Math 117: Deriving Set Theory from Axioms

John Douglas Moore

November 30, 2008

The foundations of set theory were laid by the mathematician Georg Cantor (1845-1918). His first article on the subject was published in Crelle's Journal of Mathematics in 1874. Cantor's work was so original that it engendered considerable controversy at first, but little by little Cantor's set theory was accepted because it became apparent that it was necessary for the foundations of calculus, the subject known by mathematicians as *real analysis*.

This course will present the foundations of real analysis based upon a naive version of set theory, which is close to the way that mathematicians think about sets. Using the theory of sets we will be able to define continuity of functions and show that a continuous function $f : [0, 1] \to \mathbb{R}$ achieves its maximum and minimum values at some points of $[0, 1]$. This gives a rigorous way of justifying procedures from calculus that are used to find maxima and minima of functions. Ultimately, set theory makes it possible to prove many of the key theorems of calculus, as well as theorems from differential equations and other branches of mathematics that are crucial for applications. The proofs are important because they explain exactly when the techniques one uses in an intuitive approach to calculus and these other subjects are valid, and can be applied without danger of error.

Informally, a *set* is a well-defined collection of objects. If a is one of the objects of A, we say that a is a *member* of A or an *element* of A, and we write $a \epsilon A$.

What do we mean by "well-defined" collection of objects? It is this question which was the source of much of the controversy. If we were to let A be the collection of all sets x such that $x \notin x$, in symbols

$$A = \{x : x \notin x\}, \tag{1}$$

we would have a serious problem:

$$A \in A \quad \Rightarrow \quad A \notin A, \qquad \text{but} \qquad A \notin A \quad \Rightarrow \quad A \in A.$$

Thus we would not be able to determine whether A is an element of itself. This difficulty has become known as the *Russell paradox*.

To avoid contradictions such as this, we must find some way of avoiding badly defined sets such as (1); in other words, we must restrict what kinds of statements can be used to define sets. The simplest approach would be to simply

say that (1) is not a well-defined set. But how does one specify which collections of objects are well-defined? A more careful treatment bases the theory of sets on a collection of axioms (together with logic), just like Euclidean geometry is based upon axioms. The most widely used axioms for set theory are the so-called *Zermelo-Fraenkel axioms*. A thorough presentation based upon these axioms can be found in [1].

1 The axioms for finite sets

The Zermelo-Fraenkel axioms are expressed in terms of a few undefined terms, the most important being sets and elements, which have an informal meaning but are not given formal definitions. We use the symbolism $x \epsilon A$ to denote that x is a member of the set A. Thus, for example, if $\mathbb{Z}$ is the set of integers, the symbolism $7 \in \mathbb{Z}$ means that 7 is an element which lies in the set of integers.

Our goals is to describe the Zermelo-Fraenkel axioms and sketch how they can be used to develop set theory:

Axiom of extension. *Two sets are equal if and only if they have the same elements.*

All of the other axioms simply give conditions under which sets exist.

Axiom of the null set. *There is a set which contains no elements.*

The axiom of extension implies this set is unique. We call it the *empty set* or *null set* and denote it by $\emptyset$.

Axiom schema of specification. *To every set A and every condition $P(x)$ that one could put on an element x of A, there is a set B consisting of exactly those elements $x \in A$ which satisfy $P(x)$.*

We denote this set by

$$B = \{x \in A : P(x)\}.$$

Properly speaking, the axiom schema of specification is an infinite collection of axioms, one for each choice of $P(x)$.

Definition. We say that A is a *subset* of B and write $A \subseteq B$ if

$$x \in A \quad \Rightarrow \quad x \in B.$$

Thus we introduce a new symbol $\subseteq$ to our lexicon. We can now state and prove a propostion.

Proposition. *Two sets A and B are equal if and only if*

$$A \subseteq B \quad \textit{and} \quad B \subseteq A.$$

Proof: Suppose that $A = B$. If $x \in A$ then $x \in B$, because A and B have the same elements by the axiom of extension. Hence $A \subseteq B$. Similarly, if $x \in B$ then $x \in A$, and hence $B \subseteq A$.

Conversely, suppose that $A \subseteq B$ and $B \subseteq A$. Then since $A \subseteq B$, $x \in A \Rightarrow x \in B$ and since $B \subseteq A$, $x \in B \Rightarrow x \in A$. Thus A and B have the same elements and by the axiom of extension, $A = B$.

Axiom of unions. *For every collection of sets, there is a set which consists of the elements which belong to at least one member of the collection.*

This set is called the *union* of the collection.

For example, if we have a collection consisting of the sets A and B, then the axiom of unions states that

$$A \cup B = \{x : x \in A \text{ or } x \in B\}$$

exists as a set. More generally, if $\mathcal{A} = \{A_\lambda : \lambda \in \Lambda\}$ is an arbitrary collection of sets, where Λ is an arbitrary index set, the axiom of unions states that

$$\bigcup_{\lambda \in \Lambda} A_\lambda = \{x : x \in A_\lambda \text{ for some } \lambda \in \Lambda\}$$

exists as a set.

