



THE MATHEMATICS OF INFINITY

A Guide to Great Ideas

THEODORE G. FATICONI

The Mathematics of Infinity

A Guide to Great Ideas

Theodore G. Faticoni

*Fordham University
Department of Mathematics
Bronx, NY*



A JOHN WILEY & SONS, INC., PUBLICATION

This Page Intentionally Left Blank

The Mathematics of Infinity

PURE AND APPLIED MATHEMATICS

A Wiley-Interscience Series of Texts, Monographs, and Tracts

Consulting Editor: David A. Cox

Founded by RICHARD COURANT

Editors Emeriti: MYRON B. ALLEN III, DAVID A. COX, PETER HILTON,
HARRY HOCHSTADT, PETER LAX, JOHN TOLAND

A complete list of the titles in this series appears at the end of this volume.

The Mathematics of Infinity

A Guide to Great Ideas

Theodore G. Faticoni

*Fordham University
Department of Mathematics
Bronx, NY*



A JOHN WILEY & SONS, INC., PUBLICATION

Copyright © 2006 by John Wiley & Sons, Inc. All rights reserved.

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.

Published simultaneously in Canada.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permission>.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic format. For information about Wiley products, visit our web site at www.wiley.com.

Library of Congress Cataloging-in-Publication Data is available.

Faticoni, Theodore G.

The Mathematics of Infinity: A Guide to Great Ideas

ISBN-13 978-0-471-79432-5

ISBN-10 0-471-79432-5

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

To *Professor Elliot Wolk* who taught me Set Theory.

This Page Intentionally Left Blank

Contents

Preface	ix
1 Elementary Set Theory	1
1.1 Sets	2
1.2 Cartesian Products	17
1.3 Power Sets	20
1.4 Something From Nothing	22
1.5 Indexed Families of Sets	27
2 Functions	37
2.1 Functional Preliminaries	38
2.2 Images and Preimages	52
2.3 One-to-One and Onto Functions	61
2.4 Bijections	65
2.5 Inverse Functions	68
3 Counting Infinite Sets	75
3.1 Finite Sets	75
3.2 Hilbert's Infinite Hotel	82
3.3 Equivalent Sets and Cardinality	98
4 Infinite Cardinals	103
4.1 Countable Sets	104
4.2 Uncountable Sets	117
4.3 Two Infinities	126
4.4 Power Sets	132
4.5 The Arithmetic of Cardinals	145

5 Well Ordered Sets	163
5.1 Successors of Elements	163
5.2 The Arithmetic of Ordinals	173
5.3 Cardinals as Ordinals	184
5.4 Magnitude versus Cardinality	197
6 Inductions and Numbers	205
6.1 Mathematical Induction	205
6.2 Transfinite Induction	222
6.3 Mathematical Recursion	231
6.4 Number Theory	237
6.5 The Fundamental Theorem of Arithmetic	240
6.6 Perfect Numbers	242
7 Prime Numbers	247
7.1 Prime Number Generators	247
7.2 The Prime Number Theorem	251
7.3 Products of Geometric Series	254
7.4 The Riemann Zeta Function	261
7.5 Real Numbers	265
8 Logic and Meta-Mathematics	271
8.1 The Collection of All Sets	271
8.2 Other Than True or False	274
Bibliography	283
Index	284

Preface

The most primitive of herdsman used a pouch of stones to keep track of the number of sheep he had in the field. As each sheep would enter the field the herdsman would place a stone in a pile. As the sheep would leave the field the herdsman would place the stones back into the pouch. If there were stones left on the ground then some sheep were missing. If there were no stones left and no sheep left then all was well with the herd. And if there were no more stones but there were more sheep then somehow the herdsman picked up a ewe or two.

This correspondence between pouch stones and sheep is one of the most primitive forms of counting known. In today's language this is known as a one-to-one correspondence, or a bijection between pouch stones and sheep. This kind of counting is continued today when we make an attendance sheet. Each name on the sheet corresponds to exactly one child in the class, and we know some child is missing if he/she does not respond to his/her name. A more important correspondence is found in the grocery store. There we associate a certain number called a price with each item we put in our cart. The items in the cart correspond to a number called the total price of the cart. When we compare our receipt with the objects in the cart we are imitating the sheep herdsman's pouch stones.

Believe it or not, mathematicians count like the primitive herdsmen. The number 1 is all sets that match up in an exact manner to the set $\{\bullet\}$. Thus we say that $\text{card}(\{\bullet\}) = 1$, and we say that $\text{card}(\{\ast\}) = 1$. The number 1 becomes all that we associate with one element.

We use the convenient symbol 1 to denote all possible sets that match up perfectly with $\{\bullet\}$. The symbol 1 is convenient because it is what we have been taught all these years. The number 2 is defined to be all of those sets that match up perfectly with $\{\bullet, \ast\}$.

$$\text{card}(\{\bullet, \ast\}) = 2.$$

This is 2 because we define it that way. It agrees with our training. It represents all possible sets that match up exactly with the set $\{\bullet, \ast\}$. This is exactly what you have been taught.

Next up is what we mean by *matches up perfectly*. This is the bijection we alluded to earlier. Sets A and B are called *equivalent* if there is a bijection between them. That is, they match up perfectly. In other words, there is a way of matching up elements between A and B , called a *function* or bijection

$$f : A \longrightarrow B$$

such that

1. different elements of A are mapped to different elements of B , and
2. each element of B is associated with some element of A .

