EQUIVALENTS TO THE AXIOM OF CHOICE AND THEIR USES

A Thesis

Presented to

The Faculty of the Department of Mathematics

California State University, Los Angeles

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

in

Mathematics

By

James Szufu Yang

The thesis of James Szufu Yang is approved.

Mike Krebs, Ph.D.

Kristin Webster, Ph.D.

Michael Hoffman, Ph.D., Committee Chair

Grant Fraser, Ph.D., Department Chair

California State University, Los Angeles

June 2015

ABSTRACT

Equivalents to the Axiom of Choice and Their Uses

By

James Szufu Yang

In set theory, the Axiom of Choice (AC) was formulated in 1904 by Ernst Zermelo. It is an addition to the older Zermelo-Fraenkel (ZF) set theory. We call it Zermelo-Fraenkel set theory with the Axiom of Choice and abbreviate it as ZFC.

This paper starts with an introduction to the foundations of ZFC set theory, which includes the Zermelo-Fraenkel axioms, partially ordered sets (posets), the Cartesian product, the Axiom of Choice, and their related proofs. It then introduces several equivalent forms of the Axiom of Choice and proves that they are all equivalent. In the end, equivalents to the Axiom of Choice are used to prove a few fundamental theorems in set theory, linear analysis, and abstract algebra.

This paper is concluded by a brief review of the work in it, followed by a few points of interest for further study in mathematics and/or set theory.

ACKNOWLEDGMENTS

Between the two department requirements to complete a master's degree in mathematics − the comprehensive exams and a thesis, I really wanted to experience doing a research and writing a serious academic paper. Among all the topics, set theory was my favorite and luckily I have Dr. Hoffman as my adviser.

Dr. Hoffman is very generous in helping me write this paper, in many ways. First of all, unless he has a class to teach, he always stays in his office checking my work and answering my questions, even when it passes his dinner time. Last time he worked with me from 3:30 pm to 7:00 pm! I really appreciate him being so generous in spending his time for my paper.

Secondly, when we touch a topic that needs more information, he would pick a book from his bookshelf for me to study at home. He would also bring a book from his house and loan it to me! One of the books was a book on the LaTeX document editor. I watched a tutorial video about LaTeX on YouTube and used this book as a reference and, in a couple days, I was able to do work on a LaTeX document. This saves me a lot of time, compared to using Microsoft Word on another short paper. Dr. Hoffman also offers me a rehearsal for my presentation. I can't say enough thanks to him.

The other two people to thank are also in the committee. They generously accepted my request to be in the committee the first time I asked, although they both have heavy administrative work on top of their regular teaching assignments. Dr. Krebs is the associate chair of the math department and Dr. Webster is in charge of remedial mathematics in the department.

I want to thank Dr. Krebs for his time reading my paper and giving corrections. It was such a detailed "proofread". One of the corrections on my paper was how to type quotation marks on a LaTeX document.

Dr. Webster is always busy with her duty being a mother and managing remedial math classes on top of her regular teaching assignment, but she, like Dr. Krebs, signed to be in the committee at my first request! My progress in this paper is slower, so my presentation falls on the final week, when Dr. Webster has to monitor the collective final exam paper graders from Monday through Thursday yet she still finds time on Wednesday for the presentation. Same with Dr. Krebs that he has to attend another meeting right after the presentation.

Thanks a lot to Dr. Hoffman and my thesis committee. Without them I wouldn't be able to complete this paper!

TABLE OF CONTENTS

CHAPTER 1

Introduction

In many of our mathematics classes, at some point the instructor would mention the Axiom of Choice but doesn't talk too deeply about it. For many students, the Axiom of Choice remains a mystery although almost everyone knows its basic idea. Often we are told that "pick an element from each set of a nonempty collection of nonempty sets, you can form a (choice) set". This is probably the impression a math student has about the Axiom of Choice throughout his or her math career, unless the student chooses to study set theory in a graduate program.

In fact, the Axiom of Choice has been "probably the most interesting and, in spite of its late appearance, the most discussed axiom of mathematics, second only to Euclid's axiom of parallels which was introduced more than two thousand years ago" [5]. Since being formulated by Ernst Zermelo in 1904, the Axiom of Choice has been controversial but now used without reservation by most mathematicians. "The status of the Axiom of Choice has become less controversial in recent years. To most mathematicians it seems quite plausible and it has so many important applications in practically all branches of mathematics that not to accept it would seem to be a wilful hobbling of the practicing mathematician" [15, p. 201].

The goal of this paper is hence not to question the status of the Axiom of Choice. Instead, we want to study more about the Axiom of Choice, especially about its variants and uses.

In our study of the Axiom of Choice, we should be aware that axiomatic set theory is developed in the framework of first-order predicate calculus, and there are

a few "equivalent" axiom systems in common mathematical use. By "equivalent" we mean that these axiom systems all prove the same first-order theorems, which are logical consequences of the axioms [4, 16].

Unlike other alternative axiomatic set theories such as Gödel-Bernays set theory, Zermelo-Fraenkel (ZF) set theory with Axiom of Choice (ZFC) has only one type of object: set. Therefore, in our discussion, we do not consider objects that may be elements of sets but are not sets themselves. Such objects are called atoms by some mathematicians. On the other hand, we also do not consider classes, which are objects too big to be sets.

Remark 1.1. *Cantor sensed the difference between "the collection of all sets" and and the usual sets. He called the former "inconsistent totalities" and the latter "consistent totalities". Russell pointed out the paradox, later under his name, in Gottlob Frege's set theory but was not able give a resolution. This issue had been known to the mathematics community for years. Proper classes were accepted as legitimate objects in axiomatized set theory by von Neumann in 1925. Some of his ideas were adopted by Paul Bernays in his papers published in 1937 and after. Kurt Gödel modified Bernays's axioms in 1940. This new set theory was hence known as Gödel-Bernays (GB) set theory or von Neumann-Bernays (VNB) set theory* [4, p.15]*. It is also know as von NeumannBernaysGdel (NBG) set theory. We would call it NBG because it properly gives credits to all these three great mathematicians. In NBG set theory, sets can be elements of classes, while classes can't be elements of anything. NBG and ZF prove exactly the same first-order theorems except that NBG carries a more cumber-*

some logical baggage [20, p.70] and [16].

ZFC axiomatic set theory is built upon first-order logic. There are only two binary predicates necessary in ZFC set theory, namely the membership relation, $\in$, and the "equal to" relation, $=$. Therefore there are only two atomic formulas: $x \in y$ and $x = y$. Statements in ZFC set theory are written with these two formulas and the following quantifiers and logic operators, $\forall$ (for all), $\exists$ (there exists), $\neg$ (not), $\wedge$ (and), $\vee$ (or), $\Longrightarrow$ (if), $\Longleftrightarrow$ (if and only if, iff). As we adopt ZFC axioms we treat the concepts of set and member (element) as undefined primitive notions [4, 16, 20].

CHAPTER 2

Zermelo-Fraenkel Axioms and Partially Ordered Sets

2.1 Zermelo-Fraenkel Axioms

Ernst Zermelo published his axiomatized set theory in 1908, including most of the eight Zermelo-Fraenkel (ZF) Axioms. Later other mathematicians observed that for a satisfactory theory of ordinal numbers, the axioms established by Zermelo were not sufficient. The Axiom of Replacement was then proposed by Abraham Fraenkel (in 1922) and others such as Thoralf Skolem. This system of axioms became known as the "Zermelo-Fraenkel Axioms". The idea of the Axiom of Regularity appeared in Dmitry Mirimanoff's paper in 1917, and was later explicitly included in ZF axioms by John von Neumann in 1925 [4, 20]. Note that some of the ZF axioms are just a single axiom while some are so called "axiom schema". An axiom schema is a collection of axioms, one axiom for each of a particular type of formula of first-order logic [16].

Axiom 2.1. (Axiom of Extensionality) *If X and Y have the same elements, then $X = Y$.* $\quad \forall X\, \forall Y\, [\, \forall x\, (x \in X \iff x \in Y) \Longrightarrow X = Y\,]$.

The principle of extensionality is probably the most intuitive axiom in ZFC. "It expresses the basic idea of a set: A set is determined by its elements" [20]. It says two sets are the same if and only if they have the same elements. In other words, distinct sets have at least one element that are not the same. This axiom also leads to a widely used proof technique that when we prove two sets are equal (the same),

we only need to show that they are included (or contained, $\subseteq$) by each other.

Axiom 2.2. (Axiom of Pairing) *For any u and v there exists a set $\{u, v\}$.*

$\forall u \, \forall v \, \exists X \, \forall x \, (x \in X \iff x = u \vee x = v)$.

The Axiom of Pairing is very basic yet important. Without this axiom we cannot construct ordered pairs to order, so we cannot discuss relations of arity (dimension of the domain of a relation) greater than one, and so n-tuples aren't possible when $n > 1$. Hence we cannot even define function since functions are relations. Moreover, without relations and functions defined, we cannot discuss Cartesian product either [16].

Axiom 2.3. (Axiom Schema of Specification) *If P is a property with parameter p, then for any X and p there exists a set $Y = \{u \in X \mid P(u, p)\}$ that contains all those $u \in X$ that have property P.* $\forall X \, \forall p \, \exists Y \, \forall u \, [u \in Y \iff u \in X \wedge P(u, p) \,]$.

The Axiom Schema of Specification is also known as the Axiom Schema of Separation, the Subset Axiom Schema, or sometimes by some mathematicians, the Axiom Schema of Restricted Set Comprehension. It actually got set theory out of Russell's paradox, which was under the older rule of Axiom Schema of Set Comprehension, which says that if P is a property, then there exists a set $Y = \{x \mid P(x)\}$. Russell's Paradox occurs when $P(x) = x \notin x$ [20, p. 4]. Hence it is also called the Axiom Schema of Restricted Set Comprehension.

A natural consequence of the formula above is that $Y \subseteq X$, thus this axiom helps us define subset so that we don't have subset as a undefined primitive notion. This is why it's also called the Subset Axiom Schema.

Note that the existence of the empty set, $\varnothing$, can be proved by the Axiom of Specification, and it's unique by the Axiom of Extensionality. However, some authors have an Axiom of Empty Set, which is not an original ZF axiom though. We will give a short proof on the unique existence of the empty set followed by its definition and notation.

Axiom 2.4. (Axiom of Union) *For any $\mathcal{X}$ there exists a set $Y = \bigcup \mathcal{X}$, the union of all elements of $\mathcal{X}$.* $\forall \mathcal{X}\, \exists Y\, \forall u\, [u \in Y \iff \exists X\, (X \in \mathcal{X} \wedge u \in X)]$

The Axiom of Union is saying that, for any set $\mathcal{X}$, there exists a set Y, whose elements are exactly the elements of the elements of $\mathcal{X}$. With this axiom we can define union of sets. Note that there is no corresponding "Axiom of Intersection". Instead, for any set $\mathcal{X}$, there exists a unique set Y such that for any x, $x \in Y$, x belongs to every element of $\mathcal{X}$. With union and intersection defined, by the Axiom of Extensionality, the algebra of sets are possible. For example, addition and subtraction of sets, relative complement of a set, the Commutative Laws, Associate Laws, Distributive Laws, De Morgan's Laws, etc. are all possible [4] and [16, p. 31].

Axiom 2.5. (Axiom of Power Set) *For any X there exists a set $Y = \mathcal{P}(X)$, the set of all subsets of X.* $\forall X\, \exists Y\, \forall x \in Y\, (x \in Y \iff x \subseteq X)$.

The Axiom of Power Set actually guarantees finite Cartesian products. However, for infinite Cartesian products, we need the Axiom of Choice [16, p. 41]. In addition, the Axiom of Power Set also enables us to define ordinal numbers, with which the von Neumann universe is possible. (Also see Axiom 2.8.)

Note that we have been using the binary predicates, atomic formulas, quantifiers, and operator in first-order logic in the definitions of axioms. However, in the following three axioms, we will not give a first-order logic formula in their definitions because these formulas are too cumbersome and not helpful for the context and so unnecessary.