We don't need a separate axiom to define the *intersection* of A and B by

$$A \cap B = \{x \in A \cup B : x \in A \text{ and } x \in B\},$$

because we can apply the axiom schema of specification with $P(x)$ being the statement "$x \in A$ and $x \in B$." Similarly, we can use the axiom schema of specification to define the *complement* of B in A,

$$A - B = \{x \in A : x \in A \text{ and } x \notin B\}.$$

With just the first four axioms, we can now prove many propositions and theorems regarding unions and intersections; for example:

Proposition. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

Proof: By the preceding Proposition it suffices to show that

$$A \cap (B \cup C) \subseteq (A \cap B) \cup (A \cap C) \quad \text{and} \quad (A \cap B) \cup (A \cap C) \subseteq A \cap (B \cup C).$$

To prove the first of these statements, we assume that $x \in A \cap (B \cup C)$. Then $x \in A$ and $x \in (B \cup C)$, so

$$x \in A \quad \text{and} \quad (x \in B \text{ or } x \in C)$$
$$\text{and hence} \quad (x \in A \text{ and } x \in B) \quad \text{or} \quad (x \in A \text{ and } x \in C).$$

Thus $x \in A \cap B$ or $x \in A \cap C$ or $x \in (A \cap B) \cup (A \cap C)$. The second statement is proven in a similar fashion.

Proposition. If A and B are subsets of X,

$$X - (A \cup B) = (X - A) \cap (X - B), \qquad X - (A \cap B) = (X - A) \cup (X - B).$$

Sketch of proof: If $x \in X - (A \cup B)$, then $x \in X$ and $x \notin A \cup B$. Thus x is neither an element of A nor of B. Hence x is both an element of $X - A$ and an element of $X - B$, so $x \in (X - A) \cap (X - B)$.

Conversely, if $x \in (X - A) \cap (X - B)$, then x is a member of both $X - A$ and a member of $X - B$. Thus $x \in X$ and x lies in neither A nor B. Hence x is not an element of $A \cup B$, so $x \in x \in X - (A \cup B)$.

This establishes that $X - (A \cup B) = (X - A) \cap (X - B)$. The other equality is proven in a similar fashion.

This proposition is a special case of the following more general proposition, the statement of which is known as the De Morgan laws:

Proposition. If $\{A_\alpha : \alpha \in A\}$ is a collection of subsets of X, then

$$X - \bigcup \{A_\alpha : \alpha \in A\} = \bigcap (X - A_\alpha) : \alpha \in A\},$$

$$X - \bigcap \{A_\alpha : \alpha \in A\} = \bigcup (X - A_\alpha) : \alpha \in A\},$$

You should have learned the technique for proving propositions like this in Math 8. See §5 of the text [3] for more examples of such propositions, including proofs.

For set theory to have some real content, we need a good supply of sets to work with. The next axiom enables us to construct lots of sets from the empty set.

Axiom of pairing. *If A and B are sets (not necessarily distinct), there is a set whose elements are A and B.*

In particular, this axiom implies that every set is an element of some other set. If $A \neq B$, we denote this set by $\{A, B\}$, while if $A = B$, we denote it by $\{A\}$. Starting with the empty set $\emptyset$, we can now form infinitely many distinct sets

$$\emptyset, \qquad \{\emptyset\}, \qquad \{\{\emptyset\}\}, \qquad \{\{\{\emptyset\}\}\}, \qquad \ldots,$$

and the axiom of unions allows us to construct many more from these. Indeed, we can now define zero and the *natural numbers*,

$$0 = \emptyset, \quad 1 = \{\emptyset\} = \{0\}, \quad 2 = \{\emptyset, \{\emptyset\}\} = \{0, 1\},$$
$$3 = \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\} = \{0, 1, 2\}, \ldots. \quad (2)$$

Note that the natural number n is a set consisting of exactly n elements. If A is a set, we let $A^+ = A \cup \{A\}$ and call it the *successor* to A. Then $0^+ = 1$, $1^+ = 2$, and so forth.

Axiom of powers. *For every set A, there is a set $\mathcal{P}(A) = 2^A$ whose elements are the subsets of A.*

Thus for example, if $A = \{1, 2, 3\}$, then

$$\mathcal{P}(A) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}.$$

Note that if A has n elements, then $\mathcal{P}(A)$ has 2^n elements.

2 Infinite sets

We need infinite sets in order to do calculus. The next axiom allows us to construct such sets.

Axiom of infinity. *There is a set which contains $\emptyset$ and which contains the successor of any of its elements.*

The axiom of infinity allows us to construct the set

$$\omega = \{0, 1, 2, 3, \ldots\} \quad \text{and hence its subset} \quad \mathbb{N} = \{1, 2, 3, \ldots\}.$$

If $n \in \mathbb{N}$, then n^+ is denoted by $n + 1$.

We can say that a set A is a *successor set* if it contains $\emptyset$ and contains the successor of any of its elements. Then the axiom of infinity just says that there is at least one successor set. Note that the intersection of all successor sets is just the set $\omega = \{0\} \cup \mathbb{N}$.

Theorem of Mathematical Induction. *Suppose that $P(n)$ be a statement which is either true or false for each $n \in \mathbb{N}$. If $P(1)$ is true and for each $n \in \mathbb{N}$,*

$$P(n) \text{ is true} \quad \Rightarrow \quad P(n+1) \text{ is true}, \tag{3}$$

then $P(n)$ is true for all $n \in \mathbb{N}$.

