CARDINALS AND ORDINALS

LARRY SUSANKA

Abstract. This is an appendix for a book I have (mostly) written on measure theory. It constitutes a *very* basic introduction to set theory. It contains some of the material one needs to know to understand basic set-theoretic arguments commonly employed in analysis.

It is essentially self-contained, though it does reference, in a few places, material of the rest of the text. I included Chapter One "Some Preliminaries" from the text which contains most of these references as well as "first pass" discussions of the first infinite ordinal and the Axiom of Choice.

Contents

Date: October 18, 2023.

Some Preliminaries October 18, 2023

1. Functions

A **relation**, or more precisely a **binary relation**, is a nonempty subset of some product set

$$\boldsymbol{S \times T} = \{\, (s,t) \mid s \in S \text{ and } t \in T \,\},$$

the set of **ordered pairs** formed from nonempty sets S and T.

The **domain** of a relation f consists of the set of first components of any member of f, while the **range** consists of the set of second components, and these sets can be denoted **Domain(f)** and **Range(f)** respectively.

Relations are used to model many different ideas, but three basic kinds of relations will be of interest over the next few sections. The first of these is the familiar concept of "function."

A **function** f (sometimes also called a **map**) from S to T, described by $\boldsymbol{f\colon S \to T}$, is a relation as above with domain S and range contained in (not necessarily equal to) T and which has the following property:

$$(a,b) \ \in \ f \text{ and } (a,c) \ \in \ f \Rightarrow b = c.$$

These properties insure that there is one and only one ordered pair in f having a as first coordinate for every member a of S. The range of a function is, in many sources, called the **image** of that function.

Note that the concept of function is domain dependent: the same set of pairs thought of as a subset of $R \times T$, where S is contained in but not equal to R, won't be a function. And there is a certain latitude with regard to T: it can be replaced in the description $f\colon S \to T$ by any set containing $Range(f)$.

Often, though, it is T itself under study and how $Range(f)$ sits (or could sit) in T reflects important information about T. In that context T will be called the **co-domain** of function f.

If f is a function, the notation $f(a)$ or f_a is used for $b \in T$ when $(a,b) \in f$. Occasionally a function will be said to **index** its range, and in this case the range is said to be **indexed by** the domain, whose members are called **indices**.

If A is a set[1] and $f\colon S \to T$ we define

$$\boldsymbol{f(A)} = \{\, f(s) \mid s \in A \cap S \,\} \quad \text{and} \quad \boldsymbol{f^{-1}(A)} = \{\, s \in S \mid f(s) \in A \,\}.$$

Both sets can be empty. If $f^{-1}(\{t\})$ contains at most a single member of S for each $t \in T$ we call f **one-to-one** and if $f(S) = T$ we say f is **onto** T.

If f is one-to-one and onto T then f can be used to construct a function $\boldsymbol{f^{-1}\colon T \to S}$ by defining $f^{-1}(t)$, for each $t \in T$, to be that member s of S with $f(s) = t$. This **second** definition of f^{-1} is an abuse of notation that could cause ambiguity in case, for example, both t and $\{t\}$ are members of T.

[1]We presume here that the set A is not actually an element of domain or range of function f.

If f is one-to-one but not onto T then f cannot be used to define f^{-1} as a function from T to S. However f^{-1} would be a function thought of as the set of pairs $\{\,(f(s), s) \mid s \in S\,\}$ in $f(S) \times S$.

The **restriction** of a function $f\colon S \to T$ to a nonempty set $A \subset S$ is denoted $\boldsymbol{f|_A}$ and defined to be $\{\,(a, b) \in f \mid a \in A\,\}$. $f|_A$ is a function with domain A. If g and f are functions and $f \subset g$ then $g|_S = f$ and g is called an **extension** of f.

Here are a few more items of notation:

The set of functions with domain S and co-domain T is denoted $\boldsymbol{T^S}$.

If T is the two element set $\{0, 1\}$, T^S will sometimes be denoted $\boldsymbol{2^S}$.

The collection of all subsets of a set S form a set denoted $\boldsymbol{\mathbb{P}(S)}$, called the **power set** of S.

If A and X are any sets, the notation $\boldsymbol{X - A} = \{\,x \in X \mid x \notin A\,\}$ is used. $X - A$ is called the **complement of A in X**.

If A and X are nonempty sets and $f\colon A \to \mathbb{P}(X)$ the notation $\boldsymbol{\bigcup_{a\in A} f_a}$ denotes $\{\,x \in X \mid x \in f_a \text{ for some } a \in A\,\}$. The notation $\boldsymbol{\bigcap_{a\in A} f_a}$ denotes $\{\,x \in X \mid x \in f_a \text{ for every } a \in A\,\}$.

Sets S and T with similar properties can arise from different sources. Recognizing that two sets are essentially the same in some way often comes through the presentation of a one-to-one function $g\colon S \to T$ that is onto T.

When we have this in mind, we will say that S and T are **identified** and that g identifies the element $s \in S$ with the element $g(s) \in T$. The notation $\boldsymbol{s \leftrightarrow g(s)}$ can be used to illustrate such an identification. These identifications can range in utility from a trivial convenience to something more substantial, a shift in context.

For instance, on the trivial side, if n is a positive integer, the set $\{1, \ldots, n\}$ can be identified with $\{0, \ldots, n-1\}$ through the function described by $x \leftrightarrow x-1$. More substantial examples of this vocabulary in action follow.

2^S can be identified with $\mathbb{P}(S)$ via $f \leftrightarrow \{\,a \in S \mid f(a) = 1\,\}$.

Suppose $S_0,\ \ldots,\ S_{n-1}$ are nonempty sets for some integer $n > 2$. Define $S_0 \times S_1 \times S_2$ to be $S_0 \times (S_1 \times S_2)$. More generally, $S_0 \times \cdots \times S_{n-1}$ is given by a recursive definition as $S_0 \times (S_1 \times \cdots \times S_{n-1})$. This last is called the set of all "ordered n-tuples" formed from the S_i in the specified order. Let W denote the set of all functions $f\colon \{0, \ldots, n-1\} \to \bigcup_{i=0}^{n-1} S_i$ having the property that $f(i) \in S_i$ for $i = 0, \ldots, n-1$. Then $S_0 \times \cdots \times S_{n-1}$ can be identified with W by $(a_0, \ldots, a_{n-1}) \leftrightarrow f$ where $f(k) = a_k$ for $k = 0, \ldots, n-1$.

2. Equivalence Relations

Our second use of relations is the standard method used by mathematicians to lump together objects that are manifestly different but which are similar in some way. In this context we focus on the similarities and ignore other properties.

An **equivalence relation** on S is a relation $P \subset S \times S$ that has three properties:

$(a, a) \in P \ \forall a \in S$ and (**reflexivity**)

$(a, b) \in P \Rightarrow (b, a) \in P$ and (**symmetry**)

$(a, b) \in P$ and $(b, c) \in P \Rightarrow (a, c) \in P$. (**transitivity**)

For equivalence relations, the notation $\boldsymbol{a \sim b}$ is usually used when $(a, b) \in P$.

A **partition** of any set S is a set of subsets of S whose union is S and whose **pairwise** (that is, each pair of them) intersections are **void** (that is, the empty set.) Any pair of sets whose intersection is empty is called **disjoint**, and a union of pairwise disjoint sets is called a **disjoint union**.

After presenting an equivalence relation on S, one would typically form, for each a in S, sets $[a] = \{\, b \mid a \sim b \,\}$. These sets are called **equivalence classes** and together form a partition of S denoted $\mathbf{S/P}$ or $\mathbf{S}/\sim$. Often any member of an equivalence class will be used to refer to the whole class without comment, and it is the set of classes that is of primary interest.

Alternatively, any partition of a set S could be used to form an equivalence relation on the set, where $a \sim b$ precisely when a and b are in the same partition member.

Most people have seen equivalence relations from grade school. The rational numbers constitute a very important first example.

Let S be the set of all ordered pairs of integers (c, d) indicated here by c/d where we require that $d \neq 0$. (For a discussion of the construction of the integers, see Section 5.) We say that $c/d \sim e/f$ if and only if $cf = ed$. It is easy to show this is an equivalence relation. For each c/d in S let $[c/d] = \{\, e/f \in S \mid cf = ed \,\}$. The sets $[c/d]$ form a partition of S. Any ordered pair might be called upon to represent the whole class. That is what is meant by "$2/6 = 4/12$." The collection of these classes is normally referred to as the **rational numbers**, denoted $\mathbb{Q}$.

The operations $[a/b] + [c/d] = [(ad + bc)/(bd)]$ and $[a/b][c/d] = [(ac)/(bd)]$ are **well-defined**.

In this context, "**well-defined**" means that the operations, defined here using particular representations of the equivalence classes involved, do not in fact depend on which representative is used. **Statements of this kind, wherever found in the text, require proof.** If not obvious or proved in the text, the reader should supply the proof as an exercise, or simply accept the statement as true. Since all books contain errors, oversights, mis-statements or infelicitous phrasing, the former course is the safer.

$\mathbb{Q}$ has multiplicative and additive identities $[b/b]$ and $[0/b]$ respectively, otherwise known as 1 and 0.

Another example is the usual representation, found in many beginning Physics classes, of vectors in the plane as "arrows" with a given length and direction. One takes the point of view that a vector is determined by these two quantities alone and its location is irrelevant. So a vector is really a class of arrows that are alike in these two ways. One refers to the whole class by identifying any member of the class. In the world of vector operations such as vector addition or scalar multiplication

the usual representative for a class is the arrow with tail at a specified origin, with coordinate axes centered there. With this choice the coordinates of the tip alone suffice to describe the class, and common vector operations are conveniently calculated.

The concept of equivalency, along with the companion concept of identification, can be seen throughout mathematics.

3. Order Relations

In this section we try to extract the essence of the idea of "less than" as thought of in the following three examples:

> 3 is said to be "less than" 7 on the number line because it is to the left when one represents the real numbers ordered as a line in the usual way.
>
> Consider a desk covered with many layers of paper. We might say one piece of paper is "less than or equal to" another if its distance to the table top is equal or less than the other: an ordering by "height above the table."
>
> We think of one set as "bigger than or equal to" another if it contains the other. Sets can be said to be **ordered by containment**, a very important example.

The relations we use to model these ideas are called **order relations**.

A **pre-order** on a set S is a relation $P \subset S \times S$ that has the reflexivity and transitivity properties: $((a,a) \in P \ \forall a \in S)$ and $((a,b) \in P$ and $(b,c) \in P \Rightarrow (a,c) \in P)$.

The notation $\boldsymbol{a \leq b}$ will be used to indicate that $(a,b) \in P$, while $\boldsymbol{a < b}$ will indicate that $(a,b) \in P$ but $(b,a) \notin P$.

Suppose $B \subset S$. b is called an **upper bound** for B (in the pre-ordered set S if that specificity is warranted) if $b \in S$ and $a \leq b \ \forall a \in B$.

If B has an upper bound, B is called **bounded above**. If, further, $c \in S$ and $a \leq c \ \forall a \in B \Rightarrow b \leq c$ then b is called a **least upper bound** for B.

If b is a unique least upper bound for B (that is, the only one) then b is also called the **supremum** of B, denoted $\boldsymbol{\sup B}$ or $\boldsymbol{\sup(B)}$. The supremum of a set $\{a,b\}$, if it exists, is denoted $\boldsymbol{a \vee b}$.

An element b of S is called **maximal** if $c \in S$ and $b \leq c \Rightarrow c \leq b$.

A function $f\colon J \to S$ is called **bounded above** if its range is bounded above. The **supremum of a function** f is denoted $\boldsymbol{\sup(f)}$ or $\boldsymbol{\bigvee_{\alpha \in J} f(\alpha)}$ and is defined to be $\sup\{\, f(\alpha) \mid \alpha \in J \,\}$ whenever the supremum exists.

In the case of $J = \mathbb{N}$, the non-negative integers, a function $f\colon \mathbb{N} \to S$ is called a **sequence** in S. In this case the notation $\boldsymbol{\bigvee_{i=0}^{\infty} f(i)}$ may be seen in place of $\bigvee_{i \in \mathbb{N}} f(i)$.

When $J = \{k, k+1, \ldots, n\}$ we may write $\boldsymbol{\bigvee_{i=k}^{n} f(i)}$ rather than $\bigvee_{\alpha \in J} f(\alpha)$.

The definitions of $\boldsymbol{\geq}$, $\boldsymbol{>}$, **lower bound, bounded below, greatest lower bound, infimum,** $\boldsymbol{\inf B}$, $\boldsymbol{\inf(B)}$, $\boldsymbol{a \wedge b}$, **minimal**, $\boldsymbol{\inf(f)}$, $\boldsymbol{\bigwedge_{\alpha \in J} f(\alpha)}$,

$\bigwedge_{i=0}^{\infty} f(i)$ and $\bigwedge_{i=k}^{n} f(i)$ are the obvious adaptations of the list of definitions above with ordered pairs (that is, inequalities) reversed.

A set or function as above is called **bounded** if it is bounded both above and below. Otherwise, it is called **unbounded.**

Since functions are defined to *be* sets there is potential for ambiguity in the definitions just given involving functions, which focus on the order properties in the range alone. The ordered pairs in a function will not usually have an order specified for them so this will rarely be an issue.

The pre-order P is called a **partial order** if, in addition to reflexivity and transitivity, we have:

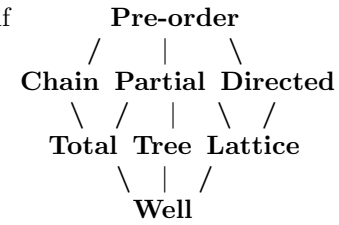
$$a \leq b \text{ and } b \leq a \Rightarrow a = b. \qquad \textbf{(antisymmetry)}$$

A partial order can be created from any pre-order on a set S by the following process. Let $a \sim b$ if and only if $a \leq b$ and $b \leq a$. Consider the set of equivalence classes $S/\!\sim$ generated by this equivalence relation. We will say $[a] \leq [b]$ precisely when $a \leq b$. This relation is well-defined (that is, it does not depend on the representatives of the classes used to define it) and a partial order on $S/\!\sim$.

Note that if P is a partial order, least upper bounds and greatest lower bounds are unique if they exist. Also, in this case, if b is maximal, $a \in S$ and $b \leq a$ then $b = a$. A similar result holds if b is minimal.

If P is a pre-order and $A \subset S$, A is called a **chain** if whenever a and b are in A then either $a \leq b$ or $b \leq a$.

If P is a pre-order and for each pair a and b in S one can find c in S such that $a \leq c$ and $b \leq c$ we say that P **directs** S, and S with P is called a **directed set**.



P, or sometimes S with P, is called a **lattice** if it is a partial order for which each pair of elements has a greatest lower bound and a least upper bound.

If S with partial order P is itself a chain, P is called a **total order**. In some contexts a total order is also called a **linear order**.

A total order on S which has the property that every nonempty subset of S contains a minimal element is called a **well-order**.

For brevity, one often refers to S as "**ordered by** P," P is said to "**order** S" and S is said to "**have the order** P." This vocabulary is extended to the various types of orders. Sometimes, when this will not cause ambiguity, the set S will be said to be ordered and the specific order P will be understood to exist without explicit mention.

As an example of this vocabulary in use we have the following interesting result:

Subsets of well-ordered S are well-ordered with the order **inherited** from S.

We define, for each α in pre-ordered S, the sets

$$I_\alpha = \{\, \beta \in S \mid \beta < \alpha \,\} \quad \text{and} \quad T_\alpha = \{\, \beta \in S \mid \beta \geq \alpha \,\}$$

are called **initial** and **terminal segments** in S, respectively.

Note that the T_α are distinct for different values of α in partially ordered S, but the I_α need not be distinct. They will be distinct if S is totally ordered.

If S is partially ordered and $\alpha, \beta \in S$ and $\beta > \alpha$ we say that β is **a successor** to α and α is **a predecessor** to β. If, further, there is no other member of S between α and β we say that β is **an immediate successor** to α and α is **an immediate predecessor** to β.

Suppose S is a generic well-ordered set. Unless there is a compelling reason to deviate, it will be standard practice to denote the first element of S as 0 and the second member by 1. For $\alpha \in S$, we will use $\boldsymbol{\alpha + 1}$ to denote the least successor to α. Unless α is the supremum of S, this immediate successor to α will always exist in well-ordered S.

$\alpha+1$ is called **the successor** to α and α is called **the predecessor** to $\alpha+1$. The vocabulary recognizes the fact that there can be at most one immediate successor or immediate predecessor in any totally ordered set.

Any member of S **except the first** that cannot be written as $\alpha + 1$ for some α in S is called a **limit member** of S. A limit member has predecessors (many of them) but no *immediate* predecessor.

If S is partially ordered and if, for each $\alpha \in S$, the initial segment I_α is well-ordered with the order inherited from S we call S with this order a **tree.**

A **branch** of a tree S is a nonempty subset B of S which is a chain with the induced order and which is maximal in the following sense:

$$\text{for each } s \in S, \text{ either } s \in B \text{ or } \{s\} \cup B \text{ is not a chain.}$$

So the well-ordered set $I_\alpha \cup \{\,\alpha\,\}$ is contained in branch B whenever $\alpha \in B$.

A **root** of a tree S is a member r of S for which $S = T_r$. A tree S is called **rooted** if it has a unique root.

3.1. ***Exercise.*** *(i) Is it true that the intersection of two or more (distinct) branches in a rooted tree is an initial segment?*

(ii) If S is a tree and $r \in S$ then T_r is a rooted tree.

If S is well-ordered and A is any union or any intersection of initial segments then A is either all of S or itself an initial segment. To see this, let α be the least member of S, if any, that is not in A. Then A is I_α. Similarly, if A is any union or intersection of terminal segments in well-ordered S, then A is itself a terminal segment or void.

3.2. ***Exercise.*** *Suppose S is a well-ordered set. Prove:*

$$S = I_\alpha \cup T_\alpha \text{ for every } \alpha \text{ in } S \qquad I_0 = \varnothing \qquad T_0 = S$$

$$T_\alpha \text{ is never empty} \qquad I_{\alpha+1} = I_\alpha \cup \{\alpha\} \text{ whenever } \alpha + 1 \text{ is defined.}$$

A function $f\colon A \to B$ between two pre-ordered sets is called **non-decreasing** if $f(\alpha) \le f(\beta)$ whenever $\alpha \le \beta$. f is called **increasing** if $f(\alpha) < f(\beta)$ whenever $\alpha < \beta$. Neither condition implies that f is one-to-one. However if the order on A is a total order, the second condition does imply that f is one-to-one.

f is called **non-increasing** if $f(\alpha) \geq f(\beta)$ whenever $\alpha \leq \beta$. f is called **decreasing** if $f(\alpha) > f(\beta)$ whenever $\alpha < \beta$.

f is called **monotone** if it is non-increasing or non-decreasing.

Suppose $f\colon A \to B$ is a function *between two partially ordered sets.* If f is non-decreasing and f^{-1} exists and is non-decreasing, f is called an **order-isomorphism** and A and B are said to be **order-isomorphic**.

3.3. ***Exercise.*** *(i) Suppose $f\colon A \to B$ is non-increasing, where A is totally ordered and B is partially ordered. In the definition of order-isomorphism applied to f the requirement that f^{-1} be non-decreasing is redundant.*

(ii) Suppose $f\colon A \to B$ is non-increasing, where A is totally ordered and B is well-ordered. Then f is eventually constant: that is, there is an $a \in A$ for which $c \geq a$ implies $f(c) = f(a)$.

(iii) If S is well-ordered, S is order-isomorphic to $\{\, I_\alpha \mid \alpha \in S \,\}$ ordered by containment.

(iv) If S is partially ordered, S is order-isomorphic to $\{\, \{\alpha\} \cup I_\alpha \mid \alpha \in S \,\}$ ordered by containment.

(v) If S is partially ordered, S is order-isomorphic to $\{\, T_\alpha \mid \alpha \in S \,\}$ ordered by ***reverse containment****: that is, $T_\alpha < T_\beta$ if $T_\alpha \neq T_\beta$ and $T_\alpha \supset T_\beta$.*

Parts (iii) through (v) of the exercise show that any partial order—and well-orders in particular—can be thought of as containment orders on families of subsets of S in several ways.

4. The Integers

We sketch in some detail the recognition of a set we will identify with the natural numbers, as you have come to know them from ordinary counting and grade-school arithmetic.

You, no doubt, have some conception of the nature of a set and readily assert the existence of sets with certain properties, combine sets in various ways, and understand what you must show to claim equality of two sets.

These conceptions were around *long before* mathematicians found it necessary to (try to) create bulletproof axiomatic structures founded on inexorable and indisputable logical chains to justify theorems. Having been burned a few times we have had a certain amount of humility forced upon us, and the various "obvious" properties of sets (which we have already used many times without remark in the preceding pages) is explored in some detail in Sections 11 and 12. These sections really do need to be examined, sooner or later.

As an example of the kind of thing that must be made explicit, and to get things started, the following assumption (to be accepted without proof) is required.

Axiom of the Empty Set:

There exists a set, denoted $\varnothing$, which has no elements.

Without this (or some similar) assumption, we cannot conclude that there are any sets whatsoever! That would mean all our discussions about sets have been about nothing, a situation tailor-made for irony if ever there was one. We accept this axiom.

If X is a set, for now we will let X^* be the set $\{X\} \cup X$.

We let $0 = \varnothing,\ 1 = 0^*,\ 2 = 1^*,\ 3 = 2^*$ and so forth. Another way of writing this is: $0 = \varnothing,\ 1 = \{0\},\ 2 = \{0, 1\},\ 3 = \{0, 1, 2\}$ and so on.

In addition to axioms justifying the obvious operations indicated above, to complete the definition of the natural numbers it is necessary to invoke another axiom of set theory, called the Axiom of Infinity. Essentially this axiom asserts that there exists at least one non-finite set and this is *not* an obvious fact; at least, it is not obvious to *everyone*.

Axiom of Infinity:

> There exists a set A with $\varnothing \in A$ and such that whenever X is a set and $X \in A$ then $X^* \in A$.

Note that the intersection of any pair of sets of the type whose existence is guaranteed by this axiom is also of this type.

Let A be one of these sets. The **natural numbers**, denoted $\mathbb{N}$, consist of the intersection of all subsets S of A for which $\varnothing \in S$ and such that whenever X is a set and $X \in S$ then $X^* \in S$. In light of the last observation, $\mathbb{N}$ does not depend on the specific choice of A, only that there is at least one such set.

It is only because of the Axiom of Infinity that we know that $\mathbb{N}$, which we might have carelessly denoted $\{\, 0, 1, 2, 3, \dots \,\}$, is actually a set, and therefore eligible to participate in the various set operations and assertions we might make about sets.

The empty set is said to **have 0 elements**. If S is a nonempty set and n is a **positive integer** (that is, $n \in \mathbb{N}$ and $n \neq 0$) we say **S has n elements** if there is a one-to-one and onto function $f\colon S \to n$. We say S is **finite** if it has n elements for some $n \in \mathbb{N}$. S is called **infinite** if it is not finite. We will discuss this concept again in Sections 16 and 19.

The natural numbers are partially ordered by containment, and it turns out that this ordering on $\mathbb{N}$ is actually a well-order.

Henceforth, if $n \in \mathbb{N}$ we will use $\mathbf{n + 1}$ in preference to n^*.

The definition of $\mathbb{N}$ is just what we need to create **Proof by Induction.**

If we have some property P which is either true or false for members of $\mathbb{N}$, let

$$S = \{\, n \in \mathbb{N} \mid P \text{ is true for } n \,\}.$$

If $0 \in S$ and if $n \in S$ implies $n + 1 \in S$ then S is a set of the kind whose existence is asserted in the Axiom of Infinity. Since $\mathbb{N}$ is the intersection of all such sets, $S = \mathbb{N}$.

In other words, **we would then be authorized to conclude that *P* is true for every member of $\mathbb{N}$.**

4.1. ***Exercise.*** *(i) As an (easy) exercise using induction, show that every member of* $\mathbb{N}$ *except* 0 *contains* 0 *among its elements, and can be written as* $n+1$ *for some* $n \in \mathbb{N}$.

(ii) Let $S = \{\, n \in \mathbb{N} \mid x \in n \Rightarrow x \in \mathbb{N} \,\}$. *Then* $S = \mathbb{N}$. *Every element of a natural number is a natural number.*

(iii) Let $S = \{\, n \in \mathbb{N} \mid x \in n \Rightarrow x \subset n \,\}$. *Then* $S = \mathbb{N}$. *Every element of a natural number is a subset of that natural number.*

(iv) $m + 1 = n + 1 \Rightarrow m = n$.

(v) For $j \in \mathbb{N}$ *let* $S_j = \{\, n \in \mathbb{N} \mid j \subset n \Rightarrow j = n \text{ or } j \in n \,\}$. *Then* $S_j = \mathbb{N}$. *(hint:* $\mathbb{N} = S_0$ *since the condition for membership in* S_0 *is trivially satisfied for all members of* $\mathbb{N}$*, and also* 0 *is in every* S_j*. Now suppose* $n \in S_j$ *and* $j \subset n + 1 = n \cup \{\, n \,\}$*. If* $n \notin j$ *then we have* $j \subset n$ *so by hypothesis* $j = n \in n + 1$ *or* $j \in n \subset n + 1$*. On the other hand, if* $n \in j$ *we have* $n \subset j \subset n + 1 = n \cup \{\, n \,\}$ *so* $j = n + 1$ *or* $j = n \in n + 1$*.)*

(vi) For $j \in \mathbb{N}$ *let* $S_j = \{\, n \in \mathbb{N} \mid j \subset n \text{ or } n \subset j \,\}$. *Then* $S_j = \mathbb{N}$. *This implies that* $\mathbb{N}$ *is a total order with containment order.*

(vii) Now define B *to be the set:*

$\{\, n \in \mathbb{N} \mid$ *if* $k \in W \subset \mathbb{N}$ *for some* $k \subset n$ *then* W *contains a least member.* $\}$

Obviously $\varnothing \in B$ *and it is not hard to show that if* $n \in B$ *then* $n + 1 \in B$*. We conclude that* $B = \mathbb{N}$ *so* $\mathbb{N}$ ***is well-ordered by containment order.***

(viii) The natural numbers have another interesting property. Every set of natural numbers that is bounded above has a least upper bound and contains this least upper bound.

We can use the facts from this exercise and induction to show that **a set C cannot have both m elements and n elements for natural numbers $n \neq m$.**

To see this, let S consist of those members s of $\mathbb{N}$ for which there is a member n of $\mathbb{N}$, $n \neq s$, and a set C which has both s elements and n elements. Obviously $\varnothing \notin S$, so every member of S has the form $k + 1$ for some natural number k. Should S be nonempty, it would contain a least member, and this leads easily to a contradiction. We conclude S is empty, and there is at most a single natural number n for which the statement "C has n elements" is true.

Note that each positive integer n is, itself, a well-ordered set which has n elements. Any natural number n is, in fact, the initial segment I_n in $\mathbb{N}$.

$\mathbb{N}$ **is an infinite set.** To see this, suppose $h\colon \mathbb{N} \to k+1$ is one-to-one and onto. Then $h(m) = k$ for some $m \in \mathbb{N}$. Define $f\colon \mathbb{N} \to \mathbb{N}$ by $f(a) = a$ when $a < m$ and $f(a) = a + 1$ if $a \geq m$. But then $h \circ f\colon \mathbb{N} \to k$ is one-to-one and onto. Since it is obvious that there can be no one-to-one function $h\colon \mathbb{N} \to 1$, the result follows.

Our next steps, left to the energetic reader, are to define negative integers and then the **integers**, denoted $\mathbb{Z}$, comprised of the natural numbers and the negative integers. The order on $\mathbb{N}$ is used to form a total order on $\mathbb{Z}$.

Though addition and multiplication of integers could be defined now, demonstrating that they have the usual properties such as commutativity, the distributive law and so on using the tools we have built to this point is a lengthy series of

applications of induction. Connecting these operations to the order relation on the integers also requires more than a bit of work.

For instance, if a, b and m are positive integers and $m = ab$ then both $a \leq m$ and $b \leq m$. And if $a > 1$ then $b < m$. But what does it take, exactly, to prove that? Laying out every last detail about basic integer arithmetic is a big project, falling under the heading of number theory and logic. Gottlob Frege and Bertrand Russell are among the luminaries who broke teeth on it. You might wish to review Exercises 17.9 and 17.10 and the more sophisticated techniques assembled in Appendix 10. In this work we will simply assume various "obvious" facts about integers.

With the integers in hand, one can define the rational numbers, $\mathbb{Q}$, as suggested in Section 2.

