal form of a non-zero ordinal α. Then, for any ordinal $\gamma > 0$, we have $\alpha \cdot \omega^\gamma = \omega^{\beta_1 + \gamma}$*

---

[9]We would like to note that this lemma can be proven without transfinite induction by simply observing that there exist unique ordinals $\beta > 0$ and θ such that $\gamma = \alpha + \beta$ and $\omega^\beta = \omega + \theta$; and hence $\omega^\alpha + \omega^\gamma = \omega^\alpha + \omega^{\alpha+\beta} = \omega^\alpha(1 + \omega^\beta) = \omega^\alpha(1 + \omega + \theta) = \omega^\alpha(\omega + \theta) = \omega^\alpha \cdot \omega^\beta = \omega^\gamma$. However, the author thinks that it would be more beneficial for the reader to see as many examples of transfinite induction as possible.

[10]**Exercise.** Prove that if $\theta \geq \omega$ is an ordinal and n is a natural number, then $n + \theta = \theta$.

Proof. It follows from Lemma 37 that $\omega^{\beta_1} \leq \alpha \leq \omega^{\beta_1} \cdot (k_1 + k_2 + \cdots + k_n)$ and hence, multiplying each side by ω^γ on the right, we get

$$\omega^{\beta_1} \cdot \omega^\gamma \leq \alpha \cdot \omega^\gamma \leq (\omega^{\beta_1} \cdot (k_1 + k_2 + \cdots + k_n)) \cdot \omega^\gamma$$

Since ordinal multiplication is associative and we have $(k_1+k_2+\cdots+k_n)\cdot\omega^\gamma = \omega^\gamma$, it follows that $\omega^{\beta_1} \cdot \omega^\gamma \leq \alpha \cdot \omega^\gamma \leq \omega^{\beta_1} \cdot \omega^\gamma$ and hence $\alpha \cdot \omega^\gamma = \omega^{\beta_1+\gamma}$. □

We now use these results to compute some sums and products of ordinals in their Cantor normal forms. One can prove similar results for ordinal exponentiation in order to express the Cantor normal form of α^β in terms of the Cantor normal forms of α and β. However, we shall not cover these identities and refer the curious reader to David Pierce's lectures notes[11].

Example. Find the Cantor normal form of the sum

$$(\omega^{\omega^\omega+\omega^2} \cdot 2 + \omega^2 \cdot 3 + 5) + (\omega^\omega + \omega \cdot 3 + 7)$$

Solution. Since ordinal addition is associative, we can ignore the parentheses. By applying Corollary 37 successively, we easily get

$$\begin{aligned}(\omega^{\omega^\omega+\omega^2} \cdot 2 + \omega^2 \cdot 3 + 5) + (\omega^\omega + \omega \cdot 3 + 7) &= \omega^{\omega^\omega+\omega^2} \cdot 2 + \omega^2 \cdot 3 + 5 + \omega^\omega + \omega \cdot 3 + 7 \\ &= \omega^{\omega^\omega+\omega^2} \cdot 2 + \omega^2 \cdot 3 + \omega^\omega + \omega \cdot 3 + 7 \\ &= \omega^{\omega^\omega+\omega^2} \cdot 2 + \omega^\omega + \omega \cdot 3 + 7\end{aligned}$$

Example. Find the Cantor normal form of the sum

$$(\omega^{\omega^\omega} \cdot 7 + \omega \cdot 6 + 1) + (\omega^{\omega^\omega} + \omega \cdot 3 + 7)$$

Solution. As in the previous example, associativity of ordinal addition and Corollary 37 imply that

$$\begin{aligned}(\omega^{\omega^\omega} \cdot 7 + \omega \cdot 6 + 1) + (\omega^{\omega^\omega} + \omega \cdot 3 + 7) &= \omega^{\omega^\omega} \cdot 7 + \omega \cdot 6 + 1 + \omega^{\omega^\omega} + \omega \cdot 3 + 7 \\ &= \omega^{\omega^\omega} \cdot 7 + \omega \cdot 6 + \omega^{\omega^\omega} + \omega \cdot 3 + 7 \\ &= \omega^{\omega^\omega} \cdot 7 + \omega^{\omega^\omega} + \omega \cdot 3 + 7\end{aligned}$$

It then follows from Lemma 37.f that

$$\begin{aligned}(\omega^{\omega^\omega} \cdot 7 + \omega \cdot 6 + 1) + (\omega^{\omega^\omega} + \omega \cdot 3 + 7) &= \omega^{\omega^\omega} \cdot 7 + \omega^{\omega^\omega} + \omega \cdot 3 + 7 \\ &= \omega^{\omega^\omega} \cdot (7 + 1) + \omega \cdot 3 + 7 \\ &= \omega^{\omega^\omega} \cdot 8 + \omega \cdot 3 + 7\end{aligned}$$

Example. Find the Cantor normal form of the product

$$(\omega^{\omega^\omega+\omega^3+1} + \omega^\omega) \cdot (\omega^{\omega^2} + \omega \cdot 2 + 3)$$

[11] One can access these notes via the link http://mat.msgsu.edu.tr/∼dpierce/Courses/320/2010/Notes/math-320-2010-text-main.pdf as of 27 April 2017.

Solution. Combining all the properties of ordinal arithmetic proven so far, we get

$$(\omega^{\omega^\omega+\omega^3+1}+\omega^\omega)\cdot(\omega^{\omega^2}+\omega\cdot 2+3)=$$
$$(\omega^{\omega^\omega+\omega^3+1}+\omega^\omega)\cdot\omega^{\omega^2}+(\omega^{\omega^\omega+\omega^3+1}+\omega^\omega)\cdot\omega\cdot 2+(\omega^{\omega^\omega+\omega^3+1}+\omega^\omega)\cdot 3=$$
$$\omega^{(\omega^\omega+\omega^3+1)+\omega^2}+\omega^{(\omega^\omega+\omega^3+1)+1}\cdot 2+(\omega^{\omega^\omega+\omega^3+1}+\omega^\omega)\cdot 3=$$
$$\omega^{\omega^\omega+\omega^3+\omega^2}+\omega^{\omega^\omega+\omega^3+2}\cdot 2+\omega^{\omega^\omega+\omega^3+1}\cdot 3+\omega^\omega$$

8. Cardinal numbers

In this section, we shall define the class of cardinal numbers, which is a special subclass of ordinal numbers. The motivation behind the concept of cardinal numbers comes from the following question: Can we find "natural representatives" for the equinumerosity classes[1] of sets?

We have established the fact that every strictly well-ordered set is isomorphic to a unique ordinal. If it is the case that every set can be well-ordered, then every set is equinumerous with some ordinal number and hence the equinumerosity class of a set can be represented by the least ordinal with which the set is equinumerous. Thus we need to show that every set can be well-ordered.