Axiom 2.6. (Axiom of Infinity) *There exists an infinite set.*

The infinite set is defined inductively, so some authors phrase this axiom as, "There exists an inductive set. $\exists X \varnothing \in X$ and X is inductive" [16]. The idea is that there is a set X with $\varnothing \in X$, and such that $x \in X$ implies $x \cup \{x\} \in X$. The set $x \cup \{x\}$ is called the successor of x. With the Axiom of Infinity, the set of natural numbers, $\mathbb{N}$, can be constructed, and hence infinite ordinals are possible.

Axiom 2.7. (Axiom Schema of Replacement or Axiom Schema of Substitution) *If a function has a set as its domain, then its range is also a set.*

There is an issue when applying a function in its traditional sense on a collection that is not a set. For example, if, with a function in its traditional sense, we take

the image of a domain that is a collection of ordinals, then the image may be too big to be a set. It could be a proper class [17, p. 93]. The Axiom Schema of Replacement asserts that the image is a set because ZF set theory cannot legally refer to a class [4, p. 179]. This axiom schema was mainly due to Fraenkel.

Axiom 2.8. (Axiom of Regularity or Axiom of Foundation or Axiom of Restriction) *Every nonempty set has a $\in$-minimal element.*

The Axiom of Regularity is rarely used by mathematicians. Most of mathematics would go on the same with or without it. It, however, produces interesting intuitive consequences such as for any set x, $x \notin x$, and for any sets x and y, $x \in y \in x$ is impossible. The axiom is equivalent to: there is no infinite descending $\in$-chain. A more important consequence of the Axiom of Regularity is the von Neumann Universe or von Neumann Hierarchy of Sets, V, which contains all sets in ZFC set theory. V is of course not a set. It is the structure of the universe of all sets. Note that the Axiom of Regularity cannot be derived from the other axioms of ZFC, if they are consistent (Bernays, 1954) [16, 21].

These are all the Zermelo-Fraenkel axioms. As a demonstration of using the most basic concepts in ZF set theory in proofs, let's try to prove the existence of the empty set, followed by defining the empty set, as an example of how the ZF set theory builds up its system.

Proposition 2.9. (The Empty Set) *There is a unique set with no elements.*

$\exists X\ \forall x,\ x \notin X.$

Proof. Since set is an undefined primitive notion, by first-order logic, there exists at least one set Z. By Axiom 2.1 (the Axiom of Extensionality), let Y be a set such that for any X, $X \in Y$ iff $X \in Z$ and $X \neq X$. Then Y does not have any elements. By Axiom 2.1, Y is unique. Hence Y is the empty set. □

Next, we can then define the empty set.

Definition 2.10. (Empty Set) *The* ***empty set*** *is a set with no elements, denoted by* $\varnothing$.

Remark 2.11. *The Zermelo-Franekel axioms are not independent – some of them are implied by the others. For example, the Axiom Schema of Specification is derived from the Axiom Schema of Replacement, and that the Axiom of Pairing is derivable from the Axiom of Power Set and the Axiom Schema of Replacement* [19, p. 237]. *This dependence issue serves as a good topic for further study after this paper.*

2.2 Partially Ordered Sets

By the Axiom of Pairing, we could define ordered pairs, and then relation and function. There are different ways to define ordered pairs. Our definition of ordered pair here is due to Kaximierz Kuratowski in 1921. In this section, we will give defi-

nitions required to prove the equivalents to the Axiom of Choice as follows.

Definition 2.12. (Ordered Pair) *For all sets x and y, we define the* ***ordered pair*** $(x, y) = \{\{x\}, \{x, y\}\}$.

Definition 2.13. (Ordered n-tuple) *For any set x_i, $i \in \mathbb{N}$, we define the* ***ordered n-tuple*** $(x_1, \ldots, x_n) = ((x_1, \ldots, x_{n-1}), x_n)$.

Definition 2.14. (Relation) *A* ***relation*** *is a set of ordered pairs* [4, p. 40].

Definition 2.15. (Function) *A* ***function*** *from a set A to a set B is a relation f from A to B with the property that for every $a \in A$ there exists one and only one (not necessarily distinct) element $b \in B$ such that $(a, b) \in f$.*

Definition 2.16. (Partial Order Relation) *The binary relation $\leqslant$ is a* ***partial order*** *on a set X if for all x, y, $z \in X$,*

(a) $x \leqslant x$. (Reflexivity)

(b) If $x \leqslant y$ and $y \leqslant z$, then $x \leqslant z$. (Transitivity)

(c) If $x \leqslant y$ and $y \leqslant x$, then $x = y$. (Antisymmetry)

If $\leqslant$ partially orders X, we call X a partially ordered set under $\leqslant$, denoted by $(X, \leqslant)$. [2, 16]

Note that the term "poset" is short for "partially ordered set" by many au-

thors, so we will also use "poset" instead of "partially ordered set" when appropriate.

Definition 2.17. (Strict Order Relation) *The binary relation $<$ is a strict order on a set X if for all x, y, $z \in X$,*

(a) $x \not< x$. (Irreflexivity)

(b) If $x < y$ and $y < z$, then $x < z$. (Transitivity)

(c) If $x < y$, then $y \not< x$. (Asymmetry)

If $<$ strictly orders X, we call X a strictly ordered set under $<$, denoted by $(X, <)$. As a shorthand, we say $x < y$ if $x \leqslant y$ and $x \neq y$. [16]

Definition 2.18. (Total or Linear Order Relation) *The binary relation $\leqslant$ is a **linear** or **total order** on a set X if $\leqslant$ is a partial order and for all x, $y \in X$, $x \leqslant y$ or $y \leqslant x$. This property is called comparability or connectedness, or trichotomy.*

Definition 2.19. (Minimal, Maximal, Minimum, Maximum) *Let $(X, \leqslant)$ be a set of sets and $x \in X$. We define*

*(a) x is **minimal** if and only if, for all $y \in X$, if $y \leqslant x$, then $x = y$.*

*(b) x is **maximal** if and only if, for all $y \in X$, if $x \leqslant y$, then $x = y$.*

*(c) x is a **minimum** if and only if for all $y \in X$, $x \leqslant y$.*

*(d) x is a **maximum** if and only if for all $y \in X$, $y \leqslant x$.* [16, p. 10]

Proposition 2.20. (Properties of Maximal, Minimal, Maximum, and Minimum)

(a) Maximum elements are maximal.

(b) Minimum elements are minimal.

(c) There can be at most one maximum element.

(d) There can be at most one minimum element.

(e) A maximal element in a linear order is a maximum.

(f) A minimal element in a linear order is a minimum.

Definition 2.21. (Chain, Maximal Chain) *Let $(X, \leqslant)$ be a set of partially ordered set.*

*(a) X is a **chain** if for all x, $y \in X$, either $x \leqslant y$ or $y \leqslant x$.*

*(b) Y is a **chain** in X if Y is a totally ordered subset of X.*

*(c) Y is a **maximal chain** in X if for any chain Z in X, $Y \subseteq Z$ implies $Y = Z$.* [19, p. 244]

Definition 2.22. (Upper Bound, Lower Bound Least Upper Bound, Greatest Lower Bound) *Let $(X, \leqslant)$ be a poset and $Y \subseteq X$.*

*(a) An element $u \in X$ is an **upper bound** for Y if $y \leqslant u$ for all $y \in Y$.*

*(b) An upper bound u_0 for Y is the **least upper bound** for Y if $u_0 \leqslant u$ for every upper bound u for Y.*

*(c) An element $v \in X$ is a **lower bound** for Y if $v \leqslant y$ for all $y \in Y$.*

*(d) A lower bound v_0 for Y is the **greatest lower bound** for Y if $v \leqslant v_0$ for every lower bound v for Y.* [13]

Note that in Definition 2.22, we assume the existence of the least upper bound

and the greatest lower bound. In addition, the least upper bound and the greatest lower bound are unique if they exist.

CHAPTER 3

The Axiom of Choice

3.1 The Cartesian Product

If we have two sets X and Y, the Cartesian product of these two sets would simply be $X \times Y = \{(x, y) \mid x \in X \text{ and } y \in Y\}$. The idea of Cartesian product is critical in understanding the Axiom of Choice. In fact, the nonempty Cartesian product of nonempty sets is equivalent to the Axiom of Choice. Here are some necessary definitions from [16, pp. 33–36].

Definition 3.1. (Finite n-ary Cartesian Product) *For all $i \in \mathbb{N}$, the* ***Cartesian product*** $X_1 \times X_2 \times \ldots \times X_n$ *is the set of all n-tuples defined by* $X_1 \times X_2 \times \ldots \times X_n = \{(x_1, \ldots, x_n) \mid \text{for each } x_i \in X_i\}$. *If all $X_i = X$ are identical, we also write* $X \times X \times \ldots \times X = X^n$.

In general, we can also define the Cartesian product of arbitrarily (finitely or infinitely, countably or uncountably) many sets.

Definition 3.2. (Generalized Cartesian Product) *Let I be an index set. Given $\{X_i \mid i \in I\}$, a family of sets indexed by I, the* ***generalized Cartesian product*** *is* $\prod_{i \in I} X_i = \{f \mid f : I \to \bigcup_{i \in I} X_i \text{ with } f(i) \in X_i \text{ for each } i \in I\}$. *If all $X_i = X$ are identical, then* $\prod_{i \in I} X_i = \{f \mid f : I \to X\} = X^I$.

Note that in the case of finite indexed family, Definition 3.1 and Definition 3.2 do not produce the same sets. For instance, $X_1 \times X_2$ and $\prod_{i\in\{1,2\}} X_i$ are different sets in ZF set theory. A typical element of $X_1 \times X_2$ in ZF set theory is an ordered pair (x_1, x_2) with $x_1 \in X_1$ and $x_2 \in X_2$. Note that $(x_1, x_2) \subseteq \mathcal{P}(X_1 \cup X_2)$ since, by our definition, $(x_1, x_2) = \{\{x_1\}, \{x_1, x_2\}\} \subseteq \mathcal{P}(X_1 \cup X_2)$. On the other hand, a typical element of $\prod_{i\in\{1,2\}} X_i$ is a function $f : \{1, 2\} \to X_1 \cup X_2$ for each $f(i) \in X_i$ with $i \in \{1, 2\}$. Also, $f \subseteq \{1, 2\} \times (X_1 \cup X_2)$ because f is a function. Since $\mathcal{P}(X_1 \cup X_2)$ and $\{1, 2\} \times (X_1 \cup X_2)$ are different sets, $X_1 \times X_2$ and $\prod_{i\in\{1,2\}} X_i$ cannot be the same set by the Axiom of Extensionality. However, $X_1 \times X_2$ and $\prod_{i\in\{1,2\}} X_i$ are isomorphic, i.e. $\phi : \prod_{i\in\{1,2\}} X_i \to X_1 \times X_2$ defined by $\phi(f) = (f(1), f(2))$ is an isomorphism. Thus we almost always treat them as the same and say $\prod_{i\in\{1,2\}} X_i = X_1 \times X_2$ [6].

To see the isomorphism, let $I = \{1, 2\}$, $X_1 = \{a, b\}$, and $X_2 = \{x, y\}$. By Definition 3.1,

$$X_1 \times X_2 = \{(a, x), (a, y), (b, x), (b, y)\} \tag{3.1}$$

However, by Definition 3.2, since $\bigcup_{i\in\{1,2\}} X_i = \{a, b, x, y\}$, each element of $\prod_{i\in\{1,2\}} X_i$ is a function $f : \{1, 2\} \to \{a, b, x, y\}$ with $f(1) \in \{a, b\}$ and $f(2) \in \{x, y\}$. Moreover, since f is a function, $f \subseteq \{1, 2\} \times \{a, b, x, y\}$. Hence we have four functions (indexed):

(1) $f_{11} = \{(1, a), (2, x)\}$

(2) $f_{12} = \{(1, a), (2, y)\}$

(3) $f_{21} = \{(1, b), (2, x)\}$

(4) $f_{22} = \{(1, b), (2, y)\}$

Therefore,

$$\prod_{i\in\{1,2\}} X_i = \{f_{11}, f_{12}, f_{21}, f_{22}\}$$

$$= \{\{(1,a),(2,x)\},\{(1,a),(2,y)\},\{(1,b),(2,x)\},\{(1,b),(2,y)\}\} \qquad (3.2)$$

Obviously the two sets in equations (3.1) and (3.2) are not the same, but there is a natural bijection between them. By applying $\phi(f) = (f(1), f(2))$, we have

(1) $\phi(f_{11}) = (f_{11}(1), f_{11}(2)) = (a, x)$, i.e. $\{(1,a),(2,x)\} \mapsto (a,x)$.