4.2. ***Exercise.*** *Till now we have had no specific well-ordered sets (other than* $\mathbb{N}$ *and its initial segments) with which to work. Now we can create examples.*

(i) Let $\mathcal{S}$ *denote the set* $\left\{ m + \frac{n}{n+1} \mid n, m \in \mathbb{N} \right\}$ *with the usual order from* $\mathbb{Q}$*. Show that* $\mathcal{S}$ *is well-ordered.*

(ii) For any $f \in \mathbb{N}^{\mathbb{N}}$ *let*

$$Support(f) = \{ n \in \mathbb{N} \mid f(n) \neq 0 \}.$$

Define $\mathcal{F}$ *to be those members of* $\mathbb{N}^{\mathbb{N}}$ *for which* $Support(f)$ *is a finite set. For* $f \in \mathcal{F}$ *let* m_f *denote the greatest member of* $Support(f)$*.*

We will define an order $\leq_R$ *on* $\mathcal{F}$ *called* ***reverse lexicographic order.***

Declare $f \leq_R f$*. Suppose* $f, g \in \mathcal{F}$ *and* $f \neq g$*. Let* j *be the* ***last*** *integer for which* $f(j) \neq g(j)$*. Declare* $f \leq_R g$ *if* $f(j) < g(j)$*.*

Show that $\mathcal{F}$ *is well-ordered with* $\leq_R$*. (hint: First show transitivity and conclude that* $\leq_R$ *is a total order. With that in hand, suppose* H *is a subset of* $\mathcal{F}$ *with at least two members. Let* n_1 *denote the least* m_f *of any* $f \in H$*. Let*

$$H_1 = \{ f \in H \mid m_f = n_1 \} \quad \text{and} \quad G_1 = \{ g \in H_1 \mid g(n_1) \leq f(n_1) \ \forall f \in H_1 \}.$$

If G_1 *contains a single member we stop: this member is the minimal member of* H*. If* G_1 *contains more than one member, let* n_2 *denote the smallest integer for which there is some* $g \in G_1$ *with* $g(n_2) \neq 0$ *but* $g(k) = 0$ *for all* k *with* $n_2 < k < n_1$*. Possibly,* $n_2 = n_1 - 1$*. Now let*

$$G_2 = \{ g \in G_1 \mid g(n_2) \leq f(n_2) \ \forall f \in G_1 \}.$$

If G_2 *contains a single element it is the least member of* H*. If* G_2 *contains more than a single member we can continue, creating by this procedure a strictly decreasing list* $n_1, n_2, \ldots$ *in* $\mathbb{N}$*. Such a list cannot be infinite in any well-ordered set. It must terminate at some least* n_k*, and the sole member of* G_k *is the minimal member of* H*.)*

(iii) Define $\mathcal{F}_1$ *to be those members* f *of* $\mathcal{F}$ *with* $m_f = 0$ *or* 1*. This set inherits the reverse lexicographic well-order from* $\mathcal{F}$*. How is this order on* $\mathcal{F}_1$ *related to that on* $\mathcal{S}$ *from part (i)?*

(iv) We will define a different order $\leq_L$ *on* $\mathcal{F}$ *called* ***lexicographic order.*** *Declare* $f \leq_L f$*. Suppose* $f, g \in \mathcal{F}$ *and* $f \neq g$*. Let* j *be the* ***first*** *integer for which*

$f(j) \neq g(j)$. Declare $f \leq_L g$ if $f(j) < g(j)$ and $g \leq_L f$ otherwise. Though $\mathcal{F}$ is totally ordered with $\leq_L$, it is not well-ordered.

The difference between $\leq_R$ and $\leq_L$ boils down to the following fact. It is impossible to create a strictly decreasing sequence of m_f values, but it is certainly possible to have a strictly increasing sequence of these values.

$$100000\ldots,\ 010000\ldots,\ 001000\ldots,\ 000100\ldots,\ 000010\ldots,\ \ldots\ldots$$

(v) Suppose A and B are disjoint well-ordered sets. Create an order on $A \cup B$ corresponding to "elements of A all follow any element of B," while retaining the given orders on A and B. Show that this order is a well-order.

*(vi) Sometimes it will be convenient in certain arguments to have a well-ordered set with a last member. In general, well-ordered sets might not **have** a last member. Suppose C is well-ordered with first element a_1 and more than one element. Give $A = C - \{a_1\}$ the inherited well-order and let $B = \{a_1\}$. Using (v) create a well-order on C that **does** have a last member.*

*(vii) Suppose A and B are well-ordered sets. Create a well-order on a **subset** of A^B analogous to the reverse lexicographic order $\leq_R$ we created for $\mathcal{F}$ in part (ii).*

5. The Real Numbers

We will now make one of the common definitions of the real numbers and discuss some important properties of this set. The following construction is due to Dedekind.

Let $\mathbb{Q}^+$ be the set of *non-negative*[2] rational numbers. We define $\mathbb{R}^+ \subset \mathbb{P}(\mathbb{Q}^+)$ to consist of exactly those sets A of non-negative rational numbers with the following three properties:

- (i) A has no largest member and
- (ii) $q \in A \Rightarrow p \in A\ \forall p \in \mathbb{Q}^+$ with $p \leq q$ and
- (iii) $A \neq \mathbb{Q}^+$.

$\mathbb{R}^+$ is (obviously) nonempty and called the set of **non-negative real numbers**. A non-negative real number, created this way, may be called a **Dedekind cut.**

If r and s are non-negative real numbers, we say $r < s$ if $r \neq s$ and $r \subset s$.

This relation is a total order on $\mathbb{R}^+$ but it is not a well-order. In fact no explicit well-order of the real numbers is known.

If r and s are nonempty (that is, "**positive**") members of $\mathbb{R}^+$ and $t \in \mathbb{R}^+$ we define binary operations "$+$" and "$\cdot$" by:

$$t + \varnothing = t \qquad \text{and} \qquad r + s = \{\, u \in \mathbb{Q}^+ \mid u < q + p \text{ for some } q \in r \text{ and } p \in s \,\},$$
$$t \cdot \varnothing = \varnothing \qquad \text{and} \qquad r \cdot s = \{\, u \in \mathbb{Q}^+ \mid u < q \cdot p \text{ for some } q \in r \text{ and } p \in s \,\}.$$

It is an exercise to show that $r + s$ and $r \cdot s$ are non-negative real numbers and the operations satisfy the commonly listed properties of addition and multiplication

[2]We include 0 in $\mathbb{Q}^+$ and $\mathbb{R}^+$. Many authors don't.

with multiplicative identity given by $\{\, [a/b] \in \mathbb{Q}^+ \mid 0 < a < b \,\}$ and additive identity $\varnothing$, which will henceforth be denoted 1 and 0, respectively. Multiplicative inverses exist for positive real numbers.

Note that this is the third usage for the symbol 1 in this section. $1 \in \mathbb{N}$ was defined to be $\{\varnothing\}$ and $1 \in \mathbb{Q}^+$ was defined as a set of ordered pairs $\{\, a/a \mid a \in \mathbb{Z} \text{ and } a \neq 0 \,\}$. We unify these disparate definitions by identifying $n \in \mathbb{N}$ with $[n/1] \in \mathbb{Q}$, and $q \in \mathbb{Q}^+$ with $\{\, p \in \mathbb{Q}^+ \mid p < q \,\} \in \mathbb{R}^+$.

Let S be any nonempty set of non-negative real numbers. If S has an upper bound in $\mathbb{R}^+$, we can show that $\bigcup_{A\in S} A \in \mathbb{R}^+$. In fact it is the supremum of S.

Let S be any nonempty set of non-negative real numbers. $\bigcap_{A\in S} A$ might actually contain a largest rational. If it does not, then $\bigcap_{A\in S} A \in \mathbb{R}^+$ and is the infimum of S. If it does contain a largest rational, remove that rational from the intersection. The result is now in $\mathbb{R}^+$ and is the infimum of S.

If $r\colon \mathbb{N} \to \mathbb{R}^+$ is a non-decreasing sequence of non-negative real numbers that is bounded above we let

$$\lim_{\mathbf{n}\to\infty} \mathbf{r_n} = \sup\{\, r_n \mid n \in \mathbb{N} \,\}$$

If $r\colon \mathbb{N} \to \mathbb{R}^+$ is a non-increasing sequence of non-negative real numbers let

$$\lim_{\mathbf{n}\to\infty} \mathbf{r_n} = \inf\{\, r_n \mid n \in \mathbb{N} \,\}$$

In either case, this number is called the **limit** of the corresponding sequence.

At this point it is an exercise to extend all of the above to a definition of the **negative real numbers** and then to the **real numbers**—consisting of both non-negative and negative real numbers. Extend the total order on the non-negative real numbers to the real numbers. Then define **multiplication, division, addition** and **subtraction** for these numbers, and show they have the familiar properties. Define **absolute value**. Define **limits of bounded monotone sequences** of real numbers.

Henceforth we let $\mathbb{R}$ denote the **real numbers.**

Define **intervals** $[a, b)$, (a, b), $(a, b]$, $(-\infty, b)$, $(-\infty, b]$, (a, ∞), $[a, \infty)$ and $[a, b]$ for real numbers a and b with $a \leq b$. The **standard topology on** $\mathbb{R}$ is that formed from a basis consisting of all intervals (a, b) with $a, b \in \mathbb{Q}$.

If $r\colon \mathbb{N} \to \mathbb{R}$ is a bounded sequence we define

$$\limsup(\mathbf{r}) = \lim_{n\to\infty} (\sup\{\, r_k \mid k \in \mathbb{N} \text{ and } k > n \,\}) \quad \text{and}$$
$$\liminf(\mathbf{r}) = \lim_{n\to\infty} (\inf\{\, r_k \mid k \in \mathbb{N} \text{ and } k > n \,\})$$

Since the supremum and infimum above are being taken over smaller and smaller sets, the sequences whose limits are referred to are monotone and the limits are defined.

When these limits are equal we refer to their common value as the **limit of the sequence** r and denote this number by $\lim_{\mathbf{n}\to\infty} \mathbf{r_n}$. When the limit exists and is L we say the **sequence converges** or, when specificity is required, **converges to** L.

In applications, it is common for sequence values r_n to be defined only for n in a terminal segment of $\mathbb{N}$. Limits, if they exist, depend only on the value of r on any terminal segment. So when considering limits, we might define r_n values in any way that is convenient or not at all for n in any particular initial segment of $\mathbb{N}$.

Show that $|\,|a| - |b|\,| \leq |\,a - b\,| \leq |\,a\,| + |\,b\,|$. This is the **triangle inequality**.

Show that the limit of a sequence r exists and is a number L exactly when the limit of the sequence $|\,r - L\,|$ exists and is 0.

Two sequences r and s are called **equivalent** if $\lim_{n\to\infty} |\,r_n - s_n\,| = 0$. The exercises above can be used to show that equivalent sequences converge or not together, and if they converge it is to the same limit.

A sequence r is called a **Cauchy sequence** if

$$\lim_{n\to\infty} (\sup\{\,|\,r_n - r_k\,|\;\;|\;k > n\,\}) \quad \text{exists and is 0.}$$

It is a fact that a sequence of real numbers converges precisely when it is Cauchy, and the definition of equivalent sequences from above forms an equivalence relation on the set of convergent sequences.

These last concepts can be used in an alternative construction of the real numbers. One examines the set of all Cauchy sequences of rational numbers, and partitions that set using the equivalence relation for sequences defined above. This does involve the creation of a preliminary definition of limit, but only for rational sequences that converge to 0. The set of these classes constitute the real numbers in this formulation.

There is a more general concept of limit that pops up sometimes. This is where the indexing set is a more general directed set and not necessarily $\mathbb{N}$, and we might as well define it here.

If J is a directed set, a function $r\colon J \to Y$ is called a **net** in Y. A net is a generalization of the idea of a sequence.

Now suppose $r\colon J \to \mathbb{R}$ is a net in $\mathbb{R}$ and $L \in \mathbb{R}$.

We call L the **limit of the net** r and write $\boldsymbol{r_\alpha \xrightarrow{\alpha} L}$ if and only if

$$\forall \varepsilon > 0\ \exists \alpha \in J \text{ so that } \alpha \leq \beta \Rightarrow |\,r_\beta - L\,| < \varepsilon.$$

Limits of nets in $\mathbb{R}$, when they exist, **depend only on the values of the net on any particular terminal segment of J.** So when considering these limits, we are free to modify or define the r_α values in any way that is convenient or not at all for α outside of any terminal segment of J.

It is possible for a directed set such as J to have a supremum, $\sigma = sup(J)$. In that case the limit is simply the number r_σ.

When the limit of a net in $\mathbb{R}$ exists we say the **net converges** or, when specificity is required, **converges to** L.

A net in $\mathbb{R}$ has at most one limit.

In case $J = \mathbb{N}$, show that $\lim_{n\to\infty} r_n$ exists and equals L if and only if r converges as a net and $r_n \xrightarrow{n} L$.

Suppose D is a nonempty subset of $\mathbb{R}$ and $c \in \mathbb{R}$. Make D into a directed set by $a \preccurlyeq b$ if and only if $|\,c-a\,| \geq |\,c-b\,|$. Now suppose that $D \subset A \subset \mathbb{R}$ and $f\colon A \to \mathbb{R}$. The function $f|_D$ is a net in $\mathbb{R}$ which might converge.

In case $c \notin D$ and if D contains a set of the form $\{\, x \in \mathbb{R} \mid x \neq c \text{ and } |\,x-c\,| < \xi \,\}$ for some $\xi > 0$ and provided $f|_D$ converges, a limit of this net is denoted

$$\lim_{\boldsymbol{x \to c}} \boldsymbol{f(x)}.$$

If there is any D satisfying the conditions above then the existence of this limit and its unique value do not depend on the particular D (satisfying the specified conditions) used in its definition, and therefore the directed set D will not usually be explicitly identified.

The various properties of $\mathbb{R}$, such as the total order on $\mathbb{R}$ and the existence of suprema and infima of bounded subsets of $\mathbb{R}$, have numerous consequences of importance here. The reader should recall, prove, look up or accept the following miscellaneous facts about the real numbers. The various topological concepts can be found in Appendix **??**.

5.1. ***Exercise.*** *(i) Suppose $f\colon (a,b) \to \mathbb{R}$ and $c \in (a_1, b_1) \subset (a,b)$. $\lim_{x\to c} f(x)$ exists and equals L if and only if $\lim_{n\to\infty} f(x_n)$ exists and equals L for every sequence $x\colon \mathbb{N} \to (a_1, c) \cup (c, b_1)$ with $\lim_{n\to\infty} x_n = c$.*

(ii) If $\lim_{x\to c} f(x)$ and $\lim_{x\to c} g(x)$ both exist and equal L and M respectively, then $\lim_{x\to c}(f(x) + g(x))$ and $\lim_{x\to c}(f(x)g(x))$ exist and $\lim_{x\to c}(f(x) + g(x)) = L + M$ and $\lim_{x\to c}(f(x)g(x)) = LM$. If $M \neq 0$ then $\lim_{x\to c}(1/g(x))$ exists and equals $1/M$.

(iii) Modify the definition of the directed set D in such a way that one-sided limits $\lim_{x\to c+} f(x)$ and $\lim_{x\to c-} f(x)$ are produced for appropriate functions f. Show that when properly restated the results of (i) and (ii) follow for your limits and that $\lim_{x\to c} f(x)$ exists exactly when both $\lim_{x\to c+} f(x)$ and $\lim_{x\to c-} f(x)$ exist and are equal.

5.2. ***Exercise.*** *(i) $f\colon (a,b) \to \mathbb{R}$ is continuous (with respect to the subspace topology on (a,b)) if and only if $\lim_{x\to c} f(x)$ exists and equals $f(c)$ for all $c \in (a,b)$.*

(ii) Constant functions are continuous, and the product and sum of continuous functions with common domain are continuous.

(iii) If $f\colon (a,b) \to (c,d)$ and $g\colon (c,d) \to \mathbb{R}$ are continuous then so is $g \circ f$.

(iv) The function $f\colon (0,\infty) \to (0,\infty)$ defined by $f(x) = x^2$ is one-to-one and onto $(0,\infty)$. Its inverse function is denoted $f^{-1}(x) = \sqrt{x}$. These functions are continuous and non-decreasing on their respective domains.

(v) The function $g\colon (0,\infty) \to (0,\infty)$ defined by $g(x) = 1/x$ is one-to-one and onto $(0,\infty)$. It is its own inverse function. It is continuous and non-increasing.

(vi) $A \subset \mathbb{R}$ is compact if and only if A is closed and bounded. This is the ***Heine-Borel Theorem****.*

(vii) Suppose $f\colon (a,b) \to \mathbb{R}$ is continuous and $[a_1, b_1] \subset (a,b)$. Let $B = \{\, f(x) \mid x \in [a_1, b_1] \,\}$. Suppose $\inf(B) \leq L \leq \sup(B)$. Then $\exists c \in [a_1, b_1]$ with $f(c) = L$. This is called the ***Intermediate Value Theorem****.*

(viii) Suppose $f\colon (a,b) \to \mathbb{R}$ is continuous. If K is a compact subset of (a,b) then $f(K)$ is compact. If J is a subinterval of (a,b) then $f(J)$ is an interval.

(ix) If $f\colon (a,b) \to (c,d)$ is one-to-one and onto and continuous then the inverse function $f^{-1}\colon (c,d) \to (a,b)$ is continuous.

(x) If $f\colon (a,b) \to \mathbb{R}$ is continuous, the values of f on $\mathbb{Q} \cap (a,b)$ determine the values of f on all of (a,b).

If a is a sequence of real numbers we define a new sequence S, called the **sequence of partial sums of a**, by $\boldsymbol{S_n = \sum_{k=0}^{n} a_k}$.

A sequence formed this way is called a **series**. Sometimes S converges. When it does its limit may be denoted $\boldsymbol{\sum_{k=0}^{\infty} a_k}$ and the **series is said to converge.**

If S_n does not converge it is said to **diverge**.

If $\sum_{k=0}^{\infty} |\, a_k \,|$ exists the series S is said to **converge absolutely**.

If the series converges but does **not** converge absolutely we say that the series converges **conditionally.**

When discussing the existence of the limit $\sum_{k=0}^{\infty} a_k$, we often say that the symbol $\sum_{k=0}^{\infty} a_k$ itself converges, diverges or converges absolutely or conditionally.

5.3. ***Exercise.*** *(i) If a series converges absolutely then it converges.*

(ii) Suppose $\sum_{k=0}^{\infty} a_k$ converges absolutely, and b is a real-valued sequence. Define for each $k \in \mathbb{N}$ the number $c_k = \sum_{i=0}^{k} a_{k-i} b_i$.

If $\sum_{k=0}^{\infty} b_k$ converges then so too does $\sum_{k=0}^{\infty} c_k$ and

$$\left(\sum_{k=0}^{\infty} a_k\right)\left(\sum_{k=0}^{\infty} b_k\right) = \sum_{k=0}^{\infty} c_k = \sum_{k=0}^{\infty}\left(\sum_{i=0}^{k} a_{k-i} b_i\right).$$

(iii) The series $E_n(x) = \sum_{k=0}^{n} \frac{x^k}{k!}$ converges absolutely for each real x. The limit is denoted e^x. The function $\mathbf{Exp}\colon \mathbb{R} \to (0,\infty)$ defined by $\mathbf{Exp}(\mathbf{x}) = e^x$ is one-to-one and onto $(0,\infty)$. Its inverse is denoted $\mathbf{Ln}$. For each real x and y, $e^{x+y} = e^x e^y$. Exp and Ln are continuous and non-decreasing on their respective domains.

(iv) The series $\sum_{k=0}^{n} x^k$ converges absolutely to $\frac{1}{1-x}$ for each $x \in (-1,1)$.

(v) The series $S_n(x) = \sum_{k=0}^{n} \frac{(-1)^k x^{2k+1}}{(2k+1)!}$ and $C_n(x) = \sum_{k=0}^{n} \frac{(-1)^k x^{2k}}{(2k)!}$ converge absolutely for each real x. Their limits are denoted $\mathbf{Sin}(\mathbf{x})$ and $\mathbf{Cos}(\mathbf{x})$, respectively, and the functions formed from these values are called the **Sine** *and* **Cosine** *functions. They are continuous.*

(vi) If a and b are real-valued sequences define $\Delta a_n = a_{n+1} - a_n$ and $\Delta b_n = b_{n+1} - b_n$ for each $n \in \mathbb{N}$. Then for $0 \le m < n$

$$\sum_{k=m}^{n} a_k\, \Delta b_k = a_{n+1} b_{n+1} - a_m b_m - \sum_{k=m}^{n} b_{k+1}\, \Delta a_k$$

which is called the ***summation by parts formula*** *for series. In case the sequence ab (defined by $(ab)_n = a_n b_n$) converges, the left sequence of partial sums converges exactly when the right sequence of partial sums does.*

(vii) Suppose a and c are real-valued sequences and we want to discover facts about the convergence of $S_n = \sum_{i=0}^{n} a_i c_i$. We define $b_k = \sum_{i=0}^{k} c_i$. Then $\Delta b_{n-1} = c_n$. The following equality of partial sums is called ***Abel's transformation*** *and is useful in several common applications.*

$$S_n = \sum_{k=0}^{n} a_k\, c_k = a_0\, c_0 + \sum_{k=1}^{n} a_k\, \Delta b_{k-1} = a_{n+1} b_n - \sum_{k=0}^{n} b_k\, \Delta a_k.$$

(viii) If the sequence a/b (defined by $(a/b)_n = a_n/b_n$) converges to a nonzero constant L then the series $\sum_{k=0}^{\infty} a_k$ converges exactly when $\sum_{k=0}^{\infty} b_k$ converges.

(ix) Suppose a is a sequence of non-zero numbers. Then $\lim_{n\to\infty} |a_{n+1}| \,/\, |a_n|$ exists exactly when $\lim_{n\to\infty} |a_n|^{1/n}$ exists. In case this common limit exists define R to be the reciprocal of the limit (if the limit is 0 let $R = \infty$.) For real x the series $\sum_{k=0}^{\infty} a_k x^k$ is called a ***power series*** *and R is called the* ***radius of convergence*** *of the series. This power series converges whenever $|x| < R$.*

(x) Suppose a is a sequence of non-zero numbers and $L = \lim_{n\to\infty} |a_n|^{1/n}$. If $L = 0$ the power series $\sum_{k=0}^{\infty} a_k x^k$ converges absolutely for all x. If $L = \infty$ (that is, if $\lim_{n\to\infty} |1/a_n|^{1/n} = 0$) the power series converges only for $x = 0$. Otherwise, let $R = 1/L$. The power series converges absolutely if $|x| < R$ and diverges if $|x| > R$. This result is called the ***Cauchy-Hadamard Theorem.***

Finally, we get to the issue of specific common representations of real numbers.

If p is an integer bigger than 1, we can represent any real number between 0 and 1 as $\sum_{k=1}^{\infty} \frac{a_k}{p^k}$ where the sequence a consists of integers with $0 \le a_n < p$ for all n.

This representation is not quite unique as stated.

Sequences a that terminate, for some n, with $a_n \neq 0$ and $a_k = 0$ for all $k > n$ and exactly one sequence b with $b_k = p - 1$ for all $k > n$ generate series for the same real number.

However this is the only duplication in the representation, so uniqueness is acquired by forbidding all representations that use sequences b that terminate in $b_k = p - 1$ for all $k > n$ for some n.

With this convention, any real number can be represented uniquely (for each p and some $k \ge 0$) as

$$\pm \left(\sum_{\boldsymbol{n=0}}^{\boldsymbol{k}} \boldsymbol{a_{-n} p^n} + \sum_{\boldsymbol{n=1}}^{\boldsymbol{\infty}} \frac{\boldsymbol{a_n}}{\boldsymbol{p^n}} \right) \qquad \text{where}$$

(*i*) $0 \le a_j < p$ for all $j \ge -k$ and

(*ii*) $a_{-k} \neq 0$ unless $k = 0$ and

(*iii*) the sequence does not terminate with $a_j = p - 1$ for all $j > m$ for any m.

The case of $p = 10$ corresponds to the ordinary **decimal representation** of numbers, while $p = 2$ and $p = 3$ generate the **binary or dyadic and ternary representations**.

We will take one further step in the progression $\varnothing \to 1 \to \mathbb{N} \to \mathbb{Q} \to \mathbb{R}$.

The **complex numbers**, denoted $\mathbb{C}$, consist of the set of all ordered pairs of real numbers with operations of addition and multiplication given by

$$(a,b)+(c,d)=(a+c,b+d) \quad \text{and} \quad (a,b)*(c,d)=(ac-bd,ad+bc).$$

An alternative way of representing an ordered pair of real numbers (a,b) thought of as a complex number is using the symbol $\boldsymbol{a+bi}$.

If $z=a+bi$ is a complex number, a is called its **real part**, and b is called its **imaginary part**. This is purely a notational device: we are associating real number a with ordered pair $(a,0)$ and bi with $(0,b)$.

$\overline{\mathbf{z}}=a-bi$ is called the **conjugate of z**. The **magnitude of z** is $\sqrt{a^2+b^2}$ and denoted $|\mathbf{z}|$. Note that if $z\neq(0,0)$ then $(1,0)=z\left(\frac{\overline{z}}{|z|^2}\right)$.

The map that associates x in $\mathbb{R}$ with $(x,0)$ in $\mathbb{C}$ preserves the arithmetic operations on $\mathbb{R}$ and sends the multiplicative identity there to the multiplicative identity in $\mathbb{C}$, so the range of this map can (and will) be identified with $\mathbb{R}$.

5.4. ***Exercise.*** *A sequence of complex numbers $z_n=a_n+b_ni$ converges to complex number $w=x+yi$ exactly when both $\lim_{n\to\infty}a_n=x$ and $\lim_{n\to\infty}b_n=y$. This happens exactly when the real sequence $|\,w-z_n\,|$ converges to 0. Adapt Exercise 5.1 wherever necessary to handle series of complex numbers. Then define the series $E_n(z)=\sum_{k=0}^{n}\frac{z^k}{k!}$ and show it converges absolutely for every complex z and define e^z to be the limit. If w and z are complex show that $e^{z+w}=e^z\,e^w$ and if $z=a+bi$ then*

$$e^z=e^a\,e^{bi}=e^a\,(\,Cos(b)+i\,Sin(b)\,).$$

Delete the negative x axis and the origin from the complex plane and define an inverse to a piece of the exponential function there. Call this inverse a logarithm. If you want this logarithm to be continuous, what choices do you have? Could you delete another half-line terminating at the origin and define another logarithm with this domain?

6. An Axiomatic Characterization of $\mathbb{R}$

Let A be any field. A may contain "a copy" $\mathbb{N}_A$ of $\mathbb{N}$. By this we mean that $\mathbb{N}_A$ is subring of A containing the identity of A and which is ring isomorphic to $\mathbb{N}$. Then since A is a field it must also contain "a copy" $\mathbb{Q}_A$ of $\mathbb{Q}$ which contains $\mathbb{N}_A$ as a subring. A is said to **have characteristic 0** when it contains a copy of $\mathbb{N}$.

Generally the additive and multiplicative identities in *any* field are denoted by 0 and 1, respectively. This could, but rarely does, lead to confusion.

If any field A is equipped with a linear order $<$ satisfying

- (i) $\quad x+y>0$ whenever $\;x,y>0$
- (ii) $\quad xy>0$ whenever $\;x,y>0$
- (iii) $\quad x+z>y+z$ whenever $\;x>y$

we call A an **ordered field**. Both $\mathbb{Q}$ and $\mathbb{R}$ are ordered fields.

If both x and $-x$ were positive (i.e. greater than 0) we would have by (i) that $0 > 0$, contradicting the assumption that our order is a linear order. So if $x \neq 0$ at most one of x or $-x$ is positive.

If $-1 > 0$ then (ii) implies $(-1)(-1) = 1 > 0$, contradicting (i). So $-1 < 0$. Using this in (iii) yields $0 + 1 > -1 + 1$ so $1 > 0$.

Define $\tilde{0} = 0$ and, having defined $\tilde{n}$ for integer n define $\widetilde{n+1} = \tilde{n} + 1$. This recursive definition gives a function from $\mathbb{N}$ to A. Repeated application of (iii) yields $\tilde{n} + 1 > \tilde{n} > 0$ for all positive integers n. In particular, we can never have $\tilde{n} = 0$. So the ordered field properties **imply** that the underlying field has characteristic 0: we do not, actually, need to assume it. There are unique copies $\mathbb{N}_A$ of $\mathbb{N}$ and $\mathbb{Q}_A$ of $\mathbb{Q}$ inside every ordered field.

An ordered field has the quality of **Dedekind completeness** or **DKC** provided that each subset which is bounded above has a least upper bound.

The real numbers as we have built them constitute a Dedekind complete ordered field. Every property of the real numbers used in analysis follows from just a few properties: those which define a field, the properties defining a linear order, (i), (ii), (iii) and DKC.

An ordered field has the **Archimedean order property** or **AOP** if, for each $x, y \in A$ with $0 < x < y$ there is an $n \in \mathbb{N}_A$ so that $y < nx$.

These properties have consequences, a few of which are explored below.

6.1. ***Exercise.*** *We will presume that A is an ordered field as above.*

(i) $x > 0$ exactly when $-x < 0$. *(ii) $x < y$ exactly when $0 < y - x$.*

(iii) $0 < x$ and $0 < y < z$ implies $0 < xy < xz$.

(iv) $x < 0$ and $y < 0$ implies $xy > 0$. *(v) $x < 0$ and $y > 0$ implies $xy < 0$.*

(vi) $x > 0$ exactly when $\frac{1}{x} > 0$. *(vii) $0 < x < y$ exactly when $0 < \frac{1}{y} < \frac{1}{x}$.*

6.2. ***Exercise.*** *Suppose A is an ordered field as above.*

Define $|0| = 0$. If x is nonzero in A, either $x > 0$ or $-x > 0$ but not both. Define $|x|$ to be x or $-x$, chosen so that $|x| > 0$.

(i) For each $x, y \in A$, show that $|\,xy\,| = |x||y|$.

(ii) For each $x, y \in A$, show that $|\,x + y\,| \leq |x| + |y|$.