8.1. Zorn's lemma, the well-ordering theorem and the axiom of choice.

In this subsection, we shall prove two important consequences of the axiom of choice, which will turn out to be equivalent to the axiom of choice assuming only the axioms of ZF.

Lemma 47 (Zorn's lemma)**.** *Let $(\mathbb{P}, \preceq)$ be a partially ordered set such that every chain has an upper bound in $\mathbb{P}$, i.e. for every $C \subseteq \mathbb{P}$, if C is a chain, then there exists $p \in \mathbb{P}$ such that $c \preceq p$ for all $c \in C$. Then there exists a maximal element in $\mathbb{P}$ with respect to $\preceq$.*

Proof. Assume to the contrary that $\mathbb{P}$ does not have a maximal element. Then, for every chain $C \subseteq \mathbb{P}$, the set $\{u \in \mathbb{P} : \forall c \in C \ c \prec u\}$ is non-empty since there are no maximal elements and there exists an upper bound for C in $\mathbb{P}$.

Let $\mathcal{C}$ be the set of chains in $\mathbb{P}$. It follows from Lemma 6 there exists a function $h : \mathcal{P}(\mathbb{P}) - \{\emptyset\} \to \mathbb{P}$ such that $h(S) \in S$ for all non-empty $S \subseteq \mathbb{P}$. Consequently, there exists a function $f : \mathcal{C} \to \mathbb{P}$ such that $c \prec f(C)$ for all $c \in C$ and $C \in \mathcal{C}$, namely the function $f(C) = h(\{u \in \mathbb{P} : \forall c \in C \ c \prec u\})$.

$\mathbb{P}$ is clearly non-empty. Fix some arbitrary element $a_0 \in \mathbb{P}$. By transfinite recursion, define

$$a_\alpha = \begin{cases} f(\{a_\beta : \beta < \alpha\}) \text{ , if } \{a_\beta : \beta < \alpha\} \text{ is a chain} \\ a_0 \text{ , otherwise} \end{cases}$$

for all ordinals $\alpha > 0$. It follows from an easy transfinite induction argument that $a_\beta \prec a_\alpha = f(\{a_\gamma : \gamma < \alpha\})$ for all $\beta < \alpha$. It follows that, for every ordinal α, there exists an injective function from α to $\mathbb{P}$, namely the function $\gamma \mapsto a_\gamma$. This contradicts the existence of the Hartogs number of $\mathbb{P}$. Thus, $\mathbb{P}$ has a maximal element. □

Lemma 48 (The well-ordering theorem)**.** *Every set can be well-ordered, i.e. for every set A there exists a well-order relation $\preceq_A$ on the set A.*

Proof. Let A be a set. Consider the set

$$\mathbb{P} = \{(B, \preceq_B) : B \subseteq A \wedge \ \preceq_B \text{ is a well-order relation on } B\}$$

and the relation $\sqsubseteq$ on $\mathbb{P}$ given by

$$(B, \preceq_B) \sqsubseteq (C, \preceq_C) \iff$$
$$B \subseteq C \wedge \preceq_B \subseteq \preceq_C \ \wedge \text{"B is an initial segment of C in } (C, \preceq_C)\text{"}$$

[1]Given a set A, the class of all sets that are equinumerous with A is called the equinumerosity class of A

It is an exercise to the reader to check that $\sqsubseteq$ is a partial order relation on $\mathbb{P}$ that satisfies the hypothesis of Zorn's lemma. Consequently, there exists a maximal element $(M, \preceq_M)$ in the partially ordered set $(\mathbb{P}, \sqsubseteq)$. If it were the case that $M \neq A$, then, for any $u \in A - M$, the pair

$$(M \cup \{u\}, \preceq_M \cup\{(m, u) : m \in M\})$$

would belong to $\mathbb{P}$ and moreover, we would have

$$(M, \preceq_M) \sqsubset (M \cup \{u\},\ \preceq_M \cup \{(m, u) : m \in M \cup \{u\}\})$$

contradicting the maximality of $(M, \preceq_M)$. Thus $M = A$ and hence A is well-ordered via the relation $\preceq_M$. □

Notice that the proof of Zorn's lemma uses ZFC and the proof of the well-ordering theorem uses only the axioms of ZF and Zorn's lemma. Consequently, if we can prove that the axioms of ZF together with the well-ordering theorem imply the axiom of choice, then Zorn's lemma, the well-ordering theorem and the axiom of choice would be equivalent over ZF.

Exercise 39. *Assuming only the axioms of ZF, prove that the axiom of choice holds if and only if for every non-empty set X there exists a function $f : \mathcal{P}(X) - \{\emptyset\} \to X$ such that $f(S) \in S$ for all non-empty $S \subseteq X$.*

Lemma 49 (ZF)**.** *If the well-ordering theorem holds, then so does the axiom of choice.*

Proof. Let X be a non-empty set. If the well-ordering theorem holds, then there exists a well-order relation $\preceq$ on the set X. It follows that the relation

$$f = \{(S, a) : S \subseteq X \ \wedge\ S \neq \emptyset \ \wedge\ a \in S \ \wedge\ \forall s \in S\ a \preceq s\}$$

is a function from $\mathcal{P}(X) - \{\emptyset\}$ to X such that $f(S) \in S$ for all non-empty $S \subseteq X$. The existence of such a function is equivalent to the axiom of choice by the previous lemma. □

8.2. **Cardinal number of a set.** Having established that every set can be well-ordered, every set can be put in a bijection with some ordinal. We would like to declare the least ordinal with which a set is equinumerous to be the cardinality of this set. For this reason, let us first give a special name to such ordinals.

Definition 58. *An ordinal number α is said to be a cardinal number if α is not equinumerous with β for all ordinals $\beta < \alpha$.*

Definition 59. *Let X be a set. The cardinal number (or simply, the cardinality) of X is the unique cardinal number which is equinumerous with X, which is denoted by $|X|$.*

Notice that our usage of the word *cardinality* and the notation $|X|$ is consistent with our earlier usage. More precisely, there exists an injection from X to Y if and only if the cardinal number of X is less than or equal to the cardinal number of Y, for both of which we use the notation $|X| \leq |Y|$.

Exercise 40. *Let A and B be sets. Prove that $|A \times B| = |\ |A| \times |B|\ |$ and $|\ {}^{B}A\ | = |\ {}^{|B|}|A|\ |$ and $|\ |A|\ | = |A|$.*

It is easily seen that the set of finite cardinal numbers are precisely the set of natural numbers and the first infinite cardinal is ω. By transfinite recursion, define the following class of ordinals.