Here is a sketch of the proof: If we let

$$A = \{0\} \cup \{n \in \mathbb{N} : \ P(n) \text{ is true } \},$$

then (3) implies that A is a successor set, so it must include $\mathbb{N}$.

Thus the axioms of set theory imply the Principle of Mathematical Induction, one of the most useful techniques for proving theorems.

As will described in more detail in the next section, we can think of inclusion as an ordering $\leq$ on ω or on $\mathbb{N}$:

$$m \leq n \quad \Leftrightarrow \quad m \subseteq n, \qquad \text{for } m, n \in \mathbb{N}.$$

One can then check that

$$\begin{aligned} m \leq m, \quad k \leq m \quad \text{and} \quad m \leq n \quad &\Rightarrow \quad k \leq n, \\ k \leq m \quad \text{and} \quad m \leq k \quad &\Rightarrow \quad k = n. \end{aligned} \tag{4}$$

We can let $P(n)$ be the statement: If S is any nonempty subset of $\{1, 2, \ldots, n\}$, then there is an element $k \in S$ such that $k \leq m$ for all $m \in S$. with a little effort, one can then show that $P(1)$ is true and that for each $n \in \mathbb{N}$,

$$P(n) \text{ is true} \quad \Rightarrow \quad P(n+1) \text{ is true,}$$

so it follows from the Theorem of Mathematical Induction that $P(n)$ is true for every $n \in \mathbb{N}$. Using this fact one can prove:

Well-ordering Theorem. *If S is any nonempty subset of $\mathbb{N}$, then there is an element $k \in S$ such that $k \leq m$ for all $m \in S$.*

In other words, each subset of $\mathbb{N}$ has a smallest element. Note that Lay [3] chooses to regard this as an axiom (see page 100), but in axiomatic set theory, the well-ordering theorem is regarded as a consequence of the axiom of infinity.

3 Partial orders and functions

In set theory, we define functions in terms of ordered pairs. We briefly sketch how this goes, referring to §6 and §7 of the text [3] for more details. The *ordered pair* of a and b can be defined to be the set

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

Proposition. *Two ordered pairs (a, b) and (c, d) are equal if and only if $a = c$ and $b = d$.*

For the proof see page 50 of [3]. If A and B are sets, we define the *Cartesian product* of A and B to be the set

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

A *relation* between A and B is a subset $R \subseteq A \times B$. Suppose that R is a relation. We then say that a is *related* to b by R, and write aRb, if $(a, b) \in R$. There are two important examples of relations, partial orders and functions.

Definition. A *partial order* on A is a relation $\leq$ lying within $A \times A$ such that

1. if $a \leq b$ and $b \leq a$, then $a = b$, and
2. if $a \leq b$ and $b \leq c$, then $a \leq c$,

for all $a, b, c \in A$. If in addition, $a \leq b$ or $b \leq a$, for all $a, b \in A$, we say that the ordering is a *total order*. A total order is said to be a *well-ordering* if in addition, every nonempty subset of A has a smallest element. If $\leq$ is a partial order on S, we write $a < b$ to mean $a \leq b$ and $a \neq b$.

Thus, for example, if X is any set, we can define a partial order on $\mathcal{P}(X)$ by

$$A \leq B \quad \Leftrightarrow \quad A \subseteq B.$$

The usual order on the set of integers $\mathbb{Z}$ is a total order. It follows from (4) and the well-ordering theorem that he usual order on the set $\mathbb{N}$ is a well-ordering.

Definition. A *function between* A and B is a relation $f \subseteq A \times B$ such that

$$(a, b), (a, b') \in f \quad \Rightarrow \quad b = b'.$$

For example, if $A = B = \mathbb{R}$, then the relation

$$R = \{(x, y) \in \mathbb{R} \times \mathbb{R} : x^2 + y^2 = 1\}$$

is not a function because $(0, -1)$ and $(0, 1)$ are both elements of R. However,

$$f = \{(x, y) \in \mathbb{R} \times \mathbb{R} : x^2 + y^2 = 1, y \geq 0\} \tag{5}$$

is a function.

We define the *domain* of a function f to be

$$\mathrm{dom}(f) = \{a \in A : (a, b) \in f \text{ for some } b \in B \}.$$

Thus, for example, the domain of the function f defined by (5) is

$$[-1, 1] = \{x \in \mathbb{R} : -1 \leq x \leq 1\}.$$

If $f \subseteq A \times B$ is a function and $\mathrm{dom}(f) = A$, we say that f is a function *from* A to B, and we write $f : A \to B$. We also use the notation

$$f(a) = b \quad \text{for} \quad (a, b) \in f.$$

Given a function $f : A \to B$, we say that

$$f(A) = \{b \in B : \text{ there exists } a \in A \text{ with } f(a) = b \}$$

is the *image* of f. More generally, if $C \subseteq A$ and $D \subseteq B$,

$$f(C) = \{b \in B : f(c) = b, \text{ for some } c \in C\},$$
$$f^{-1}(D) = \{a \in A : f(a) = d, \text{ for some } d \in D\}.$$

If $f : A \to B$ and $g : B \to C$ are functions, we can define

$$g \circ f : A \to C \quad \text{by} \quad (g \circ f)(a) = g(f(a)).$$

Thus, for example, if

$$f(x) = x^2 + 1 \quad \text{and} \quad g(y) = \sin y, \qquad (g \circ f)(x) = \sin(x^2 + 1).$$

Definition. A function $f : A \to B$ is *injective* if $f(a) = f(a') \Rightarrow a = a'$. It is *surjective* if given any $b \in B$ there is an $a \in A$ such that $f(a) = b$. It is *bijective* if it is both injective and surjective.