6.3. ***Exercise.*** *Suppose A is an ordered field as above.*

(i) AOP is equivalent to each of the following three conditions:
For each $x > 0$ there is an $n \in \mathbb{N}_A$ so that $\frac{1}{n} < x$.
For each $x > 0$ there is an $n \in \mathbb{N}_A$ so that $x < n$.
For each $x > 0$ there is a unique $n \in \mathbb{N}_A$ for which $n < x \leq n + 1$.

(ii) DKC implies AOP. (hint: If A does not have AOP then there is a y with $0 < 1 < y$ but $y \geq m$ for all $m \in \mathbb{N}_A$. So $\mathbb{N}_A$ is bounded above. If A had DKC there would be a least upper bound $p \in A$ for $\mathbb{N}_A$. Show that $p - 1$ must also be an upper bound for $\mathbb{N}_A$, a contradiction. So A cannot have DKC.)

(iii) DKC implies that sets in A which are bounded below have infima.

(iv) $\mathbb{Q}$ has AOP but not DKC so AOP does not imply DKC.

6.4. ***Exercise.*** *Suppose A is an ordered field as above with AOP.*

(i) If B is a subset of A for which $s = \sup B$ exists then for each integer $n > 0$ there is a member $t \in B$ with $s - t < \frac{1}{n}$.

(ii) If $r \in A$ then $r = \sup\{\, t \in \mathbb{Q}_A \mid t \leq r \,\}$.

6.5. ***Exercise.*** *If A is a Dedekind complete ordered field then there is a ring isomorphism between A and $\mathbb{R}$. This ring isomorphism is unique, and is an order-isomorphism.*

We have a collection of properties, axioms if you will, satisfied by the real numbers. These axioms are (some of) the axioms of ordinary set theory plus those axioms associated with a Dedekind complete ordered field. The real numbers as we have created them constitute a realization or **model** of the axioms of a Dedekind complete ordered field "inside" ordinary set theory. We have shown by our construction that these axioms are consistent (if the axioms of set theory are consistent) and that was an important finding.

However neither the "Dedekind cut" construction of the real numbers nor the "Cauchy sequence" construction correspond in a compelling way to our simple intuition about real numbers as, for example, "points on a line."

In fact, all properties of the real numbers important to analysts **follow from the axioms mentioned above, not from the details of construction employed in forming our particular realization.**

It is these axioms which capture some of our intuition about real numbers, not any particular construction. The last exercise guarantees that if someone produces a different realization of these axioms, their underlying object shares all essential features with ours. We are free, when that is convenient, to remember the axioms and forget as irrelevant their particular embodiment.

Finally, it is worth noting that the usual identification of the real numbers with **all** the points on a line is not set in stone. It is not implied by the ancient concept of a line, nor by the standard practices of the inventors of calculus who routinely employed "infinitesimals," since replaced by limits.

Practitioners of **nonstandard analysis** use a larger ordered field called the **hyperreal numbers** ${}^*\mathbb{R}$ in place of $\mathbb{R}$. The hyperreal numbers contain positive numbers smaller than any real number, and limit-taking is replaced by hyperreal arithmetic.

The main technical challenges involved in transferring nonstandard results to the standard world were overcome by Abraham Robinson, the creator of this subject, in 1960.

Though conceptually attractive, it is currently unclear if this approach offers net advantages over standard analytic technique.

7. $[-\infty,\infty]^X$ AND $\mathbb{R}^X$

$[-\infty,\infty]$ is called the set of **extended real numbers** and defined to be $\mathbb{R}\cup\{-\infty,\infty\}$, where members of $\mathbb{R}$ have their usual properties and ∞ and $-\infty$ are distinct, not real numbers and have the order, addition and multiplication properties that would seem reasonable for "infinitely large" entities.

For example, $-\infty\le a\le\infty\ \forall a\in[-\infty,\infty]$. If $a>0$ we define $a\cdot\infty=\infty$ $a\cdot(-\infty)=-\infty$ and $a+\infty=\infty$ and $(-a)\cdot\infty=-\infty$ and $(-a)+(-\infty)=-\infty$. We also define $-\infty\cdot 0=\infty\cdot 0=0$. However $-\infty+\infty$ is not defined. The symbols $\pm\infty$ have no multiplicative inverses.

Every set in $[-\infty,\infty]$ is bounded above and below by ∞ and $-\infty$ respectively. We abuse vocabulary and declare a subset of $[-\infty,\infty]$ to be **bounded above or below** if it is bounded by a real number in the specified sense. With this usage, the bounded sets in $[-\infty,\infty]$ and $\mathbb{R}$ are the same.

$[-\infty,\infty]$ is a compact topological space, where neighborhoods of a point in $\mathbb{R}$ are sets containing an open interval around that point, neighborhoods of $-\infty$ are those sets containing an interval of the form $[-\infty,a)$, and neighborhoods of ∞ are those sets containing an interval of the form $(a,\infty]$.

Suppose J is a directed set, such as $\mathbb{N}$. Recall that for each $j\in J$ the symbol T_j denotes the terminal segment of J consisting of those members n of J for which $n\ge j$.

If $r\colon J\to[-\infty,\infty]$ is a net, each $r\,(T_j)$ is a set of extended real numbers, and any such has both supremum and infimum in $[-\infty,\infty]$. So, for example, both $u(j)=\sup r\,(T_j)$ and $l(j)=\inf r\,(T_j)$ are defined for each $j\in J$ and so form extended real-valued nets l and u defined on J. u and l are are monotone: u is non-increasing while l is non-decreasing.

The reader should investigate the modifications to the definition of limits of real-valued sequences needed to make sense out of notation such as

$$\boldsymbol{\limsup(r)=L}\quad\text{or}\quad\boldsymbol{\liminf(r)=L}\quad\text{or}\quad\boldsymbol{r_\alpha\xrightarrow{\alpha}L}$$

when L is an extended real number and r is a net in $[-\infty,\infty]$. The relationship between these limits and the previously defined limits for real-valued nets (when the former limits existed) must be examined.

For sets X and Y, recall that Y^X is the set of functions from X to Y. When Y has a partial order there is a partial order induced on Y^X given by

$$f\le g\Leftrightarrow f(a)\le g(a)\ \forall a\in X.$$

This is called the **pointwise order** on Y^X.

Infima and suprema of indexed sets of functions, such as $\{\,f_\alpha\mid\alpha\in J\,\}\subset Y^X$, are themselves members of Y^X whose values on each $x\in X$ are indicated by:

$$\left(\bigvee_{\alpha\in J}f_\alpha\right)(x)=\bigvee_{\alpha\in J}f_\alpha(x)\quad\text{and}\quad\left(\bigwedge_{\alpha\in J}f_\alpha\right)(x)=\bigwedge_{\alpha\in J}f_\alpha(x)$$

provided, of course, that the "pointwise" infima and suprema exist in Y for every $x\in X$.

These definitions depend on the existence of limits in Y.

If $Y \subset W$, an infimum or supremum might exist in W^X but not in Y^X.

There is a notational issue that should be observed here. If $f_\alpha \in Y^X$, we already have a definition for the infimum and supremum of a function f_α, namely:

$$\bigvee_{x\in X} f_\alpha(x) = \sup\{\, f_\alpha(x) \mid x \in X \,\} \quad \text{and} \quad \bigwedge_{x\in X} f_\alpha(x) = \inf\{\, f_\alpha(x) \mid x \in X \,\}$$

Confusion can arise when there are functions with multiple arguments or if multiple infima and suprema are being calculated if care is not taken in specifying order and arguments. Consider, for example:

$$\bigvee_{\alpha\in J}\left(\bigwedge_{x\in X} f_\alpha(x)\right) \quad \text{and} \quad \bigwedge_{x\in X}\left(\bigvee_{\alpha\in J} f_\alpha\right)(x).$$

There is no reason to think these two limits will be equal.

Note that Y is a lattice $\Leftrightarrow$ Y^X is a lattice.

More generally, infima and suprema always exist in Y^X precisely when such always exist in Y. These always exist if $Y = [-\infty, \infty]$ but not if $Y = \mathbb{R}$.

Suppose $f\colon J \to [-\infty,\infty]^X$, where J is a directed set.

We say f **converges pointwise** to a function $\widehat{f}$ provided

$$f_\alpha(b) \xrightarrow{\alpha} \widehat{f}(b)\ \forall b \in X.$$

To describe this situation and to assert the existence of such a limit we will write

$$\boldsymbol{f_\alpha \xrightarrow{\alpha} \widehat{f}} \quad \text{or, when } J = \mathbb{N}, \text{ we may write} \quad \boldsymbol{\lim_{n\to\infty} f_n = \widehat{f}.}$$

For f and g in $[-\infty,\infty]^X$, we define $\boldsymbol{f \cdot g}$ by $(f\cdot g)(a) = f(a)g(a)$ and $\boldsymbol{f + g}$ by $(f+g)(a) = f(a) + g(a)\ \forall a \in X$.

These are called **pointwise multiplication and addition**.

The multiplication and addition defined above are commutative, and the functions that are constantly one and zero are the multiplicative and additive identities, respectively. $[-\infty,\infty]^X$ is not a real vector space, but only because addition is not defined for all pairs of functions.

For any set X define $\boldsymbol{\chi\colon \mathbb{P}(X) \to 2^X}$ by $\boldsymbol{\chi_A(a)} = \begin{cases} 0 & \text{if } a \notin A; \\ 1 & \text{if } a \in A. \end{cases}$

The map χ is an order-isomorphism.

Each χ_A is called a **step** or **characteristic function** and finite real linear combinations of these are called **simple functions**.

Note that $\chi_A \vee \chi_B = \chi_{A\cup B}$, $\chi_A \wedge \chi_B = \chi_{A\cap B} = \chi_A \cdot \chi_B$, $\chi_{A-B} = \chi_A - \chi_{A\cap B}$ and $|\,\chi_A - \chi_B\,| = \chi_{(A-B)\cup(B-A)} = \chi_{A-B} + \chi_{B-A} = \chi_A + \chi_B - 2\chi_{A\cap B}$.

When $\mathbb{G} \subset \mathbb{P}(X)$, we will use $\boldsymbol{\mathcal{S}}(\mathbb{G})$ to denote the set of simple functions constructed from the sets in $\mathbb{G}$.

A function that has constant range value t on its whole domain will sometimes be denoted t, with this usage (and the domain) taken from context. Thus, for example, χ_X is sometimes denoted by 1 and $0\chi_X$ by 0, in yet another use of each of those symbols.

When $\mathbf{H}$ is a subset of $[-\infty,\infty]^X$, we will use $\boldsymbol{\mathcal{B}}(\mathbf{H})$ to denote the bounded members of $\mathbf{H}$; $f \in \boldsymbol{\mathcal{B}}(\mathbf{H}) \Leftrightarrow f \in \mathbf{H}$ and $\exists a \in \mathbb{R}$ with $0 \le a < \infty$ and $-a \le f \le a$.

If X is a topological space, $\boldsymbol{\mathcal{C}}(X)$ denotes the continuous functions from X to $\mathbb{R}$.

$\mathbb{R}^X$, $\boldsymbol{\mathcal{B}}(\mathbb{R}^X)$ and, when X has a topology, $\boldsymbol{\mathcal{C}}(X)$, are all **vector lattices**: real vector spaces and lattices.[3] They are also **commutative rings with multiplicative identity** χ_X.

7.1. ***Exercise.*** $\boldsymbol{\mathcal{S}}(\mathbb{G})$ *is obviously a (possibly empty) vector space. Give conditions on* $\mathbb{G}$ *under which* $\boldsymbol{\mathcal{S}}(\mathbb{G})$ *is a vector lattice and a commutative ring with multiplicative identity* χ_X.

8. The Axiom of Choice

In this section we introduce another axiom of set theory, the Axiom of Choice.

Every human language has grammar and vocabulary, and people communicate by arranging the objects of the language in patterns. We imagine that our communications evoke similar, or at least related, mental states in others. We also use these patterns to elicit mental states in our "future selves," as reminder of past imaginings so that we can start at a higher level in an ongoing project and not have to recreate each concept from scratch should we return to a task. It is apparent that our brains are built to do this.

But words are all defined in terms of each other. Ultimate meaning, if there is any to be found, is derived from pointing out the window at instances in the world, or from introspection. Very often ambiguity or multiple meaning of a phrase is the point of a given communication, and provides the richness and subtlety characteristic of poetry, for instance, or the beguiling power of political speech.

Set Theory is a language mathematicians have invented to encode mathematics. But unlike most human languages, this language does everything possible to avoid blended meaning, to expose the logical structure of statements and keep the vocabulary of undefined terms to an absolute minimum. Many mathematicians believe what they do is "art." But ambiguity and internal discord is not part of our particular esthetic ensemble.

Most mathematicians believe that, though set theory may be unfinished, it serves its purpose well. Virtually all mathematical structures can be successfully modeled in set theory, to the extent that most mathematicians never think of any other way of speaking or writing.

Together, the collection of axioms (which, along with logical conventions defines the language) normally used by most mathematicians is called the **Zermelo-Fraenkel Axioms**, or simply **ZF** and the set theory that arises from these axioms is called **Zermelo-Fraenkel Set Theory**. You saw explicit mention of two axioms from ZF, the Axiom of Infinity and the Axiom of the Empty Set, in Section 5. We have used others without mention on almost every page. For example we have formed power sets.

[3] A vector lattice is called, generally, a **Riesz space**. Real function vector lattices are examples.

> **The Axiom of the Power Set** For any set A there is a set $\mathbb{P}(A)$ consisting of all, and only, the subsets of A.

Asserting the existence of a set with this feature is a dramatic and "non-constructive" thing to do, *particularly* when the underlying set is infinite. We are not told how to create this set. We just have a means of recognizing if a set we have in hand is a member of this power set, or not.

And where, exactly, did that first infinite set come from? The Axiom of Infinity brings it into existence, out of nothing, simply because mathematicians *want infinite sets* and this seems to be a consistent way to produce them.

There is another extremely useful—and arguably even less constructive—axiom which we discuss now.

We will present and presume to be true, wherever convenient, the four equivalent and useful statements below, one of which is called the Axiom of Choice. This axiom is frequently abbreviated to **AC**. The collection of the axioms of standard set theory plus this axiom is frequently denoted **ZFC**.

The discussions regarding equivalence of the Axiom of Choice and the other three statements, and the history associated with them, is a fascinating story which deserves study by every serious student of mathematics.

> **The Axiom of Choice**: If J and X are sets and $A\colon J \to \mathbb{P}(X)$ is an indexed collection of nonempty sets then there is a function $f\colon J \to X$ such that $f(\beta) \in A_\beta \ \forall \beta \in J$. A function with this property is called a **choice function** for A.

Essentially, this axiom states that given any generic set $\mathbb{S}$ of nonempty sets, there is a way of selecting one element from each member of $\mathbb{S}$. The other axioms do not imply that such a selection can be made, unless every member of $\mathbb{S}$ has an element with some unique property, which would allow it to be singled out.

> **Zorn's Lemma**: If S is a set with a partial order and if every chain in S possesses an upper bound in S, then S has a maximal member.
>
> **Zermelo's Theorem**: Every nonempty set can be well-ordered.
>
> **Kuratowski's Lemma**: Each chain in a partially ordered set S is contained in a **maximal chain** in S (that is, a chain in S not contained in any other chain in S.)

Kuratowski's Lemma is also often called **The Hausdorff Maximal Principle.**

That Zorn's Lemma implies Kuratowski's Lemma is immediate. Suppose S is a set with a partial order and and C is a chain in S. Let $\mathbb{W}$ denote the set of all chains in S which contain the chain C, ordered by containment. Any chain in $\mathbb{W}$ is bounded above by the union of the chain, so Zorn's Lemma implies that $\mathbb{W}$ contains a maximal member. That maximal member is a chain in S not properly contained in any other chain in S.

On the other hand, assuming Kuratowski's Lemma to be true, suppose S is a set with a partial order and that every chain in S possesses an upper bound in S. This time let $\mathbb{W}$ denote the set of *all* chains in S. Let X denote a maximal member of $\mathbb{W}$. So X is a chain in S not contained in any other chain. Let M be any upper

bound for X. By maximality of X, M must actually be in X and cannot be less than any other member of S: that is, M is maximal in S. So Zorn's Lemma is true.

In the last two paragraphs we have shown that Zorn's Lemma and Kuratowski's Lemma are equivalent statements.

We will now show that Zorn's Lemma implies AC. Suppose $\mathbb{S}$ is any nonempty set of nonempty sets and X is the union of all the sets in $\mathbb{S}$. Let $B = \mathbb{S} \times X$. Now let Q denote the set of all subsets of $\mathbb{P}(B)$ which are choice functions on their domains: that is, $T \in Q$ exactly when T is nonempty and there is at most one ordered pair in T whose first component is any particular member of $\mathbb{S}$, and also $s \in A$ whenever $(A, s) \in T$. These are called "**partial choice functions**." Order Q by containment. The union of any chain in Q is a member of Q so Zorn's Lemma implies that Q has a maximal member. This maximal member is a choice function on its domain, which must by maximality be all of $\mathbb{S}$.

The fact that Zermelo's Theorem implies AC is also straightforward: given any nonempty set $\mathbb{S}$ of nonempty sets, well-order the set $X = \bigcup_{S \in \mathbb{S}} S$. For each $S \in \mathbb{S}$ let $f(S)$ be the least element of S with respect to this ordering. f is the requisite choice function.

The opposite implication is a bit trickier. It involves using a choice function to create the well-order.

Suppose set A has more than one element and $f\colon \mathbb{P}(A) - \{\varnothing\} \to A$ is a choice function: that is, $f(B) \in B$ whenever $\varnothing \neq B \subset A$.

Let $\mathbb{B}$ denote the set of all nonempty containment-chains in $\mathbb{P}(A) - \{\varnothing\}$ which are well-ordered and satisfy the condition:

> Whenever I_K is an initial segment of one of these chains and if J is the union of all the sets in I_K then $J \neq A$ and $K = J \cup \{f(A - J)\}$.

$\mathbb{B}$ is nonempty: for example, $\{\,\{f(A)\}, \{\,f(A), f(A - \{f(A)\})\,\}\,\}$ is in $\mathbb{B}$.

The condition above implies that each of the chains in $\mathbb{B}$ must start with the set $\{f(A)\}$, and the successor to any set K in such a chain (if, of course, K is not the last set in the chain) has exactly one more member than does K. It also implies directly that if two different chains X and W of this kind have a common initial segment, so that $I_K \subset X$ and $I_G \subset W$ and $I_K = I_G$ then $K = G$. In other words, the least successor of an initial segment is determined by the sets in the initial segment, and *not* by the specific chain within which the initial segment sits.

Suppose that X is one of these chains. We will call S a "starting chunk" of X if $\varnothing \neq S \subset X$ and whenever $B, C \in X$ the condition $B \in S$ and $C \subset B$ implies $C \in S$. Now it might be that a starting chunk is as short as $\{f(A)\}$ or it could, possibly, be all of X. But if it is *not* all of X then because X is well-ordered there is a least member K of X not in S and so S contains all members of X less than K. That is, $S = I_K$ for some $K \in X$. So starting chunks are either initial segments or the entire chain.

Now suppose X and W are unequal chains, members of $\mathbb{B}$. Then one, say X, would contain a least set K not in the other. The initial segment I_K of X is contained in W. If there were a set in W not in I_K but less than some member of I_K then there would be a least member of W of this kind. Call that least member

G. But then the initial segment I_G of W would be a starting chunk of X and by the above remark we would have $G \in X$, contrary to its definition.

So there are no missing members of W between members of I_K, which is therefore a starting chunk of W. Since $K \notin W$ we must have $I_K = W$, and conclude that W is an initial segment of X.

To reiterate: for each pair of members of $\mathbb{B}$, one is an initial segment of the other.

Now let S be the union of all the members of $\mathbb{B}$. Each set in S comes from a member of $\mathbb{B}$ and since one of any pair of members of $\mathbb{B}$ is an initial segment of the other we conclude that S itself is a chain, and well-ordered too.

Let J denote the union of all the sets in S. If $J \neq A$ then we could extend S to $S \cup \{J \cup \{f(A - J)\}\}$ which satisfies the conditions for membership in $\mathbb{B}$ but is strictly longer than its longest member, a contradiction. We conclude that $J = A$.

So we can use S to create an order on A. If a and b are members of A there is a least member S_a of S containing a and a least member S_b containing b. Declare $a \leq b$ precisely when $S_a \subset S_b$. If J is the union of the sets in the initial segment determined by S_a then $a \notin J$ so it must be that $a = f(A - J)$. So this relation makes A into a total order. Further, if $\varnothing \neq T \subset A$ then the collection of all of the S_t with $t \in T$ has a least member, which produces a least member of T. So the order on A is a well-order.

We conclude that the existence of a choice function on $\mathbb{P}(A) - \{\varnothing\}$ implies that A can be well-ordered. So AC implies Zermelo's Theorem.

Upon accepting the Axiom of Choice, as we will do throughout this book, well-ordered sets are plentiful and can be used.

At this point we have shown the following implications among the conditions which we claim to be equivalent to the Axiom of Choice.

$$\begin{array}{ccc} \text{Zorn} & \iff & \text{Kuratowski} \\ \Downarrow & & \\ \text{AC} & \iff & \text{Zermelo} \end{array}$$

The **Principles of Induction** and **Recursive Definition** are incredibly powerful and useful techniques, extending the idea of Induction on the Integers to many more well-ordered sets and situations more varied than merely checking if an indexed set of propositions are all true. The methods are detailed in Section 14. It is important to note, and the reader should check, that the proof of the version of Recursive Definition we use here does *not* require AC.

We will now use Induction and Recursive Definition to show that Zermelo's Theorem implies Kurotowski's Lemma, thereby proving that any of the four conditions listed above implies the others. The discussion below is a typical usage of this type of argument. It uses first a recursive definition to deduce that a certain function exists, and then induction to confirm various properties of that function.

We suppose we have a chain in a partially ordered set. We will line up the members of the set not already in the chain and test them one at a time. When it is an element's turn, if it can be added to yield a bigger chain than we have up to that point we select it. Otherwise we discard it. Then we go on to the next element

and repeat until we have exhausted the possibilities. The product is a maximal chain. A rigorous justification can be produced after digesting the result in 14.3

Assume Zermelo's Theorem to be true, and suppose H is a nonempty chain in set K with partial order $\precsim$. Suppose $B = K - H$ is nonempty. There is a well-order $\leq$ for B. Since we have two orders in hand, we will use prefixes to describe which order is in use. We will let I_β stand for a $\leq$-initial segment for any $\beta \in B$.

Suppose y is a fixed element of H. For the $\leq$-first element α of B, let $P(\alpha)$ equal α if $H \cup \{ \alpha \}$ is a $\precsim$-chain, and let $P(\alpha)$ be y otherwise.

Having found $P(\beta)$ for all $\beta \in B$ with $\alpha \leq \beta < \gamma$ for some $\gamma \in B$ define $P(\gamma)$ to be γ if $H \cup \{ \gamma \} \cup P(I_\gamma)$ is a $\precsim$-chain, and let $P(\gamma)$ be y otherwise.

This serves to define $P(\gamma)$ for each $\gamma \in B$.

$H \cup P(B)$ must be a $\precsim$-chain: if not it must contain two $\precsim$-incomparable members s and t, which cannot both be in H. If one of the two, say s, is in H then there is a $\beta \in B$ with $P(\beta) = t = \beta$. But then $H \cup \{ \beta \} \cup P(I_\beta)$ is not a chain, violating the defining condition for $P(\beta)$. A similar contradiction occurs if neither s nor t are in H, by examining the point at which the *second* of the two points would have been added. So in fact $H \cup P(B)$ must be a $\precsim$-chain.

No additional members of K can be added to $H \cup P(B)$ without causing the resulting set to fail to be a $\precsim$-chain: once again, letting γ be the $\leq$-least member of B which *could* be added, if any, yields a contradiction. That member *would* have been added at stage γ.

So $A \cup P(B)$ is a maximal $\precsim$-chain in K, and Kuratowski's Theorem holds.

8.1. ***Exercise.*** *Fill in the details of a direct proof using Induction and Recursive Definition that Zermelo's Theorem implies Zorn's Lemma. We assume that K is a set with partial order $\precsim$ for which every chain has an upper bound. We assume also that K has a well-order $\leq$ with $\leq$-first member α.*

We would like to conclude that K has a $\precsim$-maximal element.

Let α denote the $\leq$-first member of K and define $G(\alpha) = \alpha$. Having defined G on $\leq$-initial segment I_β for $\beta > \alpha$ let $G(\beta) = \beta$ if β is a $\precsim$-upper bound for $G(I_\beta)$, and otherwise let $G(\beta) = \alpha$.

Show that $G(K)$ is a chain and that $G(K)$ has a $\precsim$-last member, which is $\precsim$-maximal in K.

8.2. ***Exercise.*** *(i) An axiom equivalent to our Axiom of Choice is produced if we add to that axiom the condition that $A_\alpha \cap A_\beta = \varnothing$ whenever $\alpha \neq \beta$.*

(ii) Consider the statement: "Whenever $\mathbb{B}$ is a nonempty set of nonempty pairwise disjoint sets, there is a set S for which $S \cap x$ contains a single element for each $x \in \mathbb{B}$." Show that this statement is equivalent to the Axiom of Choice.

(iii) Let $\mathbb{B}$ be a (nonempty) set of sets. $\mathbb{B}$ is said to have ***finite character*** *provided that $A \in \mathbb{B}$ if and only if every finite subset of A is in $\mathbb{B}$.* ***Tukey's Lemma*** *states that every set of sets of finite character has a maximal member: a set not contained in any other member. Show that Tukey's Lemma is equivalent to the Axiom of Choice. (hint: To prove that Tukey's Lemma implies the Axiom*

of Choice examine the set of partial choice functions and note that it satisfies the conditions of Tukey's Lemma.)

The use of AC in the formation of mathematical arguments has historically been the subject of controversy centered around the nebulous nature of the objects whose existence is being asserted in each case. In applications the axioms of set theory are usually used to affirm the existence of one precisely defined set whose elements share an explicit property. That is less obviously the case when AC is invoked.

Applications which require less than the full strength of AC are common. In an effort to control, or at least record, how the axiom is being used, weaker variants have been created. Some mathematicians award "style points" to proofs using one of these, or which avoid AC altogether. We list two of these weaker versions of AC below.

> **The Axiom of Dependent Choice**: If X is a nonempty set and $R \subset X \times X$ is a binary relation with domain all of X, then there is a sequence $r\colon \mathbb{N} \to X$ for which $(r_k, r_{k+1}) \in R\ \forall k \in \mathbb{N}$.
>
> **The Axiom of Countable Choice**: If X is a nonempty set and $r\colon \mathbb{N} \to \mathbb{P}(X)$ is a sequence of nonempty subsets of X then there is a sequence $f\colon \mathbb{N} \to X$ such that $f(n) \in A_n\ \forall n \in \mathbb{N}$.

These axioms are frequently abbreviated to **DC** and $\mathbf{AC}_\omega$, respectively.

8.3. ***Exercise.*** *(i) Prove the implications $AC \Rightarrow DC \Rightarrow AC_\omega$.*

(ii) Suppose X is infinite. For each $k \in \mathbb{N}$ let S_k denote the set of all subsets of X which have 2^k elements. ZF alone implies that S_k is nonempty for each k, and you may assume this. Let S denote the set of all the S_k. Use AC_ω twice to prove that there is a one-to-one function $f\colon Y \to \mathbb{N}$ for an infinite subset Y of X. Any set Y (infinite or not) for which there is a one-to-one function $f\colon Y \to \mathbb{N}$ is called ***countable****, and the result here may be paraphrased as "Any infinite set has an infinite countable subset in ZF+AC_ω."*

(iii) Sometimes the use of an axiom, particularly a variant of the Axiom of Choice, is hard to spot in an argument. It seems so reasonable, it is hard to see you are assuming anything. The theorem that "The union of a countable set of countable sets is countable." is an example.

Suppose A is a countable set of countable sets, and let B denote the union of all the members of A. Because A is countable, there exists one-to-one $T\colon A \to \mathbb{N}$. Because each member of A is countable, for each nonempty set $S \in A$ there is a nonempty set F_S consisting of all one-to-one functions from S to $\mathbb{N}$. Using T, this collection of sets of functions is seen to be countable, so AC_ω guarantees that we can pick a function from each. It is easy to overlook this step, and merely assert "Because each member of A is countable there exists one-to-one $G_S\colon S \to \mathbb{N}$ for each $S \in A$." and get on with the discussion using these selected functions. But it is AC_ω which endorses this selection.

To finish the argument, for each $x \in B$ we let $A_x = \{\, S \in A \mid x \in S \,\}$ and define i_x to be the least integer in $T(A_x)$. We define W_x to be that member of A_x with

$T(W_x) = i_x$. *The function* $H\colon B \to \mathbb{N}$ *given by*

$$H(x) = 2^{i_x} \cdot 3^{G_{W_x}(x)}$$

is one-to-one, so B *is countable.*

8.4. ***Exercise.*** *(i) Any chain in a tree is contained in a branch.*

(ii) Prove ***König's Tree Lemma****: If* S *is an infinite rooted tree but each* $t \in S$ *is the immediate predecessor of only finitely many members of* S *then* S *has an infinite branch. (hint: Let* K *denote those members of* S *with an infinite number of successors and for each* $t \in K$ *let* $M_t = T_t \cap K - \{t\}$*. Let* f *denote a choice function for these sets:* $f(t) \in M_t\ \forall t \in K$*. Use induction on* $\mathbb{N}$ *to create an infinite chain.)*

The next section contains another important consequence of the Axiom of Choice. Many more can be found scattered in appendices and chapters throughout this book.