For finite sets this bijection can be drawn as a picture. Let $A = \{a_1, a_2, a_3\}$ and let $B = \{b_1, b_2, b_3\}$. Then a bijection between A and B is

$$\begin{aligned} a_1 &\longmapsto b_1 \\ a_2 &\longmapsto b_2 \\ a_3 &\longmapsto b_3. \end{aligned}$$

Here is another such bijection.

$$\begin{aligned} a_1 &\longmapsto b_3 \\ a_2 &\longmapsto b_2 \\ a_3 &\longmapsto b_1. \end{aligned}$$

You see, the bijection you choose does not have to respect the subscripts. These mappings are bijections because as you can see the elements a_k are sent to different elements b_ℓ . Also each element in B is associated with an element in A . That is exactly how mathematicians count elements in sets.

An impressive extension of this idea is that we can count *infinite sets* in the same manner, but you must use different symbols to denote $\text{card}(A)$. We let

$$\text{card}(A) = \text{the cardinality of } A,$$

which is simply all sets B such that A is equivalent to B . That is, $\text{card}(A)$ is all those sets B for which there exists a bijection $f : A \longrightarrow B$. Hence $B \in \text{card}(A)$ or $\text{card}(A) = \text{card}(B)$ exactly when there is a function $f : A \longrightarrow B$ such that

1. different elements of A are mapped to different elements of B , and
2. each element of B is associated with some element of A .

Notice that the definition of *bijection* has not changed.

Since these sets are infinite we need a new symbol to denote $\text{card}(A)$ of infinite sets. It is traditional to use the Hebrew letter *aleph*

\aleph

to denote infinite cardinals. Let

$$\begin{aligned} \aleph_0 &= \{0, 1, 2, 3, \dots\}, \\ \aleph_1 &= \{x \mid x \text{ is a real number}\}. \end{aligned}$$

So \aleph_0 is the set of whole, nonnegative numbers, and \aleph_1 is the set of all real numbers. These would be decimal expansions like 1.414 and 3.14159. Then we write

$$\text{card}(\aleph_0) = \aleph_0$$

and we say *aleph naught*. It is quite a surprising mathematical (universal) truth that there is a cardinal \aleph_1 such that

$$\aleph_0 < \aleph_1.$$

Indeed there is an infinite chain of infinite cardinals

$$\aleph_0 < \aleph_1 < \aleph_2 < \aleph_3 < \dots$$

We will have a chance to expand upon this idea in the later chapters of this book.

The Mathematics of Infinity

This Page Intentionally Left Blank

Chapter 1

Elementary Set Theory

The ideal writing style ascribed to by mathematicians is that in writing mathematics, *less is more*. If we can convey the exact idea of a concept with 5 words instead of 10 then we will use 5 words. Thus we will use the statement *Cardinal numbers form a well ordered collection* over the wordier statement *The well ordered property is enjoyed by the collection of cardinal numbers*. The second statement is mathematically correct but it is more than we need to convey the idea.

I have tried to practice this ideal while writing the mathematics in this book. The only exceptions to this ideal are made on the basis of decisions on the educational value of sentence structure, the anecdotal comments, or discussions of this sort that occur between mathematical discourse. Sometimes it is good to sacrifice some mathematical austerity in the interest of getting an important point across to the reader. As the reader will clearly see, this economy of words in mathematical writings is not exercised in the text of a discussion. Discussions and intermediate anecdotes contain examples and illustrations which are the only tools we have to illustrate a concept. Since I have sacrificed a good bit of mathematical rigor in favor of clarity, examples and illustrations are necessary if I am to get some subtle ideas across to the reader. This form of personalized writing style is unavoidable when discussing advanced ideas from mathematics in the popular press.

We have a bit of a mountain to climb in this book, so be patient. Perhaps you can sit down in an overstuffed chair or at a table and

open the book. Maybe you have a pencil and paper handy. That's a good idea. Some of these topics need to be diagrammed. And certainly you have a cup of beverage, coffee would be my choice. Now turn on that lamp overhead and blend in the final ingredient: a handful of inspiration. Good luck.

1.1 Sets

Some foundation has to be laid down before any discussion of mathematics can begin. Our foundation is *Set Theory and functions* and we will discuss these notions in the first two chapters. There is one assumption that we will use implicitly and explicitly throughout this book. Our underlying assumption is that all mathematical objects considered in this book are from a Mathematical Universe in which these objects exist. Thus when we say that \aleph_0 is a cardinal it is to be understood that this cardinal lives in a Mathematical Universe and that we can examine it there. This Mathematical Universe is a classical idea due to the Greek philosopher and mathematician Plato. Thus in making our universal assumption we are following in a good classic tradition. The intent here is clear. We will write *given* x if we wish to examine an object in the Mathematical Universe. Our definition of *Set* requires us to know when an object is given. Thus our universal assumption goes to work right away. Whatever else you might believe, let us agree that this Mathematical Universe is there and that we can study the elements in it. Said assumption will not change as we work our way through this book.

Mathematical statements are explicitly or implicitly about elements in *sets*. One of the highlights of twentieth century mathematics was to show that all of mathematics can be derived from the basic concepts in *Set Theory*. There is also strong empirical evidence that a firm grounding in Set Theory gives the reader the mathematical experience necessary to understand the advanced abstract ideas covered in later chapters. Thus we begin our introduction to modern mathematics with some work in *Set Theory*. However, the Elementary Set Theory that we will cover comes with a price. It can be terribly dry and uninspiring. We ask the reader to please be patient and work through the unions, intersections, and comple-

ments contained in this chapter. Your work will pay large dividends in our later discussions.

Definition 1.1.1 A set is a collection A of objects called elements that satisfy the property

Given x then either x is an element in A
or x is not an element in A .