If $f : A \to B$ is a bijective function, it is possible to define the inverse function $f^{-1} : B \to A$ by

$$a = f^{-1}(b) \quad \Leftrightarrow \quad f(a) = b.$$

For example, if

$$f : \mathbb{R} \to (0, \infty) = \{x \in \mathbb{R} : 0 < x\} \quad \text{by} \quad f(x) = e^x,$$

then the inverse function

$$f^{-1} : (0, \infty) \to \mathbb{R} \quad \text{is given by} \quad f^{-1}(y) = \log y.$$

We can define the identity function $\mathrm{id}_A : A \to A$ by $\mathrm{id}_A(a) = a$. When $f : A \to B$ is a bijective function, $f^{-1} \circ f = \mathrm{id}_A$ and $f \circ f^{-1} = \mathrm{id}_B$.

Functions can often be defined by means of the Theorem of Mathematical Induction. This is called *mathematical recursion*.

Recursion Theorem. *Suppose that $f : X \to X$ is a function and a is an element of X Then there exists a function $u : \mathbb{N} \to X$ such that*

$$u(1) = a, \qquad u(n^+) = f(u(n)), \quad \text{for } n \in \mathbb{N}.$$

Idea of proof (see [1] §12): Let $\mathcal{C}$ denote the collection of subsets A of $\mathbb{N} \times X$ such that

$$(1, a) \in a, \qquad (n, x) \in A \quad \Rightarrow \quad (n^+, f(x)) \in A, \quad \text{for } n \in \mathbb{N}.$$

Clearly $\mathcal{C}$ is nonempty, and we let

$$u = \bigcap \{A : A \in \mathcal{C}\}.$$

We then show that $u : \mathbb{N} \to X$ is a function with the desired properties.

4 Countable and uncountable sets

Definition. Two sets A and B are *equinumerous* or have the *same cardinality* if there is a bijective function $f : A \to B$. In this case, we write $A \sim B$ or $|A| = |B|$.

Definition. A set is *finite* if it has the same cardinality as the empty set or the set $\{1, 2, \ldots, n\}$, for some $n \in \omega$; otherwise it is *infinite*. A set is *countably infinite* if it has the same cardinality as $\mathbb{N}$. A set is *countable* if it is finite or countably infinite; otherwise it is *uncountable*.

We agree to write $|A| = 0$ if A is the null set, $|A| = n$ if A has the same cardinality as the set $\{1, 2, \ldots, n\}$, and $|A| = \aleph_0$ is A is countably infinite. We write $|A| \leq |B|$ if A has the same cardinality as a subset of B. We write $|A| < |B|$

if $|A| \leq |B|$, but $|A| \neq |B|$, and say that A has strictly smaller cardinality than B in this case.

For example, if $\mathbb{Z}$ is the set of integers, we can define a bijection $f : \mathbb{N} \to \mathbb{Z}$ by

$$f(1) = 0, \quad f(2) = 1, \quad f(3) = -1, \quad f(4) = 2, \quad f(5) = -2, \quad \ldots.$$

Thus $|\mathbb{N}| = |\mathbb{Z}|$.

Remark. It might seem intuitively clear that

$$|A| \leq |B| \quad \text{and} \quad |B| \leq |A| \quad \Rightarrow \quad |A| = |B|.$$

In fact, this statement is true, but surprisingly difficult to prove. This result is known as the Schröder-Bernstein Theorem and is proven in §22 of [1].

One of the most revolutionary of Cantor's theorems on cardinality is:

Theorem. *If A is an infinite set, A does not have the same cardinality as $\mathcal{P}(A) = 2^A$.*

Proof: We prove this by contradiction. Suppose that $f : A \to \mathcal{P}(A)$ is a bijection. Let

$$S = \{x \in A : x \notin f(x)\}.$$

We claim that $S \notin f(A)$ and hence f cannot be surjective.

Indeed, if $S = f(y)$ for some $y \in A$, then

$$y \in S \quad \Rightarrow \quad y \notin f(y) \quad \Rightarrow \quad y \notin S, \qquad y \notin S \quad \Rightarrow \quad y \in f(y) \quad \Rightarrow \quad y \in S,$$

so we obtain a contradiction. Thus the assumption that $f : A \to \mathcal{A}$ is a bijection must be false, and A does not have the same cardinality as $\mathcal{P}(A) = 2^A$.

In particular, $\mathcal{P}(\mathbb{N}) = 2^{\mathbb{N}}$ is uncountable. Note how similar the argument of this theorem is to the construction of Russell's paradox. No wonder some mathematicians had difficulty accepting Cantor's arguments at first!