- $\aleph_0 = \omega$
- $\aleph_{\alpha+1} = \aleph(\aleph_\alpha)$, the Hartogs number of $\aleph_\alpha$, for all ordinals α, and
- $\aleph_\gamma = sup\{\aleph_\theta : \theta < \gamma\}$ for all limit ordinals γ.

It is easily seen that $\alpha < \beta$ implies that $\aleph_\alpha < \aleph_\beta$. Consequently, the collection of $\aleph$ numbers is a proper class. We shall next prove that the proper class of $\aleph$ numbers is a complete (transfinite) list of infinite cardinal numbers.

Lemma 50. *For all ordinals α, $\aleph_\alpha$ is a cardinal number.*

Proof. We shall prove this by transfinite induction. The claim clearly holds for $\alpha = 0$. Assume that the claim holds for an ordinal $\alpha \geq 0$. Since the Hartogs number of $\aleph_\alpha$ is the least ordinal which does not inject into $\aleph_\alpha$, no ordinal less than $\aleph_{\alpha+1}$ can be equinumerous with $\aleph_{\alpha+1}$ and hence $\aleph_{\alpha+1}$ is a cardinal. Finally, assume that the claim holds for all ordinals less than a limit ordinal γ. If $\aleph_\gamma$ were not a cardinal, then there would exist some ordinal $\delta < \aleph_\gamma$ such that δ and $\aleph_\gamma$ are equinumerous. In this case, by the definition of $\aleph_\gamma$, there would exist $\theta < \gamma$ such that $\delta < \aleph_\theta$. Since $\aleph_\theta \subseteq \aleph_\gamma$ and the latter is equinumerous with $\delta < \aleph_\theta$, we would obtain that $\aleph_\theta$ is not a cardinal number, which is a contraction. Thus $\aleph_\gamma$ is a cardinal, which completes the transfinite induction. □

Lemma 51. *Let κ be an infinite ordinal number. If κ is a cardinal number, then there exists an ordinal number α such that $\kappa = \aleph_\alpha$.*

Proof. Let κ be a cardinal number. Set λ to be the set of ordinals $\{\theta : \aleph_\theta \in \kappa\}$. It is easily seen that λ is transitive and hence is an ordinal. Observe that it follows from $\lambda \notin \lambda$ that $\aleph_\lambda \notin \kappa$ and hence $\kappa \leq \aleph_\lambda$. The proof splits into three cases.

- Case 1 ($\lambda = 0$): In this case, $\kappa = \omega = \aleph_0 = \aleph_\lambda$.
- Case 2 (λ is a successor ordinal): In this case $\lambda = \delta + 1$ for some ordinal δ. Then $\aleph_\delta \in \kappa$ and κ being a cardinal together imply that $\aleph_\lambda = \aleph(\aleph_\delta) \leq \kappa$. Consequently $\kappa = \aleph_\lambda$.
- Case 3 (λ is a limit ordinal): In this case, since $\aleph_\theta \in \kappa$ for all $\theta \in \lambda$, we have that $\aleph_\lambda = sup\{\aleph_\theta : \theta < \lambda\} \leq \kappa$. Consequently $\kappa = \aleph_\lambda$.

In all three cases, there exists an ordinal number α such that $\kappa = \aleph_\alpha$ □

Finally, before we conclude this subsection, we shall define the notions of a successor cardinal and a limit cardinal. Given a cardinal number κ, we denote the least cardinal number that is greater than κ by κ^+.

Definition 60. *Let λ be an infinite cardinal number. The cardinal λ is said to be a successor cardinal if $\lambda = \kappa^+$ for some cardinal number κ and is said to be a limit cardinal if it is not a successor cardinal.*

Equivalently, an infinite cardinal $\aleph_\alpha$ is a successor cardinal if α is a successor ordinal and is a limit cardinal if α is a limit ordinal or zero.

8.3. **Cardinal arithmetic.** In this subsection, we shall define arithmetic operations on the class of cardinal numbers. Given two cardinal numbers κ and λ, we

define

$$\kappa + \lambda = \left|\kappa \times \{0\} \bigcup \lambda \times \{1\}\right|$$
$$\kappa \cdot \lambda = |\kappa \times \lambda|$$
$$\kappa^\lambda = \left|{}^\lambda\kappa\right|$$

In other words, the sum of two cardinals is the cardinality of their "disjoint union", the product of two cardinals is the cardinality of their cartesian product and κ raised to the power λ is the cardinality of the set of functions from λ to κ.

Even though cardinal numbers are ordinal numbers of special type, cardinal arithmetic is very different than ordinal arithmetic. For example, ω^ω is a countable ordinal if the exponentiation operation is considered in ordinal arithmetic and is an uncountable cardinal if the exponentiation operation is considered in cardinal arithmetic.

Below we shall list some properties of cardinal arithmetic and prove one of these properties. The reader is expected to prove the rest. We note that, unlike properties of ordinal arithmetic, it is not necessary to use transfinite induction to prove the following.

Lemma 52. *Let κ, λ, μ be cardinal numbers. Then*

a. $\kappa + 0 = 0 + \kappa = \kappa$
b. $\kappa + \lambda = \lambda + \kappa$
c. $(\kappa + \lambda) + \mu = \kappa + (\lambda + \mu)$
d. $\kappa \leq \mu \longrightarrow \kappa + \lambda \leq \mu + \lambda$
e. $\kappa \cdot 0 = 0 \cdot \kappa = 0$
f. $\kappa \cdot 1 = 1 \cdot \kappa = \kappa$
g. $\kappa \cdot \lambda = \lambda \cdot \kappa$
h. $(\kappa \cdot \lambda) \cdot \mu = \kappa \cdot (\lambda \cdot \mu)$
i. $\kappa \leq \mu \longrightarrow \kappa \cdot \lambda \leq \mu \cdot \lambda$
j. $\kappa^0 = 1$
k. $\kappa^1 = \kappa$
l. $1^\kappa = \kappa$
m. $(\kappa^\lambda)^\mu = \kappa^{\lambda \cdot \mu}$
n. $\kappa^\lambda \cdot \kappa^\mu = \kappa^{\lambda + \mu}$
o. $\kappa^\mu \cdot \lambda^\mu = (\kappa \cdot \lambda)^\mu$
p. $\kappa \leq \lambda \longrightarrow \kappa^\mu \leq \lambda^\mu$
r. $\mu \geq 1 \wedge \kappa \leq \lambda \longrightarrow \mu^\kappa \leq \mu^\lambda$

Proof. [m.] Since $\kappa^\lambda = |{}^\lambda\kappa|$ and $\lambda \cdot \mu = |\lambda \times \mu|$, by Exercise 40, it is sufficient to prove that

$$|{}^\mu({}^\lambda\kappa)| = |{}^{\lambda \times \mu}\kappa|$$

Consider the function $g : {}^\mu({}^\lambda\kappa) \to {}^{\lambda \times \mu}\kappa$ given by

$$(g(f))(\theta_1, \theta_2) = (f(\theta_2))(\theta_1)$$

for all $f \in {}^\mu({}^\lambda\kappa)$. In other words, the function g takes a function f from μ to ${}^\lambda\kappa$ to the function from $\lambda \times \mu$ to κ whose value at (θ_1, θ_2) is the value of the function $f(\theta_1)$ at the point θ_2. We claim that g is a bijection.