(2) $\phi(f_{12}) = (f_{12}(1), f_{12}(2)) = (a, y)$, i.e. $\{(1,a),(2,y)\} \mapsto (a,y)$.

(3) $\phi(f_{21}) = (f_{21}(1), f_{21}(2)) = (b, x)$, i.e. $\{(1,b),(2,x)\} \mapsto (b,x)$.

(4) $\phi(f_{22}) = (f_{22}(1), f_{22}(2)) = (b, y)$, i.e. $\{(1,b),(2,y)\} \mapsto (b,y)$.

Clearly $\prod_{i\in\{1,2\}} X_i \cong X_1 \times X_2$ and we almost always write $\prod_{i\in\{1,2\}} X_i = X_1 \times X_2$.

3.2 The Axiom of Choice

In 1935 Kurt Gödel showed that ZFC is consistent if ZF is consistent. In 1963 Paul Cohen showed that ZF with the negation of the Axiom of Choice ($\neg$ AC) is consistent if ZF is. Therefore the Axiom of Choice is independent of ZF [16, p. 54]. As a side note, in 1940 Gödel showed the Continuum Hypothesis (CH) cannot be disproved from either ZF or ZFC. In 1963 Cohen showed that CH cannot be proved from either ZF or ZFC Axioms [1, p. 107]. Hence CH is independent of ZFC. From now on we will abbreviate the Axiom of Choice as AC when appropriate throughout this paper, especially in our proofs. To understand the Axiom of Choice, the most important concept is choice function.

Definition 3.3. (Choice Function) *Let $\mathcal{X}$ be a nonempty family of nonempty sets. A choice function on $\mathcal{X}$ is a map $f : \mathcal{X} \to \bigcup \mathcal{X}$ such that for each $X \in \mathcal{X}$, $f(X) \in X$.*

Since we think the concept of choice function is critical because of its close relationship with the concept of Cartesian product, we separate the definition of choice function from that of AC. Now we define the Axiom of Choice in terms of the concept of choice function.

Axiom 3.4. (Axiom of Choice, AC) *Any nonempty collection of nonempty sets has a choice function.*

Ernst Zermelo introduced choice function and the Axiom of Choice in 1904 to prove the Well Ordering Theorem. The idea of AC is that, given a nonempty family of nonempty sets, there exists a map (choice function) from each nonempty set to one element in itself. AC guarantees such a map exists; however, it does not guarantee that we can always construct such a map. As in Bertrand Russell's boots-and-socks metaphor, we can always *choose* what we want from "$\aleph_0$ pairs" of boots since among boots we can distinguish right and left. For example, we can define a function to *choose* the right boot from the p^{th} pair of boots, where p is prime. On the other hand, we cannot define or construct such a function to choose the right sock from the p^{th} pair of socks because we cannot distinguish right and left in socks. Nonetheless, AC guarantees such a function exists even though we cannot define it. This is where the controversy comes from.

Proposition 3.5. *Each element of the Generalized Cartesian Product is a choice function.*

Proof. Since the choice function is $f : \mathcal{X} \to \bigcup \mathcal{X}$, if we index $\mathcal{X}$ by $g : I \to \{X_i\}$ and let $f' : \{X_i\} \to \bigcup_{i \in I} X_i$. Then f' is actually the choice function f in a different notation. By indexing the domain $\mathcal{X}$ of the choice function, the choice function is the same as the composition $g \circ f' : I \to \{X_i\} \to \bigcup_{i \in I} X_i$, which is simply $g \circ f' : I \to \bigcup_{i \in I} X_i$, a typical element of the Generalized Cartesian Product of nonempty sets. □

Proposition 3.5 strongly suggests that the Generalized Cartesian Product of nonempty sets is closely related to the Axiom of Choice. Note that there are plenty various ways to define AC. Here we introduce two simple alternative forms.

Theorem 3.6. *The following statements are equivalent to AC:*

(a) Disjoint Family Form:

Suppose that $\mathcal{X}$ is a nonempty disjoint family of nonempty sets. Then there is a choice function for $\mathcal{X}$.

(b) Power Set Form:

Suppose X is a nonempty set. Then there is a function $f : \mathcal{P}(X) \setminus \{\varnothing\} \to X$ such that for all nonempty subsets $A \subseteq X$, $f(A) \in A$.

In other words, in form (a), for any distinct nonempty $A, B \in \mathcal{X}$, $A \cap B = \varnothing$. Then there exists a function $f : \mathcal{X} \to \bigcup \mathcal{X}$, such that for each $A \in \mathcal{X}$, $f(A) \in A$ [6,

p. 108].

In Zermelo's papers in 1908, he introduced a modified form of AC, which is close to form (a). He called the image of such a choice function a *transversal* or a *choice set* for a family of pairwise disjoint sets, and he asserts that any family of mutually disjoint nonempty sets has a transversal, which is the same idea of AC [18].

Proof. We will prove AC $\Longrightarrow$ Disjoint Family Form $\Longrightarrow$ Power Set Form $\Longrightarrow$ AC.

(i) AC $\Longrightarrow$ Disjoint Family Form:

Immediately, since a nonempty disjoint family of nonempty sets is a nonempty collection of nonempty sets, there is a choice function for $\mathcal{X}$. So AC implies form (a).

(ii) Disjoint Family Form $\Longrightarrow$ Power Set Form:

Let X be a nonempty set. Then $\mathcal{P}(X) \setminus \{\varnothing\}$ is a family of nonempty (not necessarily disjoint) sets. Let $\mathcal{X} = \{Y \times \{Y\} \mid Y \subseteq X, Y \neq \varnothing\}$. Then for $A \times \{A\}$, $B \times \{B\} \in \mathcal{X}$ with nonempty A, $B \subseteq X$, suppose $A \times \{A\}$ and $B \times \{B\}$ are not disjoint, i.e. $(A \times \{A\}) \cap (B \times \{B\}) \neq \varnothing$, then there exits a $(y, Y) \in (A \times \{A\}) \cap (B \times \{B\})$ such that $(y, Y) \in (A \times \{A\}) \cap (B \times \{B\}) = (A \cap B) \times (\{A\} \cap \{B\})$ implies $y \in A \cap B$ and $Y \in \{A\} \cap \{B\}$, which forces $Y = A = B$ as shown below.

Since A, B are nonempty, $\{A\} \cap \{B\} \neq \varnothing$; since $\{A\}$, $\{B\}$ are singletons and $\{A\} \cap \{B\} \neq \varnothing$, $\{A\} = \{B\}$. Furthermore, $Y \in \{A\} \cap \{B\}$ implies $Y \in \{A\} = \{B\}$ and so $Y = A = B$.

Thus if A and B are distinct nonempty subsets of X, the corresponding elements of $\mathcal{X}$, $A \times \{A\}$ and $B \times \{B\}$, are disjoint.

So we can now apply the Disjoint Family Form to $\mathcal{X}$ to get a choice function

$g : \mathcal{X} \to \bigcup \mathcal{X}$ such that for each $A \times \{A\} \in \mathcal{X}$, we have $g(A \times \{A\}) \in A \times \{A\}$. This means that $g(A \times \{A\}) = (a, A)$ for some $a \in A$. Thus we define the required function $f : \mathcal{P}(X) \setminus \{\varnothing\} \to X$ by $f(A) = \pi(g(A \times \{A\}))$ so that $f(A) = \pi(g(A \times \{A\})) = \pi(a, A) = a \in A$ for each nonempty $A \subseteq X$, where π is a function $\pi : X \times \mathcal{P}(X) \setminus \{\varnothing\} \to X$ and $(y, Y) \mapsto y$ for all $y \in Y$ with $Y \subseteq X$. Hence $f : \mathcal{P}(X) \setminus \{\varnothing\} \to X$ is the required (choice) function in the Power Set Form of AC.

(iii) Power Set Form $\Longrightarrow$ AC:

Given a family $\mathcal{A}$ of nonempty sets, define $X = \bigcup \mathcal{A}$. Note that if $A \in \mathcal{A}$, then $A \subseteq X$. Now let g be the (choice) function $g : \mathcal{P}(X) \setminus \{\varnothing\} \to X = \bigcup \mathcal{A}$ as in the Power Set Form of AC. Since $A \in \mathcal{A} \Rightarrow A \subseteq X$ and $X \in \mathcal{P}(X) \setminus \{\varnothing\}$ we have $A \in \mathcal{A} \Rightarrow A \in \mathcal{P}(X) \setminus \{\varnothing\}$ and so $\mathcal{A} \subseteq \mathcal{P}(X) \setminus \{\varnothing\}$. So if we restrict the domain of g, $\mathcal{P}(X) \setminus \{\varnothing\}$, to $\mathcal{A}$, we have $g|_{\mathcal{A}} : \mathcal{A} \to X = \bigcup \mathcal{A}$. Rewriting $g|_{\mathcal{A}}$ and call it f, we have $f : \mathcal{A} \to \bigcup \mathcal{A}$ and $A \mapsto g(A) \in A$, i.e. $f(A) \in A$ with $A \in \mathcal{A}$. Then f is the choice function for $\mathcal{A}$. Hence AC. □

Theorem 3.7. *The Cartesian product of a nonempty family of nonempty sets is nonempty. This statement is equivalent to AC.*

Proof. Let Y be a nonempty set. Let $\mathcal{X} = \mathcal{P}(Y) \setminus \{\varnothing\}$. Then $\mathcal{X}$ is a collection of nonempty sets and can be indexed. Let $\mathcal{X} = \{X_i\}_{i \in I}$, then each $X_i \subseteq Y$. Then the Cartesian product of all elements $X_i \neq \varnothing$ of $\mathcal{X}$ is $\prod_{i \in I} X_i = \{f \mid f : I \to \bigcup_{i \in I} X_i$ for each $f(i) \in X_i\}$, which is nonempty and whose elements are all choice functions

on $\mathcal{X}$ by Proposition 3.5. Hence a nonempty Cartesian product of nonempty sets implies AC.

Conversely, assuming AC, then since $\mathcal{X}$ is a family of nonempty sets, there is a choice function $g : \mathcal{X} \to \bigcup \mathcal{X}$ on $\mathcal{X}$, such that for each $X \in \mathcal{X}, f(X) \in X$. By Proposition 3.5, if we index $\mathcal{X}$ in $g : \mathcal{X} \to \bigcup \mathcal{X}$, we have $f : I \to \bigcup_{i \in I} X_i$ for each $f(i) \in X_i$. Then f is an element of the Cartesian product $\prod_{i \in I} X_i$, so $\prod_{i \in I} X_i$ is not empty with each X_i not empty. Hence AC implies that the Cartesian product of a nonempty family of nonempty sets is nonempty. $\square$

CHAPTER 4

Equivalents to the Axiom of Choice

In Chapter 3, we gave two alternative forms of AC in Theorem 3.6. They are intuitively quite close to AC. In this chapter, we will introduce other more important equivalents to AC, namely Hausdorff's Maximal Principle (HMP), Zorn's Lemma (ZL), and the Well Ordering Theorem (WOT). We will prove that they are all equivalent. Our proof sequence will be AC $\Longrightarrow$ HMP $\Longrightarrow$ ZL $\Longrightarrow$ WOT $\Longrightarrow$ AC. Before we start the proof, we need a few definitions and a theorem.