We will use the standard notation $x \in A$ when we want to say that x is an element of A . The notation $x \notin A$ is used when we want to say that x is not an element of A .

Observe that the defining condition of sets applies to objects that are given or that exist in the Mathematical Universe, and to nothing else. That must seem strange. But that little bit of philosophy must be addressed elsewhere. In this text we will assume that everything we consider exists in the Mathematical Universe. The use of the words *Given* x simply gives me a way of pointing out where that x is. This is perhaps the last time in this book that you will have the opportunity to think about this, so spend some time with it, and then read on.

Some examples of sets are in order. Consider the simplest possible set. The *empty set* is the collection that contains no elements. It is denoted by

\emptyset .

The empty set is a set since if x exists then $x \notin \emptyset$. It arises in a surprising number of places in this book. For example, the set of all living people who will celebrate their -2nd birthday today is empty, as is the set of places in the universe whose temperature is measured at *absolute zero*. Chemists and physicists define absolute zero as that tempertaure at which all atomic motion stops, while

Heisenberg's Uncertainty Principle tells us that this cannot be measured. The set of dogs with a PhD in mathematics is, at the time of this writing, also empty.

For the most part, the sets in this book are listed according to their elements. Illustrations of how this is done include

1. $\{\bullet\}$, a set with the one element \bullet
2. $\{+, =, :\}$ the set whose elements are the symbols $+, =, :$
3. $\{\text{Lori, Christian, Marcus, Gwendolyn}\}$.

The set of *natural numbers*

$$\mathbb{N} = \{0, 1, 2, 3, \dots\}$$

is given as an *implied list*. You might have called the elements of \mathbb{N} *whole numbers* in the past, but the common professional designation for $x \in \mathbb{N}$ is that x is a *natural number*. The set of *integers*

$$\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$$

is also an implied list.

An *implied list* is an infinite list of elements. An ellipsis \dots is usually used to denote the fact that the elements in an implied list continue in the same manner or under the same pattern. It is hoped that each reader sees the same pattern when they get to \dots . Implied lists can be trouble like that. While we would agree that

$$\{2, 4 \dots\}$$

begins the set of even positive numbers (so that the next listed element is 6) we might also see that $\{2, 4, \dots\}$ begins the powers of 2 so that the next element on the implied list would be $2^3 = 8$.

$$\{2, 4, 8, \dots\}$$

If we begin a set with numbers

$$2, 3, 5, x$$

then what might the next element x be? Ponder this awhile before you read on. You might say that $5 = 2 + 3$ so $x = 3 + 5 = 8$. Or you might say that 2, 3, 5 are the first three prime numbers so that $x = 7$, the fourth prime number. We will be explicit in the pattern used in an implied list, so there will be no confusion as to the next value in a list.

Sometimes the best way to denote a set is to use a *predicate* to describe the elements. A predicate is a description in words and symbols. For example, *blue* and *a temperature on Earth* are predicates. Let P be any predicate. The symbols

$$A = \{x \mid x \text{ satisfies } P\}$$

are read *A is the set of elements x such that x satisfies P*. The symbol \mid is read as *such that* or *with the property that*. For example, we would find it most difficult to give a complete list of the elements in the set $\{x \mid x \text{ is a person on the Earth}\}$. It is better by far to use the predicate. Implied lists are often better written as predicates.

$$\begin{aligned} \mathbb{N} &= \{x \mid x \text{ is a natural number}\}, \\ \mathbb{Z} &= \{x \mid x \text{ is an integer}\}, \\ \mathbb{E} &= \{x \mid x \text{ is an even natural number}\}, \\ \mathbb{P} &= \{x \mid x \text{ is a prime number}\}. \end{aligned}$$

We could not define the following sets of numbers without the use of the predicate form. Let

$$\begin{aligned} \mathbb{R} &= \{x \mid x \text{ is a real number}\}, \\ \mathbb{Q} &= \left\{ \frac{n}{m} \mid m \neq 0 \text{ and } n, m \in \mathbb{Z} \right\}. \end{aligned}$$

\mathbb{R} is called the set of *real numbers* and \mathbb{Q} is called the set of *rational numbers*. While we would have great difficulty in trying to list the real numbers (in fact it is mathematically impossible) we can make an implied list of the rational numbers \mathbb{Q} . To illustrate this point let us make an implied list of the set

$$\mathbb{Q}^+ = \left\{ \frac{n}{m} \mid m \neq 0 \text{ and } n, m \in \mathbb{N} \right\}$$

of *positive rational numbers*. We begin the implied list by listing those rational numbers with numerator 1. The next row will be those rationals with numerator 2, and then I hope the pattern is clear. We have thus constructed the following implied list.

$$\begin{array}{ccccc}
 \frac{1}{1} & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \dots \\
 \frac{2}{1} & \frac{2}{2} & \frac{2}{3} & \frac{2}{4} & \dots \\
 \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \dots \\
 \frac{3}{1} & \frac{3}{2} & \frac{3}{3} & \frac{3}{4} & \dots \\
 \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \frac{1}{6} & \dots \\
 \frac{4}{1} & \frac{4}{2} & \frac{4}{3} & \frac{4}{4} & \dots \\
 \frac{1}{4} & \frac{1}{5} & \frac{1}{6} & \frac{1}{7} & \dots \\
 \vdots & \vdots & \vdots & \vdots &
 \end{array} \tag{1.1}$$

Observe that the rows of this infinite rectangular array list the positive rationals by increasing the denominator of the fraction and holding the numerator fixed. There are repeated values on this implied list, aren't there. We could clean that up by deleting the rational numbers that are not in reduced form. The result is the implied list below.

$$\begin{array}{ccccc}
 \frac{1}{1} & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \dots \\
 \frac{2}{1} & & \frac{2}{3} & & \dots \\
 \frac{1}{2} & & \frac{1}{3} & & \dots \\
 \frac{3}{1} & \frac{3}{2} & & \frac{3}{4} & \dots \\
 \frac{1}{3} & & & \frac{1}{4} & \dots \\
 \frac{4}{1} & & \frac{4}{3} & & \dots \\
 \frac{1}{4} & & \frac{1}{3} & & \dots \\
 \vdots & \vdots & \vdots & \vdots &
 \end{array} \tag{1.2}$$

Note that among others, the fractions

$$\frac{2}{2}, \frac{3}{3}, \frac{2}{4}, \frac{4}{2}$$

have been deleted from our list.