Let $f_1, f_2 \in {}^\mu({}^\lambda\kappa)$ and assume that $g(f_1) = g(f_2)$. Let $\theta \in \mu$. Then, for any $\theta' \in \lambda$, we have

$$(f_2(\theta))(\theta') = (g(f_2))(\theta', \theta) = (g(f_1))(\theta', \theta) = (f_1(\theta))(\theta')$$

Hence $f_1(\theta) = f_2(\theta)$ for all $\theta \in \mu$. Thus $f_1 = f_2$ and hence g is an injection. To prove that g is onto, let $h \in {}^{\lambda\times\mu}\kappa$ be any function. Consider the function $f \in {}^{\mu}({}^{\lambda}\kappa)$ given by

$$(f(\theta'))(\theta) = h(\theta, \theta')$$

for any $\theta \in \lambda$ and $\theta' \in \mu$. It is easily seen that $g(f) = h$ and hence g is onto. □

Exercise 41. *Let κ be an infinite cardinal. Prove that $\kappa + \kappa = \kappa$.*
***Hint**. Notice that every ordinal number in κ can be written as $\lambda + n$ for some natural number $n \in \omega$ and $\lambda \in \kappa$. Consider the function $f : \kappa \times \{0\} \to \kappa$ given by $f(\lambda+n, 0) = \lambda+2n$ and the function $g : \kappa\times\{1\} \to \kappa$ given by $g(\lambda+n, 1) = \lambda+2n+1$.*

It turns out that the sum or product of two cardinals at least one of which is infinite is the maximum of these two cardinals. In order to prove this fact, we shall need the following theorem.

Theorem 38. *For all ordinals α, we have $|\aleph_\alpha \times \aleph_\alpha| = \aleph_\alpha$.*

Proof. We shall prove this by transfinite induction on α. By Lemma 26, the claim holds for $\alpha = 0$. Let $\alpha \geq 0$ be an ordinal number and assume that the claim holds for all ordinal $\beta < \alpha$. Consider the relation $\prec$ on $\aleph_\alpha \times \aleph_\alpha$ given by

$$(\gamma_1, \delta_1) \prec (\gamma_2, \delta_2) \longleftrightarrow max\{\gamma_1, \delta_1\} < max\{\gamma_2, \delta_2\} \vee$$
$$(max\{\gamma_1, \delta_1\} = max\{\gamma_2, \delta_2\} \ \wedge \ \gamma_1 < \gamma_2) \vee$$
$$(max\{\gamma_1, \delta_1\} = max\{\gamma_2, \delta_2\} \ \wedge \ \gamma_1 = \gamma_2 \ \wedge \ \delta_1 < \delta_2)$$

We claim that $\prec$ is a strict well-order relation on $\aleph_\alpha \times \aleph_\alpha$. We skip the details of checking that $\prec$ is a strict linear order relation[2] and only show that every non-empty subset of $\aleph_\alpha \times \aleph_\alpha$ has a minimal element with respect to $\prec$. Let $X \subseteq \aleph_\alpha \times \aleph_\alpha$ be a non-empty set. The set $\{max\{\gamma, \delta\} : (\gamma, \delta) \in X\}$ is non-empty and hence has a least element, say the ordinal θ. Then the set

$$\{\gamma : \exists\delta \ (\gamma, \delta) \in X \ \wedge \ max\{\gamma, \delta\} = \theta\}$$

is non-empty and has a least element, say the ordinal ϵ. Similarly, the set

$$\{\delta : (\epsilon, \delta) \in X \ \wedge \ max\{\epsilon, \delta\} = \theta\}$$

is non-empty and has a least element, say the ordinal ξ. We claim that (ϵ, ξ) is the least element of X. Given any $(\gamma, \delta) \in X$,

- If $max\{\gamma, \delta\} > \theta$, then $(\epsilon, \xi) \prec (\gamma, \delta)$.
- If $max\{\gamma, \delta\} = \theta$, then, by construction, we have either $\epsilon = \gamma$ or $\epsilon < \gamma$. In the former case, by construction, we have $\xi \leq \delta$ and hence $(\epsilon, \xi) \preceq (\gamma, \delta)$. In the latter case, we have $(\epsilon, \xi) \prec (\gamma, \delta)$.

Therefore $(\epsilon, \xi) \preceq (\gamma, \delta)$, which shows that (ϵ, ξ) is the least element of X with respect to $\prec$. Having shown that $\prec$ is a strict-well order relation, we shall next count the cardinality of the predecessors of an element (γ, δ). Let $(\gamma, \delta) \in \aleph_\alpha \times \aleph_\alpha$ and $(\gamma', \delta') \in pred(\gamma, \delta)$. Set $\lambda = max\{\gamma, \delta\}$. There are two possible cases.

- If $max\{\gamma', \delta'\} < max\{\gamma, \delta\}$, then $(\gamma', \delta') \in \lambda \times \lambda \subseteq (\lambda + 1) \times (\lambda + 1)$.
- If $max\{\gamma', \delta'\} = max\{\gamma, \delta\}$, then we have

$$(\gamma', \delta') \in (\lambda + 1) \times (\lambda + 1)$$

[2]The reader is expected to prove this as an exercise.