4.1 AC $\Longrightarrow$ HMP

Definition 4.1. (Self-Map) *Let X be a set. A self-map on X is a map from X to itself, $f : X \to X$.*

Definition 4.2. (Fixed Point) *Let $f : X \to Y$ be a map. Then a fixed point of X under f is a point $x \in X$ such that $f(x) = x$.*

It follows that $X \cap Y \neq \varnothing$ if f has any fixed points at all. Also, Definition 4.2 implies that the n-fold composition at x is $\underbrace{f \circ f \circ \cdots \circ f}_{n \text{ times}}(x) = f^n(x) = x$.

Definition 4.3. (Increasing Map) *Let X be a partially ordered set. $f : X \to X$ is an increasing map if for all $x \in X$ we have $x \leqslant f(x)$.*

We will use the Bourbaki-Witt Theorem to prove that AC implies HMP. To prove this theorem, we need to define chain completeness and admissible subset.

Definition 4.4. (Chain Complete) *A poset X is chain complete if every chain, including the empty chain, in X has a least upper bound in X* [14, p. 53].

Definition 4.5. (Admissible Subset) *Let $f : X \to X$ be an increasing map with X chain complete and let $B \subseteq X$. Suppose $a \in X$. Then B is an admissible subset of X that contains a if*

(i) $a \in B$.

(ii) $f(B) \subseteq B$.

(iii) If T is a chain in B, then $\sup T \in B$ [13].

Note that the admissible subset B in Definition 4.5 is chain complete by the induced ordering of X [11, p. 13]. We will have more explanation about this after Proposition 4.7.

Proposition 4.6. *The set X in Definition* 4.5 *is itself admissible containing a.*

Proof. Obviously $X \subseteq X$. (i) $a \in X$. Since f is a self-map, $f : X \to X$ implies (ii) $f(X) \subseteq X$. Let T be a chain in X. Since X is chain complete, every chain in X has a least upper bound in X, so does T. Hence, (iii) T is a chain in $X \Rightarrow \sup T \in X$. Therefore, X is admissible containing a. □

Proposition 4.7. *Let $\mathcal{A}$ be the set of all admissible subsets of X that contain $a \in X$. Let $M = \bigcap \mathcal{A}$ then M is an admissible subset of X that contains a.*

Proof. Obviously $M \subseteq X$. By Definition 4.5 (admissibility), (i) If $A \in \mathcal{A}$, then $a \in A$. Since this is true for all $A \in \mathcal{A}$, $a \in \bigcap \mathcal{A} = M$. Hence $a \in M$. (ii) We want to show that $f(M) \subseteq M$. Suppose $x \in M$, then $x \in A$ for every $A \in \mathcal{A}$. Then since A is admissible, $f(A) \subseteq A$. This is true for each $A \in \mathcal{A}$. So $f(x) \in \bigcap \mathcal{A} = M$. The same is true for every $x \in M$, so $f(M) \subseteq M$ as required. (iii) Let T be a chain in $M = \bigcap \mathcal{A}$. Then T is a chain in A for each $A \in \mathcal{A}$. Since A is admissible, $\sup T \in A$. Thus for all $A \in \mathcal{A}$, $\sup T \in \bigcap \mathcal{A} = M$ as required. Therefore M is an admissible subset of X that contains a. □

Proposition 4.7 implies that $M = \bigcap \mathcal{A}$ is the unique smallest element of $\mathcal{A}$ and is the smallest admissible subset of X such that any admissible subset of X contained in M is equal to M. We will use this fact later in our proof. Also, a, $f(a)$, and $\sup T$ are all in M and hence in every admissible subset of X. In other words, $a \in M \in \mathcal{A}$, and $S \in \mathcal{A} \Rightarrow M \subseteq S$. In our proof in the following Bourbaki-Witt Theorem, we need to prove that M is a chain.

Note that the admissible subset $B \in (\mathcal{A}, \subseteq)$ in Definition 4.5 (admissibility) is chain complete by the induced ordering of X [11, p. 13]. In other words, an admissible subset of a chain complete poset is chain complete. This is because $\mathcal{A}$ is defined by those admissible subset closure rules on X. So it is a closure system on

the complete lattice $\mathcal{P}(X)$ ordered by inclusion $\subseteq$. In fact, $(\mathcal{A}, \subseteq)$ is a poset [13, p. 114].

Proposition 4.8. *Let X be a chain complete partially ordered set and let $f : X \to X$ be an increasing self-map. Let $a \in X$. If $U = \{x \in X \mid a \leqslant x\}$, then U is an admissible subset of X that contains a.*

Proof. Clearly $U \subseteq X$ by Axiom 2.3 (the Subset Axiom Schema). (i) $a \in U$ is given. (ii) Since f is a self-map, $f|_U : U \to U$ implies $f(U) \subseteq U$. To see this, let $x \in U$. Then $a \leqslant x \leqslant f(x) \Rightarrow a \leqslant f(x)$, so $f(x) \in U$ and hence $f(U) \subseteq U$ as required. (iii) Let $T \subseteq U$ be a chain in U, then $a \leqslant t$ for all $t \in T$. So $t \leqslant \sup T$, which implies $a \leqslant \sup T$. Hence $\sup T \in U$. Thus, U is an admissible subset of X that contains a. □

We will now introduce the Bourbaki-Witt Theorem (BWT), which is crucial in proving HMP using AC. Once BWT is established, proving HMP is very easy. However, proving BWT itself takes a lot of work. Basically this task relies on the definition of admissible subset throughout the proof. In particular, in the two lemmas necessary for the proof, we will check the three conditions of admissibility in proving each lemma.

Theorem 4.9. (Bourbaki-Witt) *Let $X \neq \varnothing$ be a chain complete poset. Let $f : X \to X$ be an increasing self-map. Then for every $a \in X$, there exists a fixed point of f at*

or above a.

Proof. By Proposition 4.6, X is admissible. Let $a \in X$ and fix a throughout this proof. Let $\mathcal{A}$ be the collection of admissible subsets of X that contain a. Let $M = \bigcap \mathcal{A}$, then by Proposition 4.7, M is admissible. Note that $M \neq \varnothing$ since $a \in M$. Also, M is the smallest admissible subset of X and is the smallest element of $\mathcal{A}$, i.e. $M \in \mathcal{A}$, and $S \in \mathcal{A} \Rightarrow M \subseteq S$.

By Proposition 4.8, the set $U = \{x \in X \mid a \leqslant x\}$ is an admissible subset of X that contains a and so $M \subseteq U \in \mathcal{A}$; hence, $a \leqslant x$ for all $x \in M$. So a is the minimum of M.

Our goal now will be to prove that M is a chain because if M is a chain in the chain complete set X, then by Definition 4.4 M would have a least upper bound $p = \sup M \in X$ and then since M is admissible, $f(M) \subseteq M$ implies $f(p) \in M$. Hence, $f(p) \leqslant p$ because $p = \sup M$. But, since f is an increasing map, it is given that $p \leqslant f(p)$; hence, $f(p) \leqslant p \leqslant f(p) \Rightarrow f(p) = p$. So p is a fixed point of f.

To prove that M is a chain, we consider the following two sets, C and M_c:

$$C = \{c \in M \mid \text{for all } x \in M, \text{ either } x \leqslant c \Rightarrow f(x) \leqslant c \text{ or } x \leqslant c \Rightarrow x = c\} \tag{4.1}$$

$$M_c = \{x \in M \mid x \leqslant c \text{ or } f(c) \leqslant x\} \text{ for each extreme point } c \in M \tag{4.2}$$

The c in (4.1) is called an **extreme point** of M and C is a set of extreme points of M. Given that a is the minimum of M: if $x \leqslant a$ then $x = a$. This meets the condition of C, so $a \in C$ and $C \neq \varnothing$. Note that $C \subseteq M$ by the Subset Axiom Schema (Axiom 2.3).

Now look at the set M_c. Since a is the minimum of M, we have $a \leqslant c$ for all $c \in C$; hence $a \in M_c$ and $M_c \neq \varnothing$. By Axiom 2.3 (the Subset Axiom Schema), $M_c \subseteq M$ is the subset of M determined by $c \in C$. Some authors put $M(c)$ instead.

Recall that after the proof of Proposition 4.7, we mentioned that the set $\mathcal{A}$ is a partially ordered set under set inclusion, $(\mathcal{A}, \subseteq)$, so any admissible subset B (containing a) of the chain complete set X is chain complete. That is because $\mathcal{A} \subseteq \mathcal{P}(X)$ is defined by the admissible subset closure rules under (ii) the function f and under (iii) taking least upper bound, i.e. $f(B) \subseteq B$ and $\sup T \in B$ for any chain T in B. Thus $(\mathcal{A}, \subseteq)$ is a closure system on the complete lattice $(\mathcal{P}(X), \subseteq)$, which is bounded by $\varnothing$ and X itself with the least upper bound and the greatest lower bound given respectively by union ($\cup$) and intersection ($\cap$) of subsets of X. Note that $B \in (\mathcal{A}, \subseteq)$.

The proof of our theorem (BWT) depends on the following two lemmas. Again, our goal is to show that M is a chain in X. To do this, we need to establish the fact $M_c = M = C$ using these two lemmas. Once we know that M is a chain, with the fact that M is an admissible subset of the chain complete admissible set X, M is chain complete as well; hence $\sup M \in M$. Then eventually we will show that $\sup M$ is a fixed point using the fact that f is an increasing self-map.

Lemma 4.10. $M_c = M$ for all $c \in M$.

Proof of Lemma 4.10. We have $M_c \subseteq M$and need to show that $M \subseteq M_c$. It suffices to show that M_c is admissible because then $M \subseteq M_c$ since M is the smallest admissible

subset of X that contains a. First of all, it is clear that $M_c \subseteq M \subseteq X$.

Now check the conditions of admissibility.

(i) $a \in M_c$:

We already know that $a \in M_c$.

(ii) $f(M_c) \subseteq M_c$:

Let $x \in M_c \subseteq M$, then $x \in M$ and by (4.2), $x \in M_c \Rightarrow x \leqslant c$ or $f(c) \leqslant x$ for each $c \in C$. Note that $c \in C \Rightarrow c \in M$ by (4.2) and (4.1) for all $c \in C$. In summary, if $x \in M_c$, then we have $x \leqslant c$ or $f(c) \leqslant x$ with $x, c \in M$.

Case I: $x \leqslant c$.

If $x \leqslant c$ then by (4.1), $f(x) \leqslant c$ or $x = c$ with $x, c \in M$.

(a) If $f(x) \leqslant c$, then $f(x) \in M$ and immediately by (4.2) $f(x) \in M_c$.

(b) If $x = c$, then $f(c) \leqslant c \Rightarrow f(x) \leqslant c$ because f is increasing. Thus we have $f(x) \in M_c$.

Case II: $f(c) \leqslant x$.

On the other hand, if $f(c) \leqslant x \in M_c \subseteq M$, then $f(c) \in M_c \subseteq M$. Also, $f(c) \leqslant x \leqslant f(x)$ because f is increasing. Then $f(c) \leqslant f(x)$ and by (4.2) we have $f(x) \in M_c$.

Therefore, we conclude that $f(x) \in M_c$ for all $x \in M_c$; hence $f(M_c) \subseteq M_c$.

(iii) T is a chain in $M_c \Rightarrow \sup T \in M_c$:

Let T be a chain in M_c, and let $\sup T$ be the least upper bound of T in X. We will show that $\sup T \in M_c$.

Since M is admissible, we have $\sup T \in M$. Since $x \notin M_c$ when x falls in between c and $f(c)$. We only consider two cases. If $x \leqslant c$ for all $x \in T$, then

$\sup T \leqslant c$ and so, by (4.2), $\sup T \in M_c$. Otherwise, if there are some $x \in T$ such that $f(c) \leqslant x$, then $f(c) \leqslant x \leqslant \sup T \Rightarrow f(c) \leqslant \sup T$, and so $\sup T \in M_c$ by (4.2). Hence, T is a chain in $M_c \Rightarrow \sup T \in M_c$. Therefore M_c is admissible subset of X that contains a.