It follows from Cantor's theorem that if A is an infinite set,

$$|A| < |\mathcal{P}(A)| < |\mathcal{P}(\mathcal{P}(A))| < |\mathcal{P}(\mathcal{P}(\mathcal{P}(A)))| < \cdots.$$

Theorem. *Any subset of $\mathbb{N}$ is countable.*

Proof: Suppose that $S \subset \mathbb{N}$. We can assume that S is infinite; otherwise there is nothing to prove. We define a function $f : \mathbb{N} \to S$ as follows. By the well-ordering theorem S has a smallest element, $f(1)$. Having constructed $f(1), \ldots, f(k)$, we let $f(k+1)$ be the least element of

$$S - \{f(1), \ldots, f(k)\}.$$

In this way, we obtain a function $f : \mathbb{N} \to S$ which is injective by construction. If $n \in S$, then there is an $m \in \mathbb{N}$ with $m < n$ such that n is the smallest element of

$$S - \{f(1), \ldots, f(m)\}.$$

But then $n = f(m+1)$, so f is surjective.

Corollary. *Any subset of a countable set is countable.*

Proof: Suppose that T is a countable set and $S \subset T$. We can assume that T is infinite, otherwise the corollary is trivial. Thus there is a bijection $g : T \to \mathbb{N}$ and g restricts to a bijection from S to $g(S)$. But then $|S| = |g(S)|$ and $g(S)$ is countable by the theorem, so S must be countable.

Theorem. *If A and B are countably infinite, then so is their Cartesian product $A \times B$.*

Once we have the natural numbers $\mathbb{N}$, we could prove some theorems about such numbers, such as the unique factorization of a natural number into primes. Suppose that we have done that. Then one proof of this theorem would go like this: If A and B are countably infinite, then $A \times B$ is clearly infinite, and their exist bijections

$$f : A \to \mathbb{N}, \qquad g : B \to \mathbb{N}.$$

Define

$$h : A \times B \to \mathbb{N} \quad \text{by} \quad h(a,b) = 2^{f(a)}3^{g(b)}.$$

By unique factorization of integers, we see that h is injective. Thus $|A \times B| = |h(A \times B)|$, a subset of $\mathbb{N}$. Therefore it follows from the previous theorem that $A \times B$ is countable.

Theorem. *The set $\mathbb{Q}$ of rational numbers is countably infinite.*

Sketch of proof: Any rational number x can be written as a ratio $x = m/n$, where $m \in \mathbb{Z}$ and $n \in \mathbb{N}$, where we can assume that m and n have no common factors. (We agree to write $0 = 0/1$.) This representation is unique, so we have an injection

$$f : \mathbb{Q} \to \mathbb{Z} \times \mathbb{N} \quad \text{defined by} \quad f(m/n) = (m,n).$$

Thus $|\mathbb{Q}| = |f(\mathbb{Q})|$, where $f(\mathbb{Q})$ is a subset of the countably infinite set $\mathbb{Z} \times \mathbb{N}$, and hence $\mathbb{Q}$ is countably infinite.

Theorem. *The set $\mathbb{R}$ of real numbers is uncountable.*

Sketch of proof: Note that if $(0,1) = \{x \in \mathbb{R} : 0 < x < 1\}$, the function

$$f : (0,1) \to \mathbb{R}, \quad \text{defined by} \quad f(x) = \tan\left(\pi x - \frac{\pi}{2}\right),$$

is bijective. Therefore, it suffices to show that the "open interval" $(0,1)$ is uncountable.

We argue by contradiction. Suppose that $f : \mathbb{N} \to (0,1)$ were a bijection. Then since every element in $(0,1)$ has a decimal expansion, we could write

$$\begin{array}{rcl} f(1) & = & 0.a_{11}a_{12}a_{13}\cdots, \\ f(2) & = & 0.a_{21}a_{22}a_{23}\cdots, \\ f(3) & = & 0.a_{31}a_{32}a_{33}\cdots, \\ \vdots & \vdots & \qquad \cdots, \end{array}$$

where each $a_{ij} \in \{0, 1, \ldots, 9\}$. We could then let

$$y = 0.b_1b_2b_3 \cdots, \quad \text{where} \quad b_i = \begin{cases} 1 & \text{if } a_{ii} \neq 1, \\ 0 & \text{if } a_{ii} = 1, \end{cases}$$

and y would not be in the image of f, contradicting the fact that f was assumed to be surjective. (This is called Cantor's diagonalization argument.)

We write $|A| = c$ if A has the same cardinality as $\mathbb{R}$. To summarize,

$$|\mathbb{N}| = |\mathbb{Z}| = |\mathbb{Q}| = \aleph_0 < |\mathbb{R}| = c.$$

Theorem. *The set $\mathbb{R} \times \mathbb{R}$ of ordered pairs of real numbers has the same cardinality as the set $\mathbb{R}$ of real numbers.*

Sketch of proof (following [2], page 997): Since $A = \{x \in \mathbb{R} : 0 < x < 1\}$ and $\mathbb{R}$ have the same cardinality and $A \times A$ has the same cardinality as $\mathbb{R} \times \mathbb{R}$, it suffices to show that $A \times A$ has the same cardinality as A.