Therefore, the cardinality of $pred(\gamma, \delta) \subseteq (\lambda + 1) \times (\lambda + 1)$ is at most

$$|(\lambda + 1) \times (\lambda + 1)| = | \ |\lambda + 1| \ \times \ |\lambda + 1| \ | = | \ |\lambda| \ \times \ |\lambda| \ |$$

Since $|\lambda| < \aleph_\alpha$, there exists $\nu < \alpha$ such that $|\lambda| = \aleph_\nu$ and hence, by the induction assumption, we have $| \ |\lambda| \ \times \ |\lambda| \ | = |\lambda| < \aleph_\alpha$. Therefore, any element in $\aleph_\alpha \times \aleph_\alpha$ has less than $\aleph_\alpha$ many predecessors in the strictly well-ordered set $(\aleph_\alpha \times \aleph_\alpha, \prec)$. It follows that we have $ot(\aleph_\alpha \times \aleph_\alpha, \prec) \leq \aleph_\alpha$. Consequently, $|\aleph_\alpha \times \aleph_\alpha| \leq |\aleph_\alpha|$. Finally, observe that we have $|\aleph_\alpha| \leq |\aleph_\alpha \times \aleph_\alpha|$ via the injection $\xi \mapsto (\xi, 0)$ and hence $|\aleph_\alpha \times \aleph_\alpha| = \aleph_\alpha$ by Cantor-Schröder-Bernstein theorem. This completes the transfinite induction. □

Corollary 39. *Let κ and λ be infinite cardinals. Then $\kappa + \lambda = \kappa \cdot \lambda = max\{\kappa, \lambda\}$.*

Proof. Let κ and λ be infinite cardinals. Without loss of generality, assume that $\kappa = \aleph_\alpha$ and $\lambda = \aleph_\beta$ for some ordinals $\alpha < \beta$. It follows from the previous theorem and Lemma 52 that

$$\lambda = \aleph_\beta \leq \aleph_\beta + \aleph_\alpha \leq \aleph_\beta + \aleph_\beta = \aleph_\beta \cdot 2 \leq \aleph_\beta \cdot \aleph_\beta \leq \aleph_\beta = \lambda$$

and that

$$\lambda = \aleph_\beta \leq \aleph_\beta \cdot \aleph_\alpha \leq \aleph_\beta \cdot \aleph_\beta = \aleph_\beta = \lambda$$

□

Exercise 42. *Let κ be an infinite cardinal. Prove that $\kappa^n = \kappa$ for all $n \in \omega - \{0\}$.*

Recall that a countable union of countable sets is countable. It turns out that, in this fact, one can change the word "countable" to "of size κ" for any infinite cardinal κ, using Corollary 39.

Lemma 53. *Let κ be an infinite cardinal and $\{X_\alpha\}_{\alpha \in \kappa}$ be an indexed system of sets such that $|X_\alpha| \leq \kappa$. Then $|\bigcup_{\alpha \in \kappa} X_\alpha| \leq \kappa$.*

Proof. For each $\alpha \in \kappa$, choose a surjection $f_\alpha : \kappa \to X_\alpha$. Consider the map $g : \kappa \times \kappa \to \bigcup_{\alpha \in \kappa} X_\alpha$ given by $g(\alpha, \beta) = f_\alpha(\beta)$ for all $\alpha, \beta \in \kappa$. It is straightforward to check that g is a surjection. By Lemma 25, there exists an injection from $\bigcup_{\alpha \in \kappa} X_\alpha$ to $\kappa \times \kappa$. Since $|\kappa \times \kappa| = \kappa$, we have that $|\bigcup_{\alpha \in \kappa} X_\alpha| \leq \kappa$. □

Contrary to cardinal addition and multiplication, determining the result of an exponentiation in cardinal arithmetic is "difficult" in a certain sense that will be explained in the next subsection.

8.4. **Continuum Hypothesis and Generalized Continuum Hypothesis.** After Cantor proved that the cardinality of real numbers is greater than the cardinality of natural numbers, he tried to find a set whose cardinality is strictly between that of real numbers and that of natural numbers.

After failed attempts, he conjectured that there exists no such set. The statement that there exists no set X such that $|\mathbb{N}| < |X| < |\mathbb{R}|$ has been known as the *continuum hypothesis*. Since the cardinality of the set of real numbers $\mathbb{R}$, which is usually called the *continuum* and denoted by $\mathfrak{c}$, is equal to the cardinal $2^{\aleph_0}$, the continuum hypothesis (CH) is equivalent to the assertion that

$$\aleph_1 = 2^{\aleph_0}$$

By transfinite recursion, define the following class of cardinal numbers called the $\beth$ numbers.

- $\beth_0 = \aleph_0$
- $\beth_{\alpha+1} = 2^{\beth_\alpha}$ for all ordinals α, and
- $\beth_\gamma = sup\{\beth_\theta : \theta < \gamma\}$ for all limit ordinals γ.

In this notation, the continuum hypothesis can be restated as

$$\aleph_1 = \beth_1$$

One can generalize this statement in the obvious way. The statement that

$$\aleph_\alpha = \beth_\alpha \text{ for all ordinals } \alpha$$

is known as the *generalized continuum hypothesis*. Equivalently, the generalized continuum hypothesis (GCH) is the assertion that $\aleph_{\alpha+1} = 2^{\aleph_\alpha}$ for all ordinals α.

One of the most fascinating achievement of the last century in set theory is the *independence* of these statement from the axioms of ZFC. It follows from the work of Kurt Gödel in 1940 and the work of Paul Cohen in 1963 that if the axiom of ZFC are consistent[3], then one cannot[4] prove or disprove CH or GCH using the axioms of ZFC.

One should perhaps mention that Paul Cohen was awarded the Fields Medal in 1966 for his ground breaking work, which remains the only Fields Medal awarded for a work in mathematical logic.

It should be obvious at this point that the axioms of ZFC are insufficient to provide an answer to some basic questions about cardinal exponentiation. Nevertheless, there are some results that can proven in ZFC regarding cardinal exponentiation, some of which we are going to learn in the next subsection.

8.5. **More on cardinal exponentiation.** We begin by restating Cantor's theorem in the terminology of cardinal numbers.

Theorem 40 (Cantor's theorem revisited)**.** *For all cardinals κ, $\kappa < 2^\kappa$.*

An immediate consequence of the results we have proven so far is that the set of functions from κ to λ has the same cardinality as the power set of κ provided that λ is not "too big".

Lemma 54. *Let κ and λ be cardinal numbers such that $\omega \leq \kappa$ and $2 \leq \lambda \leq \kappa$. Then $2^\kappa = \lambda^\kappa$.*

Proof. It follows from Cantor's theorem, Exercise 40 and Corollary 39 that

$$2^\kappa \leq \lambda^\kappa \leq \kappa^\kappa \leq (2^\kappa)^\kappa \leq 2^{\kappa\cdot\kappa} = 2^\kappa$$

and hence $2^\kappa = \lambda^\kappa$. □

Next we turn our attention to the following problem. We know that, given an infinite cardinal κ, we have $\kappa^n = \kappa$ for all $n \in \omega$ and $\kappa^\kappa > \kappa$. One can ask the following question: What is the least cardinal λ such that $\kappa^\lambda > \kappa$?

This question is "difficult" in the sense that its answer is generally independent of ZFC for arbitrary κ, as was the case with the continuum hypothesis[5]. Nevertheless,

[3]A set of axioms is said to be consistent if no proof of a contradiction can be derived from these axioms.