Note that M_c is admissible implies $M \subseteq M_c$ since M is the smallest admissible subset of X. On the other hand, (4.2) implies $M_c \subseteq M$. Hence, $M_c = M$ as required. □

Now we present the second lemma required to prove BWT and prove it.

Lemma 4.11. $C = M$

Proof of Lemma 4.11. This lemma says that every element of M is an extreme point. It will suffice to prove that C is an admissible subset of X that contains a. Note that $C \subseteq M$. So we only need to show that $M \subseteq C$. If C is admissible, than since M is the smallest admissible subset of X that contains a, we have $M \subseteq C$. It is clear that $C \subseteq M \subseteq X$.

Now check the conditions of admissibility.

(i) $a \in C$:

We already showed that $a \in C$.

(ii) $f(C) \subseteq C$:

Let $c \in C \subseteq M$, so $c \in M$ and $f(c) \in M$ since M is admissible. We will show that $f(c) \in C$. Let $x \in M$ and suppose $x \leqslant f(c)$. According to (4.1) we need to show

$f(c) \in C$ by showing $f(x) \leqslant f(c)$ or $x = f(c)$. Since $x \in M = M_c$ (Lemma 4.10), by (4.2) we have $x \leqslant c$ or $f(c) \leqslant x$.

Case I: $x \leqslant c$.

Then by (4.1) $f(x) \leqslant c$ or $x = c$.

(a) If $f(x) \leqslant c$, then $f(x) \leqslant c \leqslant f(c)$ since f is increasing, so $f(x) \leqslant f(c)$. Thus we have $x \leqslant f(c) \Rightarrow f(x) \leqslant f(c)$ with $c \in C$ and so $f(c) \in C$ by (4.1).

(b) If $x = c$, then $f(x) = f(c)$ and by plugging $f(x)$ and $f(c)$ into (4.1) we have $x \leqslant f(c) \Rightarrow f(x) = f(c) \Rightarrow f(x) \leqslant f(c)$, hence $f(c) \in C$.

Case II: $f(c) \leqslant x$.

Then $x \leqslant f(c) \Rightarrow f(c) \leqslant x$, so $x = f(c)$. So $x \leqslant f(c) \Rightarrow x = f(c)$; hence $f(c) \in C$ by (4.1). Thus, by Case I and II, we have $f(C) \subseteq C$.

(iii) T is a chain in $C \Rightarrow \sup T \in C$:

Let T be a chain in C. Let $b = \sup T$ be the least upper bound of T in X. We must prove that $b \in C$. Considering (4.1), let $x \in M$ and let $x \leqslant b$. We must show that $f(x) \leqslant b$ or $x = b$. Since $x \in M = M_c$ for all $c \in M$ (Lemma 4.10), by (4.2) we have $x \leqslant c$ or $f(c) \leqslant x$ for every $c \in T \subseteq C$.

Case I: $f(c) \leqslant x$ for all $c \in T$.

Then since $c \leqslant f(c)$, we have $c \leqslant x$ for all $c \in T$, thus x is an upper bound of T. Hence $b \leqslant x$ since $b = \sup T$. But we assumed that $x \leqslant b$, so $x = b$. Hence $x \leqslant b \Rightarrow x = b$ for all $x \in M$ and so by (4.1) $b = \sup T \in C$.

Case II: $x \leqslant d$ for some $d \in T$.

Let $d \in T \subseteq C$ be an extreme point of M and let $x \leqslant d$. Since d is an extreme

point, we have $f(x) \leqslant d$ or $x = d$ by (4.1).

(a) If $f(x) \leqslant d$, since $b = \sup T$, $f(x) \leqslant d \leqslant b$, and so $x \leqslant b \Rightarrow f(x) \leqslant b$. So $b = \sup T \in C$ by (4.1).

(b) If $x = d$, since $b \in M = M_d$ and $d \in C$, by (4.2) we have $b \leqslant d$ or $f(d) \leqslant b$.

(1) If $b \leqslant d$, since $b = \sup T$, $d \leqslant b$. So $b = d = x$. Hence $x \leqslant b \Rightarrow x = b$. Thus $b = \sup T \in C$.

(2) If $f(d) \leqslant b$, then since we assumed $x \leqslant b$, we have $f(x) = f(d) \leqslant b$ and so $x \leqslant b \Rightarrow f(x) \leqslant b$. Hence $b = \sup T \in C$.

Therefore T is a chain in $C \Rightarrow \sup T \in C$. It follows that C is an admissible subset of X that contains a.

(4.1) shows that $C \subseteq M$ by Axiom 2.3 (the Subset Axiom Schema). But since M is the smallest admissible subset of X that contains a , we have $C = M$. □

With the two lemmas we can now complete the proof of the theorem.

Next we will show that M is a chain in X. Let $x, y \in M$. Since $M = M_c = C$ for all $c \in M$, we can say that $x \in C$ and $y \in M_c = M_x$ if we take $c = x$. Then we have $y \leqslant x$ or $f(x) \leqslant y$. For $f(x) \leqslant y$, since $x \leqslant f(x)$, $x \leqslant y$. Thus we have $y \leqslant x$ or $x \leqslant y$ for all $x, y \in M$ and so M is a chain in X.

Let $p = \sup M$. Since M is a chain in X and M is an admissible subset of X, M is chain complete and so $\sup M = p \in M \subseteq X$. Since M is admissible, $f(M) \subseteq M \Rightarrow f(p) \in M$. Hence $f(p) \leqslant p$ because $p = \sup M$. But, since f is an increasing map, $p \leqslant f(p)$; hence, $f(p) \leqslant p \leqslant f(p) \Rightarrow f(p) = p$. Therefore, p is a fixed point of f [7, 11, 13]. □

Next we will prove that AC implies HMP. Hausdorff's Maximal Principle is an earlier formulation of Zorn's Lemma proved by Felix Hausdorff in 1914. Its basic idea is that *every chain is contained in a maximal chain.* We will prove Zorn's Lemma using Hausdorff's Maximal Principle later after this section. For now let us define Hausdorff's Maximal Principle.

Theorem 4.12. Hausdorff's Maximal Principle

Let the set $\mathcal{C}$ of all chains of a poset $(X, \leqslant)$ be partially ordered by set inclusion, $\subseteq$. Then $(\mathcal{C}, \subseteq)$ has a maximal element.

Proof. Suppose on the contrary that $\mathcal{C}$ has no maximal element. Then to each $C \in \mathcal{C}$, there is associated a nonempty set $C^* = \{C' \in \mathcal{C} \mid C \subsetneq C'\}$, a collection of strict super sets of C. Note that here we don't allow $C = C'$ because this would allow the existence of maximal elements.

By AC, there is a function g with domain $\{C^* \mid C \in \mathcal{C}\}$ satisfying $g(C^*) \in C^*$. In other words, there is a $C' \in C^*$ such that $g(C^*) = C'$. Consequently, for each $C \in \mathcal{C}$ there is a function $f : \mathcal{C} \to \mathcal{C}$ defined by $f(C) = g(C^*) = C'$ with $C \subsetneq g(C^*) = f(C) = C'$ for all $C \in \mathcal{C}$. Note that f is an increasing self-map since $C \subsetneq f(C) = g(C^*)$.

Let $(\mathcal{B}, \subseteq) \subseteq (\mathcal{C}, \subseteq)$ be a sub-collection of chains in X. Then since each $\mathcal{B} \subseteq \mathcal{C}$ has a least upper bound $\bigcup_{B \in \mathcal{B}} B = \bigcup \mathcal{B} \in \mathcal{C}$, $(\mathcal{C}, \subseteq)$ is a nonempty chain complete poset. Then by Theorem 4.9 (Bourbaki-Witt), for every chain $C \in \mathcal{C}$ there exists

a fixed "point" chain $\bar{C}$ at or above C with $f(\bar{C}) = \bar{C}$. But $\bar{C} \subsetneq f(\bar{C})$ because $f : \mathcal{C} \to \mathcal{C}$ is an increasing self-map. We have a contradiction. Thus $(\mathcal{C}, \subseteq)$ has a maximal element. [13] □

4.2 HMP $\Longrightarrow$ ZL

Theorem 4.13. (Zorn's Lemma) *Let $(X, \leqslant)$ be a poset in which every chain has an upper bound in X. Then X has a maximal element.* [4, p. 151]

Proof. By HMP, there is a maximal chain C in X. By hypothesis C has an upper bound $u \in X$. We shall prove that u is a maximal element of X.

If there is an $x \in X$ with $u \leqslant x$, then $C \cup \{x\}$ is a chain that contains C since $C \cup \{x\}$ is a comparable subset of X. Since C is a maximal chain, $C \cup \{x\} = C$ and so $x \in C$. Thus $x \leqslant u$ and hence $u = x$. Therefore u is a maximal element of X. □

4.3 ZL $\Longrightarrow$ WOT

Definition 4.14. (Well-Ordered Set, Well-Ordered Relation) *A linearly ordered set $(X, \leqslant)$ is said to be **well-ordered** if and only if every nonempty subset B of X contains a least element; that is, if there exists an element $b \in B$ such that $b \leqslant x$ for every $x \in B$. Such an element b is called the minimum or least element of B. If X is a well-ordered set then $\leqslant$ is a **well-order relation**.*

Theorem 4.15. (Well-Ordering Theorem or Well-Ordering Principle) *Every set can*

be well-ordered.

Proof. Let X be a nonempty set. Let $X^* = \{(X_k, \leqslant_k) \mid X_k \subseteq X\}$ where $(X_k, \leqslant_k)$ are well-ordered subsets of X indexed by $k \in I$. In other words, X^* is the collection of all subsets $X_k \subseteq X$ for which there is a well ordering $\leqslant_k$ on X_k, i.e., it's the collection of all subsets of X which can be well-ordered. Note that $X^* \subseteq \mathcal{P}(X)$, where $\mathcal{P}(X)$ is partially ordered by set inclusion, $(\mathcal{P}(X), \subseteq)$, and X^* is partially ordered by $\leqslant^*$ as defined below.

We partially order X^* by $\leqslant^*$ as follows: $(X_i, \leqslant_i) \leqslant^* (X_j, \leqslant_j)$ for $i, j \in I$ if and only if

(a) $X_i \subseteq X_j$.

(b) $\leqslant_i \subseteq \leqslant_j$. In other words $x \leqslant_i y \Rightarrow x \leqslant_j y$, for all $x, y \in X_i$.

(c) If $x \in X_i$ and $y \in X_j \setminus X_i$, then $x \leqslant_j y$.

We write $(X^*, \leqslant^*)$.

In order to apply Zorn's Lemma, we show that any chain C in $(X^*, \leqslant^*)$ has an upper bound in X^*. Note that $C = \{(X_c, \leqslant_c) \in X^* \mid X_c \subseteq X$ can be well-ordered and $\leqslant_c$ is a linear order$\}$, where $C \subseteq X^*$ and $(C, \leqslant^*)$.

The natural candidate for this upper bound is $(\bigcup_{X_c \in C} X_c, \leqslant')$ or $(\bigcup C, \leqslant')$, where, for any $(X_c, \leqslant_c) \in C$, $(X_c, \leqslant_c) \leqslant^* (\bigcup C, \leqslant')$ with:

(a) $X_c \subseteq \bigcup C$, which is true since with C a chain under set inclusion, $\subseteq$, $X_c \in C \Rightarrow X_c \subseteq \bigcup C$.

(b) $\leqslant_c \subseteq \leqslant'$, which is true since $X_c \subseteq \bigcup C$ and for any $x, y \in X_c$, $\{(x, y) \mid x \leqslant_c y\} \subseteq \{(x, y) \mid x \leqslant' y\}$. i.e. for any $x, y \in X_c$, $x \leqslant_c y \Rightarrow x \leqslant' y$.