Let $(x, y) \in A \times A$, and write x and y as infinite decimal expansions in which we replace an infinite sequence of 0's by an infinite sequence of 9's so that the decimal representation is unique. We now divide the digits in the expansions of x and y into groups, each group ending with the first nonzero integer that is encountered. For example,

$$x = .4\ 003\ 02\ 6\ 7\ 0004\ \cdots, \qquad y = .06\ 2\ 007\ 4\ 09\ \cdots,$$

where neither expansion can have a nonending sequence of zeros. We then let $\psi(x, y) = z \in A$, where z is constructed from x and y by taking the first group from x, then the first group from y, then the second group from x, then the second group from y, and so on:

$$z = \psi(x, y) = .4\ 06\ 003\ 2\ 02\ 007\ \cdots.$$

Given $z \in A$, one can construct elements $x, y \in A$ by reversing this process. We thus check that ψ is the desired bijection.

We remark that it follows from the above theorem that if $\mathbb{C}$ is the the set of complex numbers, then $|\mathbb{C}| = c$.

5 The axiom of choice

The axioms we have described so far are essentially those introduced by Zermelo in 1908. They form the basis for Zermelo set theory, often denoted by Z. For more advanced results in the cardinality of sets, we need an additional axiom, the axiom of choice:

Axiom of choice. *Let $\mathcal{C}$ be a collection of nonempty sets. Then there is a function*

$$f : \mathcal{C} \longrightarrow \bigcup \{X : X \in \mathcal{C}\}$$

such that $f(X) \in X$ for every $X \in \mathcal{C}$.

It can be proven that the following special case is actually equivalent to the axiom of choice:

Special Case. *Let X be any set. Then there is a function*

$$f : \mathcal{P}(X) - \{\emptyset\} \longrightarrow X$$

such that $f(A) \in A$ for every $A \in \mathcal{P}(X) - \{\emptyset\}$.

Such a function is called a *choice function*. This theorem is often used informally in mathematics, sometimes without explicit mention. Here is a theorem which uses the axiom of choice in its proof:

Theorem. *Any infinite set contains a countable subset.*

We sketch the argument following [1], §15, where one will find a more detailed presentation. Suppose that X is an infinite set and let $f : \mathcal{P}(X) - \{\emptyset\} \to X$ be a choice function for $\mathcal{P}(X) - \{\emptyset\}$. We then define a function $g : \mathbb{N} \to X$ as follows: Let

$$g(1) = f(X) = x_1, \quad g(2) = f(X - \{x_1\}) = x_2, \quad \ldots,$$
$$g(n) = f(X - \{x_1, \ldots, x_{n-1}\}) = x_n, \cdots .$$

(Each set $X - \{x_1, \ldots, x_{n-1}\}$ is nonempty because X is finite.) We then check that g is injective. If B is the image of g, then B is a countable subset of X.

Here is another theorem which uses the axiom of choice, but not in such an obvious way:

Theorem. *The union of a countable collection of countable sets is countable.*

Proof: It is relatively easy to prove that a finite union of countable sets is countable, so we can suppose that our countable collection is $\mathcal{A} = \{A_n : n \in \mathbb{N}\}$. We need to show that

$$S = \bigcup \{A_n : n \in \mathbb{N}\}$$

is countable. Since subsets of countable sets are countable, it suffices to construct an injective function $f : S \to \mathbb{N} \times \mathbb{N}$. Since each A_λ is countable, there is an injective function $g_n : A_n \to \mathbb{N}$. (It is in choosing countably many injections that we use the axiom of choice.) We define $f : S \to \mathbb{N} \times \mathbb{N}$ as follows: If $a \in S$, there is a smallest $n \in \mathbb{N}$ such that $a \in A_n$. We define

$$f : S \to \mathbb{N} \times \mathbb{N} \quad \text{by} \quad f(a) = (n, g_n(a)) \in \mathbb{N} \times \mathbb{N}.$$

If $f(a) = f(b)$, then both a and b must lie in the same set A_n and $g_n(a) = g_n(b)$. Since g_n is injective, $a = b$. We thus conclude that f is injective, which finishes the proof.

For another proof of this theorem, see [3], page 83.

Definition. A element $x \in \mathbb{R}$ is *algebraic* if it is a root of some polynomial

$$P(x) = a_n x^n + \cdots + a_1 x + a_0, \quad \text{where} \quad a_n, \ldots, a_1, a_0 \in \mathbb{Z}. \tag{6}$$

If it is not algebraic it is said to be *transcendental*.

Thus, for example, $\sqrt{2}$ is an algebraic since it is a root of the polynomial $P(x) = x^2 - 2$.

We now give a proof due to Cantor (see [2], pages 996-997) that the set of algebraic numbers is countable. We say that the polynomial $P(x)$ defined by (6) has *degree* n and has *index*

$$n + |a_n| + \cdots + |a_1| + |a_0|.$$

If we let $\mathcal{P}$ denote the collection of all polynomials with integer coefficients, then

$$\mathcal{P} = \bigcup \{\mathcal{P}_m : m \in \mathbb{N}\},$$

where $\mathcal{P}_m$ is the collection of polynomials which have index m. As an exercise one can show that each $\mathcal{P}_m$ is finite. It then follows from the preceding theorem that $\mathcal{P}$ is countable. Using the algebraic fact that any polynomial of degree n has at most n roots, one can then conclude that the set $\mathbb{A}$ of algebraic numbers is countable.

Since $\mathbb{R}$ is uncountable, there are lots of transcendental numbers. It is usually harder to show that a given number is transcendental; Hermite proved that e is transcendental in 1873, while Lindemann proved that π is transcendental in 1882 (see [2] for a discussion of the history).