We will return to this list several times in this book. You might ask: Why is he doing that? Why would he present a list like that over and over again? The answer is that by referring to this list several times I hope to pique your curiosity. This list will be a friendlier mathematical object when next you see it.

Given a general mathematical collection X , like a set, there are several ways that mathematicians will examine X . We will define operations between X and sets like it. We will consider subsets in X that have the same properties that X has. We will also compare sets like X by considering functions between them. In this section we will consider operations on sets and subsets of sets, and in the next chapter we will consider functions between sets.

We operate on sets in the following ways.

Definition 1.1.2 Let A and B be sets.

1. $A \cap B = \{x \mid x \in A \text{ and } x \in B\}$.
2. $A \cup B = \{x \mid x \in A \text{ or } x \in B\}$. Our use of the word or is the inclusive or. That is, $x \in A$, $x \in B$, or x is in both A and B .

Because this book is being directed at people with at least a high school education, the assumption is that you are familiar with the operations \cap and \cup for sets. Thus we will make a minimal number of concrete examples of what these operations can do. Let $A = \{\bullet, :\}$ and $B = \{\bullet, *\}$. Then $A \cap B = \{\bullet\}$, $A \cup B = \{\bullet, :, *\}$.

Recall that \mathbb{P} = the set of prime numbers. These are the numbers whose proper natural divisors are precisely 1 and themselves. The first few primes are 2, 3, 5, 7, 11, while 1, 4, 9, 12, and 15 are not prime. We say that 4, 9, 12, and 15 are *composite numbers*. Even numbers other than 2 are composite numbers.

A positive natural number p is a prime number if given a set of p dots, \bullet , then the only rectangle that we can form from those dots is a line of dots. For example, 2 and 3 are prime because the only rectangles we can form with those dots are

$$\bullet\bullet \quad \text{and} \quad \bullet\bullet\bullet$$

while 4 and 6 are not primes since we can draw rectangles with 4 and 6 dots.



If we let \mathbb{E} denote the even numbers, then $\mathbb{E} \cap \mathbb{P} = \{2\}$ since 2 is the only even prime. $\mathbb{E} \cup \mathbb{P} = \{x \mid x \text{ is even or } x \text{ is prime}\}$. Some but not all of the elements in $\mathbb{E} \cup \mathbb{P}$ are

$$2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 101, 102.$$

Now that we have defined *set* we can define what we call a *subset*. The definition of subset represents a common technique in mathematics. Once we have defined an object X that satisfies some properties then we can define a similar subset in X that satisfies the properties of X . If we consider these subsets as slices of X then the subsets of X can be used to study X .

Definition 1.1.3 Let A and B be sets. We say that A is a subset of B if

given $x \in A$ then we can show that $x \in B$.

We write

$$A \subset B$$

if A is a subset of B .

For example, $\{3\} \subset \{1, 2, 3, 4\}$ because 3 is an element of $\{1, 2, 3, 4\}$, $\{2, 3\} \subset \mathbb{P}$ because 2 and 3 are primes, and the set of children on Earth is a subset of the set of people on the Earth. If $A \subset B$ then we also say that A is *contained in* B . The containments

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}$$

are clear since each $x \in \mathbb{Z}$ can be written as $x = \frac{x}{1}$, and each fraction is a real number. Other containments are the set of your family contained in the set of people on Earth. The land making up the country of Peru is contained in the land that makes up the

South American continent. College professors form a set that is a subset of the set of educated people on Earth, and the ages of people on this Earth in years forms a subset of the natural numbers \mathbb{N} .

When you are writing proofs please do not take shortcuts. Use as much pencil lead, pen ink, and blank paper as possible when proving theorems. Fill in the details. Don't be afraid to use some paper. The trees that were used to manufacture that paper have since departed. You are not saving trees by using less paper in your mathematical deliberations. Get a feel for the flow of your argument. Let the mathematics guide your argument. It is unlikely that we can find a logically consistent and convincing proof by writing down an argument with many gaps. Your hard work will be rewarded when you can follow some of the subtler arguments presented in later chapters. For now, to give you a running start in your reading, we present our arguments in all their baroque detail.

The identities in the next theorem are not hard. By including the proofs we are improving the reader's intuition of what constitutes a proof. This could be the first mathematics you have ever met that did not reduce to a formula or that did not require you to plug in some kind of numbers. I therefore recommend that you carefully read the following theorem as it consists of elementary examples. For instance, to show that $A \subset B$ we start with an element $x \in A$ and then show that $x \in B$. There is no formula and there is no shorter way to do this.

Theorem 1.1.4 *Let A , B , and C be sets.*

1. $A \subset A$.
2. $A \cap B \subset A$.
3. *If $A \subset B$ and if $B \subset C$ then $A \subset C$.*

Proof: 1. Proofs like this are mathematical double-talk. $A \subset A$ because given $x \in A$ then $x \in A$.