(c) If $x \in X_c$ and $y \in \bigcup C \setminus X_c$, then $x \preccurlyeq' y$. True because $x \in X_c \subseteq \bigcup C$ implies $x \in \bigcup C$. $y \in \bigcup C \setminus X_c$ implies $y \in \bigcup C$. So either $x \preccurlyeq' y$ or $y \preccurlyeq' x$ since $\preccurlyeq'$ is a liner order (see "proof" below). Since $y \notin X_c$, $y \not\preccurlyeq' x$ and so $x \preccurlyeq' y$.

Note that $\preccurlyeq'$ is a linear ordering on $\bigcup C$. Here is an informal proof.

Let $x, y \in \bigcup_{X_c \in C} X_c = \bigcup C$. Then $x \in X_x$ for some $X_x \in C$ and $y \in X_y$ for some $X_y \in C$. Since C is a chain, it is linearly ordered by set inclusion, so either $X_x \subseteq X_y$ or $X_y \subseteq X_X$. WLOG suppose that $X_x \subseteq X_y$, then $x, y \in X_y$ with $(X_y, \preccurlyeq_y)$. So either $x \preccurlyeq_y y$ or $y \preccurlyeq_y x$ since X_y is well-ordered under the linear ordering $\preccurlyeq_y$. Since $\preccurlyeq_y \subseteq \preccurlyeq'$, we have $x \preccurlyeq' y$ or $y \preccurlyeq' x$ and so $\bigcup C$ is linearly ordered under $\preccurlyeq'$.

This applies to all subsets $X_c \subseteq \bigcup C$ for all $X_c \in C$, where $\preccurlyeq_c \subseteq \preccurlyeq'$. WLOG if $x, y \in \bigcup C$, then either $x \preccurlyeq' y$ or $y \preccurlyeq' x$ since $\bigcup C$ is bigger than or equal to any $X_c \in C$. Thus $\preccurlyeq'$ is a linear order relation on $\bigcup C$.

To apply Zorn's Lemma, we need to show that $(\bigcup C, \preccurlyeq') \in X^*$. Obviously $(\bigcup C, \preccurlyeq')$ is an upper bound for C if $(\bigcup C, \preccurlyeq')$ is well-ordered. We shall prove that $(\bigcup C, \preccurlyeq')$ is well-ordered and hence $(\bigcup C, \preccurlyeq') \in X^*$. We don't know whether $\bigcup C$ is well-ordered yet, but we do know a nonempty intersection of a nonempty subset of $\bigcup C$ with an element of the chain C, e.g. $X_i \in C$, is well-ordered.

Let $S \neq \varnothing$, $S \subseteq \bigcup C$ with $(S, \preccurlyeq')$ since S by property (b) inherits the ordering from $\bigcup C$. Then there exists $(X_i, \preccurlyeq') \in C$ such that $X_i \cap S \neq \varnothing$ and $S \cap X_i \in X^*$ with $(S \cap X_i, \preccurlyeq')$. Note that by properties (a) and (b) the order on X_i is $\preccurlyeq'$ because $X_i \in C \Rightarrow X_i \subseteq \bigcup C$ and so X_i inherits the order $\preccurlyeq'$ on $\bigcup C$.

Since $(X_i, \preccurlyeq') \in X^*$ is well-ordered, $S \cap X_i \subseteq X_i$ is also well-ordered and

contains a unique least element, say $x_0 \in S \cap X_i$. We want to show that x_0 is the least element of $\bigcup C$ under $\leqslant'$.

Let $x \in S \subseteq \bigcup C$. Then $x \in X_j$ for some $X_j \in C$. We want to show that $x_0 \leqslant' x$. Since C is a chain, we know $X_i \subseteq X_j$ or $X_j \subseteq X_i$.

Case I: $X_i \subseteq X_j$.

If $x \in X_i \subseteq X_j$, then $x_0 \leqslant' x$ since they are both in $S \cap X_i$ with x_0

being the least element. If $x \in X_j \setminus X_i$, then $x_0 \leqslant' x$ by property (c).

Case II: $X_j \subseteq X_i$.

Then $x \in X_i$ as well. We know that $x \in S \cap \bigcup C$, so $x \in S \cap X_i$. Since x_0 is the least element of $S \cap X_i$, we have $x_0 \leqslant' x$.

Hence x_0 is the least element of $S \subseteq \bigcup C$ for all arbitrary nonempty subset S of $\bigcup C$. So x_0 is the least element of $\bigcup C$ since it is a nonempty subset of itself. Thus $\bigcup C$ is under the linear ordering $\leqslant'$ and has a least element. We conclude that $\bigcup C$ is well-ordered.

Thus $(\bigcup C, \leqslant') $ is well-ordered and so $(\bigcup C, \leqslant') \in X^*$.

Since any chain C in X^* has an upper bound $(\bigcup C, \leqslant'$ in X^*, by Zorn's lemma, $(X^*, \leqslant^*)$ has a maximal element $(X_M, \leqslant_M)$.

We claim that $X_M = X$ and hence $(X, \leqslant_M)$ is well-ordered, because if $X_M \neq X$, take any $\tilde{x} \in X \setminus X_M$ and extend $\leqslant_M$ to $X_M \cup \{\tilde{x}\}$, where $(X_M \cup \{\tilde{x}\}, \leqslant_M) \in X^*$, by defining $x \leqslant_M \tilde{x}$ for all $x \in X_M$; then $(X_M, \leqslant_M) \lneq^* (X_M \cup \{\tilde{x}\}, \leqslant_M)$ [strictly less under $\leqslant^*$], which contradicts the maximality of $(X_M, \leqslant_M)$. So $X = X_M$ and so any arbitrarily chosen set X or $(X, \leqslant_M)$ can be well-ordered. [13] $\square$

4.4 WOT $\Longrightarrow$ AC

Theorem 4.16. *WOT implies AC.*

Proof. Let $\mathcal{X}$ be any nonempty set whose elements are nonempty sets. By WOT there exists a linear order relation $\leqslant$ such that $(\bigcup_{X\in\mathcal{X}} X, \leqslant)$ or $(\bigcup \mathcal{X}, \leqslant)$ is well-ordered.

Consequently, each set $X \in \mathcal{X}$ contains a least element x. Therefore, the function $f : \mathcal{X} \to \bigcup \mathcal{X}$, defined by $f(X) = x \in X$, for all $X \in \mathcal{X}$, is a choice function. This proves AC. [13] □

CHAPTER 5

Uses of the Axiom of Choice in Mathematics

The Axiom of Choice (AC) has many more equivalents than the three we have studied, including some weak forms such as Axiom of Dependent Choice, Principle of Finite Choice, Axiom of Countable Choice, etc. Among those, HMP, ZL, and WOT are probably the most frequently seen variants of AC. Now that we have shown their equivalences, we should also study their applications in mathematics.

An interesting fact is that AC itself is hardly ever used directly to prove things in mathematics. The main purpose of AC is to appear intuitive, so as to disguise its strangeness. HMP, ZL, and WOT are not so obviously true but are more frequently used in proofs [16].

We will give a few theorems each proved by some equivalents of AC.

5.1 Application in Set Theory

Definition 5.1. (Transitive Set) *A set X is **transitive** if every element of X is a subset of X.*

In other words, if X is a transitive set and $x \in X$, then $x \subseteq X$. A transitive set has the property that for all sets x, y, if $x \in y$ and $y \in X$ then $x \in X$. In abbreviation, we put $x \in y \in X$. [2, 9]

Definition 5.2. (Ordinal Number or Ordinal, ON) *A set α is an **ordinal number***

or, in short, ***ordinal*** *if*

(a) α *is transitive.*

(b) α *is well-ordered by* $\in$*.* [10]

We denote **ON** *to mean the collection of all ordinal numbers.* [16]

Note that the collection of ordinals, ON, is not a set. In our subsequent proofs, we assume ordinals in ON start with 0 when we take ON as not infinitely descending. We also have to define ordinal ordering, $<$, successor ordinal, and limit ordinal.

Definition 5.3. ($<$, Successor Ordinal, Limit Ordinal) *Let* α, β *be ordinals.*

(a) For all ordinals α, β*,* $\alpha < \beta$ *if and only if* $\alpha \in \beta$*. We write* $\alpha \leq \beta$ *to mean* $\alpha \in \beta$ *or* $\alpha = \beta$*.*

(b) For all ordinals α*, define the* ***successor ordinal*** $\alpha^+ = \alpha \cup \{\alpha\}$*. We also denote* $\alpha^+ = \alpha + 1$ *to mean the successor ordinal of* α*.*

(c) A nonzero ordinal α *is a* ***limit ordinal*** *if for all ordinals* $\beta, \alpha \neq \beta^+$*.* [16]

Definition 5.4. (Initial Segment) *Let the set* X *be linearly ordered by* $\leqslant$ *and* $A \subseteq X$*.* A *is an* ***initial segment*** *if* $x \in X, y \in A$ *then* $x \leqslant y$ *implies* $x \in A$*.*

Trivially, $\varnothing$ and X are initial segments.

Proposition 5.5. *The union of a family of initial segments is an initial segment. The intersection of a nonempty family of initial sections is an initial segment.*

Proposition 5.6. *Suppose $I \subseteq X$ and $J \subseteq X$ are initial segments. Then either $I \subseteq J$ or $J \subseteq I$.*

Note that Proposition 5.5 and Proposition 5.6 can actually be considered as properties of initial segments. They are important concepts for our next proof yet they are quite intuitive, so we won't give them proofs since their proofs are not important to the context of this paper.

Definition 5.7. (Cofinal Subset) *If X is linearly ordered by $\leqslant$, we say A is **cofinal** in X iff $A \subseteq X$ and for every $x \in X$ there is $y \in A$ such that $x \leqslant y$.*

Theorem 5.8. *Every linearly ordered set has a well-ordered cofinal subset.*

Proof using WOT. Let $(X, \leqslant)$ be a nonempty linearly ordered set. Then by WOT, X can be well-ordered by a well-order relation $\preceq$ and so every nonempty subset of $(X, \preceq)$ has a least element.

Since the collection of ordinals (ordinal numbers, ON) is well-ordered, we can well-order X by indexing its elements with ordinals α. Suppose the well-ordered set is $(X, \preceq)$ with $X = \{x_\alpha\}_{\alpha \preceq \beta} = \{x_\alpha \mid \alpha \preceq \beta \text{ for some } \beta \in \text{ON}\}$. Note that the original ordering $\leqslant$ has nothing to do with the new well-ordering $\preceq$. In other words, $\preceq$ scrambles X and put it in well-order.

We want a cofinal subset of X. Let the well-ordered cofinal subset of X be $A = \{x_\gamma \mid \delta \preceq \gamma \Rightarrow x_\delta \leqslant \gamma \text{ for some } \delta \in \text{ON}\}$. Then $A \neq \varnothing$ since $x_0 \in A$ with

$0 \in \text{ON}$ being the initial ordinal. This is because the condition for $x_0 \in A$, i.e., $\delta \preceq 0 \Rightarrow x_\delta \leqslant x_0$, is vacuously true.

Now we want to show that A is cofinal.

Suppose that $x_\alpha \in (X, \preceq)$. Let γ be the least ordinal such that $\alpha \preceq \gamma$ with $x_\gamma \in A$. Then $\delta \preceq \gamma \Rightarrow X_\delta \leqslant x_\gamma$. There are two cases.

Case I: $x_\alpha \in A$. Then $\delta \preceq \alpha \Rightarrow x_\delta \leqslant x_\gamma$. Since x_α, x_γ and $\alpha \preceq \gamma$, we have $\alpha \preceq \gamma \Rightarrow x_\alpha \leqslant x_\gamma$. Thus for any arbitrary $x_\alpha \in X$, there exists an $x_\gamma \in A$ such that $x_\alpha \leqslant x_\gamma$ and so A is cofinal in X.