6 Ordinal numbers*

Fraenkel extended Zermelo's axioms for set theory by introducing two additional axioms in 1922:

Axiom schema of substitution. *If $P(x, y)$ is a statement such that for each x in some set A, $\{y : P(x, y)\}$ is a set, then there is a function f with domain A such that $f(x) = \{y : P(x, y)\}$, for $x \in A$.*

This is actually a strengthening of the axiom schema of specification, so we could in fact eliminate the axiom schema of specification from the collection of axioms.

Axiom of regularity. *Any nonempty set A contains an element x such that $\{x\} \cap A = \emptyset$.*

This axiom is also known as the axiom of foundation. It implies that there is no set B such that $B \in B$. Indeed if there were such a set, we could let $A = \{B\}$. Then A contains only one element B and it is not the case that $\{B\} \cap A = \emptyset$.

The axiomatization of set theory obtained from the axioms introduced so far, except for the axiom of choice, is called Zermelo-Fraenkel set theory, and is

denoted by ZF. If the axiom of choice is included, the resulting theory is denoted by ZFC. It is ZFC that most mathematicians regard as the foundation for all of mathematics. The axiom of choice is the most controversial of the axioms, and a small minority of mathematicians would prefer to develop mathematics (insofar as possible) using ZF rather than ZFC.

The two new axioms are useful in extending the process of counting beyond ω to "transfinite numbers."

Suppose that S is a well-ordered set and let $x \in S$. The *initial segment* of S determined by x is

$$x_< = \{y \in S : y < x\}.$$

Recall that $y < x$ means that $y \leq x$ and $y \neq x$. With this preparation in place, we can now give von Neumann's definition of ordinal number.

Definition. An *ordinal number* is a set S well-ordered by inclusion (so that $\leq = \subseteq$) such that for each $x \in S$, $x_< = x$.

For example, zero and the natural numbers defined by (2) are ordinal numbers when they are ordered by inclusion. Thus if

$$3 = \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\} = \{0, 1, 2\},$$

the initial segment determined by 2 is just

$$2_< = \{0, 1\} = \{\emptyset, \{\emptyset\}\} = 2.$$

Another ordinal number we have already encountered is $\omega = \{0\} \cup \mathbb{N}$. Yet other ordinal numbers can be constructed by taking successors of ω,

$$\omega + 1 = \omega^+, \omega + 2 = (\omega^+)^+, \cdots .$$

Without the axiom schema of substitution, it is not possible to construct the set

$$\omega \cdot 2 = \omega + \omega = \{0, 1, 2, \ldots, \omega, \omega + 1, \omega + 2, \ldots\},$$

in spite of the fact that for each $n \in \omega$, we can construct the set

$$y = \omega + n = \{0, 1, 2, \ldots, \omega, \omega + 1, \omega + 2, \ldots, \omega + n\}.$$

Until we know that $\{\omega + n : n \in \omega\}$ is a set, we cannot construct $\omega + \omega$ by the axiom of unions. Let $P(n, y)$ be the statement $y = \omega + n$. Applying the axiom schema of substitution, we can then construct a function f with domain ω such that $f(n) = \omega + n$, for each $n \in \omega$. The range of this function is then the desired set $\omega \cdot 2$, a set which one can then show is an ordinal number.

From here, we can construct yet other ordinals

$$(\omega \cdot 2) + 1 = (\omega \cdot 2)^+, (\omega \cdot 2) + 2 = ((\omega \cdot 2)^+)^+, \cdots, \omega \cdot \omega = \omega^2, \ldots,$$

and so forth. This process goes on forever.

Once one has the ordinal numbers, one can extend the principle of mathematical induction.

Theorem of Transfinite Induction. *Suppose that $P(\alpha)$ be a statement which is either true or false for each ordinal number α. If $P(0)$ is true and for each ordinal number α,*

$$P(\beta) \text{ is true whenever } \beta < \alpha \quad \Rightarrow \quad P(\alpha) \text{ is true}, \tag{7}$$

then $P(\alpha)$ is true for every ordinal α.

To prove this, we simply note that if $P(\alpha)$ were not true for some ordinal α, let

$$A = \{\beta \in \alpha : \ P(\beta) \text{ is not true}\}.$$

Since α is well-ordered, A has a least element β and for this element (7) is contradicted.

One can extend ordinals indefinitely, obtaining an ordinal with the same cardinality as $\mathbb{R}$, for example. Since any ordinal number is well-ordered, there is a first uncountable ordinal, which is useful for constructing examples in topology. However, the collection of all ordinals does not exist as a set. Originally, this was thought of as a paradox:

Burali-Forti Paradox. *There is no set which has all the ordinals as its elements.*

Indeed, if there were such a set Ω it would be well-ordered by the Theorem of Transfinite Induction, and would therefore be an ordinal itself. But then $\Omega \in \Omega$, contradicting the axiom of regularity. One says that the collection of all ordinals forms a *class* but not a set.

One can define functions on the class of all ordinals by the process of *transfinite induction*. The idea is the same as for ordinary recursion. Given a set x_0 and a method of defining x_α for any ordinal α in terms of x_β for ordinals β such that $\beta < \alpha$, one can define a *transfinite sequence* (x_α). This could be regarded as a function from the class of all ordinals to the class of all sets, except that it goes from one class to another class, not from one set to another set.