[4]Here we do not merely mean that we do not currently know whether CH or GCH is true or not. What is meant is that it is a *theorem* that CH and GCH cannot be proved or disproved using the axioms of ZFC, provided that ZFC is consistent.

[5]For example $(2^{\aleph_0})^{\aleph_1} = 2^{\aleph_1} = \aleph_2 > 2^{\aleph_0}$ under GCH, whereas, $(2^{\aleph_0})^{\aleph_1} = 2^{\aleph_1} = 2^{\aleph_0}$ under the assumption $2^{\aleph_0} = 2^{\aleph_1}$, which is known as Luzin's hypothesis.

it happens to be the case that one can provide an upper bound for λ which is sometimes better than κ. In order to learn this upper bound, we shall need the concept of *cofinality*.

Definition 61. *Let $(\mathbb{P}, \preceq)$ be a partially ordered set. A subset $A \subseteq \mathbb{P}$ is said to be cofinal if for all $p \in \mathbb{P}$ there exists $q \in A$ such that $p \preceq q$.*

Definition 62. *Let α be an ordinal. The cofinality of α is the least ordinal λ such that there exists a function $f : \lambda \to \alpha$ whose range is cofinal in $(\alpha, \leq)$. We denote the cofinality of α by $cf(\alpha)$.*

In other words, the ordinal $cf(\alpha)$ is the least ordinal such that there exists a sequence over α indexed by $cf(\alpha)$ whose elements are unbounded in the well-ordered set $(\alpha, \in_\alpha)$. The next exercise easily follows from the definition of cofinality.

Exercise 43. *Let α be an ordinal number. Prove that $cf(\alpha)$ is a cardinal number.*

For example, $cf(S(\alpha)) = 1$ for any ordinal α since the function $f : 1 \to S(\alpha)$ with $f(0) = \alpha$ satisfies the requirements. On the other hand, $cf(\omega_1) = \omega_1$ since the supremum of the range of any function from ω to ω_1, countable union of countable sets being countable, is a countable ordinal and hence not unbounded in ω_1. An infinite cardinal whose cofinality is strictly smaller than itself is $\aleph_\omega$. We have that $cf(\aleph_\omega) = \omega$ since the function $f : \omega \to \aleph_\omega$ given by $f(n) = \aleph_n$ for all $n \in \omega$ satisfies the requirements.

We next show that it is sufficient for one to consider only strictly increasing maps while considering cofinalities.

Lemma 55. *Let α be a non-zero ordinal number. Then there exists a strictly increasing function $f : cf(\alpha) \to \alpha$ whose range is cofinal in $(\alpha, \leq)$.*

Proof. If α is a successor ordinal, then the function $f : 1 \to \alpha$ with $f(0) = sup(\alpha)$ satisfies the requirements. Assume that α is a limit ordinal. Let $g : cf(\alpha) \to \alpha$ be a function whose range is cofinal in $(\alpha, \leq)$. By transfinite recursion, define the function $f : cf(\alpha) \to \alpha$ by

$$f(\theta) = max\{g(\theta), sup\{f(\xi) + 1 : \xi < \theta\}\}$$

for all $\theta \in cf(\alpha)$. It follows from the definition that f is strictly increasing. Moreover, the range of f is cofinal in $(\alpha, \leq)$ since the range of g is cofinal and $f(\theta) \geq g(\theta)$ for all $\theta \in cf(\alpha)$. □

Exercise 44. *Let α be a limit ordinal. Prove that $cf(\aleph_\alpha) = cf(\alpha)$.*

Having defined the notion of cofinality, our next goal is to show that $\kappa^{cf(\kappa)} > \kappa$ for any infinite cardinal κ. In order to prove this, we shall need a remarkable fact known as König's theorem. We first define the sum of an indexed system of cardinals. Given an indexed system $\{\kappa_i\}_{i \in I}$ of cardinals, we define their sum to be

$$\sum_{i \in I} \kappa_i = \left| \bigcup_{i \in I} \kappa_i \times \{i\} \right|$$

In other words, $\sum_{i \in I} \kappa_i$ is the cardinality of the disjoint union of the cardinals κ_i. Similarly, given an indexed system $\{\kappa_i\}_{i \in I}$ of cardinals, we define their product to be the cardinality of the product $\prod_{i \in I} \kappa_i$ of the indexed system of sets. Abusing the notation, we shall also denote this cardinal by $\prod_{i \in I} \kappa_i$.

Theorem 41 (König's theorem)**.** *Let $\{\lambda_i\}_{i\in I}$ and $\{\kappa_i\}_{i\in I}$ be indexed systems of cardinals such that $\lambda_i < \kappa_i$ for all $i \in I$. Then we have*

$$\sum_{i\in I} \lambda_i < \prod_{i\in I} \kappa_i$$

Proof. Consider the function $f : \bigcup_{i\in I} \lambda_i \times \{i\} \to \prod_{i\in I} \kappa_i$ given by

$$(f(\theta, j))(i) = \begin{cases} 0 \text{ if } i \neq j \\ \theta + 1 \text{ otherwise} \end{cases}$$

for all $i, j \in I$ and $\theta \in \lambda_j$. Assume that $f(\theta, j) = f(\theta', j')$ for some $j, j' \in I$ and $\theta \in \lambda_j$ and $\theta' \in \lambda_{j'}$. Since the sequences $f(\theta, j)$ and $f(\theta', j')$ with the index set I are non-zero only at the indices j and j' respectively, we have that $j = j'$. On the other hand, since $\theta + 1 = (f(\theta, j))(j) = (f(\theta', j))(j) = \theta' + 1$, we have $\theta = \theta'$. Therefore f is one-to-one and hence $\sum_{i\in I} \lambda_i \leq \prod_{i\in I} \kappa_i$. In order to finish the proof, it suffices to show that there exists no surjection from $\bigcup_{i\in I} \lambda_i \times \{i\}$ to the cartesian product $\prod_{i\in I} \kappa_i$ of the indexed system $\{\kappa_i\}_{i\in I}$.