Those who want a slightly more detailed look at the ZF axioms can find them listed in Sections 11 and 13. The discussions there are rudimentary but, I hope, a practical guide providing a taste of modern set theory.

9. Nets and Filters

Suppose $r\colon J \to X$ is a net. Recall that this means that J is pre-ordered and there is an upper bound in J for each two-element subset of J.

If $A \subset X$, r is said to be **in** A if $r(J) \subset A$. r is said to be **eventually in** A if there is a terminal segment $T_\alpha \subset J$ such that $r(T_\alpha) \subset A$.

r is said to be **frequently in** A if $r(T_\alpha) \cap A \neq \varnothing\ \forall$ terminal segments T_α in J.

Obviously, if r is eventually in A then r is frequently in A.

A **subnet of** r is a net $s\colon K \to X$ such that $\exists f\colon K \to J$ for which $s = r \circ f$ and $\forall m \in J\ \exists n \in K$ such that $f(T_n) \subset T_m$. Note that f is not presumed to be non-decreasing. It is simply eventually in any terminal segment of J. A subnet of a subnet is also a subnet of the original net.

A net in a set X is called **universal** if the net is eventually in A or eventually in A^c for all $A \in \mathbb{P}(X)$.

9.1. ***Proposition.*** *Each net* $r\colon D \to X$ *has a universal subnet.*

Proof. Let $\mathbb{M} \subset \mathbb{P}^2(X) = \mathbb{P}(\mathbb{P}(X))$ be the set of all those $\mathbb{G} \subset \mathbb{P}(X)$ such that r is frequently in each member of $\mathbb{G}$ and also if $A, B \in \mathbb{G}$ then $A \cap B \in \mathbb{G}$.

Obviously $\{X\} \in \mathbb{M}$ so $\mathbb{M} \neq \varnothing$, and chains in $\mathbb{M}$ ordered by inclusion have upper bounds in $\mathbb{M}$ (the union of the chain) so $\mathbb{M}$ contains a maximal member $\mathbb{K}$.

If r is eventually in A or eventually in A^c then the maximality of $\mathbb{K}$ guarantees that one or the other is in $\mathbb{K}$. It remains to consider the case where r is frequently in A but $A \notin \mathbb{K}$. By maximality of $\mathbb{K}$ there must be some $S \in \mathbb{K}$ so that r is not frequently in $A \cap S$: that is, r is eventually in $(A \cap S)^c$. Now, if T is any member

of $\mathbb{K}$, r is frequently in $T \cap S = (T \cap S \cap A) \cup (T \cap S \cap A^c)$ so r must be frequently in $T \cap S \cap A^c \subset T \cap A^c$. This is true for any $T \in \mathbb{K}$ so by maximality of $\mathbb{K}$, $A^c \in \mathbb{K}$.

We have just shown that either A or $A^c \in \mathbb{K}\ \forall A \in \mathbb{P}(X)$.

Now let $E = \{\, (\alpha, B) \mid \alpha \in D,\ B \in \mathbb{K} \text{ and } r(\alpha) \in B \,\}$ directed by $(\alpha, B) \leq (\beta, C)$ precisely when $\alpha \leq \beta$ and $B \supset C$. The net $s\colon E \to X$ defined by $s((\alpha, B)) = r(\alpha)$ is a subnet of r and universal by construction. □

We now move on to the next idea of this section.

A nonempty subset $\mathbb{F}$ of $\mathbb{P}(X)$ is called a **filterbase on** X if

$$(a) \qquad \varnothing \notin \mathbb{F} \qquad \text{and}$$
$$(b) \qquad A, B \in \mathbb{F} \Rightarrow A \cap B \in \mathbb{F}.$$

If the additional condition

$$(c) \qquad A \in \mathbb{P}(X),\ B \in \mathbb{F} \Rightarrow A \cup B \in \mathbb{F}$$

holds, $\mathbb{F}$ is called a **filter on** X .

Given any nonempty subset $\mathbb{F}$ of $\mathbb{P}(X)$ for which finite intersections of members of $\mathbb{F}$ are nonempty there is a unique smallest filterbase containing $\mathbb{F}$. Each filterbase is contained in a unique smallest filter. This filterbase and this filter are said to be **generated by** $\mathbb{F}$.

The most common example of a filterbase is the collection of all open sets containing a particular point of a topological space. A filter containing this filterbase would be the set of all neighborhoods of that point.

Another filterbase would be $\{\, (0, a) \subset (0, \infty) \mid a > 0 \,\}$.

Yet another example is given by the following: Let $r\colon D \to X$ be a net in X. Let $\mathbb{F} = \{\, r(T_d) \mid d \in D \,\}$, where each T_d is a terminal segment of D. $\mathbb{F}$ is a filterbase. The collection of all sets containing any terminal segment, $\mathbb{G} = \{\, A \in \mathbb{P}(X) \mid r(T_d) \subset A \text{ for some } d \in D \,\}$, is the smallest filter containing $\mathbb{F}$.

Let $\boldsymbol{\mathcal{F}}$ denote the set of filters on X. $\boldsymbol{\mathcal{F}}$ is partially ordered by containment. If $\{\, \mathbb{F}_\alpha \mid \alpha \in J \,\}$ is any chain of filters then $\bigcup_{\alpha \in J} \mathbb{F}_\alpha$ is also a filter and an upper bound for the chain. So $\boldsymbol{\mathcal{F}}$ possesses maximal members called **ultrafilters.**

When $\mathbb{G}$ is a filterbase, $\{\, \mathbb{F} \in \boldsymbol{\mathcal{F}} \mid \mathbb{F} \supset \mathbb{G} \,\}$ is nonempty and possesses maximal members, which are maximal in $\boldsymbol{\mathcal{F}}$ as well. So any filterbase is contained in an ultrafilter.

It is not hard to show that a filter $\mathbb{F}$ on X is an ultrafilter if and only if whenever $A \in \mathbb{P}(X)$ then $A \in \mathbb{F}$ or $A^c \in \mathbb{F}$.

This provides a link between universal nets and ultrafilters.

If $\mathbb{F}$ is the filterbase on X formed from the net r as above, let $s\colon E \to X$ be a universal subnet of r. Let $\mathbb{K} = \{\, A \in \mathbb{P}(X) \mid s(T_d) \subset A \text{ for some } d \in E \,\}$. Since s is universal, either A or A^c is in $\mathbb{K}\ \forall A \in \mathbb{P}(X)$. $\mathbb{K}$ is an ultrafilter containing $\mathbb{F}$.

Alternatively, suppose $\mathbb{F}$ is any filter and $\mathbb{G}$ is the ultrafilter generated by $\mathbb{F}$. Let $J = \{\, (x, A) \mid x \in A \in \mathbb{F} \,\}$ and $K = \{\, (x, A) \mid x \in A \in \mathbb{G} \,\}$. Direct J and K by $(x, A) \leq (y, B)$ if and only if $A \supset B$. We define $r\colon J \to X$ by $r((x, A)) = x$ and $s\colon K \to X$ by $s((x, A)) = x$. The filters $\mathbb{F}$ and $\mathbb{G}$ are precisely the sets formed from

terminal segments of J and K by r and s, respectively. If we define $f\colon K \to J$ by $f((x, A) = (x, X)$, then $s = r \circ f$ and it follows that s is a subnet of r. Moreover, s is a universal net.

Suppose X is a nonempty set and $p \in X$. Let $\mathbb{F}_p$ denote the collection of all subsets of X containing p. $\mathbb{F}_p$ is an ultrafilter, and ultrafilters of this type are called **principal**. Other kinds of ultrafilters are called **free**.

If X is nonempty let $\mathbb{K}$ denote the set of **cofinite** subsets of X: that is, all subsets S of X for which $X - S$ is a finite set. If X is finite, $\mathbb{K} = \mathbb{P}(X)$. But if X is infinite, $\mathbb{K}$ is a filter on X, the **filter of cofinite subsets of X**.

9.2. ***Exercise.*** *(i) If $V_1, V_2, \ldots, V_n$ is a finite partition of X and $\mathbb{F}$ is an ultrafilter on X then $\mathbb{F}$ contains exactly one of the V_i.*

(ii) If an ultrafilter on X contains a finite set it contains a one point set, and is principal.

(iii) Suppose $\mathbb{U}$ is an ultrafilter on infinite X. $\mathbb{U}$ is free exactly when $\mathbb{U}$ contains all cofinite subsets of X.

(iv) There is a free ultrafilter $\mathbb{U}$ on $\mathbb{N}$ containing the set of even natural numbers. There is another containing the set of odd natural numbers. In fact if A is any infinite subset of $\mathbb{N}$ there is a free ultrafilter on $\mathbb{N}$ containing A.

10. Rings and Algebras of Sets

Consider $\mathbb{P}(X)$, the power set on the set X. When there is no danger of ambiguity and $A \subset X$, the notation A^c is often seen in place of $X - A$. $\mathbb{P}(X)$ is partially ordered by containment. $\mathbb{P}(X)$ is a lattice, with $A \wedge B = A \cap B$ and $A \vee B = A \cup B$. There is also additional structure on subsets of $\mathbb{P}(X)$.

$\mathbb{G} \subset \mathbb{P}(X)$ will be referred to as a **ring in X** if

(i) $\quad \varnothing \in \mathbb{G}$

(ii) $\quad A, B \in \mathbb{G} \Rightarrow A - B \in \mathbb{G}$ $\quad$ and

(iii) $\quad A, B \in \mathbb{G} \Rightarrow A \cup B \in \mathbb{G}$.

The last two items can be rephrased by saying that $\mathbb{G}$ is **closed** with respect to the operations $\cup$ and $-$.

$\mathbb{G} \subset \mathbb{P}(X)$ will be referred to as an **algebra on X** if, in addition

(iv) $\quad X \in \mathbb{G}$

Items (ii) and (iv) imply that $\mathbb{G}$ is **closed** with respect to the operations $\cup$ and ^{c}. In the presence of (iii), this last statement implies (ii).

It is apparent that item (i) is redundant in the presence of (ii) and (iv). Also, if $\mathbb{G}$ is a ring in X and A and B are in $\mathbb{G}$ then so is $A \cap B$. In fact, item (iii) could be replaced by "if A and B are in $\mathbb{G}$ then so is $A \cap B$" to yield equivalent definitions for a ring in a set.

10.1. ***Exercise.*** *(i) Show that the smallest algebra on X containing a topology for X consists of all sets $A \cap B$ or $A \cup B$ where A is an open set and B is a closed set.*

(ii) If $\mathbb{G}$ is a ring in X then both $\{ A \in \mathbb{P}(X) \mid A \in \mathbb{G}$ or $A^c \in \mathbb{G} \}$ and $\{ A \in \mathbb{P}(X) \mid A \cap B \in \mathbb{G}$ whenever $B \in \mathbb{G} \}$ are algebras on X.

*(iii) The set of "**clopen sets**" (that is, sets that are both open and closed) in a topology for X constitutes an algebra on X.*

10.2. ***Exercise.*** *(i) Suppose $\mathbb{G}$ is a ring in X. Define multiplication in $\mathbb{G}$ by $A \cdot B = A \cap B$ and addition by **symmetric difference**:*

$$A \triangle B = (A - B) \cup (B - A).$$

Show that $\mathbb{G}$ is a commutative (algebraic) ring with these operations. This ring has identity when the union of all sets in $\mathbb{G}$ is a member of $\mathbb{G}$. An additive subgroup $\mathbb{S}$ of this ring is an ideal exactly when $A \subset B$, $A \in \mathbb{G}$ and $B \in \mathbb{S}$ imply $A \in \mathbb{S}$.

(ii) Suppose X is infinite and $\mathbb{G}$ is a ring in X. Let $\mathbb{S}$ denote the finite members of $\mathbb{G}$. Then $\mathbb{S}$ is an ideal.

When thinking of rings and algebras of sets, bear in mind the following two.

All finite unions of bounded subintervals of $\mathbb{R}$ constitute a ring in $\mathbb{R}$.

An algebra on $\mathbb{R}$, obviously closely related to this ring, would be all finite unions of subintervals (bounded or not) of $\mathbb{R}$.

Apart from its raw defining qualities, an algebra on a set has useful properties which will be used throughout this work. Turn to Section **??** on Boolean Algebras and Rings to find out how some of these properties, when extracted and studied on their own, necessarily return to their roots as an algebra on a set.

Because of the way we handle the material of later chapters, rings in a set will be less common than algebras on a set.

Ordinals and Cardinals October 18, 2023

Might not a mouse
in iron grip of owl, review
his forest world
in wonder 'midst his fear?

And see his meadow home below,
and tree and stream as new,
and think
"how beautiful from here?"

The history of the subject of this appendix, aspects of set theory, began in its modern form with the daring genius of George Cantor in the late nineteenth century. The subject was vigorously promoted by David Hilbert in the early twentieth century, who sought a solid foundation for mathematics, free of **antinomies**, or internal contradiction. This subject has had more than its fair share of detractors and champions and schisms and creative personalities.

In addition to Cantor and Hilbert, these subjects were organized and formulated by Frege, Skolem, Brouwer, Hausdorff, Zermelo, Fraenkel, Bernays, von Neumann and many others.

Kurt Gödel's astonishing incompleteness results in 1931 and Paul Cohen's invention of the technique of forcing in 1963 deserve special mention.

Constraints of space prevent more than a cursory inspection of mathematical logic. We will, however, consider quite a few topics used by mathematicians which have a "set theoretic aspect."

Axiomatic set theory is the study of the consequences of the unproven assumptions about undefined primitive objects and how the existence of sets with certain properties may be inferred. It constitutes the underpinnings of virtually all modern mathematics.

An excellent source for the history of set theory and variations on these topics may be found in Fraenkel, Bar-Hillel and Levy *Foundations of Set Theory* [**?**].

11. The Axioms of ZF Set Theory: Part One

We list and discuss in the next few sections the collection of axioms and the vocabulary normally employed by most mathematicians to communicate news of, and justify claims about, their mathematical creations.

Set theory is meaningless to humans without some very clear ideas, in advance, of the structures set theory is to model. On the most simplistic level, I think of sets as bags of objects with some common characteristic, and elements as the objects in the bags. Set theory does not tell me this. I just use this mental image to help me organize my thinking about sets.

On its own, set theory is nothing more (or less) than a juxtaposition of symbols satisfying rules. There is no *reason* for a language without something to communicate. That is why the study of subject areas—groups, rings, real numbers, measures, topological spaces, bags and so on—must, practically, come first. We will then find that set theory helps us feel more confident that inconsistency has not maneuvered its way into our forest, at least in an obvious way, while our attention was devoted to trees.

The point must be made that set theory does not "construct" anything, even if that vocabulary is ubiquitous. We *presume*, at the outset, that there is a collection of "objects-of-the-mind," a universe of discourse we will call $\mathcal{S}\mathbf{et}$. Each individual object is called "a" **set**. We will assign names to some of these sets in the course of a mathematical argument, usually letters or a combination of letters. After much debate, mathematicians have largely agreed on some properties that the "objects-of-their-collective-mind" must possess. These properties are called axioms. Axioms are expressed by reference to the names of sets. They represent assumed-truths about how $\mathcal{S}\mathbf{et}$ works, and though they cannot be *proven*, these properties could potentially be contradictory.

Contradictions don't present much of a problem for us in most aspects of our lives but for the intended purpose of $\mathcal{S}\mathbf{et}$ they would be deadly, and the ease with which they can be formed in ordinary human language worrisome.

Whatever axioms we adopt, we insist that it is possible to name any set without contradiction. We do *not* intend to imply that we can unambiguously describe each set—only that the act of assigning a name to a set with a given property must not produce, alone, contradiction.

If we find such a contradiction we have decisions to make.

First, if the axioms we use guarantee the existence of this set we must abandon or change one or more axioms. The choice here is only about which we modify.

Second, (and this is the usual situation) we could simply accept that the set we thought to name was an illusion. The "name" we gave it actually names nothing. There are no sets with the given property.

If we cannot bear to give up this set we must, instead, give up or change one or more of the axioms. Perhaps we were mistaken in our perception of one or more of these elemental features of $\mathcal{S}\mathbf{et}$.

As a variation on the theme we might find not a contradiction but instead strange and horrifying sets, thrust forward perhaps by the left hand of AC, stumbling and blinking into the bright lights of mathematical center stage. We might be willing to abandon cherished axioms to expunge these stains.

The latter paths would not be taken lightly. They would represent a bifurcation in the collective vision of $\mathcal{S}\mathbf{et}$. Heresy is a serious matter for any primate, with consequences.

We will ask critical questions about the axioms. Among them:

- Is our list of axioms consistent, or do they conflict with each other? If they conflict, we would interpret that to mean that we had misunderstood

some properties of **Set**, through wishful thinking or some other human propensity.

- Properties of **Set** are intended to mirror properties of interest to mathematicians. We want to deduce, from obvious properties of **Set**, those that are much less obvious. Is our list of axioms rich enough to allow us to make logical inferences about mathematically interesting topics? About *any* mathematical topic? Decide *any* interesting mathematical question?
- Does our exploration of these consequences suggest previously unconsidered qualities which **Set** could have? These are properties that might have been placed among the axioms, but their relevance was unknown at the outset so no one thought to include them. Might there be consequences of these properties so compelling as to *change our perception* of **Set**?

Setting aside most of the ontological and epistemological issues as beyond the author, let us proceed to outline set theory itself. If you are troubled by the lack of referents for some of the words below, or want more precise or detailed formulations of these statements, your option is to study logic and set theory until the feeling passes. We note that though the *mathematics* in this appendix follows the standard path, our philosophical ruminations take one tack among many. Such matters can be the source of *spirited* debate.

Our axioms are called the **Zermelo-Fraenkel Axioms** and the set theory associated with these axioms is called **Zermelo-Fraenkel Set Theory**. The ten axioms are divided into two groups. The set theory associated with first nine is denoted ZF while that associated with all ten, including the Axiom of Choice, is denoted ZFC. Standard set theory includes all ten.

We have used these axioms on virtually every page of this work.

Let's get specific. The first line of Section 1 refers to a product set $S \times T$, consisting of **ordered pairs** (a, b) where $a \in S$ and $b \in T$. This ordered pair is defined to be $\{\{a\}, \{a, b\}\}$ which is a subset of $\mathbb{P}(S \cup T)$. So $S \times T$ is a particular element of $\mathbb{P}^3(S \cup T)$. We know what an ordered pair is. We knew before we ever started this discussion. This is simply one way (the most common way) to *model* the idea in a universe consisting of nothing but sets.

A function f from S to T would also be an element of $\mathbb{P}^3(S \cup T)$. A set of such functions is a member of $\mathbb{P}^4(S \cup T)$. If $A \subset S$ then $f(A)$ is a subset of T. An order on S is an element of $\mathbb{P}^3(S)$. The very existence of these sets and their properties can only be inferred from axioms, the axioms of ZF.

Not all of the ten axioms below are independent of each other. Some combinations imply others. We will not worry here about assembling the minimal list of axioms, just a useful one, free of obvious inconsistencies and able to generate structures rich enough to model most of mathematics.

These axioms make reference to the names of sets and the binary "element of" relation between sets, usually denoted $\in$. All other relations between sets, such as $\subset$ or $=$ or $\neq$ or $\notin$ are defined in terms of this relation. Set theory does not presume to tell us what this relation means. We provide the meaning ourselves, part of our vision of **Set**.

11.1. **Axiom of the Empty Set**
There exists a set, denoted $\varnothing$, which has no elements.

This axiom gets us started. At least, we're talking about *something*, even if it is nothing. A set without members is called an **individual.** Our vision of **Set** calls for one. $x \in \varnothing$ is always false, no matter what set x represents.

11.2. **The Axiom of Extensionality**
Sets are equal exactly when they have the same elements.

This axiom constitutes the *definition* of the equality relation in ordinary set theory. An implication of this axiom is that sets have properties of any kind solely by virtue of the elements (also called members) they contain. To show a set A equals a set B you need only show that $A \subset B$ and $B \subset A$. Sets with the same elements cannot be distinguished: they are the same set. There is, for instance, but a single individual in our set theory.

11.3. **The Axiom of Pairing**
If A and B are sets there is a set whose elements are, exactly, A and B.

It follows immediately that ordered pairs with specified first and second elements can be created. Also, letting $A = B$ we can create the nested sets $\{A\}$, $\{\{A\}\}$, $\{\{\{A\}\}\}$ and so on.

11.4. **Axiom of Infinity**
There exists a set A with $\varnothing \in A$ and such that whenever X is a set and $X \in A$ then $X \cup \{X\} \in A$.

This axiom makes the structure we are creating **much** richer by capturing one view of an infinite set.

11.5. **The Axiom of Union**
If S is a set of sets there is a set whose elements are exactly the elements of the members of S.

Rephrasing, this axiom allows us to infer that the union of any family of sets is itself a set *so long as this family is indexed by a set.* This condition is a "size" restriction, and in conjunction with other size restrictions in the two axiom schema (below) *seems* to have struck a balance between the desire to infer the existence of—or "select" or "create"—sets based on any clear criterion and the need to avoid contradictory axioms.

11.6. **The Axiom of the Power Set**
For any set A there is a set $\mathbb{P}(A)$ consisting of all, and only, the subsets of A.

The Axiom of the Power Set is not constructive. It has nothing to say about which sets these are, or how many there are. It only tells you how a set can be identified as a member of the class $\mathbb{P}(A)$, and that this class is a *set*, eligible to participate as a first-class citizen in building other sets.

The effect of including this axiom is to guarantee that if, *by any means*, we ever find ourselves in possession of a member of **Set** all of whose members are in a set A then that set is an element of the *set* $\mathbb{P}(A)$.

In combination with other axioms several of which provide methods of identifying subsets we will see that $\mathbb{P}(A)$ can be **huge** in comparison to A. Exactly *how huge* is an interesting question.[4]

12. Properties of Sets

We are going to tiptoe around basic issues involving **logic**, with apologies.

First, we will use but not justify certain principles of mathematical logic such as the realization of the statement $A \Rightarrow B$ as $B \vee \neg A$ and the like.

Second, we will allude to "explicit properties of sets." We mentioned above that "sets have properties solely by virtue of the elements they contain" but we never codified *how* such properties might be specified.

We can't leave it to common sense and assume everyone will know what to do.

In ordinary language, for instance, we can consider the old chestnut concerning the Spanish village in which the barber (a man) shaves every man who doesn't shave himself. This sentence certainly *seems* meaningful. But who shaves the barber?

Call an adjective "anti-adjectival" if it does not describe itself. For instance "polysyllabic" describes itself and so is *not* anti-adjectival, while "monosyllabic" *is* anti-adjectival. It certainly seems that this adjective should either apply to itself or not. Is "anti-adjectival" anti-adjectival?

These are essentially "one-step" contradictions, but one can imagine nested statements whose contradictory nature is far less obvious. In an attempt to protect ourselves from fallacy we must severely limit how we form meaningful statements when those statements are to be used in conjunction with axioms to deduce the existence of sets.

A statement of the type we will use in this context will be an ordered collection of symbols, called a **formula**. These symbols are ordered left-to-right on a page, and there is an implication of temporal order: some symbols are to be examined before others when they are interpreted. Our formulae can be reduced to a finite list of symbols involving the names of sets, the logical symbols "and" ($\wedge$) and "not" ($\neg$), together with the quantifier "there exists" ($\exists$) and the "element of" relation ($\in$), which is a relation between pairs of *sets*. We include parentheses to enforce temporal order in complicated expressions.

As with the "element of" relation, set theory does not tell us the meaning of the quantifier or in fact any of these symbols. Meaning comes from *us*, from our vision

[4]The size of a set can be defined by its cardinality, as in Section 16. One has the feeling that the statement "$cardinal(A) < cardinal(B) \Longrightarrow cardinal(\,\mathbb{P}(A)\,) < cardinal(\,\mathbb{P}(B)\,)$" should be obviously true, yet it is known to be independent of ZFC.

By this we mean that it can neither be proven nor refuted using the axioms of ZFC.

This is just the kind of unsettling situation that the circa-1900 Logicians and Set Theorists were working to banish. Interesting but independent statements of this kind are all too common, and many Set Theorists claim that this is an invitation to change our notion of set by adding more axioms.

of $\mathcal{S}\mathbf{et}$. It is part of the "meta-mathematical" world, our a priori understanding of how mathematics must work, and what it means.

The logical symbols "or" ($\vee$), "implies" ($\Rightarrow$), "implies and is implied by" ($\Leftrightarrow$) and "equals" ($=$) and the quantifier "for all" ($\forall$) are defined in terms of these more basic symbols. For instance $\forall x(x$ has a certain property) is defined to be $\neg(\,\exists x(x$ has the negation of this property)$\,)$. The logical statement $A \vee B$ is defined as $\neg((\neg A) \wedge (\neg B))$.

In practice, statements involving the existential quantifier offer opportunities to name sets. A statement involving the universal quantifier asserts that sets with certain properties do not exist.

A name for a set in our formulas can be "**bound**," which means associated with a quantifier, or "**free**," which means available for a substitution with any set or, for convenience, "**constant**" by which we will mean that it is a previously uniquely identified set. Examples of constants are $\varnothing$ and $\mathbb{N}$, but generally a name for a set can be re-used to represent a different set in a different part of a mathematical discussion.

$$\forall x(\,(\varnothing \in y) \wedge (y \in x)\,)$$

has one constant, one free variable and one bound variable.

If all variables in a formula are bound it is called a **sentence**. A properly formed sentence always has, as truth value, one of "true" or "false." This is a restriction on the type of logic we use. For instance

$$\forall x(\,\exists y(\,(\varnothing \in y) \wedge (y \in x)\,)\,)$$

is false, as can be seen by substituting $\varnothing$ for x in $\exists y(\,(\varnothing \in y) \wedge (y \in x)\,)$. On the other hand, the following sentence is true:

$$\forall x(\,(x = \varnothing) \vee (\exists y(y \in x))\,).$$

Sentences are either true or false, and that value can be determined in a mechanical (albeit laborious) way by "truth tables" when you are provided with an assignment of truth values for all of the elementary assertions from which the sentence is formed. There are eight possible truth assignments to the elementary parts of the logical sentence $A \vee (B \wedge C)$, and any one is possible. But the sentence $A \vee (B \wedge \neg A)$, which has similar structure, has only four permitted combinations of truth value. A and $\neg A$ *must* end up with opposite truth values. By asserting, as an axiom, that $C = \neg A$ we cut the number of permitted assignments in half.

In our context, only those truth assignments compatible with the simultaneous truth of all axioms are allowed. As you add axioms you weed out, from the collection of those truth assignments which were permitted by the raw structure of the sentence, those which are incompatible with these axioms. If a system is inconsistent, this might leave no permitted truth assignments at all. For instance, the two axioms $C = A$ and $C = \neg A$ conflict in this way. If there is even *one* properly formed sentence which we have discovered lacks a permitted truth assignment, then we recognize the collection of axioms to be contradictory and unusable.

Much of mathematics might be described (by logicians and set theorists) as the effort to determine which truth assignments are permitted to mathematically interesting properly formed sentences in conjunction with the ZF axioms. It is an

unavoidable, possibly lamentable, fact that *it might not be possible to deduce, by the rules of basic logic in conjunction with any given collection of axioms, if one or more truth assignments to a properly formed sentence is allowed.* It has been proven that it is not even possible to determine if the simultaneous truth of the specific collection of axioms under discussion here is permitted. One day these axioms might be shown to be inconsistent; there might be no way of consistently assigning truth values in ZF.

We must (and will) simply assume that consistent truth values can be assigned, that we perceive correctly these properties of sets, that our axioms are all true with their intended meaning which is drawn from our prior understanding of **Set**.

If a formula contains one or more free variables it is called a **predicate formula**. It is important that a properly formed predicate formula always turn into a properly formed sentence whenever all free variable are replaced by constants or bound by quantifiers. Only formulas which produce such sentences can be used to define our "explicit properties."

If a formula containing one free variable is "true" when a specific set is substituted in place of that free variable, we say the set "has the property" corresponding to the formula. If the formula is "false," this set "does not have the property."

We need to emphasize that just because we can't *determine* if a sentence is true or false—even if it is *impossible* to determine this—doesn't mean it is neither, according to doctrine. All it means is that we don't know which: it is still one or the other. (If the intended meaning of such a sentence is sufficiently important some mathematicians would say that this situation calls for new axioms to resolve the question.)

Let $P(x)$ denote a properly formed predicate formula with free variable x. The sentence $\exists x P(x)$, if consistent with the axioms you assume, is your license to name a set with the property given by $P(x)$. In conjunction with the Axiom of Extensionality and other axioms, you might conclude that there is one and only one set with this property. You may then declare the set with this property a constant, and assign it a fixed name.

The Axiom of the Empty Set could be rephrased as

$$\exists x(\,\forall w(w \notin x)\,).$$

$\varnothing$ satisfies the property given by $\forall w(w \notin x)$, and no other set does.