7 Zorn's Lemma*

Suppose that X is a set with a partial order $\leq$. A *chain* in X is a subset of X which is totally ordered.

Zorn's Lemma. *If every chain in X has an upper bound, then X has a maximal element.*

Here is a rough idea of how Zorn's lemma could be proven from the axiom of choice (following the Wikipedia article on Zorn's lemma): Suppose that Zorn's

lemma is false. Then there is a partially ordered set $(X, \leq)$ with no maximal element such that every chain in X has an upper bound. If T is a chain in X, we can let $f(T)$ be an element of X larger than the the upper bound of T. Such an element exists by the axiom of choice.

Using the function f, we can now construct a transfinite sequence in X as follows. We choose x_0 to be an arbitrary element of X. If α is an ordinal number and we have defined x_β for all $\beta < \alpha$ in such a way that $\{x_\beta : \beta < \alpha\}$ is a chain, we can define

$$x_\alpha = f\left(\{x_\beta : \beta < \alpha\}\right) \quad \text{and} \quad \{x_\beta : \beta \leq \alpha\}$$

is also a chain. This process continues indefinitely and exhibits the class of all ordinals as a subset of X. This implies that the class of all ordinals is a set which contradicts the Burali-Forti paradox.

Zorn's lemma is actually the way in which the axiom of choice is applied in many branches of mathematics. For example, we would like to show that any vector space has a basis. This is a standard theorem from the theory of finite-dimensional vector spaces, but what about infinite-dimensional vector spaces?

Theorem. *Any vector space has a basis, even if it is infinite-dimensional.*

Idea of proof: Let V be a vector space and consider the collection $\mathcal{C}$ of all subsets B of V such that the elements of B are linearly independent, partially ordered by inclusion. If $\{B_\alpha : \alpha \in A\}$ is a chain in $\mathcal{C}$, then $\bigcup\{B_\alpha : \alpha \in A\}$ is an upper bound of the chain. Hence $\mathcal{C}$ must have a maximal element and that maximal element is a basis.

Remark. It can be shown that in the presence of the other axioms introduced so far, the axiom of choice is equivalent to Zorn's lemma. It is also equivalent to the following theorem:

Well-ordering Theorem. *Every set X can be well-ordered.*

Sketch of proof (following [1], §17): Consider the collection $\mathcal{C}$ of all pairs, consisting of a subset A of X together with a well-ordering $\leq$ of the subset. Given two such elements $(A, \leq)$ and $(B, \leq')$ of $\mathcal{C}$, we let

$$(A, \leq) \leq (B, \leq') \quad \Leftrightarrow \quad A \subseteq B \quad \text{and} \quad \leq' \text{ extends } \leq.$$

One then checks that given any chain in $\mathcal{C}$, the union of the chain possesses a well-ordering which makes it into an upper bound. Thus Zorn's lemma implies that $\mathcal{C}$ must contain a maximal element. This maximal element must be a pair $(A_m, \leq)$ where $\leq$ is a well-ordering of A_m. If A_m were not equal to X one could extend the well-ordering to a larger subset of X, contradicting maximality.

8 Consistency of the axioms*

Of the axioms introduced so far, it is the axiom of choice which is most controversial. However, it would be difficult to prove many of the familiar theorems

needed for analysis without the axiom of choice, or equivalently, Zorn's lemma. Fortunately, it was proven by Kurt Gödel (1940) that if ZF (Zermelo-Fraenkel set theory without the axiom of choice) is consistent (that is does not lead to any contradictions), so is ZFC (Zermelo-Fraenkel set theory with the axiom of choice). Thus most mathematicians are comfortable with accepting the axiom of choice as one of the axioms. It was later proven by Paul Cohen (1962) that the axiom of choice is not implied by the axioms of ZF.

In addition to proving that the set of real numbers is uncountable, Georg Cantor proposed the *continuum hypothesis*: There is no set whose cardinality is strictly larger than that of the natural numbers and strictly less than that of the real numbers. Kurt Gödel showed that if ZFC is consistent, then it remains consistent when the continuum hypothesis is added. Paul Cohen showed that the continuum hypothesis is not implied by the axioms of ZFC.

Given the axioms for set theory, one can build up the real numbers from the theory of sets. This is done in stages. From the natural numbers $\mathbb{N}$ one constructs the integers $\mathbb{Z}$, from the integers $\mathbb{Z}$ one then constructs the rational numbers $\mathbb{Q}$, and from the rational numbers $\mathbb{Q}$ one finally constructs the real numbers $\mathbb{R}$. After that, one must define addition and multiplication so that the usual rules of arithmetic for $\mathbb{R}$ are satisfied.

Needless to say, this is an arduous process, and we will not carry it out in this course. Instead, we will use a collection of axioms for the real numbers themselves as the foundation for analysis.

References

[1] Paul R. Halmos, *Naive set theory*, Springer, New York, 1998.

[2] Morris Kline *Mathematical thought from ancient to modern times*, Oxford University Press, New York, 1972.

[3] Steven R. Lay, *Analysis: with an introduction to proof*, Pearson Prentice Hall, Upper Saddle Riven, NJ, 2005.