Let $h : \bigcup_{i\in I} \lambda_i \times \{i\} \to \prod_{i\in I} \kappa_i$ be any function. For each $i \in I$, consider the function $h_i : \lambda_i \to \kappa_i$ given by $h_i(\theta) = (h(\theta, i))(i)$ for all $\theta \in \lambda_i$. Since we have $\lambda_i < \kappa_i$ for all $i \in I$, the function h_i cannot be a surjection and hence, using the axiom of choice, we can choose $\delta_i \in \kappa_i$ such that $\delta_i \notin ran(h_i)$ for each $i \in I$. Then the sequence $(\delta_i)_{i\in I} \in \prod_{i\in I} \kappa_i$ is not in the range of h and hence h is not a surjection. □

Corollary 42. *For any infinite cardinal κ, we have $\kappa < \kappa^{cf(\kappa)}$.*

Proof. Let κ be an infinite cardinal and $f : cf(\kappa) \to \kappa$ be a strictly increasing function whose range is cofinal. Then, since $f(\xi) < \kappa$ for all $\xi \in cf(\kappa)$, it follows from König's theorem that

$$\kappa = sup\{f(\xi) : \xi \in cf(\kappa)\} \leq \sum_{\xi\in cf(\kappa)} |f(\xi)| < \prod_{\xi\in cf(\kappa)} \kappa = \left| {}^{cf(\kappa)}\kappa \right| = \kappa^{cf(\kappa)}$$

□

Corollary 43. *For any infinite cardinal κ, $cf(2^\kappa) > \kappa$.*

Proof. Let κ be an infinite cardinal. It follows from the previous corollary that

$$(2^\kappa)^{cf(2^\kappa)} = 2^{\kappa \cdot cf(2^\kappa)} = 2^{max\{\kappa, cf(2^\kappa)\}} > 2^\kappa$$

and hence it cannot be the case that $cf(2^\kappa) \leq \kappa$. Thus $cf(2^\kappa) > \kappa$.

□

Before concluding this subsection, we shall give a special name to those infinite cardinals that are equal to their own cofinality since they are of importance.

Definition 63. *Let κ be an infinite cardinal. The cardinal κ is said to be a regular cardinal if $cf(\kappa) = \kappa$ and is said to be a singular cardinal if it is not a regular cardinal.*

Using Lemma 55, one can show that the cofinality of an infinite cardinal is always a regular cardinal.

Exercise 45. *Prove that if α is a limit ordinal, then $cf(\alpha)$ is regular.*

An important consequence of Lemma 53 is that successor cardinals are regular.

Lemma 56. *Successor cardinals are regular.*

Proof. Let κ be a successor cardinal, say $\kappa = \lambda^+$ for some infinite cardinal λ. Let $f : cf(\kappa) \to \kappa$ be a function whose range is cofinal. Then we have

$$\kappa = sup\{f(\theta) : \theta \in cf(\kappa)\} = \bigcup\{f(\theta) : \theta \in cf(\kappa)\}$$

by the definition of a cofinal subset. Observe that $|f(\theta)| \leq \lambda$ since κ is a cardinal number. If it were the case that $cf(\kappa) < \kappa = \lambda^+$, then Lemma 53 would imply that $\kappa = \bigcup\{f(\theta) : \theta \in cf(\kappa)\} \leq \lambda$ which is a contradiction. Therefore $cf(\kappa) = \kappa$ and hence κ is regular. □

8.6. **Cardinal exponentiation under GCH.** Even though some most basic questions regarding cardinal exponentiation are independent of ZFC as was pointed out before, it turns out that cardinal exponentiation trivializes if one additionally assumes the generalized continuum hypothesis.

Theorem 44 (ZFC+GCH)**.** *Let κ and λ be infinite cardinals. Then*

$$\kappa^\lambda = \begin{cases} \lambda^+ & \text{if } \kappa \leq \lambda \\ \kappa^+ & \text{if } cf(\kappa) \leq \lambda < \kappa \\ \kappa & \text{if } \lambda < cf(\kappa) \end{cases}$$

Proof. We prove each case separately.

- If $\kappa \leq \lambda$, then it follows from Lemma 54 that $\kappa^\lambda = 2^\lambda$ and hence $\kappa^\lambda = \lambda^+$ by GCH.
- If $cf(\kappa) \leq \lambda < \kappa$, then it follows from the corollary of König's theorem that $\kappa^\lambda \geq \kappa^+$. Moreover, $\kappa^\lambda \leq \kappa^\kappa = 2^\kappa = \kappa^+$ by GCH and hence $\kappa^\lambda = \kappa^+$.
- If $\lambda < cf(\kappa)$, then the range of any function from λ to κ is not cofinal and hence is contained in some $\theta < \kappa$. Consequently, we have ${}^\lambda\kappa = \bigcup_{\theta\in\kappa} {}^\lambda\theta$. It follows that

$$\begin{aligned} \kappa \leq \kappa^\lambda = \left| \bigcup_{\theta\in\kappa} {}^\lambda\theta \right| \leq \sum_{\theta\in\kappa} {}^\lambda\theta \leq \sum_{\theta\in\kappa} {}^\lambda(2^\theta) \\ \leq \sum_{\theta\in\kappa} 2^{\theta\lambda} \leq \sum_{\theta\in\kappa} max\{\theta,\lambda\}^+ \leq \kappa \end{aligned}$$

  Observe that, for the last two inequalities, we use GCH and Lemma 53 respectively.

□

For example, assuming GCH, we have

$$\begin{aligned} \aleph_\omega^{\aleph_{\omega_1}} = \aleph_{\omega_1}^+ = \aleph_{\omega_1+1} \\ \aleph_\omega^{\aleph_1} = \aleph_\omega^+ = \aleph_{\omega+1} \\ \aleph_2^{\aleph_0} = \aleph_2 \end{aligned}$$

As a concluding remark for this section, we would like to note that, while most statements that are not trivial consequences of König's theorem turn out to be independent of ZFC, some non-trivial identities regarding cardinal exponentiation are provable using the axioms of ZFC. For example, the following surprising inequality is proven by Saharon Shelah.

$$\aleph_\omega^{\aleph_0} \leq 2^{\aleph_0} + \aleph_{\aleph_4}$$

9. The von Neumann hierarchy of sets

In this section, we shall describe a construction which systematically exhausts the universe of sets $\mathbf{V}$. By transfinite recursion, define the following class of sets

- $\mathrm{V}_0 = \emptyset$
- $\mathrm{V}_{\alpha+1} = \mathcal{P}(\mathrm{V}_\alpha)$ for all ordinals α, and
- $\mathrm{V}_\gamma = \bigcup_{\beta<\gamma} \mathrm{V}_\beta$ for all limit ordinals γ.

The class of sets

$$\{x : \exists\alpha \text{ “}\alpha \text{ is an ordinal”} \ \wedge \ x \in \mathrm{V}_\alpha\}$$

is called the *von Neumann hierarchy of sets*[1]. Our goal is to show that the von Neumann hierarchy equals the universe of sets $\mathbf{V}$. More precisely, we would like to prove that for every set x there exists an ordinal α such that $x \in \mathrm{V}_\alpha$. In order to prove this, we shall need the following lemma.