2. As with every proof that attempts to prove $X \subset Y$ we start with an element of X and show that it is Y . So let $x \in A \cap B$. By the definition of $A \cap B$, $x \in A$ and $x \in B$. Specifically, $x \in A$. Then $A \cap B \subset A$ by the definition of subset.

3. Suppose that $A \subset B$ and that $B \subset C$. Let $x \in A$. We must show that $x \in C$. Using the definition of subset, $x \in A$ and $A \subset B$ implies that $x \in B$. And $x \in B$ and $B \subset C$ implies that $x \in C$. We have just proved that

$$x \in A \text{ implies that } x \in C$$

so we can conclude that $A \subset C$. This completes the proof.

Admittedly we have included a great deal of detail in these proofs. We do this to make a point. When you construct a proof you are trying to tell the reader what you are thinking. Thus in your argument every reason for the given steps must be explained. Since the reader does not know your mind, even the simplest of arguments can be confusing to them. Furthermore, the ideas flow one to the next and *every statement in the proof is accompanied by a reason*. This flow helps the reader follow the argument. A jumble of loosely connected statements will not form a convincing argument.

The next type of argument, called a proof by contradiction, can be kicked around and explored. It is probably the first time you have met such an argument. Each proof by contradiction proceeds as follows. Suppose that we want to prove that P is a true statement. The proof by contradiction begins with the assumption that P is a false statement. We then follow our logical noses until we arrive at a mathematical mistake called a contradiction. This contradiction shows us that we began our proof with a false statement. That is, it is false that P is false. Therefore P must be a true statement. The rest of this book contains many proofs by contradiction, so try to work your way through this one.

Theorem 1.1.5

$$\emptyset \subset A.$$

Proof: Suppose for the sake of contradiction that $\emptyset \subset A$ is a false statement. Then $\emptyset \not\subset A$. Consider for a moment what it means when $B \not\subset A$. It means that some element of B is not in A . Since $\emptyset \not\subset A$ there is some element $x \in \emptyset$ such that $x \notin A$. But $x \in \emptyset$ contradicts the fact that \emptyset has no elements. This is the desired contradiction, the mathematical mistake. We conclude

that our initial assumption $\emptyset \not\subset A$ is false. Therefore $\emptyset \subset A$, which completes the proof.

Have you ever wondered what it means for objects x and y to be equal, $x = y$? If x and y are numbers then we have an intuitive understanding of what it means for $x = y$. For example, would you have recognized the following equations on your own?

$$\begin{aligned}\frac{1}{2} &= .5 \\ &= .499\bar{9} \\ &= \frac{1}{2^2} + \frac{1}{2^3} + \frac{1}{2^4} + \dots\end{aligned}$$

Equality is not always such a simple concept. For our purposes we will accept that *if objects x and y exist, there is an understanding of the identification $x = y$.*

The equality of sets is not intuitive at all. It is mathematically definite, mathematically rigorous. Thus for sets A and B there is a precise definition of equality.

Definition 1.1.6 *Let A and B be sets. We say that A equals B if $A \subset B$ and $B \subset A$. The traditional notation is*

$$A = B.$$

Thus to prove that $A = B$ we must first prove that $A \subset B$ and then that $B \subset A$. There are no shortcuts in this kind of proof. There are two things to show and we must be prepared to prove them. Some examples will show you what we mean.

If we let

$$\mathbb{F} = \text{the set of finite or repeating decimal numbers}$$

then we can prove that

$$\mathbb{Q} = \mathbb{F}$$

as follows. To prove that $\mathbb{Q} \subset \mathbb{F}$ let $x \in \mathbb{Q}$. Then x is a fraction and your training in converting fractions into decimals makes x a finite or a repeating decimal. Thus $\mathbb{Q} \subset \mathbb{F}$. Conversely, we show that

$\mathbb{F} \subset \mathbb{Q}$. Let $x \in \mathbb{F}$. You may remember some training in arithmetic that shows you how to convert the finite or repeating decimal x into a fraction. We will not pursue that arithmetic here. Then $x \in \mathbb{Q}$ and we conclude that $\mathbb{F} = \mathbb{Q}$.

Let

$$A = \{x \mid x \text{ is a solution to } x^2 - 1 = 0\}.$$

We will show that

$$A = \{-1, 1\}.$$

The numbers -1 and 1 are solutions to $x^2 - 1 = 0$ since

$$(-1)^2 - 1 = 0 \text{ and } 1^2 - 1 = 0.$$

Thus $\{-1, 1\} \subset A$. Conversely, let us suppose that $x \in A$. Then x is a real number such that $x^2 - 1 = 0$. Our high school algebra gives us the following equations.

$$\begin{aligned} x^2 - 1 &= 0, \\ (x - 1)(x + 1) &= 0, \\ x - 1 = 0 &\quad x + 1 = 0, \\ x = 1 &\quad x = -1. \end{aligned}$$

Thus $A \subset \{-1, 1\}$ from which we conclude that $\{-1, 1\} = A$.

Other examples come from the operations \cup and \cap on sets A , B , and C .

Theorem 1.1.7 *Let A , B , and C be sets.*

1. $A \cup B = B \cup A$.
2. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

Proof: 1. We begin by showing that $A \cup B \subset B \cup A$. The reader will note the interplay between the definitions of \cup and \subset . Because of its elementary nature the proof is going to sound like mathematical doublespeak.

Let $x \in A \cup B$. By definition of \cup , $x \in A$ or $x \in B$, so that $x \in B$ or $x \in A$. By the definition of \cup , $x \in B \cup A$, and hence $A \cup B \subset B \cup A$.