Case II: $x_\alpha \notin A$. Then $\delta \preceq \alpha$ but $x_\alpha \leqslant x_\delta$, WLOG assuming the negation of $x_\delta \leqslant x_\alpha$ is $x_\alpha \leqslant x_\delta$. So $x_\alpha \leqslant x_\gamma$ since $x_\delta \leqslant x_\gamma$. Note that $x_\alpha \in X$, still. So we have an arbitrary $x_\alpha \in X$ and there exists an $x_\gamma \in A$ such that $x_\alpha \leqslant x_\gamma$. Thus A is cofinal in X as required.

Now we show that A is well-ordered. We want to show that every nonempty subset of A has a least element.

Let S be a nonempty subset of A. Let $\Delta = \{\lambda \mid x_\lambda \in S \text{ for some } \lambda \in \text{ON}\}$. Then $\Delta \neq 0$ since $S \neq \varnothing$. Note that all the x_λ are also in A since $S \subseteq A$.

Let β be the smallest ordinal in Δ. Then $\beta \preceq \lambda$ for all $\beta \neq \lambda \in \text{ON}$. Since $\beta \in \Delta$, there exists $x_\beta \in S \subseteq A$ and so $x_\beta \in A$.

Now since $\beta \preceq \lambda$ and $x_\beta, x_\lambda \in A$, by the condition of A, we have $\beta \preceq \lambda \Rightarrow x_\beta \leqslant x_\lambda$ for all $x_\lambda \in S$. Hence x_β is the smallest element of S and so A is well-ordered.

Therefore A is a well-ordered cofinal subset of X as required. [16] □

Proof using ZL. Let $(X, \leqslant)$ be a nonempty linearly ordered set. We want a well-

ordered cofinal subset of X. To use Zorn's Lemma, we need a nonzero poset.

Let $\mathcal{A}$ be the nonempty collection of all well-ordered subsets of X ordered by "end-extension": $A \subseteq B$ if A is an initial segment of B for any $A, B \in \mathcal{A}$.

Then by Proposition 5.6, $\mathcal{A}$ is linearly ordered under set inclusion $\subseteq$, so $(\mathcal{A}, \subseteq)$ is a poset. Note that elements in $\mathcal{A} \subseteq \mathcal{P}(X)$ are subsets of X, i.e. $W_k \subseteq X$ if $W_k \in \mathcal{A}$.

We claim that $\mathcal{A}$ has no infinite descending chains. To see this, suppose $\mathcal{A}$ has an infinite descending chain $A_0 > A_1 > A_2 > \cdots$ and $x_i \in A_i \setminus A_{i+1}$ for each $A_i \in \mathcal{A}$ with $i \in \mathbb{N}$, then $x_0 > x_1 > \cdots$, hence A_0 dosn't have a least element and so would not be well-ordered. A contradiction, so $\mathcal{A}$ has no infinite descending chains.

To apply ZL we must show each chain in $\mathcal{A}$ has an upper bound in $\mathcal{A}$. Let C be a chain in $\mathcal{A}$. Then $\bigcup_{W \in C} W = \bigcup C$ is an upper bound for C. Then we want to show that $\bigcup C$ is a well-ordered subset of X ordered by "end-extension", i.e., $\bigcup C \in \mathcal{A}$. Obviously $\bigcup C \subseteq X$ is linearly ordered since X is linearly ordered. We will show that $\bigcup C$ has a smallest element by showing it does not have an infinite descending chain.

Suppose $\bigcup C$ did not have a least element such that it had an infinite descending chain $x_0 > x_1 > \cdots$, then x_0 must be contained in a subset $W \in C$, so $x_0 \in W$. W would have an infinite descending chain starting at x_0. We have a contradiction. So $\bigcup C$ has a smallest element and is well-ordered, hence $\bigcup C \in \mathcal{A}$.

So every chain in $\mathcal{A}$ has an upper bound in $\mathcal{A}$. By Zorn's Lemma, $\mathcal{A}$ has a maximal element $W_M \subseteq X$. Since $W_M \in \mathcal{A}$ is maximal and $\mathcal{A}$ is linearly ordered, W_M is the largest well-ordered subset of X. So all other well-ordered subsets of X are subsets of W_M, i.e., $W_k \subseteq W_M$ for all $W_k \in \mathcal{A}$.

If $x \in X$, there is a well-ordered initial segment W containing x, $x \in W$. Since

$W \subseteq W_M$, there is $y \in W_M$ with $x \leqslant y$. So W_M is a cofinal subset of X. $\square$

5.2 Application in Linear Analysis

Definition 5.9. (Vetor Space). *If $\mathbb{F}$ is a field, then the **vector space** over $\mathbb{F}$ is a set $\mathcal{V}$ of vectors with operations of addition, $+ : \mathcal{V} \times \mathcal{V} \to \mathcal{V}$ and scalar multiplication $\cdot : \mathbb{F} \times \mathcal{V} \to \mathcal{V}$ which satisfy the following properties for all vectors u, v, $w \in \mathcal{V}$ and scalars $\lambda, \mu \in \mathbb{F}$.*

1. $(u + v) + w = u + (v + w)$. (Associativity)
2. $v + w = w + v$. (Commutativity)
3. *There is a zero vector $\vec{0} \in \mathcal{V}$ which has the property that $v + \vec{0} = v$ for every $v \in \mathcal{V}$.*
4. *For each v in $\mathcal{V}$ there is a vector $-v \in \mathcal{V}$ such that $v + (-v) = \vec{0}$.*
5. $\lambda(v + w) = \lambda v + \lambda w$.
6. $(\lambda + \mu)v = \lambda v + \mu v$.
7. $(\lambda\mu)v = \lambda(\mu v)$.
8. $1 \cdot v = v$. [8, p. 50]

Definition 5.10. (Spanning Set). *A subset $\mathcal{S}$ of a vector space $\mathcal{V}$ is said to **span** $\mathcal{V}$ or to be a **spanning set** for $\mathcal{V}$ if* span $(\mathcal{S}) = \mathcal{V}$. *That is, each vector in $\mathcal{V}$ can be written as a finite linear combination of the vectors in $\mathcal{S}$.* [8, p. 70]

Definition 5.11. (Linearly Dependent/Independent) *A subset $\mathcal{X}$ of a vector space $\mathcal{V}$ over a field $\mathbb{F}$ is said be* ***linearly dependent*** *if there is a finite subset $\{v_1, v_2, \ldots, v_m\}$ of distinct elements of $\mathcal{X}$ and scalars $\lambda_1, \lambda_2, \ldots, \lambda_m \in \mathbb{F}$, not all 0, such that $\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_m v_m = \vec{0}$. A set of vectors is called* ***linearly independent*** *if it is not linearly dependent.* [8, p. 74]

Proposition 5.12. *Suppose $\mathcal{S}$ is a subset of a vector space $\mathcal{V}$ over a field $\mathbb{F}$, then*

(a) $\mathcal{S}$ is linearly independent if and only if every finite subset of $\mathcal{S}$ is linearly independent.

(b) $\mathcal{S}$ is linearly independent if and only if $\lambda_1 \vec{v}_1 + \cdots + \lambda_m \vec{v}_m = \vec{0}$ for scalars $\lambda_1, \ldots, \lambda_m \in \mathbb{F}$ and distinct vectors $\vec{v}_1, \ldots, \vec{v}_m \in \mathcal{S}$ implies $\lambda_1 = \lambda_2 = \cdots = \lambda_m = 0$.

Definition 5.13. (Basis). *A linearly independent spanning set for a vector space $\mathcal{V}$ is called a* ***basis*** *for $\mathcal{V}$.*

Proposition 5.14. *Suppose $\mathcal{B}$ is a subset of vector space $\mathcal{V}$. Then the following are equivalent.*

(a) $\mathcal{B}$ is a basis for $\mathcal{V}$.

(b) $\mathcal{B}$ is a maximal linearly independent set in $\mathcal{V}$.

(c) $\mathcal{B}$ is a minimal spanning set for $\mathcal{V}$. [8]

Theorem 5.15. *Every vector space has a basis.*

Proof. Let $\mathcal{V}$ be a vector space over some field $\mathbb{F}$. We want to show that $\mathcal{V}$ has a basis. Note that every vector space contains at least the zero vector $\vec{0}$. If $\mathcal{V}$ contains only $\vec{0}$, $\mathcal{V} = \{\vec{0}\}$, then the empty set $\varnothing$ is a basis for $\mathcal{V}$. Then we are done. Otherwise, we consider the case where $\mathcal{V}$ contains at least one nonzero vector as follows.

Let L be the set of all linearly independent subsets of $\mathcal{V}$, i.e. $L = \{B \subseteq \mathcal{V} \mid B$ is linearly independent$\}$. Then L is a poset under set inclusion, i.e. $(L, \subseteq)$. Note that $L \neq \varnothing$. Since $\mathcal{V}$ contains at least one nonzero element, $\mathcal{V}$ will have a linearly independent subset. To see this, suppose $\mathcal{V} = \{\vec{v}\}$ has only one element. Then the singleton set $\{\vec{v}\}$ is in L and so $L \neq \varnothing$.

To use Zorn's Lemma, we need a poset. Here L is the poset we need.

Let $C = \{B_k\}_{k \in I} \subseteq L$ be a chain in L. Then a natural upper bound of C would be $\bigcup_{B_k \in C} B_k = \bigcup C \subseteq \mathcal{V}$. Note that all elements of C, i.e. all B_k, are linearly independent subsets of $\mathcal{V}$, so $B_k \subseteq \mathcal{V}$.

In order to make use of Zorn's Lemma, we need to show that $\bigcup C$ is linearly independent, i.e. $\bigcup C \in L$.

Let $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n \in \bigcup_{B_k \in C} B_k = \bigcup C$, where $\vec{v}_i$ are distinct vectors in $\bigcup C$ for each i, $1 \leq i \leq n$. To show linear independence, let $\lambda_1 \vec{v}_1 + \lambda_2 \vec{v}_2 + \cdots + \lambda_n \vec{v}_n = \vec{0}$, where $\lambda_i \in \mathbb{F}$, $1 \leq i \leq n$, are scalars. Then for each $\vec{v}_i$, there is a $B_i \in C$ with $\vec{v}_i \in B_i$. Since C is a chain, one of the linearly independent $B_1, \ldots, B_n \in C$ is largest. Call it B_M. Then $\vec{v}_1, \ldots, \vec{v_n}$ are all in B_M. Since B_M is linearly independent, $\lambda_1 = \lambda_2 = \cdots = \lambda_n = 0$ by Proposition 5.12.

Since $\{\vec{v}_1, \ldots, \vec{v}_n\} \subseteq \bigcup C$ is linearly independent, $\bigcup C$ is linearly independent

by Proposition 5.12 and so $\bigcup C \in L$. Then by Zorn's Lemma, L has a maximal element, i.e. a maximal linearly independent subset of $\mathcal{V}$, which is a basis of $\mathcal{V}$ by Proposition 5.14. □

5.3 Application in Abstract Algebra

Definition 5.16. (Ring, Commutative Ring, Identity) *A **ring** R is a set together with two binary operations $+$ and $\cdot$ (called addition and multiplication) satisfying the following axioms:*

(a) $(R, +)$ is an abelian group,

(b) $\cdot$ is associative: $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ for all $a, b, c \in R$,

(c) The distributive laws hold in R: for all $a, b, c \in R$,

$$(a + b) \cdot c = a \cdot c + b \cdot c \text{ and } c \cdot (a + b) = c \cdot a + c \cdot b.$$

*The ring R is **commutative** if multiplication is commutative. R is said to have an **identity** (or **contain** a 1) if there is an element $1 \in R$ with*

$$1 \cdot a = a \cdot 1 = a \text{ for all } a \in R.$$

We also denote the ring R with its binary operations by $(R, +, \cdot)$. [3]

We shall write ab instead of $a \cdot b$ for $a, b \in R$. The additive identity of R will be denoted by 0 and the additive inverse of the ring element a will be denoted by $-a$.