Lemma 57. *Let $\varphi(x)$ be a formula in the language of set theory with one free variable such that the class $\{x : \varphi(x)\}$ is non-empty. Then there exists a set y such that $\varphi(y)$ holds and $\neg\varphi(z)$ for any $z \in y$.*

Proof. Assume to the contrary that the claim does not hold, i.e for any set y, if $\varphi(y)$ holds, then there exists $z \in y$ such that $\varphi(z)$ holds. Let x be a set such that $\varphi(x)$ holds. By recursion, we can construct a sequence of sets $(x_i)_{i\in\mathbb{N}}$ such that $x = x_0$, $x_{i+1} \in x_i$ and $\varphi(x_i)$ holds for all $i \in \mathbb{N}$. This contradicts the axiom of foundation since the membership relation $\in_X$ is not well-founded on the set $X = \{x_i : i \in \mathbb{N}\}$. □

The following basic properties of the von Neumann hierarchy are left to the reader as an exercise.

Exercise 46. *Let $\alpha < \beta$ be ordinals. Prove that $V_\alpha \subseteq V_\beta$.*

Exercise 47. *Prove that V_α is transitive for all ordinals α.*

We are now ready to show that

$$\mathbf{V} = \bigcup_{\alpha\in\mathbf{ON}} \mathrm{V}_\alpha$$

where $\mathbf{ON}$ denotes the class of ordinal numbers, that is, the von Neumann hierarchy equals the universe of sets. More precisely, we have the following.

Theorem 45. *For every set x, there exists an ordinal α such that $x \in V_\alpha$.*

Proof. Assume to the contrary that there exists a set x such that $x \notin \mathrm{V}_\alpha$ for any ordinal α. Then, by the previous lemma, there exists a set w such that $w \notin \mathrm{V}_\alpha$ for any ordinal α and for all $y \in w$ there exists an ordinal γ such that $y \in \mathrm{V}_\gamma$. For each $y \in w$, let γ_y be the least ordinal such that $y \in \mathrm{V}_{\gamma_y}$. It follows from the axiom of replacement that the class $\{\gamma_y : y \in w\}$ is a set. Let θ be the ordinal $sup\{\gamma_y : y \in w\}$. Then $w \subseteq \mathrm{V}_\theta$ and hence $w \in \mathcal{P}(\mathrm{V}_\theta) = \mathrm{V}_{\theta+1}$, which is a contradiction. □

Having shown that every set belongs to some level of the von Neumann hierarchy, we are now going to assign a rank to each set. For any set x, we define the *rank* of x to be the least ordinal α such that $x \in V_{\alpha+1}$ and denote it by $rank(x)$.

[1]This class of sets is sometimes called the cumulative hierarchy or the von Neumann universe.

Exercise 48. *Prove that $rank(\alpha) = \alpha$ for any ordinal α.*

One should think of the class of ordinals as the "skeleton" of the universe of sets since the whole universe is created via a transfinite recursion on the class of ordinals using the power set operation and the union operation. The rank of a set measures at which point a set is created.

An important consequence of Lemma 57 is that if a property holds for the empty set and is inherited from the elements of a set to the set itself, then this property indeed holds for all sets. This principle is known as the principle of $\in$-induction and may be considered as a generalization of the principle of transfinite induction.

Theorem 46 (The principle of $\in$-induction)**.** *Let $\psi(x)$ be a formula in the language of set theory with one free variable such that*

- *For every set x, if $\psi(y)$ holds for every $y \in x$, then $\psi(x)$ holds.*

Then $\psi(x)$ holds for every set x.

Proof. Assume towards a contradiction that there exists x such that $\neg\psi(x)$. By Lemma 57, there exists a set y such that $\neg\psi(y)$ and $\neg\neg\psi(z)$ for all $z \in y$. However, by the assumption, that $\neg\neg\psi(z)$ for all $z \in y$ implies that $\psi(y)$ holds, which leads to a contradiction. □

One can prove a variant of this principle which works on an arbitrary transitive class[2] not necessarily the class of all sets. For the general statement, the reader is referred to [2, Theorem 6.4]. Just like one can generalize transfinite induction to $\in$-induction, one can also generalize transfinite recursion to what is called $\in$-recursion. We shall state the $\in$-recursion theorem for the class of all sets and refer the reader to [2, Theorem 6.5] for a general statement and its proof, which works on arbitrary transitive classes.

Theorem 47 (The $\in$-recursion theorem)**.** *Let F_φ be a class function. Then there exists a class function F_ψ such that $F_\psi(x) = F_\varphi(F_\psi \restriction x)$ for all sets x.*

The proof of this generalized recursion principle resembles the proof of the transfinite recursion principle: We define $F_\psi(x) = y$ if and only if $f(x) = y$ for some "computation" f, which is a function such that $dom(f)$ is a transitive set and $f(z) = F_\varphi(f \restriction z)$ for all $z \in dom(f)$. That such a class function F_ψ indeed satisfies the required property is proven by $\in$-induction on the class of all sets.

[2]A class C is said to be transitive if it contains the elements of its elements. More formally, if a class C is defined by the formula $\varphi(x)$, then C is transitive if $\forall x\ \varphi(x) \to (\forall y\ (y \in x \to \varphi(y)))$.

10. Coda

This course is intended to serve as an introductory course in axiomatic set theory. The author sincerely hopes that the reader enjoyed the course and benefited as much as possible. Those who are struck by the intrinsic beauty of the subject and who wish to learn deeper and more beautiful set theoretic topics are strongly suggested to read the graduate level textbooks of Jech [2] and Kunen [3], both of whom the author thinks are terrific expositors. The author also believes that the reader who enjoyed what he or she has learned so far would undoubtedly find more of the "supreme beauty" that Russell referred to, in studying advanced set theory.

> "Mathematics, rightly viewed, possesses not only truth, but supreme beauty - a beauty cold and austere, like that of sculpture, without appeal to any part of our weaker nature, without the gorgeous trappings of painting or music, yet sublimely pure, and capable of a stern perfection such as only the greatest art can show. The true spirit of delight, the exaltation, the sense of being more than Man, which is the touchstone of the highest excellence, is to be found in mathematics as surely as poetry."
>
> Bertrand Russell.

References

[1] Karel Hrbacek and Thomas Jech, *Introduction to set theory*, third ed., Monographs and Textbooks in Pure and Applied Mathematics, vol. 220, Marcel Dekker, Inc., New York, 1999. MR 1697766

[2] Thomas Jech, *Set theory*, Springer Monographs in Mathematics, Springer-Verlag, Berlin, 2003, The third millennium edition, revised and expanded. MR 1940513

[3] Kenneth Kunen, *Set theory*, Studies in Logic (London), vol. 34, College Publications, London, 2011. MR 2905394

Middle East Technical University, Ankara, Turkey,
Email address: burakk@metu.edu.tr