Conversely we must show that $B \cup A \subset A \cup B$. Let $x \in B \cup A$. Then $x \in B$ or $x \in A$ so that $x \in A$ or $x \in B$. Thus, by definition of \cup , $x \in A \cup B$ and hence $B \cup A \subset A \cup B$.

Therefore $A \cup B = B \cup A$.

The proof of part 2 is left as an exercise for the reader.

Now that was deadly dull. You might complain that the whole idea of the proof was to say that “ P or Q ” is the same as saying “ Q or P ,” and you’d be right. This is a tedious example of what it means to prove that two sets are equal. But the essence of the proof is the important thing. The introduction of all that detail is an attempt to get you to fill in more detail when you think and write. People tend to make intellectual leaps when they argue. I suppose that the person arguing feels that he can fill in the detail later or perhaps the arguer feels that the detail is unimportant. When original arguments avoid the details they are usually hard to follow and they are often incorrect. We cannot allow ourselves the luxury of gaps in our train of thought when we argue mathematically.

Think of a mathematical argument as a maze in which you want to get from point A to point B . There are many paths we can choose. Because this is a maze, we cannot lift our pencils off the paper, thus ignoring the boundaries of the paths, and then drop the pencil point on B . That would be contrary to the spirit of playing with mazes. Many paths from A lead to a blind alley. We cannot find B down these paths. However, we can learn something from these blind alleys. They indicate the paths that we should avoid in trying to get to B . Perhaps there is some general principle that we can learn from these blind alleys.

Let us apply the maze analogy to the construction of a proof. We know what we have to start with (the hypotheses) and we know where we want to finish (the conclusion). Try repeatedly to construct a logical path from the hypotheses to the desired conclusion. Don’t give up. Keep at it. Refine your argument but do not leave out essential details. Mathematical proofs do not come cheaply. They require a good deal of hard work. Eventually, we will find that path from the hypotheses to the conclusion. Then, if we apply what we have learned about the problem, we will rewrite the proof and perhaps find an elegant or beautiful path through the logical maze from the point of origin represented by the hypotheses to the

ending point represented by the conclusion.

We want to investigate what it means when an element x is not in a set A , $x \notin A$. This is also a good opportunity to stretch our logical experience with the word “not.”

Let me introduce a new notion here. Sometimes we will find that several sets A , B , and perhaps C are contained in one larger set \mathcal{U} . This set \mathcal{U} plays the role of the universe for A , B , C in that all operations on A , B , and C are relative to \mathcal{U} . For this reason \mathcal{U} is called a *universal set*. For example, the set of real numbers \mathbb{R} is a universal set for the natural numbers \mathbb{N} and the rational numbers \mathbb{Q} . On the other hand, \mathbb{N} is the universal set for \emptyset , $\{1\}$, and $\{2, 4, 6, \dots\}$.

Definition 1.1.8 Let A and B be sets in some universal set \mathcal{U} . Then

$$A \setminus B = \{x \mid x \in A \text{ and } x \notin B\}$$

and

$$A' = \{x \mid x \in \mathcal{U} \text{ and } x \notin A\}.$$

We say that A' is the complement of A .

Intuitively, to form $A \setminus B$ we take the elements of A and we throw out the elements of B . To form A' we take those elements in the universal set \mathcal{U} that are not in A .

Some examples will help us see what we are talking about.

Let us begin with $A = \{a, b, c, d, e\}$, $B = \{c, d, e, f, g\}$, $C = \{a, b, c, d, e, f, g, h\}$. Then $A \setminus B$ is the set of elements in A that are not in B . Thus

$$A \setminus B = \{a, b\}$$

as the reader can easily verify. It is somewhat of a surprise to see that

$$A \setminus C = \{\}$$

since each element of A is in C , $A \subset C$. Do this one element by element. We throw out a since $a \in A$ and $a \in C$. We throw out b since $b \in A$ and $b \in C$. We throw out c since $c \in A$ and $c \in C$. Keep going until you run out of elements from A to test. So $A \setminus C$ has no elements, $A = \emptyset$. Let us find $C \setminus B$. Write down the elements of C that are not in B . Delete the elements of C that are in B .

$$C \setminus B = \{a, b, h\}$$

Let \mathcal{P} be the set of people on Earth and let A be the set of male people. Then $\mathcal{P} \setminus A$ is the set of female people. This should be enough examples of a finite nature.

Recall that \mathbb{E} = the set of even integers and let $\mathcal{U} = \mathbb{N}$. Then \mathbb{E}' is the set of integers that are not even. That is, \mathbb{E}' is the set of odd numbers. If \mathbb{T} is the set of natural numbers that are divisible by 3 then \mathbb{T}' is the set of natural numbers not divisible by 3. That is, \mathbb{T}' is the set of natural numbers that have a nonzero remainder when divided by 3.

Recall that \mathbb{P} = the set of prime numbers and let $\mathcal{U} = \mathbb{N}$. Because primes cannot be properly factored while composite numbers can be properly factored, \mathbb{P}' is the set whose elements are 0 and 1, and the composite numbers. A number > 1 not in \mathbb{P} but in \mathbb{N} is divisible by at least two numbers $a, b \neq 1$. Then such a number is a composite number, and hence the elements of \mathbb{P}' are 0, 1 and the composite numbers.

Let \mathbb{Q}^- be the set of negative rational numbers,

$$\begin{aligned}\mathbb{Q}^- &= \left\{ -\frac{n}{m} \mid n, m \neq 0 \in \mathbb{N} \right\} \\ &= \{ -q \mid q \in \mathbb{Q}^+ \}.\end{aligned}$$

Let $\mathcal{U} = \mathbb{Q}$ and let $A = \mathbb{Q}^- \cup \{0\}$. Then A' is the set \mathbb{Q}^+ of positive rational numbers.