Definition 5.17. (Field) *A **field** F is a commutative ring $(F, +, \cdot)$ if and only if:*

(a) $(F, +)$ *is an abelian group,*

(b) $(F \setminus \{0\}, \cdot)$ *is an abelian group,*

(c) multiplication, $\cdot$*, distributes over addition,* $+$*.*

Note that F is a commutative ring with identity 1, where $1 \neq 0$. In F, every nonzero $a \in F$ has a multiplicative inverse, i.e., there exists $b \in F$ such that $ab = ba = 1$.

Definition 5.18. (Ideal of a Commutative Ring) *Let R be a commutative ring and let $r \in R$. A subset I of R is an* ***ideal*** *of R if and only if*

(a) $rI = \{ra \mid a \in I\} = \{ar \mid a \in I\} = Ir$*,*

(b) I is closed under $\cdot$ by elements from R, i.e. $rI = Ir \subseteq I$ for all $r \in R$.

Proposition 5.19. *Let R be a commutative ring. Then R is a field if and only if its only ideals are $\{0\}$ and R.*

Definition 5.20. (Trivial Ideal, Proper Ideal, Maximal Ideal) *Let R be a commutative ring. Then R and $\{0\}$ are ideals. $\{0\}$ is called the* ***trivial ideal*** *and can be denoted by 0. An ideal I of R is* ***proper*** *if $I \neq R$. An proper ideal M in a ring R is a* ***maximal ideal*** *of R if $M \neq R$ and the only ideals containing M are M and R.*

Theorem 5.21. *Every commutative ring with identity has a maximal ideal.*

Proof. Let R be a commutative ring with identity 1. We want to show that R contains a maximal ideal.

If R is a field, then the only ideals are $\{0\}$ and R. Since R is not a proper ideal, the trivial ideal $\{0\}$ is the maximal ideal. We are done.

If R is not a field, then there are other proper ideals between $\{0\}$ and R. Particularly if R is not a field, then there exists a non-invertable $a \in R$ with no multiplicative inverse in R.

If $I \subseteq R$ is a proper ideal of R, then $1 \notin I$ since otherwise $I = R$ and I is not proper. With the same non-invertable $a \in R$, let $I = \{ar \mid a \in I, r \in R \text{ and } ar \neq 1\}$. Then I is a proper ideal. To see that I is an ideal, let $x, y \in I$, then $x = ar_1, y = ar_2$ for some $r_1, r_2 \in R$. Then $x + y = a(r_1 + r_2) \in I$ and $ar_1 \cdot ar_2 = a \cdot (ar_1r_2) \in I$. Also, $0 \in I$ since $0 \cdot r = 0 \in I$ and $-a \in I$ since $-a + a = 0 \in I$. To see that I is proper, we know that $I \neq R$ because $1 \notin I$. Thus, I is a proper ideal.

To use Zorn's Lemma, we need a nonempty poset.

Since ideals are partially ordered by set inclusion $\subseteq$, let P be the collection of all proper ideals of R partially ordered by set inclusion, i.e., $P = \{I \subseteq R \mid I \text{ is an ideal}\}$. Then $P \neq \varnothing$ since the above-mentioned ideal $I \in P$. Note that all such proper ideals $I \in P$ must contain a non-invertable a as shown above.

To apply Zorn's Lemma, we need to show that every chain C in P has an upper bound in P.

Let $C = \{I_k\}_{k \in J} = \{I_k \in C \mid I_i \subseteq I_j \text{ or } I_j \subseteq I_i, i, j, k \in J\}$ be a chain in $(P, \subseteq)$. A natural upper bound of C is $\bigcup_{I_k \in C} I_k = \bigcup C$. Then $\bigcup C$ contains all the ideals I_k in C. Note that $\bigcup C \subseteq R$ and that $I_k \in C \subseteq P$ for all proper ideals $I_k \in C$.

Next, we want to show that $\bigcup C \in P$, i.e., we want to show that $\bigcup C$ is a proper ideal.

To see that $\bigcup C$ is an ideal, first we know that $0 \in \bigcup C$ since $0 \in I_k$ for all $k \in J$. Also, let $r = -1$, then we have $-a \in \bigcup C$.

Now, suppose that $a, b \in \bigcup C$. Then there exists some $I_i, I_j \in C$ with $a \in I_i$ and $b \in I_j$. Since C is a chain, either $I_i \subseteq I_j$ or $I_j \subseteq I_i$. WLOG suppose that $I_j \subseteq I_i$. Then both $a, b \in I_i$, hence $a + b \in I_i$, and so $a + b \in \bigcup C$. Hence $\bigcup C$ is closed under addition.

Finally, suppose that $a \in \bigcup C$ and $r \in R$. Then $a \in I_k \in \bigcup C$ for some $I_k \in C$. Thus $ar \in I_k$ and so $ar \in \bigcup C$. So $\bigcup C$ is closed under multiplication by arbitrary ring elements.

So $\bigcap C \in P$ is a proper ideal, i.e., $\bigcup C \in P$.

Therefore $\bigcup C \in P$ is an upper bound of a chain C in R. By Zorn's Lemma P has a maximal element, i.e. R has a maximal ideal. $\square$

CHAPTER 6

Conclusion

In Chapters 1 and 2, we start with a review of the history of (ZF) axiomatic set theory, exploring how mathematics can be built up with only a few axioms. During the review, we learned the nature of the foundation of mathematics by seeing how mathematicians discover problems that challenge the consistence of theoretic mathematics and how mathematicians come up with a solution. In particular we saw how Russell's Paradox was found and overcome by modifying one axiom in ZF and then brought in a new concept, class, to handle problems that the original ZF axiomatic set theory could not manage. The system of mathematics was hence expanded to a broader universe.

One moral from this is that the foundation of mathematics is made of ideas and thoughts expressed with symbols and logic. It is not the mathematics most people would think is, i.e., the mathematics that is composed of numbers, arithmetic, and applications in sciences. In a deeper sense we learn that mathematics isn't always so "certain". It depends. Especially when we learn how AC was assumed and used almost unconsciously by some mathematicians, formally brought up to discussion, been challenged and later accepted by most mathematicians. From the nature of AC, we also see that, for many mathematicians, mathematics doesn't always have to be "constructible" to produce good mathematics. This part of modern mathematics can be very counter-intuitive. One famous such example is the Banach-Tarski paradox. We thereafter explore the fundamentals necessary to AC − partially ordered set and its related theories.

In Chapter 3, we start with Cartesian product and see how it is related to AC. Then we study AC by exploring the relationship between Cartesian product and AC, studying the concept of "choice function", and eventually AC itself. We then studied some simpler forms of AC, i.e., the disjoint form, the power set form, and then Cartesian product. These simpler forms of AC don't really apply to more complicated cases in mathematics. So we study the more advanced forms in Chapter 4 − HMP, ZL, and WOT.

In Chapter 4, we prove the equivalences between AC, HMP, ZL, and WOT. The most difficult part in this chapter is to prove the Bourbarki-Witt Theorem. We use Bourbaki-Witt to prove that AC implies HMP and experience a cumbersome task. Instead, we could have used transfinite induction, which seems less complicated than the Bourbaki-Witt approach. Anyway, it is a good experience and we actually experience AC and its equivalent theorems. This gives us a better sense especially when we apply them in proofs.

In Chapter 5, we use the equivalents to AC to prove three theorems in set theory, linear analysis, and abstract algebra. Here we experience the very fundamental theorems we almost always assume true in linear analysis and abstract algebra. This gives us a feel of how set theory supports other branches of mathematics. It seems that they all find theoretical and logical sources of reason in set theory. This makes our effort writing this paper worthwhile, especially in understanding how other theories in mathematics are rooted in set theory.

This study brings a few points of interest in further studies. In experiencing how mathematics was expanded by strengthening its axiom systems, we see there are

other "mathematics" to study, i.e., the non-ZFC part of mathematics. We also see that mathematics is not "constant" or "certain". It is dynamic − it is expanding and growing. The reasons of this expansion and growth are in the studies in the foundation of mathematics. One such interest of study would be Kurt Gödel's and Paul Cohen's theories. In set theory itself, the nearest topics to study after this can be combinatorial set theory, measure theory, Borel and analytic sets, and models of set theory, etc. There are a lot more interesting topics in set theory to study, such as constructible sets, forcing, (very) large cardinals, etc. However, these topics of interest to study would take years in one's time in a graduate program if not self-learning.

Anyway, this paper serves as a beginning interest in set theory. We deal with the most fundamental ideas in mathematics in the beginning and end in applications in the very fundamental and important theorems in linear analysis and abstract algebra that we learned in undergraduate mathematics. Our work here accomplishes the goal of this paper although, if time permits, we would also like to prove the Thychonoff theorem in general topology using WOT and ZL.

REFERENCES

[1] Paul J. Cohen, *Set Theory and the Continuum Hypothesis*, The Benjamin/Cummings Publishing Company, Inc., Reading, Massachusetts, 1966.

[2] F. R. Drake and D. Singh, *Intermediate Set Theory*, John Wiley and Sons, 1996.

[3] David S. Dummit & Richard M. Foote, *Abstract Algebra*, $3^r d$ ed., John Wiley & Sons, Inc., Hoboken, NJ, 2004.

[4] Herbert Enderton, *Elements of Set Theory*, Academic Press, inc., New York, 1977.

[5] Abraham Fraenkel, Y. Bar-Hillel, and A. Levy, *Foundations of Set Theory*, Amsterdam: North-Holland, 2^{nd} edition, 1973.

[6] Derek Goldrei, *Classic Set Theory: For guided independent study*, 1^{st} ed., Chapman & Hall, London, UK, 1996.

[7] Seymour Hayden & John F.Kennison, *Zermelo-Fraenkel Set Theory*, Charles E. Merrill Publishing Company, Columbus, Ohio, 1968.

[8] Michael Hoffman, *Linear Analysis for Applications - Notes for MATH 502*, Cal. State Univ. - Los Angeles, September, 2013.

[9] M. Holz, K. Steffens, E. Weitz, *Introduction to Cardinal Arithmetic*, Basel Switzerland, Birkhäuser Verlag, 1999.

[10] Karel Hrbacek, Thomas Jech, *Introduction to Set Theory: Third Edition, Revised and Expanded*, Marcel Dekker, Inc., New York · Basel, 1999.

[11] Serge Lang, *Real and Functional Analysis*, 3^{rd} ed., Graduate Texts in Mathematics Vol. 142, Springer-Verlag Berlin Heidelberg, 1993.

[12] Kam-tim Leung & Doris L. Chen, *Elementary Set Theory, Part I/II*, Hong Kong University Press (printed by Condor Production Ltd.), Hong Kong, 1967.

[13] You-Feng Lin/Shwu-Yeng T. Lin, *Set Theory: An Intuitive Approach*, Houghton Mifflin Company, Boston, 1974.

[14] George Markowsky, *Chain-complete posets and directed sets with applications*, Algebra Universalis, **6**, No 1, (1976), 53-68.

[15] Elliot Mendelson, *Introduction to Mathematical Logic*, New York: Van Nostrand Reinhold, 1964.

[16] Judith Roitman, *Introduction to Modern Set Theory*, Revised Edition, 2011. Available at `http://galois.math.ku.edu/~roitman/stb3fullWeb.pdf`, or at `http://www.math.ku.edu/~roitman/`, 2011. (Earlier publication by John Wiley & Sons, Inc., 1990)

[17] A. Shen and N.K. Vereshchagin, *Basic Set Theory*, Student Mathematical Library Vol. 17, American Mathematical Society, 2002. (Translated by Shen from Russian)

[18] Standford Encyclopedia of Philosophy, *The Axiom of Choice*, 2015, `http://plato.stanford.edu/entries/axiom-choice/`

[19] Patrick Suppes, *Axiomatic Set Theory*, Dover Publications, Inc., New York, 1972.

[20] Thomas Jech, *Set Theory: The Third Millennium Edition, Revised and Expanded*, Springer-Verlag Berlin Heidelberg, 2003.

[21] Robert L. Vaught, *Set Theory: An Introduction*, 2^{nd} ed., Birkhäuser Boston, 1995.