The Axiom of Extensionality tells us that the relation of equality $x = y$ is shorthand for

$$\forall z\Big(((z \in x) \vee (z \notin y)) \wedge ((z \in y) \vee (z \notin x))\Big).$$

You will notice three variables in the formula above, only one of which is bound by a quantifier. To apply this formula to a set y the variable x must first be bound. For instance the formula $P(y)$ defined as

$$\exists x\Big(\forall w(w \notin x) \wedge \forall z\Big(((z \in x) \vee (z \notin y)) \wedge ((z \in y) \vee (z \notin x))\Big)\Big)$$

has one free variable. $\exists y P(y)$ is a true statement: the formula $P(y)$ is true when $\varnothing$ is inserted in place of y and not for any other set. So $\varnothing$ has the property defined by formula P and other sets do not.

The intelligibility of expressions in mathematics is heavily dependent on the use of shorthand. The predicate formula $P(y)$ above can be reduced to $y = \varnothing$ which seems to be significantly easier to understand.

This is typical. At its root, a properly formed sentence is discovered to be true or not true for no reason other than that a collection of axioms coupled with the rules of our chosen logic asserts it is true or not true. *Recognizing* the link to those axioms without extensive reliance on shorthand is both very awkward and unnecessary. Everyone uses the abbreviated versions of the formulae.

For descriptions of basic logic and the formation of formulae I recommend Enderton *A Mathematical Introduction to Logic* [**?**] and Hedman *A First Course in Logic* [**?**].

Any explicit property of sets can be used to identify what we will call the **class** consisting of the sets with that property. To say that a set is in a certain class is merely shorthand for saying it has the explicit property which defines the class.

For instance, the formula $x = x$ says, in words, that x has the same elements as does x. Every set satisfies this property. So $\boldsymbol{\mathcal{S}}\mathbf{et}$ is a class.

If we assert $A \in \mathfrak{C}$, where $\mathfrak{C}$ is a class, we intend nothing more than that A is (the name of) a set with the property of the class. Strictly, it is a manner of speaking and not a reference to the $\in$ relation, unless we can confirm that $\mathfrak{C}$ is a set.

Our axioms will be seen in later remarks to require that $\boldsymbol{\mathcal{S}}\mathbf{et}$ is not (the name of) a set. A class such as this which is not a set is called a **proper class** and will never be involved in inferences based on axioms about the existence of sets. For instance $\mathbb{P}(\boldsymbol{\mathcal{S}}\mathbf{et})$ and $\{\, \boldsymbol{\mathcal{S}}\mathbf{et} \,\}$ are not meaningful. Proper classes are never elements of anything, even in a metaphorical sense. The axioms of set theory never apply or refer to any objects other than sets or properties of sets.

Suppose $P(x)$ is a properly formed predicate formula with one free variable x. We can assign a name A to a member of the class $\mathfrak{C}$ defined by $P(x)$. Any meaningful sentence Q involving A really corresponds to the assertion

$$\exists x \big(\ P(x) \wedge \widetilde{Q}(x)\ \big)$$

where we assume here that the symbol x is not used elsewhere in the sentence Q, and the predicate formula $\widetilde{Q}(x)$ is the formula Q in which x has replaced A wherever it occurs. We may find this sentence—in conjunction with the axioms of set theory—has only a single allowable truth value. We conclude in that event that either all or none of the sets in $\mathfrak{C}$ have property $\widetilde{Q}(x)$, depending on whether that single truth value is "true" or "false." As a working mathematician using the natural shorthand one would never insert the steps leading to $\widetilde{Q}(x)$, but instead deduce that Q was "true" or "false" directly and then draw the proper conclusion about all members of the class $\mathfrak{C}$.

13. The Axioms of ZF Set Theory: Part Two

13.1. **The Axiom Schema of Subset Selection**
Suppose P is an explicit property of sets. If A is any set then the collection of all $y \in A$ for which $P(y)$ is true is a set.

It was Skolem in 1922 who proposed that the "explicit properties" mentioned in the Axiom Schema (here and in the Axiom Schema of Replacement below) be drawn from the predicate formulas described in the previous section.

This is a "restricted" selection axiom, because it requires the elements it is gathering to be inside a set which has already been created.

The Axiom Schema of Subset Selection asserts that the class determined by an explicit property of sets is, when restricted to a set, itself a set. It is called a "schema" to indicate it is not really a single axiom. Rather it is a different axiom for each property P.

It has several other names in common use, such as **Axiom Schema of Specification**, the **Axiom Schema of Separation** and the **Axiom Schema of Restricted Comprehension**.

This axiom can be used, for instance, to deduce directly that if $\mathfrak{C}$ is a proper class then no set can contain all the elements of $\mathfrak{C}$. So classes, if they are not sets, are "larger" than any set.

As another example of the usage of this axiom, suppose it is a known that a member of nonempty class $\mathfrak{C}$ defined by property P is a subset of a known set A. Then the intersection of all the sets in class $\mathfrak{C}$ is a set. To see this define predicate formula Q by

$$\forall y \ (\ (P(y) \text{ is true}) \Rightarrow (x \in y)\ ).$$

The Axiom Schema of Subset Selection using Q and requiring $x \in A$ yields the desired intersection. It is easy to show that this definition does not depend on *which* known set A is used, only that there is one such set in hand.

It might not be common to encounter a proper class without being able to deduce that some member of the class is a subset of a known set. But if this were to occur, ZFC provides no means to deduce that the elements shared by all members of the class comprise a set.

We mention next another restricted set selection process. Sets can be formed as the "range" of a "function-like" relationship, even if that "range" is not contained in a class known to be a set.

13.2. **The Axiom Schema of Replacement**
Suppose P is an explicit property of ordered pairs of sets for which $P(x, y)$ is true for at most one set y for each set x. If A is any set then the class of all y for which $P(x, y)$ is true for *some* $x \in A$ is a set.

The word "replacement" in the name of this axiom schema comes from one of its applications. Suppose P provides a way of associating a single set $f(x)$ to set x for certain sets x. There is no need to insist that f is defined for all $x \in \mathbf{Set}$. The property P we have in mind is given by:

$$P(x, y) \text{ is true if } y = f(x) \text{ and } P(x, y) \text{ is false otherwise.}$$

For a set A let B be the class formed by "replacing" those $x \in A$ upon which $f(x)$ is defined by $f(x)$. A set z is in the class B precisely if z satisfies the property

$$(\ z \in A \text{ and } f(z) \text{ is not defined }) \text{ or } (\ z = f(x) \text{ for some } x \in A\ ).$$

We can conclude that B is a set by invoking the Axiom Schema of Replacement.

This axiom implies the Axiom Schema of Subset Selection.

Classes identified as sets by the Axiom Schema of Replacement are not, necessarily, subsets of any previously known set. However the elements of the new set are *associated* with elements of another set so in this sense the newly formed class is not "larger" than some previously defined set. This loose size restriction *should be* enough for this class to be a set. This axiom declares that it is.

It was the inclusion of this axiom in 1922 by Adolf Fraenkel (along with independent contributions of Thoralf Skolem) that finished the job begun in 1908 by Ernst Zermelo, leading to the modern formulation of set theory shortly thereafter.

Last among the axioms of ZF is the Axiom of Foundation.

13.3. **Axiom of Foundation**

Every nonempty set A has an element which contains no element of A.

If a set has the property of this axiom it is called **well founded**. This axiom asserts that all sets are well founded.

You will also see this axiom called the **Axiom of Regularity.**

To set the stage for an example of this axiom in action, we first discuss the reason for the careful distinction between classes and sets.

Consider the following argument, associated with "**Russell's Paradox**" from a critique by Bertrand Russell of Frege's pre-ZF attempts to reform set theory on a more rigorous basis.

The property of sets "$x \notin x$" is either true or not for each set x, and is quite explicit. Let $\mathcal{C}$ denote the class defined by this property.

Russell noted that $\mathcal{C}$ cannot be a set. For if $\mathcal{C}$ **were** a set then it would either contain itself or not, and both cases lead to a contradiction. The act of assigning a set name to the class leads to this contradiction. If $\forall x(x \in C \Leftrightarrow x \notin x)$ is true then $\mathcal{C} \in \mathcal{C} \Leftrightarrow \mathcal{C} \notin \mathcal{C}$ must be true, which it obviously is not.

To see how the Axiom of Foundation dispatches Russell's class $\mathcal{C}$ from the ranks of $\mathcal{S}\mathbf{et}$, consider $\{\mathcal{C}\}$. If $\mathcal{C}$ is a set the Axiom of Foundation tells us that $\mathcal{C} \notin \mathcal{C}$ which implies $\mathcal{C} \in \mathcal{C}$, a contradiction.

As a side effect, if x is a set and $x \in x$ we could form the set $\{x\}$ which is not well founded. So no set can be an element of itself. For instance, $\mathcal{S}\mathbf{et}$ cannot be a set.

We should point out that classes of this kind, formed by a property including self-reference, can be shown (by contradiction) to be proper classes using the Axiom of Subset Selection. Though efficient at dealing with these classes, banishing contradiction is not the function of the Axiom of Foundation in set theory.

The earlier axioms asserted that we may speak, in our mathematical arguments, of reasonable or necessary objects from our vision of $\mathcal{S}\mathbf{et}$. This last axiom seems

unnatural in that it *restricts* our purview directly. We will discuss implications of this axiom again and a very natural related condition when we introduce the Zermelo hierarchy of sets in Section 18.

Finally, we get to what is (along with the Power Set Axiom) the least "constructive" of the selection axioms of ZFC.

13.4. **The Axiom of Choice**

> If J and X are sets and $A\colon J \to \mathbb{P}(X)$ is an indexed collection of nonempty sets then there is a function $f\colon J \to X$ such that $f(\beta) \in A_\beta\ \forall \beta \in J$.
> A function with this property is called a **choice function** for A.

See Section 8 for an rather thorough initial discussion of AC.

Our vision of **Set** allows us to reach around in a known nonempty set B, without peeking, and grab one member. AC provides for this through a choice function f on $\{B\}$. Membership in B is the only known property of $f(B)$, which is quite different from the identification of a single element by some property of that element, possessed by it alone. This feature of AC warrants some consideration, but this "choice" axiom goes far beyond this (very) finite case. It calls for simultaneous selection of elements from *any set of nonempty sets.* The function values of a choice function can then be gathered to form a set whose elements have no linking property except through this mysterious function.

AC cannot be proven in ZF. It is something "extra" and in the past its acceptance in the formation of mathematical arguments was *quite* controversial, particularly when its use was not concealed by less-controversial but equivalent reformulations. To this day many mathematicians regard it as a victory if a proof can be concocted which avoids the use of this axiom on infinite sets of sets.

However AC cannot be dispensed with, without sacrificing lots of very compelling mathematics. Without going into the interesting details (and using words some of which are defined elsewhere) we list here just a few theorems of ZFC that cannot be proven in ZF:

(*i*) An infinite set has a countable subset.

(*ii*) A vector space always has a basis.

(*iii*) The union of countably many countable sets is countable.

(*iv*) Any pair of sets have comparable cardinality.

(*v*) Any set has the same cardinality as an ordinal.

(*vi*) Every net has a universal subnet.

(*vii*) The product of any set of compact topological spaces is compact.

All of these—and *many* more—simply must go without the Axiom of Choice *or some other powerful axiom to replace it.*

But AC also implies that certain monstrous objects must exist such as non-measurable sets or bizarre decompositions of a sphere whose (finite number of) pieces can be rotated and translated to create two spheres, each of the same volume as the original sphere. This kind of infestation disturbs people. You can't escape results like this in ZFC.

Some objections to AC were essentially esthetic in nature, boiling down to folks asking themselves "Do I want to spend my professional life debating the properties of objects whose existence can *only* be inferred by appeal to this axiom?" Good question. Others focussed on the practical, expecting that AC would prove inconsistent with ZF, or conflict with other aspects of our vision of **Set** so central that they simply could not be abandoned.

A portion of this controversy subsided when Gödel proved that if ZFC contained an internal contradiction then an inconsistency could be found in ZF alone.

Gödel has also shown that the structure of set theory which we have discussed or *any other structure complicated enough to model the integers with their arithmetic* cannot be proven to be consistent[5] from within that structure. Some people find this to be unsettling. His result means that the statement "If the ZF axioms are consistent and ..." is an ineradicable but unstated premise to every theorem in ZFC set theory, and that means virtually every theorem of modern mathematics. This problem cannot be cured—even for integer arithmetic—by throwing ZF away and trying a new group of axioms.

As mentioned earlier and repeated here for emphasis, there is, in fact, the potential for a presentation of a direct internal contradiction in ZF. It is at least conceivable though regarded as unlikely that one day using these axioms someone will prove a statement to be both true and false. That would be an exciting day to be a mathematician.

Taking a slightly different tack, Gödel's proof that there are sensible statements in **any** system (not just ours) which could model integer arithmetic and which cannot be proven to be true or false from within that system implies that either the statement or its negation could be added as a new axiom without introducing new internal contradiction. This put a definitive end to Hilbert's original goal for set theory which was, in part, to invent a provably consistent axiomatization of mathematics within which every meaningful mathematical statement could be decided.

Some mathematicians prefer a different organization of the ideas inherent in ZFC, called the **von Neumann-Bernays-Gödel Axioms**, abbreviated **NBG,** which among other things gives a more prominent role to the concept of class. The theorems of ZFC and those theorems of NBG **which refer only to sets** are coincident. ZFC and NBG set theories are identical: statements about sets are true or false together in these two theories. Model theorists refer to this situation by saying that **NBG is a conservative extension of ZFC.** A statement P involving only sets has a proof using the NBG axioms exactly when P has a proof using the ZFC axioms.

The axioms of NBG prove a theorem that can sometimes be used to shorten proofs in ZFC and we formulate this theorem as an axiom, the **Axiom of Global Choice.** It can be phrased for our purposes in the following way.

A **class function** is a means by which a member $\tau(A)$ of class Y can be produced for every member A of class X. If X is a proper class this is not a function: the

[5]**Gödel's Incompleteness Theorems** state, roughly that (i) "Any formal language that is rich enough to model integer arithmetic contains statements that cannot be proven or refuted from within the language." and (ii) "This formal system is consistent." is one of those statements.

domain of any function is a set. We have used, implicitly, class functions in our statements of various axioms in ZFC: the Axioms of Pairing, Union and Power Set for instance.

A class function τ is called a **global choice function for $\mathfrak{Set}$** if $\tau(A) \in A$ for every nonempty member A of $\mathfrak{Set}$. It is a theorem of NBG that:

13.5. **The Axiom of Global Choice**
$\mathfrak{Set}$ has a global choice function.

This is equivalent to the statement that $\mathfrak{Set}$ can be well-ordered or, more explicitly, there is a one-to-one association between $\mathfrak{Set}$ onto the class of ordinal numbers $\mathfrak{Ord}$, which we will discuss in Section 15.

Positing the existence of a global choice function, or a way of lining up every possible set in a "class version" of a well order, sometimes allows one to dodge circumlocutions which would be required of an argument carried out purely in ZFC.

On a different note, philosophers of mathematics who refuse to accept AC often also refuse to accept the so-called "law of the excluded middle."

In this work, we will accept that the existence of a set with certain properties is proven if we can create a false statement by assuming such a set does not exist. Proving that a statement cannot be false is, for us, the same as proving it to be true.

If you reject this, another large fraction of modern mathematics goes away. One such school of mathematicians, who are said to use **Intuitionistic Logic** and restrict set formation to those sets which can be created using an explicit algorithm, are called **Constructivists.** L. E. J. Brouwer was one prominent constructivist, but his version is but one among many. There are numerous schisms and variant set theories, espousing different collections of allowable logical principles and axioms. One recent and particularly intriguing version is that of the school founded by Errett Bishop, whose followers are explicit in regarding much of modern mathematics as an elaborate hallucination.

It is safe to say that working mathematicians who refuse to accept the **Classical Logic** (an expansion and clarification of the ancient **Aristotelian Logic**) or the ZFC (or NBG) axioms constitute a small but interesting minority among mathematicians.

Finally, I would like to comment about several attitudes or belief systems available to users of languages.

Platonists believe or act as if the mental structures they create and talk about, using the language, have explicit reality. Mathematicians conceive of reaching out and manipulating elements of a group, of checking if an infinite number of open sets in a topology have some property and the like. Axiomatic set theory is useful to a staunch Platonist because it helps him or her avoid erroneous statements about these real things, and as a tool to assist in unambiguous communication. The steps used to build a group are just as real as—or more real than—pounding nails into wood to build a house. Harkening back to Plato himself, we could say that is because the creation of a group uses perfect ideal concepts, whereas housing

construction suffers from ambiguity involving the nature of a wall, a nail, and what it means "to pound." The *concept* of a house is more real than an actual house, which can be no more than a shadow of the "ideal house."

Formalists, on the other hand, take no position on the existence of anything beyond finite lists of several types of symbols set down according to a short list of rules. Upon assumption of a list of axioms some of these become true sentences. That is the content of all of mathematics: a listing of these true sentences. All else is window dressing, self-delusion or convenient fiction. A very fast computer could generate and check untold trillions of these lists of symbols and ultimately, given time and clever parsing, compile a record of *many* true sentences. Starting with the ZFC axioms that listing might contain, scattered about, most of current mathematics and much more.

The "literal reality" of many of the objects manipulated by mathematicians with their "Platonist" hat on is subject to easy criticism. The utility of an enormous listing of tautologies generated by a Formalist programmer, upon which is scattered a virtually invisible trace of true sentences about which a Platonist might care, is questionable.

In fact most mathematicians as far as I can tell act as if they are Platonists but sometimes write as if they are Formalists. Most humans (I am including mathematicians here) seem to have very little trouble reconciling this kind of contradiction.

One feature of these philosophical positions of relevance to us involves the conundrums generated by self-reference. Sometimes an object, a "whole," is defined by prescribing its parts. **Impredicativity** is a quality which is present in such definitions when these parts are prescribed by means which reference the whole. Definitions of this kind have a bad reputation, leading to the contradictions which were the bane of nineteenth century foundation studies. It is present in set theory if we assume that $\mathbf{Set} \in \mathbf{Set}$, leading to Russell's paradox.

It is present in our identification of sets by means of properties given as lists of symbols. What is such a list if not a certain kind of set? And how much of set theory itself must be developed to work with the lists we use to identify sets and make assertions about them?

It is also apparently present in our Axiom Schema if we think of ourselves as *constructing* sets rather than *identifying* sets through properties. The sets involved in an explicit property leading to the definition of a set might only exist by a feature of the totality to which that set itself belongs. The Axiom Schema are, it would seem, fraught with opportunity for impredicativity.

The Axiom of the Power Set is often complicit in issues involving self-reference. For instance suppose we have a set A and let K be the class of subsets of A with property P. K is a set because $\mathbb{P}(A)$ is a set. In a typical application we might let X denote the intersection of the members of the set K and go on to deduce that $X \in K$. So X is defined in terms of a set of which it is a member; we have a case of impredicativity.

We sidestep these problems by positing the existence of $\mathbf{Set}$ *a priori,* which puts us solidly in the Platonist camp. The axioms form a structure we use to give names to sets that are there already. The statement "They are the three fastest runners on our team." may be self referential ("they" are identified by reference to a whole

of which "they" are a part) but is not problematic—the team was there, including its fastest members, all along.

Some of the sting is removed from criticism of this position by the observation that no (sufficiently complex) axiomatic structure *can* be proven to be consistent from within that structure, so all hope of "building up" $\mathcal{S}\mathbf{et}$ resides in a choice: the decision to believe in whatever axioms you assume to carry out the procedures. There is no easy way out. To talk about interesting mathematics, one simply *must* take quite a lot on faith, either abstract axioms as properly formed symbol patterns or pre-existing intended meaning to which the sentences refer.

Our choice, to believe in $\mathcal{S}\mathbf{et}$ straightaway, reduces some of our worries about impredicativity and changes our interpretation of the meaning of contradiction should that be found. Also, we are less delicate about adding axioms to the structure. Any property of $\mathcal{S}\mathbf{et}$ about which we (collectively) have no doubt, which we agree $\mathcal{S}\mathbf{et}$ *absolutely must possess*, can be added as an axiom without harm or fear of producing contradiction. If we possess such an insight we *should* add it to our system. If it is independent of the other axioms it could reveal to us other new and more subtle truths about $\mathcal{S}\mathbf{et}$.

One belief system notably absent from the (conscious) musings of most recent mathematicians is **Empiricism**. That is the point of view that knowledge comes from experience of the world, a central tenet of the sciences. Whatever the facts of the matter may be, mathematicians just do not conceive of themselves as "running experiments." Rather, introspection is seen as the source of truth.

If we posit that the human mind is the result of eons of experiences of animals running "survival experiments" with the "winners" sometimes determined by better fidelity of an internal model to parts of an external reality, perhaps introspection is only Empiricism once removed. That would make more reasonable Eugene Wigner's observations in his famous commentary concerning *"The Unreasonable Effectiveness of Mathematics in the Natural Sciences."* Mathematics becomes a consciously created mirror image of a natural mirror image of parts of an external reality.

It would also give new meaning to the point of view of the ancient Greeks: that the study of pure mathematics gives insight to the true nature of the world, of which our daily experiences reveal only a shadow.

14. Induction and Recursion

In Section 8 on page 28 we used the Principles of Induction and Recursive Definition to show that Zermelo's Theorem implies Kuratowski's Lemma. Here, we reproduce two common variations of these ideas in 14.1 and 14.2. Following that we prove several results which justify common practice as represented in these two variations. We then utilize Induction and Recursive Definition in typical ways to prove several results of independent interest.

The **Principle of Induction** holds on propositions $\{ P_\beta \mid \beta \in A \}$ indexed by a well ordered set A:

14.1. Suppose that P_α is true for the first element α of A. Suppose too that we can establish P_γ to be true whenever P_β is true $\forall \beta$ with $\alpha \leq \beta < \gamma$. Let F be the subset of A containing those γ for which P_γ is false. Were F found to be nonempty in well ordered A it would contain a least member in violation of our two suppositions. We may conclude that F is void and hence P_γ is true $\forall \gamma \in A$.

This principle can be used as it stands to verify a well ordered list of explicit true-or-false properties of mathematical objects known to exist. Something a bit more subtle is required to **"construct"** sets using a **Recursive Definition.**

14.2. Suppose A is a well ordered set and we select, for the first member α of A, an element G_α of a set X. Suppose we can choose, find or build an element $G_\gamma \in X$ whenever we are in possession of elements $G_\beta \in X$, for $\alpha \leq \beta < \gamma$.

Then there is a function $G\colon A \to X$ for which each G_β is a possible result of the construction process.

If the construction process produces a unique value at each β then the function G is unique.

Though we may *think* of this as *constructing* G and use that vocabulary, what is really going on is somewhat different. Rather, we merely deduce that a function G with the necessary properties must exist if our axioms are to be consistent. We reel in the net, from the ocean of **Set**, to find a nonempty set of compatible functions. We can then assign a name to a member of the set, an embodiment of the construction procedure. Once we know it exists, we might apply 14.1 to determine some of its properties.

We now prove 14.3 and 14.4 which clarify and imply 14.2.

The existence of a well ordered set big enough to do whatever it is you are trying to do *might* require AC. The production of the element G_γ when you possess G_β for $\alpha \leq \beta < \gamma$ could also require AC, depending on the specifics of the construction. However the proof of the first result below, which implies 14.2, does not use any form of AC.

The proof of the second result involves a choice function. A function of the requisite kind might exist in specific cases without appeal to AC, but in general its existence is not assured without that axiom.

Suppose set A is well ordered with more than one member. Let α be the first member of A and tack on an extra last member z to the end of A for later convenience. Call the expanded well ordered set $\overline{A}$. Suppose B is a nonempty set.

Let G be the set of all possible functions whose domain is one of the initial segments I_b of $\overline{A}$ for some $b \in \overline{A}$, and whose range is B. Note: $I_b \subset A$ for all $b \in \overline{A}$ and in particular $I_z = A$. We specifically include the empty function $\varnothing$ as a member of G, corresponding to empty segment I_α.

Suppose given any function $P\colon G \to B$. In examples this will correspond to the thing we imagine to be the "inductive construction step." It is a means by which we look at a function defined on an initial segment and produce from that a member of B.

Let $X|_{I_b}$ denote the restriction of function X defined on I_c to I_b for $b \leq c$.

14.3. There exists a unique $X\colon A \to B$ such that $X(b) = P\left(X|_{I_b}\right)\ \forall b \in A$.

To see why, we define S to be

$$\{X \in G \mid X(b) = P\left(X|_{I_b}\right) \text{ whenever } b \text{ is in the domain of } X\}.$$

S is nonempty: the empty function is in S. So is $\{(\alpha, P(\varnothing))\}$.

Suppose X and Y are in S and I_c is the intersection of their domains. Let b denote the first member of I_c where $X(b) \neq Y(b)$, if there is such a value. But then $X|_{I_b} = Y|_{I_b}$ and because both functions are in S we must have $X(b) = Y(b)$, contradicting the definition of b. Therefore X and Y agree on the intersection of their domains, so one is an extension of the other. That means S is a chain with containment order.

The union of the sets (functions) in S is itself a function Y. The domain of this function is a union of initial segments contained in I_z and so itself is an initial segment I_c for some $c \leq z$. This function is in S.

If $c < z$ then $\{\,(c, P\,(Y))\,\} \cup Y$ is a member of S containing Y but not equal to Y, contradicting the fact that Y is the union of all members of S.

So $c = z$ and Y is the unique function we were looking for.

We now consider the case where there is more than one choice at each step of the induction.

Suppose given any function $Q\colon G \to \mathbb{P}(B)\text{–}\{\varnothing\}$. Our only condition is that each $Q(X)$ is a nonempty subset of B.

14.4. There exists a function $X\colon A \to B$ such that $X(b) \in Q\left(X|_{I_b}\right)\ \forall b \in A$.

Let f denote a choice function on $\{\,Q(X) \mid X \in G\,\}$. In other words, $f(Q(X)) \in Q(X)$ whenever $X \in G$.

Defining $P(X)$ to be $f(Q(X))$ for each $X \in G$ and applying the previous result yields the existence of a function with the necessary property.

Often, the intent of a mathematician using a Recursive Definition is to create a function by a certain procedure at each stage. We want to consider a case where sometimes the thing we want to do cannot be done. A member $*$ of B will be called a **cemetery point** for the function Q of the previous result if, first, whenever $* \in Q(X)$ then $Q(X) = \{*\}$ and, second, whenever $*$ is in the range of X then $Q(X) = \{*\}$. The proof of the fact given below is immediate.

14.5. Consider the situation of the previous result where B possesses a cemetery point $*$ for Q. Let T denote those members of S with domain I_z. So T is nonempty and if $*$ is in the range of a member X of T there is a member $b \in \overline{A}$ with $* \notin X\,(I_b)$ but $X(c) = *$ for every $c \geq b$.

In the last result below we expand the domain of validity of Recursive Definition one more time. We don't require that the construction step place the result in a fixed set known in advance.

This time we will let G denote the class of all possible functions whose domain is one of the initial segments I_b of $\overline{A}$ for some $b \in \overline{A}$. G is a class and *not* a set because the range of these functions could be any member of **Set**.

Suppose given a class function P on G. By this we mean a method of associating a unique set to each member of G. We have met class functions before in our discussion of the Axiom of Global Choice, and earlier. (The power set operation is a class function. The function that returns the union of all the elements in a set is another.)

14.6. There is a set B and a unique $X\colon A \to B$ such that $X(b) = P\left(X|_{I_b}\right)\ \forall b \in A$.

We define S to be the class of functions X in the class G for which $X(b) = P\left(X|_{I_b}\right)$ whenever b is in the domain of X.

Again, the empty function is in the class S. So is $\{(\alpha, P(\varnothing))\}$.

Exactly as before, any pair of members X and Y of the class S agree on the intersection of their domains, so one is an extension of the other.

That means that there is at most one of these functions with domain I_b for each $b \in \overline{A}$. The Axiom Schema of Replacement implies directly that the class S is a set. We can let B denote the union of all the ranges of the functions in S and apply 14.3 for the necessary conclusion.

14.7. ***Exercise.*** *(i) Justify the validity of the principles expressed in 14.1 and 14.2.*

(ii) Justify 14.5.

(iii) Re-examine the arguments from page 28 that Zermelo's Theorem implies Kurotowski's Lemma. Both recursion and induction are used there.

(iv) Prove the result from 8.1: that Zermelo's Theorem implies Zorn's Lemma.

We can use the recursion technique to extend the domains of functions ("construct" functions) in useful ways. The construction of the isomorphism below provides an example. In this example, not only is existence inferred but certain properties are shown to project from earlier to later stages.

Suppose A and B are two well ordered sets. If α denotes the first member of A, define $f(\alpha)$ to be the first member of B. Now suppose $\beta > \alpha$ and we have defined the function f on the initial segment I_β and that $f(I_\beta)$ is B or an initial segment of B. Define $f(\beta)$, if possible, to be the least member of the set $B - f(I_\beta)$.

If this can be done for a certain β then we have f defined on $\{\beta\} \cup I_\beta = I_{\beta+1}$ and $f(I_{\beta+1})$ is an initial segment of B, namely $I_{f(\beta)+1}$, or B itself.

There are two possibilities:

First, this might be possible for each $\beta \in A$ and serve to define an increasing function f on all of A. $f(A)$ is an initial segment—or all of—B and $f(I_\beta)$ is an initial segment of B for each $\beta \in A$. So A is order isomorphic to B or an initial segment of B.

Second, there might come a point when the set used above to define $f(\beta)$ is empty, so the construction cannot proceed. In this case f is only defined on some largest I_β and is an order isomorphism between I_β and B: that is, B is order isomorphic to an initial segment of A. By construction, $f(I_\gamma) = I_{f(\gamma)}$, an initial segment of B, whenever $\gamma < \beta$. Conclusion:

14.8. If A and B are well ordered sets then either they are order isomorphic to each other or one is order isomorphic to an initial segment of the other.