Another example is found by setting $\mathcal{U} = \{x \in \mathbb{R} \mid x \geq 0\}$ = the set of *nonnegative real numbers*. Then $\mathbb{Q}^+ \subset \mathcal{U}$ and $(\mathbb{Q}^+)'$ is the set of *positive irrational numbers*.

Here's a challenge for the reader. Without peeking below, try to describe the symbols $x \in (A)'$ using words. This might help you see the power of this symbolism over the language in dealing with

double negatives. As you work on this exercise note the interplay between the symbols x , A , and $()'$ and the word *not*.

The next theorem is another example of a *proof by contradiction*.

Theorem 1.1.9 Assume that there is some universal set \mathcal{U} that contains the set A . Then $A \cap A' = \emptyset$.

Proof: As with each proof by contradiction, we begin by assuming for the sake of contradiction that $A \cap A' \neq \emptyset$. We now seek that mathematical mistake, that contradiction. Then $A \cap A'$ must contain an element, say, $x \in A \cap A'$. By the definition of \cap , $x \in A$ and $x \in A'$ or equivalently $x \in A$ and $x \notin A$. This conclusion contradicts the fact that A is a set. (Reread Definition 1.1.1 to see that if A is a set then $x \in A$ or $x \notin A$ but not both.) Thus our initial assumption is false. That is, $A \cap A' = \emptyset$, which completes the proof.

At this point I hope you have convinced yourself that

$$(A')' = A$$

for each set A . Here is how I hope you argued. Let $x \in A$. By the definition of A' , $x \notin A'$, and hence x is in the complement of A' . The complement of A' is $(A')'$ so $x \in (A')'$ and so $A \subset (A')'$.

On the other hand, suppose that $x \in (A')'$. Then by definition of complement, $x \notin A'$. Either $x \in A$ or $x \notin A$. Assume to the contrary that $x \notin A$. Then $x \in A'$, contrary to $x \notin A'$. Therefore $x \in A$, which implies that $(A')' \subset A$. Consequently $A = (A')'$, and this completes the proof.

These two proofs were instructive exercises in mathematical double talk. The level of detail I used was strictly an educational device. No professional includes the attention to minutiae that we have given here. However, in an introductory work like the present one, it is necessary for us to observe this obsessive level of detail in a proof. This level of detail allows you very little freedom of thought when you read it, and that is the key. By giving this level of detail you and I are certain to be thinking the same thing, thus giving me knowledge of how much further I can push the level of understanding.

To hone the reader's writing skills we present a series of elementary exercises. The reader is encouraged to try these exercises and to practice the completeness and clarity of a rational argument that characterizes a well constructed proof. Let A , B , and C be sets contained in some larger universal set \mathcal{U} .

1. $A \cap B \subset A \cup B$.
2. $A \cap B = B \cap A$.
3. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.
4. $A \cup A' = \mathcal{U}$.
5. **DeMorgan's Law**
 - (a) $(A \cap B)' = A' \cup B'$.
 - (b) $(A \cup B)' = A' \cap B'$.
6. $(A \setminus B) \cap B = \emptyset$.
7. $(A \setminus B) \cap (B \setminus A) = \emptyset$.
8. $(A \setminus B) \cup (B \setminus A) = (A \cup B) \setminus (A \cap B)$.

1.2 Cartesian Products

The next operation on sets allows us to make a larger dimensional set from a number of smaller dimensional sets.

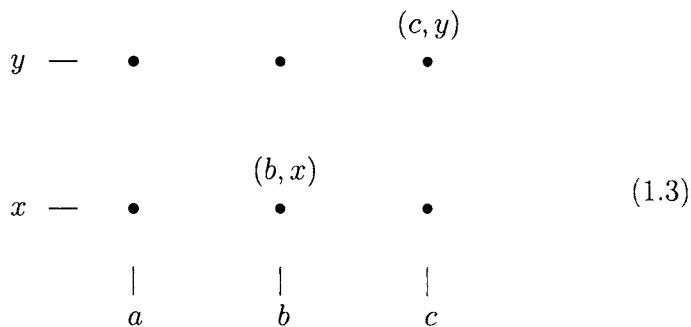
Definition 1.2.1 *Let A and B be sets. The Cartesian product of A and B is the set of pairs*

$$A \times B = \{(x, y) \mid x \in A \text{ and } y \in B\}.$$

To form $A \times B$ we take the elements of A and B and pair them up. For example, if we let $A = \{a, b, c\}$ and let $B = \{x, y\}$ then