We find another example of this kind of reasoning in the proof of the following statement.

14.9. Suppose f and g are increasing functions from well ordered A onto either initial segments of well ordered B or possibly onto all of B. Then $f = g$.

Consider the following argument.

If f and g ever differ, it would happen for some earliest $\beta \in A$. β cannot be the first member α, since both $f(\alpha)$ and $g(\alpha)$ must be the first member of B. Suppose $f(\beta) < g(\beta)$. Then $f(\beta)$ is missing from $g(A)$ so $g(A)$ cannot be an initial segment or B, contrary to assumption. If $f(\beta) > g(\beta)$ a similar problem occurs with $f(A)$. So f and g must be equal on A.

14.10. ***Exercise.*** *Suppose $f\colon A \to B$ and the sets A and B are well ordered.*

(i) Adapt the argument from above to observe that if f is nondecreasing onto an initial segment of B or onto B itself then f must carry initial segments in A onto initial segments in B.

(ii) If f is increasing then there is an increasing function $g\colon A \to B$ for which $g(A) = B$ or $g(A)$ is an initial segment of B.

(iii) If nonempty subset A of well ordered set B is given the well order inherited from B then A is order isomorphic to B or to an initial segment of B.

(iv) One might speculate that if f is one-to-one then there is an increasing function $g\colon A \to B$. This is false. For example, let $A = B = \mathbb{N}$ and let f be the identity function. Give the range the usual order on $\mathbb{N}$. Depart from the usual order on the domain by declaring $0 > n$ for all integers $n \neq 0$. The domain order is modified to have first element 1 and last element 0. Whatever $g(0)$ is chosen to be, the initial sequence $\{0, 1, \ldots, g(0)\}$ in the range is finite and so cannot contain a one-to-one image of infinite $\mathbb{N}$.

15. Distinguishing Among Well Ordered Sets

In this section we use the notation $(A, \leq_1)$ to indicate that A is a set and $\leq_1$ is an order on A. The empty well ordered set, and the well ordered set with but a single member, constitute trivial examples. In this section our orders will all be well orders, and from time to time we will refer to the class of all well ordered sets by $\boldsymbol{\mathcal{W}}$**ell**.

15.1. ***Exercise.*** *Use the Axiom Schema of Replacement to show that* $\mathbf{Well}$ *is not a set. (hint: If* $(A, \leq) \in \mathbf{Well}$ *define* $f((A, \leq)) = A$*. If* $\mathbf{Well}$ *were a set, the Axiom Schema of Replacement and Zermelo's Theorem would imply that* $\mathbf{Set}$ *is a set.)*

For $(A, \leq_1), (B, \leq_2) \in \mathbf{Well}$, we will use the notation

$$ordinal(A, \leq_1) \leq ordinal(B, \leq_2)$$

to mean that $(A, \leq_1)$ is order isomorphic to $(B, \leq_2)$ or to an initial segment of $(B, \leq_2)$ or, to take care of the trivial case, $A = \varnothing$. The order isomorphism and initial segment, if such exist at all, are unique, as we saw in the last section.

The notation $ordinal(A, \leq_1) = ordinal(B, \leq_2)$ will mean that $A = B = \varnothing$ or $(A, \leq_1)$ is order isomorphic to $(B, \leq_2)$.

As usual with this type of notation, we write $ordinal(A, \leq_1) < ordinal(B, \leq_2)$ if $ordinal(A, \leq_1) \leq ordinal(B, \leq_2)$ and $ordinal(A, \leq_1) \neq ordinal(B, \leq_2)$.

Obviously:

15.2. $ordinal(A, \leq_1) \leq ordinal(A, \leq_1)$ and

15.3. $ordinal(A, \leq_1) \leq ordinal(B, \leq_2)$ and $ordinal(B, \leq_2) \leq ordinal(C, \leq_3)$
$$\implies \quad ordinal(A, \leq_1) \leq ordinal(C, \leq_3).$$

It is also true that:

15.4. $ordinal(A, \leq_1) \leq ordinal(B, \leq_2)$ and $ordinal(B, \leq_2) \leq ordinal(A, \leq_1)$
$$\implies \quad ordinal(A, \leq_1) = ordinal(B, \leq_2).$$

To see this last item, we suppose $f\colon A \to B$ is an order isomorphism from A onto B or an initial segment of B and $g\colon A \to B$ is an order isomorphism onto A or an initial segment of A. Then $h = g \circ f$ is an order isomorphism from A onto A or an initial segment of A. The identity function on A is another such function. We saw in Section 14 that this means h is the identity so g and f are inverse (to each other) order isomorphisms.

The construction in Section 14 also shows that for any two well ordered sets A and B:

15.5. Either $ordinal(A, \leq_1) \leq ordinal(B, \leq_2)$
$$\text{or } ordinal(B, \leq_2) \leq ordinal(A, \leq_1).$$

This has widespread consequences. For instance, it implies immediately that if $(A, \leq)$ is a well order on infinite A then either A or an initial segment of A is order isomorphic to $\mathbb{N}$ with usual order. Every well ordered infinite set starts with a copy of $\mathbb{N}$.

We make the following important observation regarding the relationship between a well ordered set and its initial segments with induced well order (indicated by the same symbol) from the larger set.

15.6. Let $(I_s, \leq)$ be an initial segment, with induced well order, of $(S, \leq)$.
Then $ordinal(I_s, \leq) < ordinal(S, \leq)$.

This fact follows from the observation at the end of the last section regarding increasing functions. We found that there is at most one function $f\colon I_s \to S$ which is increasing and takes initial segments of I_s onto initial segments of S or S itself. Since the function $f(t) = t$ for all $t \in I_s$ obviously satisfies these conditions, and this function is not onto S, it is not an order isomorphism.

We now turn to a different issue. Let $\mathbb{S}$ be any nonempty set of well ordered sets, and suppose $(A, \leq_1) \in \mathbb{S}$. Let C be the set of those members c of A for which there is a member of $\mathbb{S}$ which is order isomorphic to the initial segment $(I_c, \leq_1)$ of $(A, \leq_1)$. If C is empty, then $ordinal(A, \leq_1) \leq ordinal(B, \leq_2)$ for every $(B, \leq_2)$ in $\mathbb{S}$. In other words, $(A, \leq_1)$ is a least member of $\mathbb{S}$.

On the other hand, if C is nonempty then it contains a least member c. So $ordinal(I_c, \leq_1) \leq ordinal(B, \leq_2)$ for every $(B, \leq_2)$ in $\mathbb{S}$ with equality for at least one member of $\mathbb{S}$.

We come to the following very important conclusion.

15.7. Suppose $\mathbb{S}$ is a nonempty set of well ordered sets. Then $\mathbb{S}$ contains a least member: that is, a member $(A, \leq_1)$ with $ordinal(A, \leq_1) \leq ordinal(B, \leq_2)$ for every $(B, \leq_2)$ in $\mathbb{S}$.

16. Cardinality

The order discussions above don't address well the specific idea of "bigger" or "smaller" in the sense of "raw size." Why do we think a pile with three marbles is smaller than one with five marbles? Why do we think there are more real numbers than there are integers?

The **cardinality** of a nonempty set A is said to be less than or equal to that of a set B if $\exists$ one-to-one $g\colon A \to B$. We declare the cardinality of $\varnothing$ to be less than or equal to that of B for all $B \in \mathbf{Set}$.

The notation for this situation is $cardinal(A) \leq cardinal(B)$.

The cardinality of a nonempty set A is said to be equal to that of a set B if $\exists\ g\colon A \to B$ that is one-to-one and onto B. The notation for this situation is $cardinal(A) = cardinal(B)$. We declare $cardinal(\varnothing) = cardinal(B)$ only when $B = \varnothing$. Two sets with equal cardinality are called **equipotent.**

We say that the cardinality of a set A is less than that of a set B, and write $cardinal(A) < cardinal(B)$, if $cardinal(A) \leq cardinal(B)$ but $cardinal(A) \neq cardinal(B)$.

16.1. ***Exercise.*** *(i) Show that if $f\colon A \to B$ is onto then $cardinal(B) \leq cardinal(A)$. (The Axiom of Choice is used in an essential way here.)*

(ii) Show that if A and B are nonempty and there is no function $f\colon A \to B$ which is onto B, then $cardinal(A) < cardinal(B)$. (This time use Zermelo's Theorem and 14.8.)

It is obvious that:

16.2. $cardinal(A) \leq cardinal(A)$ and

16.3. $cardinal(A) \leq cardinal(B)$ and $cardinal(B) \leq cardinal(C)$

$$\Rightarrow \quad cardinal(A) \leq cardinal(C).$$

It is also true that:

16.4. $cardinal(A) \leq cardinal(B)$ and $cardinal(B) \leq cardinal(A)$

$$\Rightarrow \quad cardinal(A) = cardinal(B).$$

The following proof of this result, known as the **Schröder-Bernstein Theorem**, is due to Birkhoff and MacClane.

Suppose $f\colon A \to B$ and $g\colon B \to A$ are both one-to-one. To avoid labeling circumlocution in the proof, we will presume that A and B do not share any members. Suppose x and y are members of A or B. We will call x an **ancestor** of y if y can be obtained from x by various compositions of f and g.

For each point x in A or B, one of three mutually exclusive conditions hold: (i) x has an even (possibly 0) number of ancestors. (ii) The set of ancestors of x is infinite. (iii) x has an odd number of ancestors.

Let $A_E,\ A_I$ and A_O stand for the members of A for which conditions (i), (ii) and (iii) hold respectively. Define $B_E,\ B_I$ and B_O similarly as subsets of B. Then $f(A_E) = B_O,\ f(A_I) = B_I$ and $g^{-1}(A_O) = B_E$. The function $h\colon A \to B$ defined to be f on $A_E \cup A_I$ and g^{-1} on A_O is one-to-one and onto B. The existence of this h establishes that $cardinal(A) = cardinal(B)$.

Any set can be well ordered and, for any two well ordered sets, one is order isomorphic to the other or an initial segment of the other. Since an order isomorphism on a well ordered set is one-to-one, we can conclude that for any sets A, B:

16.5. Either $cardinal(A) \leq cardinal(B)$ or $cardinal(B) \leq cardinal(A)$.

Though we don't prove it here, without AC there is no way to guarantee that cardinalities are always comparable. A function needed to decide the matter might not exist.

It is a fact that for any set A the cardinality of A is less than and not equal to the cardinality of $\mathbb{P}(A)$. This might seem obvious when A is finite, but it is not obvious when A is an infinite set.

The proof of this result involves a construction leading to a contradiction. Suppose A is nonempty. It is clear that $cardinal(A) \leq cardinal(\mathbb{P}(A))$, as demonstrated by the one-to-one function $f\colon A \to \mathbb{P}(A)$ defined by $f(x) = \{x\}\ \forall x \in A$. Now suppose the cardinalities were equal. Then there would be a one-to-one function $g\colon A \to \mathbb{P}(A)$ onto $\mathbb{P}(A)$. Let $B = \{\, b \in A \mid b \notin g(b) \,\}$. Since g is presumed to be onto $\mathbb{P}(A)$, there must be an a in A with $g(a) = B$. But then one asks if $a \in B$, and comes to the conclusion that $a \in B \Rightarrow a \notin B$ and $a \notin B \Rightarrow a \in B$, a contradiction. Therefore no such g can exist and $cardinal(A)$ is strictly less than $cardinal(\mathbb{P}(A))$ as advertised.

We have just shown that for any set A there is always a set with larger cardinality, and its power set $\mathbb{P}(A)$ provides an example.

16.6. **Cantor's Theorem:** $\quad cardinal(A) < cardinal(\mathbb{P}(A))$.

We are contemplating here the idea of a "hierarchy of distinguishable infinities," without end.

One might ask if there are sets with intermediate cardinality between A and $\mathbb{P}(A)$ when A is an infinite set. That none exist between the cardinalities of $\mathbb{N}$ and $\mathbb{P}(\mathbb{N})$ (shown to have the cardinality of $\mathbb{R}$, historically called "The Continuum," in Section 19) is called the **Continuum Hypothesis.** The great pioneer Cantor, working at the end of the nineteenth century, spent enormous effort attempting to prove this hypothesis.

The statement that, generally, there are no sets with intermediate cardinalities between infinite A and $\mathbb{P}(A)$ is known as the **Generalized Continuum Hypothesis**. These hypotheses are often abbreviated as **CH** and **GCH**, respectively. They have been the subject of much wrangling.

16.7. **Generalized Continuum Hypothesis:**
If A is an infinite set then there is no set B with
$cardinal(A) < cardinal(B) < cardinal(\mathbb{P}(A))$

GCH turns out to be independent of ZF: that is, it can neither be proven nor refuted using these axioms. Both truth values are permitted if ZF itself is consistent.

Quite a few modern set theorists seem to think the hypothesis "should" be false. The existence of the power set on $\mathbb{N}$ requires its own axiom, but in the presence of this set of larger cardinality, the formation of sets of (possibly) intermediate cardinality between that of $\mathbb{N}$ and $\mathbb{P}(\mathbb{N})$ requires (see 16.8 below) invocation of the unrelated Axiom of Choice in its guise as Zermelo's Theorem applied to $\mathbb{P}(\mathbb{N})$. Why *should* intermediate cardinality be forbidden? What intuitive or esthetic principle, what vision of **Set**, demands it? Among mathematicians other than set theorists, those who think about this hypothesis at all seem usually (by my estimate) to be "Continuum Hypothesis agnostic" or, at least, write as if they are.

Even so, the hypothesis can be interesting. For example if, after assuming ZF+GCH and an additional hypothesis "A," you obtain a contradiction, you can conclude either "A" is false or that "A" implies GCH to be false. Most mathematicians would regard a hypothesis that *might* be true, but only if GCH is false, to be highly suspect, and of dubious utility even if found to be true in that case. So your proof could be useful to others as a warning beacon: "Steer clear of hypothesis A."

It has also been shown that ZF+GCH implies the Axiom of Choice, so ZFC+GCH and ZF+GCH are the same set theory.

There are numerous results in analysis and elsewhere that are true in ZF+GCH but false or unknown if GCH is false. For instance, it is known that the ultrapower construction of the hyperreal numbers is independent of the initial choice of free ultrafilter (up to order preserving field isomorphism) in the presence of CH. It is unknown if this is true if CH is false.

On another note, if A and B are nonempty sets with $cardinal(A) < cardinal(B)$ then A with any well order is order isomorphic to an initial segment of B and not to B itself where B has any well order. This is because the existence of an order isomorphism would imply that A and B are equipotent.

16.8. Suppose A and B are nonempty sets.

> $cardinal(A) \leq cardinal(B)$ precisely when **there is at least one well order** $\leq_1$ on A **and** $\leq_2$ on B for which $ordinal(A, \leq_1) \leq ordinal(B, \leq_2)$.
>
> $cardinal(A) < cardinal(B)$ if and only if **for every well order** $\leq_1$ on A **and** $\leq_2$ on B we have $ordinal(A, \leq_1) < ordinal(B, \leq_2)$.

The connection implied in these last two items between the ordinal and cardinal relations is very useful. For example consider any nonempty set $\mathbb{T}$. Let $\mathbb{S}$ denote the set of well orders on any of the sets in $\mathbb{T}$. We saw in the last section that $\mathbb{S}$ has a least member $(A, \leq_1)$. The set A has least cardinality among all sets underlying any member of $\mathbb{T}$.

16.9. Every set of sets contains a set of least cardinality.

Suppose given any set A and let $\mathbb{B} \subset \mathbb{P}(\mathbb{P}(A))$ consist of all members of $\mathbb{P}(\mathbb{P}(A))$ of cardinality exceeding that of A. Let C have minimal cardinality among members of $\mathbb{B}$.

C is of minimal cardinality exceeding that of A among a specified class of sets, and this leaves open the prospect that there could be a set B, formed by a different process, with $cardinal(A) < cardinal(B) < cardinal(C)$.

Let $\leq_3$ be a minimal well order on C. We found that each initial segment I_s of C with induced order from C satisfies $ordinal(I_s, \leq_3) < ordinal(C, \leq_3)$. By the minimality property of $(C, \leq_3)$, and the fact that each of these initial segments is in $\mathbb{P}(\mathbb{P}(A))$, we must have $cardinal(I_s) \leq cardinal(A)$.

If $cardinal(B) < cardinal(C)$ then B with any well order is order isomorphic to one of these initial segments of C with order $\leq_3$. It follows that $cardinal(B) \leq cardinal(A)$. We have proved the following fact.

16.10. For every set A there is a member C of $\mathbb{P}(\mathbb{P}(A))$ so that

> (i) $cardinal(A) < cardinal(C)$ and
>
> (ii) whenever B is a set with $cardinal(B) < cardinal(C)$ then $cardinal(B) \leq cardinal(A)$.

We can do more with the argument from above.

16.11. Suppose $(C, \leq_3)$ is a minimal well ordered set among well ordered sets with cardinality exceeding that of nonempty set A.

> Then there is unique member $c \in C$ for which $(I_c, \leq_3)$ is order isomorphic to A with its minimal well order. This member c has the following additional properties.
>
> For each well order $(A, \leq_1)$ there is a unique s in the terminal segment $T_c \subset C$ for which $(I_s, \leq_3)$ is order isomorphic to $(A, \leq_1)$.
>
> Each $(I_s, \leq_3)$, for $s \in T_c$, is order isomorphic to A with some well order.

The proof is left as an exercise.

If C is infinite, the minimality assumption requires that C with minimal order $\leq_3$ is without last element. We form a proof by contradiction, supposing that C

with order $\leq_3$ is both minimal **and** has a last element L. Let α be the first element of C and let I_b be the initial segment of C order isomorphic to $\mathbb{N}$. Define the function $g\colon C \to I_L$ by $g(L) = \alpha$, $g(s) = (s+1)$ for all other $s \in I_b$, and $g(s) = s$ for $s \in T_b - \{L\}$. This function $g\colon C \to I_L$ is one-to-one and onto I_L contradicting the fact, implied by minimality of $(C, \leq_3)$, that $cardinal(I_L) < cardinal(C)$.

16.12. If C is infinite and $(C, \leq_3)$ is minimal among all well orders on C then C is the union of its initial segments, none of which has cardinality equal to that of C itself.

We gather some implications of the remarks from above into the following.

16.13. Suppose C has at least two elements and $(C, \leq_3)$ is minimal among all well orders on C. Define K to be the set

$$\{\, c \in C \mid I_c \subset I_s \text{ whenever } cardinal(I_c) = cardinal(I_s) \text{ for some } s \in C \,\}.$$

Suppose also that A is nonempty and $cardinal(A) < cardinal(C)$.

(i) There is a unique member c of K for which $(I_c, \leq_3)$ is order isomorphic to A with its least well order.

(ii) If K has a largest member, and c (from above) is that largest member, then **any** well order on A is order isomorphic to $(I_s, \leq_3)$ for a unique $s \in C$, which must lie in the terminal segment T_c.

(iii) If c (from above) is not the largest member of K let d be the first member of K beyond c. Then **any** well order on A is order isomorphic to $(I_s, \leq_3)$ for a unique $s \in C$, which must lie in the interval

$$T_c - T_d = T_c \cap I_d$$

As an example, if $C = \mathbb{N}$ then one minimal well order on C is the natural order, and $K = \mathbb{N}$. K has no last member.

If, on the other hand, C is a set of smallest cardinality among sets in $\mathbb{P}(\mathbb{N})$ whose cardinality exceeds that of $\mathbb{N}$, then C with minimal well order does have an initial segment of largest cardinality: that of $\mathbb{N}$ itself. So K does have a largest member b, and I_b is order isomorphic to $\mathbb{N}$ with usual order.

17. Ordinal and Cardinal Numbers

In our discussions of ordinality and cardinality we have used vocabulary as if there were well defined objects $ordinal(A, \leq)$ and $cardinal(A)$ and well orderings on the class of these objects. However well orderings must be defined on sets, and all elements are sets as well. There can be no order of any kind on either **Set** or **Well** in ZFC set theory because these classes are not sets.

Our vocabulary was merely a "manner of speaking," shorthand for assertions about the existence or nonexistence of certain types of functions between the named sets and orders.

We would like to identify specific sets with $ordinal(S, \leq)$ or $cardinal(A)$ so that the metaphorical inequalities become actual relations of some kind among "cardinal numbers" and "ordinal numbers."

First, the Axioms of the Empty Set, Pairing, Infinity, Extensionality and Union imply that the class $\boldsymbol{\mathcal{W}}\textbf{ell}$ of well ordered sets is nonempty, and contains many different sets of sets well ordered by containment. $0 = \varnothing \in \boldsymbol{\mathcal{W}}\textbf{ell}$. So is $1 = \{0\}$ and $2 = \{0, 1\}$. So is $\mathbb{N}$ and $\mathbb{N} \cup \{\mathbb{N}\}$.

In general, if S is any chain of sets each of which is itself a set well ordered by containment and T is the union of the chain, then both T and $T \cup \{T\}$ are well ordered by containment. Using the Axiom of Foundation the latter, which has a last element T, is definitely longer than any member of the chain S. These members appear as initial segments of $T \cup \{T\}$. Guided by these thoughts, we create "standard" well ordered sets below.

Suppose $(A, \leq)$ is a nonempty member of $\boldsymbol{\mathcal{W}}\textbf{ell}$.

Let $\mathbb{V}_\alpha = \{\varnothing\}$, where α is the first element of A. Let $\Theta(\alpha) = \varnothing$.

Suppose we have found $\mathbb{V}_\beta$ and $\Theta(\beta) \in \mathbb{V}_\beta$ for all $\beta \in A$ with $\alpha \leq \beta < \gamma$.

The next step could go two ways, depending on if γ is a limit member of A or not. If it *is* a limit member we define

$$\mathbb{V}_\gamma = \bigcup_{\alpha \leq \beta < \gamma} \mathbb{V}_\beta \quad \text{and} \quad \Theta(\gamma) = \bigcup_{\alpha \leq \beta < \gamma} \Theta(\beta)$$

If γ *is not* a limit member then it is the successor to a member $\zeta \in A$, with $\gamma = \zeta + 1$. We define

$$\mathbb{V}_\gamma = \mathbb{P}(\mathbb{V}_\zeta) \quad \text{and} \quad \Theta(\gamma) = \Theta(\zeta) \cup \{\Theta(\zeta)\}.$$

In either case, if $\beta < \gamma$ then $\Theta(\beta) \subset \Theta(\gamma)$ and $\Theta(\beta) \neq \Theta(\gamma)$.

If γ is the successor to ζ, $\Theta(\gamma)$ is called a **successor ordinal**, specifically **the successor ordinal to $\boldsymbol{\Theta}(\zeta)$**. If γ is a limit member, $\Theta(\gamma)$ is called a **limit ordinal**.

Successor ordinals have one more element than the union of all ordinals built before. That extra element is the union itself. Limit ordinals are the union of all ordinals built before.

In any case, we conclude as an application of Recursive Definition (see 14.6) that Θ is uniquely defined on all of A. Moreover, each $\Theta(\beta)$ is in the *set* $\mathbb{V}_A$, where $\mathbb{V}_A = \bigcup_{\beta \in A} \mathbb{V}_\beta$. This version of Recursive Definition required the Axiom of Replacement.

We have created an increasing function $\Theta\colon A \to \mathbb{V}_A$ where the image $\Theta(A)$ of Θ is totally ordered by containment.

In fact, it is easy to see that $\Theta(A)$ is well ordered by containment. Because Θ is one-to-one, each $\Gamma \subset \Theta(A)$ is of the form $\Theta(S)$ for $S \subset A$. S has a least member s. $\Theta(s)$ is the least member of Γ.

So $(A, \leq)$ is order isomorphic to $\Theta(A)$ with containment order.

As a chain, $\Theta(A)$ is maximal in the following sense. Suppose $B \subset \bigcup_{\beta \in A} \Theta(\beta)$ and for every $\gamma \in A$ either $B \subset \Theta(\gamma)$ or $\Theta(\gamma) \subset B$.

It may be that $B = \bigcup_{\beta \in A} \Theta(\beta)$, but if not let γ be the least member of A for which $B \subset \Theta(\gamma)$. So

$$\bigcup_{\alpha \leq \beta < \gamma} \Theta(\beta) \subset B \subset \Theta(\gamma).$$

If $\Theta(\gamma)$ is a limit ordinal, then it equals the union on the left and $B = \Theta(\gamma)$. If, on the other hand, $\gamma = \zeta + 1$ then $\bigcup_{\alpha \leq \beta < \gamma} \Theta(\beta) = \Theta(\zeta)$, which differs from $\Theta(\gamma)$ by one element. So $B = \Theta(\gamma)$ or $B = \Theta(\zeta)$.

There is no room between members of the chain $\Theta(A)$ to fit any more sets.

We note that if A is a positive integer n, or if $A = \mathbb{N}$, then $A = \Theta(A)$.

We now suppose $(B, \leq_2)$ is any other well ordered set and $f\colon A \to I_b$ is an order isomorphism onto the initial segment I_b of $(B, \leq_2)$.

We can create, by an identical process to the one above for A, a function $\Xi\colon B \to \mathbb{W}$. It is easily verified that:

(i) The set $\Theta(A)$ is actually equal to the initial segment $\Xi(I_b)$ of $\Xi(B)$.

(ii) The "universe" $\mathbb{V}_A$ for A is the universe $\mathbb{W}_b$ at stage b for B if b is a limit member of B. If b is a successor, $b = k + 1$, then A has a last member δ and $\mathbb{V}_A = \mathbb{V}_\delta = \mathbb{W}_k$.

(iii) $\Theta(\beta) = \Xi(f(\beta))$ for every $\beta \in A$.

So if A is infinite, we have created an extension of the integers and that extension goes to any **specific** height into the hierarchy of well ordered sets. If later we decide we want to go higher yet, there is no change to the previously created structure other than to append **more** at the "end." Of course, we can *always* add at least one more ordinal beyond any *set* of ordinals we have identified: there is no stopping point to this process.

$\varnothing$ together with any set which can be formed as $\Theta(A)$ for some nonempty well ordered $(A, \leq)$ will be called an **ordinal number**. The specific ordinal number associated with and order isomorphic to $(A, \leq)$ will be denoted $\mathfrak{Ord}(A, \leq)$ and we define $\mathfrak{Ord}(\varnothing) = \varnothing$. Sometimes an ordinal number is simply called "an ordinal."

The class of all of these ordinal numbers will be denoted $\mathfrak{Ord}$.

17.1. Suppose β and γ are two ordinal numbers.

$\beta \leq \gamma$ if and only if $\beta \subset \gamma$.

$\beta < \gamma$ if and only if $\beta \in \gamma$.

Note the following important fact.

Suppose W is any *set* of ordinals. Let S denote the union of all the members of W. $\mathbb{P}(S)$ has cardinality exceeding that of any of the ordinals in W. Let $A = \mathbb{P}(\mathbb{P}(S))$ and let $(A, \leq)$ be a well ordering of A and build the ordinal numbers using $(A, \leq)$ as above.

$$cardinal(S) < cardinal(\mathbb{P}(S)) < cardinal(A).$$

There is an initial segment I_μ of A with $cardinal(I_\mu) = cardinal(\mathbb{P}(S))$. So each member of W is order isomorphic to an initial segment of $\Theta(I_\mu)$: that is, $W \subset \Theta(I_\mu) \subset \Theta(A)$.

In particular, $\Theta(I_\mu)$ is an ordinal exceeding any ordinal in W.

17.2. Any set of ordinal numbers has a least upper bound in $\mathfrak{Ord}$, the union of all the ordinals in the set.

17.3. ***Exercise.*** *Suppose γ is an ordinal and $\varnothing \in A \subset \gamma$.*

(i) A is an ordinal if and only if $\beta \in A \Rightarrow \beta \subset A$.

(ii) A is a limit ordinal if and only if

$$\beta \in A \quad \Rightarrow \quad (\quad \beta \subset A \quad \textit{and} \quad \beta \cup \{\beta\} \in A \ ).$$

(iii) The union of any set of limit ordinals is a limit ordinal.

17.4. **Ord** is not a set, for if it were you could let S be the union of all of its members. So $S \cup \{S\}$ would be an ordinal larger than any ordinal, an instant contradiction.

17.5. ***Exercise.*** *Suppose A is a set which is well ordered by containment, $\varnothing \in A$ and whenever $C \in A$ then $C \cup \{C\} \in A$. Then A is an ordinal. This characterization can be used to create a predicate formula satisfied by, and only by, ordinals.*

We now proceed to define cardinal numbers.

The **cardinal number** of a set A, denoted $|\mathbf{A}|$, is now defined to be the least ordinal number of any well ordering on the set A. Each infinite cardinal number is a limit ordinal number.

$|A|$ is order isomorphic to A with a least well order.

Any "least" ordinal number of a given cardinality is called a **cardinal number**. Sometimes a cardinal number is called, simply, "a cardinal" or "an **aleph**" due to the notational convention in some sources of using letters of the Hebrew alphabet such as ℵ (aleph), ד (dalet) or ב (bet) to denote infinite cardinals.

Since the various cardinals are drawn from **Ord**, any set of cardinals inherits a well ordering.

As with ordinals, the class **Card** of all cardinal numbers is not a set. The proof of that fact and the other two below are left to the reader.

17.6. The union of any set of cardinal numbers is a cardinal number. This union is the least upper bound of this set of cardinal numbers.

17.7. **Card** is not a set.

17.8. The inequalities developed for the concepts of ordinality and cardinality in Sections 15 and 16 apply to ordinal and cardinal numbers, with only minor notational adjustments.

The cardinal number of a set A with n elements for $n \in \mathbb{N}$ is usually simply denoted $|A| = n$. Since there is only a single ordinality class for each finite A, n can be used unambiguously for the ordinal number of A with any well order too.

Also, the cardinal number $|\mathbb{N}|$, which is $\mathbb{N}$ itself **together with its usual order**, is usually indicated by $\aleph_{\mathbf{0}}$, read "aleph nought." The first cardinal after $\aleph_0$ is denoted $\aleph_1$. The cardinal number of the continuum, $|\mathbb{R}|$, which is $\aleph_1$ in the presence of the Continuum Hypothesis, is usually denoted by the letter $\mathbf{c}$.

On the other hand, Greek symbols are frequently used to denote infinite ordinal numbers. The symbol ω is reserved in most sources for $|\mathbb{N}|$ when one wishes to emphasize its ordinal nature.

Similarly the first aleph beyond $\aleph_0$, the cardinal $\aleph_1$, is usually denoted Ω when one wants to emphasize its ordinal nature.

$$\aleph_0 = \omega = |\mathbb{N}| \qquad \aleph_1 = \Omega = \text{ (if CH is true) } \mathbf{c}.$$

17.9. ***Exercise.*** *Suppose α and β are two ordinals. They induce the obvious well orders on the sets $\alpha \times \{0\}$ and $\beta \times \{1\}$. Create the well order $\leq$ on the set $A = (\alpha \times \{0\}) \cup (\beta \times \{1\})$ which corresponds to the phrase "the points in $\beta \times \{1\}$ follow the points in $\alpha \times \{0\}$" while preserving the order in $\alpha \times \{0\}$ and $\beta \times \{1\}$ separately.*

Define ***ordinal addition*** *by $\alpha + \beta = \mathfrak{Ord}(A, \leq)$.*

(i) $n + \omega = \omega$ for any integer n but $\omega + 1 > \omega$.

(ii) $\beta + \omega = \bigcup_{n \in \mathbb{N}} \beta + n$ for any ordinal β.

(iii) More generally, for any β and any limit ordinal γ, $\quad \beta + \gamma = \bigcup_{\delta < \gamma} \beta + \delta$.

17.10. ***Exercise.*** *Suppose $\aleph$ and $\beth$ are two cardinals. Define* ***cardinal multiplication*** *by declaring $\aleph \cdot \beth = 0$ if either $\aleph$ or $\beth$ is 0, and $|\aleph \times \beth|$ if neither are 0.*

Define ***cardinal addition*** *by letting $\aleph + \beth = |\aleph + \beth|$ where the "plus" on the right is the ordinal addition from Exercise 17.9. Since cardinals* ***are*** *ordinals, this is an abuse of notation. The use of Hebrew symbols for cardinals can indicate that cardinal addition is intended, while Greek symbols would imply that ordinal operations are intended. However this should not be taken as a rule. The intended operation must be taken from context.*

(i) Show that cardinal addition is commutative and associative with identity 0.

(ii) Also, cardinal multiplication is commutative and associative with identity 1.

(iii) Show that cardinal multiplication distributes over cardinal addition.

(iv) If m, n and k are integers show that $m + n = m + k$ exactly when $n = k$. If $m \neq 0$ show that $m \cdot n = m \cdot k$ exactly when $n = k$.

Suppose γ is any ordinal number bigger than 0. We have already defined $\aleph_0$ to be the cardinal number of $\mathbb{N}$. Suppose we have defined $\aleph_\beta$ for all $0 \leq \beta < \theta \leq \gamma$.

Let $\aleph_\theta$ denote the first cardinal containing $\aleph_\beta$ for all $0 \leq \beta < \theta$.

If θ is a successor ordinal, $\theta = \zeta + 1$, then $\aleph_\zeta = \bigcup_{0 \leq \beta < \theta} \aleph_\beta$ and we find that $\aleph_\theta$ is simply the first cardinal exceeding $\aleph_\zeta$. A cardinal of this kind is called a **successor cardinal**, specifically the successor to $\aleph_\zeta$, and the notation $\aleph_\theta = \aleph_{\zeta+1} = \aleph_\zeta^+$ indicates this situation.

On the other hand, if θ is a limit ordinal, then $\aleph_\theta$ has no immediate predecessor cardinal, and we define $\aleph_\theta = \bigcup_{0 \leq \beta < \theta} \aleph_\beta$. The cardinal $\aleph_\theta$ is called a **limit cardinal** in this case.

Either way, we have defined $\aleph_\theta$ and this procedure serves to define $\aleph_\gamma$ for any ordinal number γ.

17.11. ***Exercise.*** *Show that $\gamma \leq \aleph_\gamma$ for any ordinal number γ.*

We note that *all* infinite cardinals are limit ordinals: the vocabulary "successor cardinal" and "limit cardinal" refers to the relationship the cardinal in question has with the set of smaller cardinals in this hierarchy of cardinals we have identified.

Any infinite cardinal can be found in this hierarchy *somewhere*. For if there were any missing cardinals we could assign one of these a name. Among the set of cardinals less than or equal to this one, there would be a least cardinal ℸ not in this hierarchy. ℸ is the first cardinal beyond the union of all smaller cardinals, and would have shown up in the hierarchy constructed using any ordinal number beyond the upper bound of the set of ordinals used to form the hierarchy of cardinals smaller than ℸ: this is a contradiction. The act of naming a missing cardinal led to this state. We conclude that none are missing.

The containment hierarchy of the alephs is closely wedded to the hierarchy of all ordinals, among which they are sparsely mingled. This is hardly surprising. The set of all cardinals bounded above by some cardinal forms a well ordered set and so is order isomorphic to some ordinal. Our construction of the alephs merely makes the order isomorphism explicit.

Suppose γ is any ordinal number bigger than 0. Defined $\beth_0$ to be the cardinal number of $\mathbb{N}$. Suppose we have defined $\beth_\beta$ for all $0 \leq \beta < \theta \leq \gamma$.

If θ is a successor ordinal, $\theta = \zeta + 1$, define $\beth_\theta$ to be $|\mathbb{P}(\beth_\zeta)|$.

If θ is a limit ordinal, then define $\beth_\theta = \bigcup_{0 \leq \beta < \theta} \beth_\beta$.

As with the alephs, this procedure serves to define "the bets," the class $\mathcal{B}\mathbf{et}$ of all $\beth_\gamma$ for any ordinal number γ.

GCH can be rephrased as "$\aleph_\gamma = \beth_\gamma$ for all ordinals γ" or, equivalently,

17.12. $$\mathbf{GCH} \iff \mathcal{C}\mathbf{ard} = \mathcal{B}\mathbf{et}.$$

18. The Zermelo Hierarchy and the Constructible Universe

During the process of recognizing the occupants of $\mathcal{O}\mathbf{rd}$ we created, for each ordinal number γ, a "universe" for γ which we will here denote $\mathbb{V}_\gamma$. This set was defined by recursion, using a version that required the Axiom of Replacement. It was built (or could be rebuilt) by using γ itself to guide repeated application of the Axiom of Power Set or, at limit ordinals, the Axioms of Union and Pairing. The Axiom of Infinity requires us to consider limit ordinals. We started with the empty set and ended as soon as we arrived at a set containing γ. During this procedure we invoked neither the Axiom of Foundation nor the Axiom of Choice.

We gather together the class $\mathbb{V}$ of all sets in $\mathbb{V}_\gamma$ for any $\gamma \in \mathcal{O}\mathbf{rd}$. $\mathbb{V}$ is a truly huge class of sets, called the **Zermelo hierarchy**.

By abuse of notation ($\mathcal{O}\mathbf{rd}$ is not a set) we can write

$$\mathbb{V} = \bigcup_{\gamma \in \mathcal{O}\mathbf{rd}} \mathbb{V}_\gamma.$$

To say $A \in \mathbb{V}$ means that A satisfies property $Q(x)$ given by:

$$\exists\gamma(P(\gamma) \wedge x \in \mathbb{V}_\gamma)$$

where $P(x)$ is the predicate formula from 17.5 which determines the class $\mathfrak{Ord}$.

Clearly, $\mathbb{V}_0 = \{\varnothing\}$ must be in $\mathfrak{Set}$ if the ZF axioms are to hold. And if each $\mathbb{V}_\beta$ is a set for every β with $0 \leq \beta < \gamma$ then $\bigcup_{\beta<\gamma} \mathbb{V}_\beta$ is a set and so $\mathbb{V}_\gamma = \mathbb{P}\Big(\bigcup_{\beta<\gamma} \mathbb{V}_\beta\Big)$ is a set. We conclude that every member of $\mathbb{V}$ *must* be in $\mathfrak{Set}$ if the ZF axioms are to be consistent.

Nowhere do the axioms of ZF forbid, explicitly, the existence of sets outside the Zermelo hierarchy.

Let us suppose we have a set x_0 in our ZF world, obtained perhaps by fiat through some additional axiom, and which is not in the Zermelo hierarchy but whose existence is presumed to be compatible with the ZF axioms sans the Axiom of Foundation. By the Axiom of Extensionality x_0 must itself have members.

Some members of x_0 may be in the Zermelo hierarchy, but we will now show that not *all* of its members can be in the Zermelo hierarchy.

For each $y \in x_0$, if y is in the Zermelo hierarchy $\mathbb{V}_\lambda$ at some "level" λ, let λ_y be the least ordinal for which this is true.

Let us further suppose that at least one such y is in the Zermelo hierarchy. By the Axiom Schema of Replacement, the class of these least ordinals is a set. This set of ordinals has an upper bound δ. It follows that $\mathbb{V}_\delta$ contains every member of x_0 which is in the Zermelo hierarchy. If $x_0 - \mathbb{V}_\delta = \varnothing$ then $x_0 \in \mathbb{V}_{\delta+1}$, contradicting our initial premise that x_0 is outside the hierarchy.

We conclude that x_0 must itself have a member not in the Zermelo hierarchy. Let x_1 be the name of a member of $\{\, x_0 - \mathbb{V}_\delta \,\}$. Iterating, we can produce by this process a sequence $x_0, x_1, x_2, ...$ of sets each of which is outside the Zermelo hierarchy and for which $x_{i+1} \in x_i$ for each integer $i \in \mathbb{N}$.

Consider the class $X = \{\, x_0, x_1, x_2, ... \,\}$. Once again, this class is a set by the Axiom Schema of Replacement. Each member of X contains at least one member of X, violating the Axiom of Foundation.

Therefore, assigning a name to a set outside the Zermelo hierarchy is inconsistent with the axioms of standard set theory, which includes the Axiom of Foundation. Our vision of $\mathfrak{Set}$ does not include sets which it is contradictory to name. In the ZF universe, we may presume that $\mathfrak{Set} \subset \mathbb{V}$. Coupled with our earlier remark, we have $\mathfrak{Set} = \mathbb{V}$.

Let us now consider a putatively distinct set theory obtained by replacing the Axiom of Foundation in ZF by one of its consequences, $\mathfrak{Set} = \mathbb{V}$.

$\mathbb{V}$ itself can be built, unaltered, using this new collection of axioms.

If a non-well founded set is in one of the levels of the Zermelo hierarchy $\mathbb{V}_\beta$, there would be such a set $X \in \mathbb{V}_\gamma$ for some least ordinal γ. Every member of X contains at least one member of X.

Select $a_0 \in X$. So $a_0 \in \mathbb{V}_\delta$ for some $\delta < \gamma$.

Having found nonempty a_i for integer i select a member a_{i+1} of $a_i \cap X$. This procedure creates a sequence $a_0, a_1, \dots$ where each succeeding set is an element of the previous set.

The set $\{ a_1, a_2, \dots \}$ is in $\mathbb{V}_\delta$ and is not well founded. This contradicts the minimality condition of γ.

We conclude that every set in $\mathbb{V}$ is well founded.

So we have produced a condition, that the universe of all possible sets can be created by iterated application of the unrestricted power set axiom on the union of previously created levels starting with the empty set, which is both natural and equivalent to the Axiom of Foundation, at least **in the presence of the other axioms of ZF.**

18.1. $$\textbf{Set} = \mathbb{V} \quad \Longleftrightarrow \quad \textbf{Axiom of Foundation}$$

This helps explain what the Axiom of Foundation means and provides motive for its inclusion.

We can now characterize proper classes in a different and potentially useful way.

18.2. A class is a proper class exactly when there are sets satisfying the property of the class at arbitrarily high levels of the Zermelo hierarchy.

18.3. ***Exercise.*** *Without AC we cannot know that a given set can be well ordered so our means of defining the cardinal number $|A|$ as the least ordinal number of any well ordering of set A fails. In ZF, the Axiom of Foundation allows us to define cardinality in a different way.*

An obvious choice for the cardinal number of a set A is to define it as the class of all sets B for which there is a one-to-one and onto function $f\colon A \to B$. However this class is not a set, and therefore is not a first-class citizen in ZF: a theory of sets.

If A is any set let δ be the least ordinal for which there is a one-to-one and onto function $f\colon A \to B$ for some $B \in \mathbb{V}_\delta$. We then define the cardinal number of A, denoted $|A|$, to be the class of all those $B \in \mathbb{V}_\delta$ for which such a function exists. Since any set A appears in $\mathbb{V}_\gamma$ for some smallest γ, and the identity function is one-to-one and onto, we know that there is such a δ, and in fact $\delta \leq \gamma$ when $A \in \mathbb{V}_\gamma$. Being contained in a set, $|A|$ is itself a set.

*Cardinals defined in this way are called **Scott's Trick Cardinals**, after their inventor Dana Scott. These cardinals illustrate a means by which we can carve out unique sets from corresponding proper classes, and the technique has general utility.*

In our next discussion we will insist that sets be "built" by use of the Axiom of Union and a restricted version of the Axiom Schema of Subset Selection.

We start with $\mathbb{L}_0 = \{\varnothing\}$, just as in the formation of the Zermelo hierarchy.

Having found $\mathbb{L}_\delta$ for all ordinals $\delta < \gamma$ we proceed in one of two ways, depending on if γ is a limit ordinal or not. If γ is a limit ordinal, we simply let $\mathbb{L}_\gamma = \bigcup_{\delta<\gamma} \mathbb{L}_\delta$ just as when we created $\mathbb{V}$.

The difference is found when γ is a successor, $\gamma = \mu + 1$. In that case we form members of $\mathbb{L}_\gamma$ using two methods. First, any union of a set of sets contained in $\mathbb{L}_\mu$ will be included in $\mathbb{L}_\gamma$. Second, all sets that can be formed by the Axiom Schema of Subset Selection for properties that *only refer to members of* $\mathbb{L}_\mu$ are included in $\mathbb{L}_\gamma$. That is it.

The statement that the property $P(y)$, with one free variable y, involved in application of the Axiom Schema of Subset Selection must include only reference to members of $\mathbb{L}_\mu$ requires amplification. Any constant set involved in $P(y)$ must of course be from $\mathbb{L}_\mu$. But also any bound variables also refer only to sets in $\mathbb{L}_\mu$. So a statement $\forall x$ in a formula must be conjoined with the condition $x \in \mathbb{L}_\mu$ and, similarly, $\exists x$ in a formula occurs with restriction $x \in \mathbb{L}_\mu$. It is only y itself that need not be in $\mathbb{L}_\mu$.

Including a similarly restricted method of forming sets using the Axiom Schema of Replacement would be redundant. That is because $\mathbb{L}_\mu$ is a set, not a proper class, so both the "domain" and "range" of an allowable relation are sets. The desired set can be formed instead by an application of the Axiom Schema of Subset Selection for an appropriately formulated property which involves only members of $\mathbb{L}_\mu$.

It should be noted that, though we formed each $\mathbb{V}_\gamma$ from the Zermelo hierarchy along with the ordinal γ, we did not need all of $\mathbb{V}_\gamma$ to locate γ. We simply needed to form $\{\mu\}$ if $\gamma = \mu + 1$ or $\bigcup_{\delta<\gamma} \delta$ if γ was a limit ordinal, and both of these can be recognized with our restrictions. $\mathfrak{Ord}$ is the same class here as before.

We now define the **constructible hierarchy** to be

$$\mathbb{L} = \bigcup_{\gamma \in \mathfrak{Ord}} \mathbb{L}_\gamma.$$

If we augment ZF with the axiom

18.4. **Axiom of Constructibility**: $\mathfrak{Set} = \mathbb{L}$

we obtain a set theory called **constructible set theory.**

No sets outside of $\mathbb{L}$ were used to build $\mathbb{L}$. The Axiom of the Power Set applied to a member A of $\mathbb{L}$ yields those members of $\mathfrak{Set}$ which are subsets of A. Under the assumption of the Axiom of Constructibility all those subsets are in $\mathbb{L}$ and none of the other axioms of ZF, when applied to members of $\mathbb{L}$, produce sets outside of $\mathbb{L}$ so the Axiom of Constructibility is consistent with ZF.

These changes are very strong in their effect on mathematical questions. The Axiom of Choice is a theorem in constructible set theory. There is always a constructible choice function.

Seen from the *outside*, there are fewer ways to form sets in constructible set theory compared to ZFC, which could imply that there are fewer sets in the constructible hierarchy than in the Zermelo hierarchy. If that is the case, application of the Axiom of the Power Set in constructible set theory produces "smaller" power sets. However cardinality is determined by the existence of certain functions between sets, and there would be fewer of these functions in the constructible hierarchy as well. It is not obvious how these competing influences on the number and types of cardinals *as seen by an observer confined to* $\mathbb{L}$ will balance, but in fact GCH is provable in constructible set theory.

It follows that GCH and AC are consistent with ZF, assuming only that ZF itself is consistent. There is at least one universe of discourse in which GCH and ZF hold: constructible set theory.

Constructible set theory is a valuable source of examples, consistency checks of other axioms and so on. Though many mathematicians would be completely happy paddling around in this pond, most set theorists regard constructible set theory as proscriptive—many of the theorems about which they like to think are trivial or do not apply in constructible set theory.

A subclass $\mathbb{M}$ of the Zermelo hierarchy is called **transitive** if $A \in \mathbb{M} \Rightarrow A \subset \mathbb{M}$. Let S denote a set of our allowable statements, augmented with some or all of the ZF axioms, and suppose $\mathbb{M}$ is a transitive subclass of the Zermelo hierarchy. When the scope of applicability of each statement in S is restricted by augmenting S with the axiom $\boldsymbol{\mathcal{S}}\mathbf{et} = \mathbb{M}$ we say the resulting system, taking these statements as axioms, is the **relativization** of S to the class $\mathbb{M}$. If the relativization of S is consistent we say that $\mathbb{M}$ is a **transitive model** for S.

Rephrasing the results of the last few paragraphs, we see that the constructible hierarchy is the relativization of ZF to the transitive class $\mathbb{L}$ and $\mathbb{L}$ is a transitive model for ZF (if ZF itself is consistent) within which GCH (and so AC) is not only consistent with ZF but *true*. By this we mean that if the axioms of this model all are true then there is one and only one consistent truth value that can be assigned to GCH in the model: it must be true. Gödel used this fact to conclude that GCH and AC were consistent with ZF, if ZF itself is consistent.

Paul Cohen made a huge breakthrough in technique related to consistency results with his principle of **forcing**. This principle states (in original form) that if P is one of our allowable sentences and if *every* countable (in ZFC) transitive model $\mathbb{M}$ of ZFC is contained in a countable transitive model $\mathbb{W}$ of ZFC+P then P is consistent with ZFC.

If you accept that the constructible hierarchy is a model of ZFC+GCH then GCH is consistent with ZFC. So if we want to prove GCH to be independent of ZFC we can do it, according to Cohen, by showing that any countable transitive model of ZFC is contained in a countable transitive model for ZFC+$\neg$GCH.

How Cohen carried this program out, showing GCH to be independent of ZFC, is a pretty long story. Ciesielski *Set Theory for the Working Mathematician* [**?**] will get you started and Jech *Set Theory* [**?**] fills in more detail.

18.5. ***Exercise.*** *If M is a set let $A_0(M)$ denote the union of M with all the members of M. For $k \in \mathbb{N}$ with $k > 0$ let $A_k(M)$ be the union of A_{k-1} with all the members of A_{k-1}.*

*The class $T(M) = \cup_{k\in\mathbb{N}} A_k$ is called the **transitive closure of** $\mathbf{M}$. It is a set. Also, it is the smallest among sets S containing M for which $A \in S$ implies $A \subset S$.*

If M is a proper class, can the above steps be modified to produce a least transitive class containing M?

19. Some Specific Cardinal Relationships

Here are some interesting and perhaps surprising facts about the cardinalities of some common sets.

19.1. $|\mathbb{N}| = \left|\mathbb{Q}^+\right| = |\mathbb{N} \times \mathbb{N}|$.

19.2. $\left|\mathbb{R}^+\right|$ is the cardinal number of the open interval $(0, 1)$.

19.3. $|\mathbb{P}(\mathbb{N})|$ is the cardinal number of the open interval $(0, 1)$.

19.4. $\left|(0, 1)^{\mathbb{N}}\right|$ is the same as the cardinal number of the open interval $(0, 1)$.

We now recall some old definitions and make some new ones. Sets with cardinalities no more than that of $\mathbb{N}$ are called **countable**. Sets with cardinality less than that of $\mathbb{N}$ are called **finite**. Sets that are not finite are called **infinite**. Sets with cardinality equal to that of $\mathbb{N}$ are called **countably infinite**, and a specific one-to-one and onto function exhibiting this equality is called an **enumeration** of the set. If a countable set is finite, a one-to-one map from an initial segment of $\mathbb{N}$ onto the set is called a **listing** of the set. Sets that are not countable are called **uncountable** or **uncountably infinite**.

Together, items 19.1-19.3 imply that $\mathbb{Q}$ is countable and $\mathbb{R}$ is uncountable.

One proof of 19.1 follows a procedure invented by Cantor called "diagonalization." We start by listing all positive rationals grouped as follows :

$$\frac{1}{1}, \qquad \frac{1}{2}, \frac{2}{1}, \qquad \frac{1}{3}, \frac{2}{2}, \frac{3}{1}, \qquad \frac{1}{4}, \frac{2}{3}, \frac{3}{2}, \frac{4}{1}, \qquad \frac{1}{5}, \frac{2}{4}, \frac{3}{3}, \frac{4}{2}, \frac{5}{1} \qquad \text{etc.}$$

These groups consist of the fractions whose numerator and denominator add to a specific value. The first time a specified numerator p appears it is in a fraction with 1 as a denominator, and this happens in the "pth" grouping. The second time it shows up is in the "$p+1$" grouping and 2 is the denominator. So $\frac{p}{q}$ will show up as the "pth" item in the "$p+q-1$" grouping.

For each positive rational number r define $g(r)$ to be the place on the list where r in lowest terms is found.

$$\text{If } r = \frac{p}{q} \text{ (lowest terms)} \qquad \text{then} \qquad g(r) = \frac{(p+q-2)(p+q-1)}{2} + p.$$

g is a one-to-one function from $\mathbb{Q}^+$ to $\mathbb{N}$. So the cardinality of $\mathbb{Q}^+$ is no more than $\mathbb{N}$. Since the opposite inequality is obvious, we have proved the cardinalities of $\mathbb{N}$ and $\mathbb{Q}^+$ are the same.

We use an extremely similar diagonalization procedure to show that the cardinalities of $\mathbb{N}$ and $\mathbb{N} \times \mathbb{N}$ are the same.

Let $\mathbb{N} \times \mathbb{N} = \bigcup_{n \in \mathbb{N}} S_n$ where

$$S_n = \{\, (a, b) \in \mathbb{N} \times \mathbb{N} \mid a + b = n \,\} = \{\, (0, n),\ (1, n-1), \ldots,\ (n, 0) \,\}.$$

Each S_n has $n+1$ elements and is disjoint from S_m if $m \neq n$. Define $f(n, 0) = \frac{n(n+1)}{2}$ for each $n \in \mathbb{N}$. Note that $f((n+1, 0)) - f((n, 0)) = n + 1$, so there is just enough room between these consecutive values to fit the rest of S_n. This suggests that we define, for $0 \leq k \leq n$, $f((n-k, k)) = \frac{n(n+1)}{2} + k$. This function $f \colon \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ is one-to-one and onto, proving 19.1.

The proof of 19.2 relies on some presumptions about the definitions of the real numbers. Any positive real number r can be given a binary representation

$$r = \sum_{j=k}^{\infty} \frac{a_j}{2^j} \text{ where } a_j \text{ is 0 or 1 for each } j.$$

We will make two assumptions about this representation for each r. First we presume that k, the first subscript, has been chosen so that a_k is nonzero. Second, to avoid redundancy in the representation, series with $a_j = 1$ for all j beyond a certain point are replaced by the equivalent series that terminates in all zeroes.

The representation to which we refer, for example, of two and three fourths is:

$$1 \cdot 2^1 + 0 \cdot 2^0 + \frac{1}{2^1} + \frac{1}{2^2} \text{ where } k = -1$$

and the equivalent sum

$$1 \cdot 2^1 + 0 \cdot 2^0 + \frac{1}{2^1} + \frac{0}{2^2} + \frac{1}{2^3} + \frac{1}{2^4} + \frac{1}{2^5} + \dots$$

is avoided.

Returning to the proof:

$$\text{For any real number } r = \sum_{j=k}^{\infty} \frac{a_j}{2^j},$$

$$\text{when } k \leq 0 \text{ let } f(r) = \frac{1}{2^{2-k}} + \frac{1}{2^{3-2k}} \sum_{j=k}^{\infty} \frac{a_j}{2^j}$$

$$\text{and when } k > 0 \text{ let } f(r) = \frac{1}{2} + \frac{1}{2} \sum_{j=k}^{\infty} \frac{a_j}{2^j}.$$

Real numbers between 0 and 1 are squeezed into the second half of the unit interval. Real numbers between 1 and 2 are squeezed into the second quarter of the unit interval. Real numbers between 2 and 4 are squeezed into the second eighth of the unit interval, and so on. f is one-to-one so the cardinality of the positive real numbers cannot exceed that of the unit interval. Since the opposite inequality between cardinalities is obvious, item 19.2 is proven.

It remains to show that the cardinality of the real numbers between 0 and 1 is the same as that of $\mathbb{P}(\mathbb{N})$.

The definition of the real numbers between 0 and 1 characterizes them as a collection of subsets of rationals. So the cardinality of these real numbers does not exceed that of the power set on the rationals. It follows from 19.1 that this is the same as the cardinality of $\mathbb{P}(\mathbb{N})$.

To finish the proof of 19.3 we need the converse.

Define a function h as follows:

$$\text{If X is a finite (or empty) set of positive integers}$$
$$\text{let } h(X) = \frac{1}{2} + \sum_{j\in X} \frac{1}{2^{j+2}}$$
$$\text{and if X is infinite}$$
$$\text{let } h(X) = \sum_{j\in X} \frac{1}{2^{j+2}}.$$

h is a one-to-one function from the power set of $\mathbb{N}$ into the unit interval, so the cardinality of this power set does not exceed that of these real numbers.

To prove 19.4 we first note that there is obviously a one-to-one function from $(0,1)$ to $(0,1)^{\mathbb{N}}$ so $cardinal((0,1)) \leq cardinal\left((0,1)^{\mathbb{N}}\right)$. To show equality we need to produce a one-to-one $h\colon (0,1)^{\mathbb{N}} \to (0,1)$.

Suppose $y\colon \mathbb{N} \to (0,1)$ is a member of $(0,1)^{\mathbb{N}}$. Each $y(k)$ has a binary expansion of the form $0.y_{k1}y_{k_2}\cdots y_{kn}\cdots$ where we forbid, as above, binary expansions that terminate in all trailing 1s—that is, use $0.010000\cdots$ rather than $0.001111\cdots$.

For each k let P_k be the $(k+1)$st prime. So $P_0 = 2$, $p_1 = 3$, $p_3 = 5$ and so on. We define $x = h(y)$ to be the real number whose binary expansion contains 0 *except* at locations in the expansion of the form P_k^{n+1} for exponents $n \in \mathbb{N}$. At location P_k^{n+1} x has value $y_{k,n}$. Thus

$$\begin{array}{rcccccccccc} \text{binary place:} & 1 & 2 & 3 & 2^2 & 5 & 6 & 7 & 2^3 & 3^2 & \cdots \\ x = 0.0 & y_{11} & y_{21} & y_{12} & y_{31} & 0 & y_{41} & y_{13} & y_{22} & \cdots \end{array}$$

This function is one-to-one.

20. Some General Cardinal Relationships

There is no difficulty proving that if $\varnothing \neq A \subset B$ and $C \neq \varnothing$ then

$$|A \cup C| \leq |B \cup C| \qquad |A \times C| \leq |B \times C| \qquad \left|A^C\right| \leq \left|B^C\right| \quad \text{and} \quad \left|C^A\right| \leq \left|C^B\right|$$

and we leave verification to the reader. The following are less obvious.

20.1. If $|S| \leq |T|$ and T is infinite then $|S \cup T| = |T|$.

So if A and B are nonempty sets at least one of which is infinite $|A \cup B|$ is the greater of $|A|$ or $|B|$.

20.2. If A is infinite, $|A| = |A \times A|$.

So if A and B are nonempty sets at least one of which is infinite, $|A \times B|$ is the greater of $|A|$ or $|B|$.

This also implies that if C is nonempty, both $\left|(A \times B)^C\right|$ and $\left|A^C \times B^C\right|$ are the greater of $\left|A^C\right|$ or $\left|B^C\right|$ and are therefore equal to each other.

20.3. If A, B and C are nonempty sets, $\left|\left(A^B\right)^C\right| = \left|A^{B\times C}\right|$.

20.4. If A is infinite and $|2| \leq |B| \leq |A|$ then $\left|B^A\right| = |\mathbb{P}(A)|$.

> On the other hand if A is infinite and $|A| \leq |B|$ all we can say here is that $\left|B^A\right| \leq |\mathbb{P}(B)| = \left|2^B\right|$. See 21.8 for one extra case.

We create some preliminary results.

When C is infinite, since $\mathbb{N}$ is the smallest infinite cardinal there is one-to-one function $f\colon \mathbb{N} \to C$. We will show that $|\{a\} \cup C| = |C|$ whenever $a \notin C$. Define $g\colon \{a\} \cup C \to C$ by

$$g(x) = \begin{cases} f(1), & \text{if } x = a; \\ x, & \text{if } x \neq a \text{ and } x \notin f(\mathbb{N}); \\ f(n+1), & \text{if } x = f(n) \text{ for some } n \in \mathbb{N}. \end{cases}$$

So g is one-to-one and onto C. Our conclusion: any two infinite sets which differ by a single element are equipotent. This extends, obviously, to infinite sets which differ by finitely many elements.

We now consider the set $(A \times \{c\}) \cup (\{c\} \times A)$ for infinite A and $c \notin A$.

Let F consist of the set of all one-to-one and onto functions

$$g\colon (B \times \{c\}) \cup (\{c\} \times B) \to B$$

where B is an infinite subset of A. F is nonempty, because every infinite set contains a copy of $\mathbb{N}$ and an explicit function $g\colon (\mathbb{N} \times \{c\}) \cup (\{c\} \times \mathbb{N}) \to \mathbb{N}$ of the required type is easy to create, a task left to the reader.

We now order F by declaring $f \leq g$ if g extends f. A function in F is a set of ordered pairs, a subset of $[(A \times \{c\}) \cup (\{c\} \times A)] \times A$. Chains in F with this order have upper bounds in F, namely the union of the chain. So F has a maximal member. Let $m\colon (B \times \{c\}) \cup (\{c\} \times B) \to B$ be a maximal member of F.

So $|(B \times \{c\}) \cup (\{c\} \times B)| = |B|$.

Suppose $A - B$ is infinite. Then there is a subset $H \subset A - B$ equipotent to $\mathbb{N}$. So there is a one-to-one function $g\colon (H \times \{c\}) \cup (\{c\} \times H) \to H$ which could be used to extend m. This is impossible by the assumed maximality of m.

So we know that $A - B$ is finite. But then $(B \times \{c\}) \cup (\{c\} \times B)$ differs from $(A \times \{c\}) \cup (\{c\} \times A)$ by a finite number of elements, and A differs from B by a finite number of elements, so all four sets are equipotent.

It should be checked that what we have shown, in fact, implies 20.1.

20.5. ***Exercise.*** *Suppose A is infinite and $(A, \leq)$ is order isomorphic to $|A|$, and $s \in A$. Because $A = I_s \cup T_s$ and $|I_s| < |A|$, 20.1 implies that $|T_s| = |A|$.*

We now have what we need to prove 20.2. As above, let F be the set of one-to-one and onto functions $g\colon B \times B \to B$ where B is an infinite subset of A. By 19.1 and the fact that A is infinite we find that F is nonempty. Ordering F by extension, as above, Zorn's lemma implies that there is a maximal member $m\colon B \times B \to B$ in F, and from this $|B \times B| = |B|$.

If $|B| = |A|$ we can use a one-to-one and onto map $g\colon A \to B$ to show that $|A \times A| = |A|$ and we are done.

To finish, we need to consider the possibility of $|B| < |A|$. In this case the fact that $A = (A - B) \cup B$ and 20.1 imply that $|A - B| = |A|$.

So $|B| < |A - B|$ and there is a one-to-one map from B onto subset C of $A - B$. Since $|C| = |B|$, we find that each of the three disjoint component sets of

$$S = (C \times C) \cup (B \times C) \cup (C \times B)$$

has cardinality of C so there is one-to-one function $h\colon S \to C$ onto C.

This presents a contradiction, because h can be used to extend m, assumed maximal in F. So 20.2 is proved.

The proof of 20.3 is easy and left to the reader.

To see 20.4 we use 20.2 to justify the following equalities for infinite A and any set B with $|2| \leq |B| \leq |A|$.

$$|\mathbb{P}(A)| \;=\; \left|2^A\right| \;\leq\; \left|B^A\right| \;\leq\; \left|A^A\right| \;\leq\; |\mathbb{P}(A \times A)| \;=\; |\mathbb{P}(A)|.$$

So equality holds throughout and we have 20.4.

The exercises with which we finish this section provides some important practice with cardinal numbers. They also offer results about cardinals of the type which are most used by most mathematicians outside of set theory.

20.6. ***Exercise.*** *(i) A set is infinite if and only if it is equipotent to a proper subset of itself.*

(ii) Suppose A is an infinite set and B_n is the set of all subsets of A which have n elements for $n \in \mathbb{N}$. Then for each $n > 0$,

$$|A| = |B_n| = \left| \bigcup_{n \in \mathbb{N}} B_n \right|.$$

(iii) $\left| \left(2^{\mathbb{N}}\right)^{\mathbb{N}} \right| = \left|2^{\mathbb{N} \times \mathbb{N}}\right| = \left|2^{\mathbb{N}}\right|$. *Since* $\left|2^{\mathbb{N}}\right| = |\mathbb{R}| = \mathbf{c}$, *we have shown that the set of all sequences in $\mathbb{R}$ has cardinality $\mathbf{c}$.*

With a different emphasis, we have shown that the set of all countable subsets of $\mathbb{R}$ has cardinality $\mathbf{c}$.

(iv) If $\mathfrak{C}$ denotes the set of continuous real valued functions on $\mathbb{R}$ then $|\mathfrak{C}| = |\mathbb{R}|$. So the continuous functions constitute a very "sparse" subset of $\mathbb{R}^{\mathbb{R}}$, which has the cardinality of $\mathbb{P}(\mathbb{R})$. (hint: A member f of $\mathfrak{C}$ is determined by $f|_{\mathbb{Q}}$.)

(v) Show that if A and B are both infinite the set $Fin\left(B^A\right)$ consisting of members of B^A with finite range has the cardinality of B or $\mathbb{P}(A)$, whichever is larger.

(vi) The usual topology on $\mathbb{R}$ has cardinality $\mathbf{c}$.

20.7. ***Exercise.*** *Suppose V is a vector space over the field F with basis B. We suppose B is infinite. Let V^d denote the algebraic dual of V: that is, the set of all F-linear functions from V to F. Let C be a basis of V^d. We will show in this exercise that the dimension of V^d exceeds the dimension of V itself.*

(i) $|V| = |F \times B|$. This is the greater of $|F|$ or $|B|$.

(ii) Every member of V^d is uniquely determined by its effect on the members of B. It follows that $\left|V^d\right| = \left|F^B\right|$.

(iii) On the other hand by (i) we have $\left|V^d\right| = |F \times C|$*, which is the greater of* $|F|$ *or* $|C|$*.*

(iv) Suppose $|B| = \aleph_0$*. Let* $b_0, b_1, \ldots$ *denote an enumeration of the members of* B*. For each nonzero* $t \in F$ *define* $\phi_t \in V^d$ *to be the linear map taking* b_i *to* t^{i+1} *for* $i \in \mathbb{N}$*.*

Select $t_1, \ldots, t_k$*,* ***distinct nonzero*** *members of* F *with* $k > 1$*, and suppose that for certain members* $c_1, \ldots, c_k$ *of* F *we find*

$$0 = \sum_{j=1}^{k} c_j \phi_{t_j}.$$

Evaluating the right side at b_i *for* $i = 0, \ldots, k-1$ *we find that for each of these* i *choices* $0 = \sum_{j=1}^{k} c_j t_j^{i+1}$ *or, in matrix form,*

$$\begin{pmatrix} t_1 & t_2 & \cdots & t_k \\ t_1^2 & t_2^2 & \cdots & t_k^2 \\ \vdots & \vdots & \ldots & \vdots \\ t_1^k & t_2^k & \cdots & t_k^k \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

The matrix on the left is invertible. In fact its determinant is

$$t_1 t_2 \cdots t_k \left(\prod_{1 \le i < j \le k} (t_j - t_i) \right)$$

and is (closely) related to the Vandermonde determinant of linear algebra. So we find that $c_1 = c_2 = \cdots = c_k = 0$*. In other words, the set*

$$E = \{\, \phi_t \mid t \in F,\ t \neq 0 \,\}$$

is a linearly independent subset of V^d *and since* $|E| = |F|$ *we have the cardinal number of any basis of* V^d *at least that big.*

If $|B| > \aleph_0$ *the dimension of* V^d *can only increase, so* $|C| \ge |F|$ *in all cases.*

(v) Parts (ii),(iii) and (iv) combine to show that $|C| = \left|F^B\right|$*: in words, the dimension of the algebraic dual is* $\left|F^B\right|$*. This will be at least* $|\mathbb{P}(B)|$ *and, in any case, will* ***exceed the dimension of*** $\mathbf{V}$*.*

The multiplicative commutativity property of fields was used in the last exercise. Jacobson in *Lectures in Abstract Algebra Volume II Linear Algebra* [**?**] shows that identical results hold for unitary modules over any division ring.

20.8. ***Exercise.*** *(i) Suppose given sets* S *and* H_s *for each* $s \in S$ *and an infinite set* V*. Suppose that* $|S| \le |V|$ *and, for each* s*,* $|H_s| \le |V|$*. Then:*

$$\left| \bigcup_{s \in S} H_s \right| \le |V|.$$

In particular, the countable union of countable sets is countable.
(hint: Identify each H_s *with a subset of* $\{s\} \times V$ *in* $S \times V$*. See also 21.5 for a related result.)*

(ii) ***König's Theorem*** *Suppose S is nonempty and for each $s \in S$ there are nonempty sets H_s and V_s with $|H_s| < |V_s|$. The product $\Pi_{s\in S}V_s$ consists of all those functions $f\colon S \to \bigcup_{s\in S} V_s$ for which $f(s) \in V_s$ for all $s \in S$. Then:*

$$\left| \bigcup_{s\in S} H_s \right| < \left| \prod_{s\in S} V_s \right|.$$

(hint: Suppose $g\colon \bigcup_{s\in S} H_s \to \Pi_{s\in S}V_s$. For each $t \in S$, $|g(H_t)| \leq |H_t| < |V_t|$, so the set $\{ v \in V_t \mid v \neq g(h)(t)$ for any $h \in H_t \}$ must be nonempty. Define $f(t)$ to be a selection from this nonempty subset of V_t, and carry out this selection for each $t \in S$. The function $f \in \Pi_{s\in S}V_s$ but $f \notin g(H_t)$ for any $t \in S$. So $f \notin g\left(\bigcup_{s\in S} H_s\right)$. So g is not onto $\Pi_{s\in S}V_s$: that is, there can be no function from $\bigcup_{s\in S} H_s$ onto $\Pi_{s\in S}V_s$.)

This theorem generalizes Cantor's Theorem. If $H_s = \{s\}$ and $V_s = 2$ for each $s \in S$ then König's Theorem reduces to $|S| < 2^S$.

(iii) $|S \times H| < \left|V^S\right|$ *if* $|H| < |V|$.

We have defined addition of ordinals and addition and multiplication of cardinals. It is traditional in the presentation of set theory to develop more completely the implied arithmetics of ordinals and of cardinals (they are different) which extend multiplication, addition and exponentiation of ordinary integers. In this "higher arithmetic," if $\beth$ and $\aleph$ are cardinals, $\beth^{\aleph}$ will be closely related to the set of functions from $\aleph$ to $\beth$. It will be what we have denoted $\left|\beth^{\aleph}\right|$. Exponentiation of ordinals, however will be quite different.

We have assembled all the facts needed to make the study of the properties of these operations proceed smoothly. However we will stop here, referring the reader to Devlin's *Joy of Sets* [**?**] or Halmos *Naive Set Theory* [**?**] or Hrbacek and Jech *Introduction to Set Theory* [**?**] for this material.

21. Regular and Singular Cardinals

In this section we extract elements of the presentation of Hrbacek and Jech *Introduction to Set Theory* [**?**] and arguments to be found in Ciesielski *Set Theory for the Working mathematician* [**?**].

If β is any nonzero ordinal number, we say a set of ordinals $S \subset \beta$ is **cofinal in** β if for each $\gamma \in \beta$ there is a $\zeta \in S$ with $\zeta \geq \gamma$.

If $\mu, \beta \in \mathfrak{Ord}$, a function $f\colon \mu \to \beta$ is called **cofinal in** β if $f(\mu)$ is cofinal in β.

The **cofinality of** $\beta \in \mathfrak{Ord}$ is the least ordinal μ for which there is a cofinal function $f\colon \mu \to \beta$. This least ordinal μ will be denoted $\mathbf{cf}(\beta)$.

The identity function on β is cofinal in β, so $cf(\beta)$ cannot exceed β.

There are a couple of cases to consider.

First, if β is the successor to ordinal μ then $cf(\beta) = 1$ and $\{\mu\}$ is cofinal in β and $f\colon 1 \to \beta$ defined by $f(0) = \mu$ is a cofinal function.

Second, β might have no immediate predecessor: it could be a limit ordinal. Suppose $f\colon \mu \to \beta$ is cofinal in β. If $\zeta \in \beta$ then there is $\gamma \in \mu$ for which $f(\gamma) \geq \zeta+1$ so $\zeta \in f(\gamma)$. That means

$$\beta = \bigcup_{\gamma\in\mu} f(\gamma).$$

21.1. ***Exercise.*** *(i)* $cf(cf(\beta)) = cf(\beta)$.

(ii) We noted above that $cf(\beta) = 1$ *if* β *has an immediate predecessor. In particular, this is true if* $\beta \in \omega$ *or* $\beta = \omega + 1$ *or* $\omega + 2$ *and so on. (Recall* $\omega = \aleph_0$ *is the set* $\mathbb{N}$ *with the usual order.) Only limit ordinals have interesting cofinality.*

(iii) $cf(\omega) = \omega$.

(iv) If $f\colon cf(\beta) \to \beta$ *is a cofinal function, there is an increasing cofinal function* $g\colon cf(\beta) \to \beta$. *(hint: Suppose to avoid triviality that* β *is a limit ordinal. For each* $\gamma \in \beta$ *let* S_γ *be the set of those members of* $cf(\beta)$ *taken by* f *to an ordinal at or beyond* γ. *Define* $h\colon \beta \to cf(\beta)$ *by selecting* $h(\delta)$ *to be the least member of* S_δ. *The image of* h, $A = h(\beta)$, *is well ordered subset of* $cf(\beta)$. *Define* $W : A \to \beta$ *by letting* $W(t)$ *equal the least member of the set of those* $\delta \in \beta$ *for which* $h(\delta) = t$. W *is increasing on* A, *which is itself order isomorphic to* $cf(\beta)$ *or to an initial segment of* $cf(\beta)$. *But by minimality of* $cf(\beta)$ *the later case is impossible, and the order isomorphism composed with* W *is the required increasing function* g.*)*

(v) If $f\colon \mu \to \beta$ *is any one-to-one and onto function (not necessarily increasing) between ordinals then* $cf(\beta) \leq \mu$. *This means that* $|cf(\beta)| = cf(\beta)$*: that is,* $cf(\beta)$ *is always a cardinal number and cannot exceed* $|\beta|$.

(vi) If β *is a limit ordinal and* $\aleph_0 \leq \beta < \aleph_1$ *then* $cf(\beta) = \aleph_0$.

Suppose $\mu \in \mathfrak{Ord}$ and $\aleph \in \mathfrak{Card}$ and $\mu < cf(\aleph)$. Suppose further that $f\colon \mu \to \aleph$.

Let μ_f be the greatest member of $f(\mu)$, if there is one. $\mu_f \in \aleph$, so $\mu_f < \aleph$.

On the other hand, suppose $f(\mu)$ does not contain a greatest member. Let μ_f be the least ordinal exceeding every member of $f(\mu)$. Clearly $\mu_f \leq \aleph$. However if $\mu_f = \aleph$ then f is a cofinal map for $\aleph$ from a set of cardinality smaller than $cf(\aleph)$, a contradiction. So $\mu_f < \aleph$ here too. We have shown:

21.2. Suppose $\mu \in \mathfrak{Ord}$ and $\aleph \in \mathfrak{Card}$ and $\mu < cf(\aleph)$. Suppose further that $f\colon \mu \to \aleph$. The least upper bound of $f(\mu)$ has cardinality strictly less than $\aleph$.

21.3. ***Exercise.*** *Suppose* $\aleph$ *and* $\beth$ *are cardinals. Show that*

$$\beth^{\aleph} = \bigcup_{\gamma<\aleph} \beth^{\gamma} \quad \textit{when } \aleph < cf(\beth).$$

We now address a different topic. Suppose $\aleph \in \mathfrak{Card}$ is infinite and $\daleth = cf(\aleph)$. Let $f\colon \daleth \to \aleph$ be a cofinal map and let $g\colon \aleph \to \aleph^{\daleth}$ be any function.

Define for each $\beta \in \daleth$ the set $A_\beta = \{\, g(\gamma)(\beta) \mid \gamma < f(\beta) \,\} \subset \aleph$. The cardinality of A_β cannot exceed that of $f(\beta)$ which is less than that of $\aleph$. So the set $\aleph - A_\beta$ is never empty.

Define $h(\beta)$ for each $\beta \in \daleth$ to be the least member of $\aleph - A_\beta$.

So $h \in \aleph^{\urcorner}$ but $h(\beta) \neq g(\gamma)(\beta)$ for any $\gamma \in \aleph$. That means g cannot be onto $\aleph^{\urcorner}$, whose cardinality must therefore exceed that of $\aleph$. Our conclusion:

21.4. $\aleph < \left|\aleph^{cf(\aleph)}\right|$ for any infinite $\aleph \in \mathfrak{Card}$.

$\aleph \in \mathfrak{Card}$ is said to be a **regular** cardinal if $cf(\aleph) = \aleph$ and **singular** if $cf(\aleph) < \aleph$.

21.5. ***Exercise.*** *Suppose S, I and S_i for $i \in I$ are all sets, and $S \neq \varnothing \neq I$.*

(i) Suppose $|S|$ is a regular cardinal. If $|I| < |S|$ and also $|S_i| < |S|$ for each $i \in I$ then

$$\left|\bigcup_{i\in I} S_i\right| \;<\; |S|.$$

(hint: We saw in Exercise 20.8 that the cardinality of the union cannot exceed that of S, so we need only show that equality is forbidden. Suppose that the S_i are all disjoint and nonempty and, to obtain contradiction, that $h\colon \bigcup_{i\in I} S_i \to |S|$ is one-to-one and onto. For convenience presume that $I = |I|$. For each $i \in I$ let $f(i)$ be the least member of $|S|$ for which the terminal segment $T_{f(i)} \subset |S| - h(S_i)$. Then f is a cofinal map, contradicting regularity of $|S|$.)

(ii) If $|S|$ is a singular cardinal there are sets I and S_i for $i \in I$ with $|I| < |S|$ and $|S_i| < |S|$ for all $i \in I$ and

$$\left|\bigcup_{i\in I} S_i\right| \;=\; |S|.$$

(hint: Let $f\colon cf(|S|) \to |S|$ be cofinal. Then $|S| = \bigcup_{\beta\in cf(|S|)} \beta$, and each $|\beta| < |S|$.)

Apparently, based on this exercise, it is important to know which cardinals are regular and which are singular. We enshrine the conclusions of this exercise in the two items below.

21.6. A cardinal is regular if and only if it cannot be created as the union of a smaller number of smaller sets.

21.7. An infinite cardinal is singular if and only if it can be created as the union of a smaller number of smaller sets.

We noted above that $\omega = \aleph_0$ is regular.

Also, any infinite successor cardinal (such as $\aleph_1$, obviously) is regular.

To see this, suppose $\aleph^+ = \left|\bigcup_{\beta\in\beth} S_\beta\right|$ where $\beth$ is a cardinal and $\aleph$ is infinite and each $|S_\beta| \leq \aleph$. Because $\aleph^+$ is the first cardinal beyond $\aleph$, we can identify each S_β with a subset of $\{\beta\} \times \aleph \subset \beth \times \aleph$. So

$$\aleph^+ = \left|\bigcup_{\beta\in\beth} S_\beta\right| \leq |\beth \times \aleph|$$

and the last cardinal is the maximum of $\beth$ or $\aleph$. The only way the inequality can be satisfied is if $\beth \geq \aleph^+$.

An interesting fact concerning infinite successor cardinals is the following.

If $\aleph$ is infinite, the range of any function $f\colon \aleph \to \aleph^+$ is bounded above by some ordinal $\gamma < \aleph^+$. Therefore

$$\left(\aleph^+\right)^{\aleph} = \bigcup_{\gamma<\aleph^+} \gamma^{\aleph}.$$

But when $\aleph^+ > \gamma > 0$ we know that $\left|2^{\aleph}\right| = \left|\gamma^{\aleph}\right| = \left|\aleph^{\aleph}\right|$. So

$$\left|\left(\aleph^+\right)^{\aleph}\right| = \left|\bigcup_{\gamma<\aleph^+} \gamma^{\aleph}\right| \leq \left|\aleph^+ \times 2^{\aleph}\right| = \left|2^{\aleph}\right|.$$

It is obvious that $\left|2^{\aleph}\right|$ cannot exceed $\left|\left(\aleph^+\right)^{\aleph}\right|$ and we have proven:

21.8. $\left|2^{\aleph}\right| = \left|\left(\aleph^+\right)^{\aleph}\right|$ for any infinite cardinal $\aleph$.

There are infinite singular cardinals too.

For instance, consider $\aleph_{\aleph_0}$. The function $f\colon \aleph_0 \to \aleph_{\aleph_0}$ defined by $f(n) = \aleph_n$ is cofinal in $\aleph_{\aleph_0}$, and we conclude that $cf(\aleph_{\aleph_0}) = \aleph_0$. The cardinal number $\aleph_{\aleph_0}$ is singular. It is the least infinite singular cardinal.

The same argument works with any limit ordinal γ. The function $f\colon \gamma \to \aleph_\gamma$ defined by $f(\beta) = \aleph_\beta$ for each $\beta \in \gamma$ is cofinal in $\aleph_\gamma$. We conclude that unless $|\gamma| = \aleph_\gamma$, the cardinal $\aleph_\gamma$ is singular.

Any uncountable regular limit cardinal $\beth$ (if, in fact, there are any) must be **rather big** because of this odd condition $\beth = \aleph_\beth$, which it must satisfy. It is a place where an ordinal "catches up" to the cardinal to which it correspond through the aleph-subscript order isomorphism. It is hard to imagine how this can happen.

The condition is not, however, uncommon and in fact this condition is not enough to ensure regularity.

To see this suppose γ is any infinite ordinal. Let $\beth_0 = \aleph_\gamma$. Having found $\beth_n$ define $\beth_{n+1}$ to be $\aleph_{\beth_n}$. Now let $\beth = \bigcup_{n\in\mathbb{N}} \beth_n$. Obviously, the cofinality of $\beth$ is $\aleph_0$, and it is not (too) hard to show that $\aleph_\beth = \beth$.

So there are infinite singular cardinals as big as we like which satisfy this weird condition.

In ZFC, it is known to be impossible to prove[6] that there is a regular limit cardinal beyond $\aleph_0$, but no proof has been found that they are forbidden. Most set theorists seem to believe it never will be shown that their existence is incompatible with ZFC.

An uncountable regular limit cardinal is also known as a **weakly inaccessible cardinal**.

[6]The existence of an uncountable regular limit cardinal implies the consistency of ZFC. So their existence cannot be proven within ZFC (assuming ZF itself to be consistent) by Gödel's famous Incompleteness Theorem.

The regularity of $\mathbf{c}$ is independent of ZFC. Several popular additions to ZF, such as CH, imply regularity of $\mathbf{c}$. It is, however, known that $cf(\mathbf{c}) > \aleph_0$. This is because, from 21.4, for any infinite $\aleph$,

$$\left| \left(2^{\aleph}\right)^{cf\left(2^{\aleph}\right)} \right| > \left|2^{\aleph}\right| = \left|2^{\aleph\times\aleph}\right| = \left| \left(2^{\aleph}\right)^{\aleph} \right|.$$

We conclude:

21.9. $cf\left(2^{\aleph}\right) > \aleph$ for infinite cardinal $\aleph$.

As a special case we have

$$\left|\mathbf{c}^{cf(\mathbf{c})}\right| > \mathbf{c} = \left|2^{\aleph_0}\right| = \left|2^{\aleph_0\times\aleph_0}\right| = \left| \left(2^{\aleph_0}\right)^{\aleph_0} \right| = \left|\mathbf{c}^{\aleph_0}\right| \quad \text{so} \quad cf(\mathbf{c}) > \aleph_0.$$

Cohen showed that this imposes, essentially, the only restriction on the equation $c = \left|2^{\aleph_0}\right| = \aleph_\gamma$. It is consistent in ZFC, but unprovable, to assume that $\mathbf{c} = \aleph_\gamma$ for any "reasonably" defined $\gamma > 0$ for which $cf(\gamma) \neq \aleph_0$. For instance $\gamma = 1$, $\gamma = 2$ or $\gamma = \omega + 1$ are consistent choices. γ could be a singular or a weakly inaccessible cardinal.

The word "reasonably" is needed to forbid choices of the type "select $\gamma > \left|2^{\aleph_0}\right|$."

So, in the end, Cantor's hope that the Continuum Hypothesis would follow from ZFC has failed with a peculiar twist: it is consistent with ZFC that $\mathbf{c} = \aleph_1$. But it is also consistent that there is a vast chasm filled with a myriad of alephs between $\aleph_0$ and $\mathbf{c}$.

An uncountable regular cardinal $\beth$ is known as a **strongly inaccessible cardinal** provided $\left|2^{\aleph}\right| < \beth$ whenever the cardinal $\aleph$ is less than $\beth$. This defining condition implies that $\beth$ *cannot* be a successor, so strongly inaccessible cardinals (if any) must be weakly inaccessible. So the existence of these cardinals cannot be derived from the axioms of ZFC. It is believed that their existence is not incompatible with ZFC.

One important quality of these cardinals stems from the following fact. Consider a mathematical universe created from ZFC+SI, where SI is the assertion that a strongly inaccessible cardinal exists. SI is called a **Large Cardinal Axiom**. With this assumption, $\mathfrak{Ord}$ goes *much* higher than we thought.

The addition of an axiom of this kind has the same flavor as when the Axiom of Infinity was added so that $\aleph_0$ would be a *set*, very much bigger than all the cardinals before it. Without the Axiom of Infinity, the Zermelo hierarchy is stunted, rising only to what *we*, with that axiom, would call $\mathbb{V}_{\aleph_0}$. Without the Axiom of Infinity the proper class $\mathfrak{Ord}$ of all ordinals is what *we* would call the set $\aleph_0$. A Large Cardinal Axiom has a similar effect. The Zermelo hierarchy now rises to undreamed-of heights, with qualitatively distinct properties.

Why should $\aleph_0$ be the only cardinal that dominates $\left|2^{\lambda}\right|$ whenever λ is a smaller cardinal? This is a very powerful property. Proponents of SI claim that this kind of uniqueness would be unnatural and do various kinds of violence to their vision of $\mathfrak{Set}$. They say SI *must* be true.

Let ℸ be the least strongly inaccessible cardinal in ZFC+SI, and consider the Zermelo hierarchy $\mathbb{V}_{ℸ}$ up to height ℸ. $\mathbb{V}_{ℸ}$ is transitive.

Let ZFCℸ denote the **relativization** of ZFC to $\mathbb{V}_{ℸ}$: these are the usual axioms of ZFC which are restricted in scope so as to act on and refer to elements of the *set* $\mathbb{V}_{ℸ}$ alone. *Nothing* you can do with these relativized axioms will allow you to deduce that there are sets other than the elements of $\mathbb{V}_{ℸ}$. A resident of this universe would perceive $\mathbb{V}_{ℸ}$ to be a proper class. It is a fact that ZFCℸ is consistent if ZFC+SI itself is, and $\mathbb{V}_{ℸ}$ is said, therefore, to be a **transitive model** of ZFC in ZFC+SI.

The discussion of various types of cardinals and their properties can proceed much deeper. A **large cardinal** is any cardinal whose existence is known to be unprovable in ZFC, but whose existence is not known to be inconsistent with ZF.

There are many delicate theorems regarding these large cardinals, of which there may be many types such as strongly inaccessible cardinals, measurable cardinals, Ramsey cardinals, Mahlo cardinals, huge cardinals, supercompact cardinals and so on. These cardinals acquire the virtue of existence upon inclusion of an additional axiom to (or replacing an axiom of) ZFC set theory.

Though large cardinals cannot be built in ZFC their properties have explicit implications for numerous questions such as the measurability or analyticity of objects which *can* be built in ZFC. The study of large cardinals is currently an extremely active research area.

This is not work for the **faint of heart.** Seekers who aspire to understand the inner mysteries should consult Jech *Set Theory* [**?**].

Index