$$A \times B = \{(a, x), (b, x), (c, x), (a, y), (b, y), (c, y)\}.$$

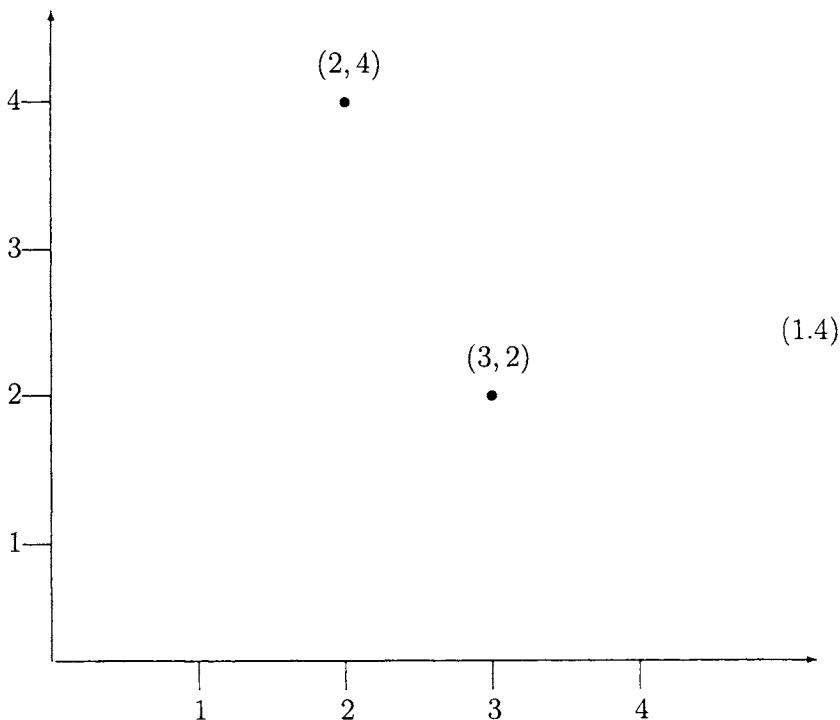
It is natural to picture $A \times B$ as a two dimensional lattice of points.



The *Cartesian plane*

$$\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$$

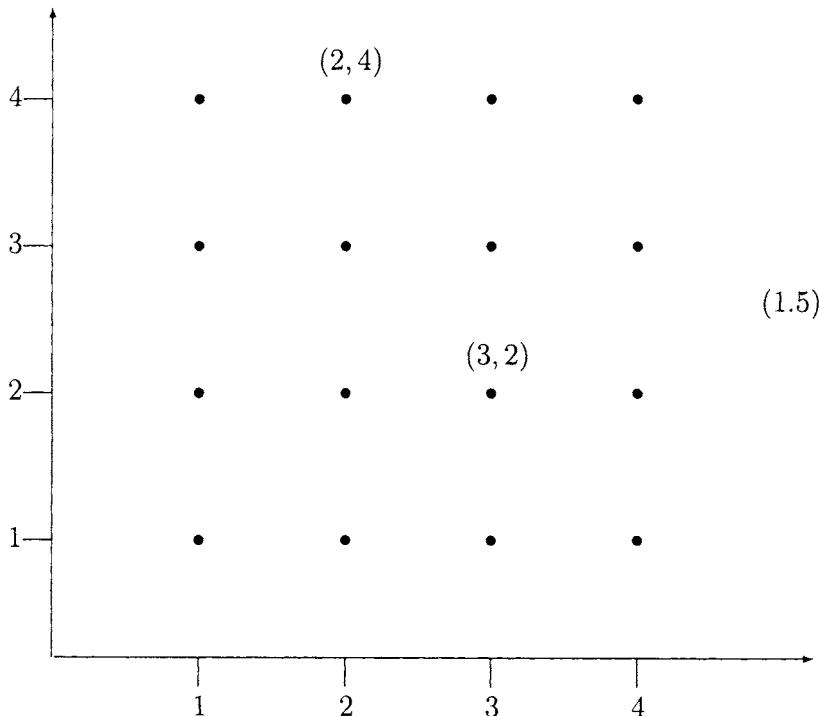
is the plane equipped with x and y axes with which you are familiar.



The Cartesian Plane $\mathbb{R} \times \mathbb{R}$

This is the plane on which you graphed points, lines, and parabolas in an early algebra course. The reason that the Cartesian plane \mathbb{R}^2 and the Cartesian product $\mathbb{R} \times \mathbb{R}$ are equal is that *each point in the Cartesian plane can be represented by a unique pair of real numbers (x, y)* . So the point labelled by $(2, 4)$ is found by moving 2 units in the x direction and then 4 units in the y direction.

We then have a picture that we can use to diagram a number of Cartesian products. Such an example is *the pairs of natural numbers* $\mathbb{N} \times \mathbb{N}$.



The Array of Natural Numbers $\mathbb{N} \times \mathbb{N}$

$$\mathbb{N} \times \mathbb{N} = \{(n, m) \mid n, m \in \mathbb{N}\}.$$

Here is how we picture $\mathbb{N} \times \mathbb{N}$ as a subset of the Cartesian plane. This array extends indefinitely upward and to the right. Thus $\mathbb{N} \times \mathbb{N}$ contains $(1, 1)$ and $(101, 102)$ but it does not contain $(\frac{1}{2}, 3)$ or $(\sqrt{2}, 0)$. We will use an array similar to this one when we count the set

$$\mathbb{Q}^+ = \{x \in \mathbb{Q} \mid x > 0\}.$$

The next result may strike you as a bit odd. It is another example of the care with which the empty set must be treated. The proof is one by contradiction. Briefly, $A \times \emptyset$ is empty because it does not contain any *pairs*.

Example 1.2.2 If A is a set then

$$A \times \emptyset = \emptyset.$$

Proof: Assume for the sake of contradiction that there is an element $p \in A \times \emptyset$. Then $p = (x, y)$ for some $x \in A$ and $y \in \emptyset$. This contradicts the fact that \emptyset has no elements. Thus our initial assumption is false, and it must be true that $A \times \emptyset = \emptyset$. This concludes the proof.

For instance, $\mathbb{N} \times \emptyset$, $\mathbb{R} \times \emptyset$, and, for sets A , $\emptyset \times A$ are the empty set. You can match these mind stretchers by thinking of multiplication by 0. Given any number x , $x \cdot 0 = 0$ and in much the same way $X \times \emptyset = \emptyset$ for sets X .

This is not the last time we will make an analogy between operations on sets and operations on numbers. In the later portion of this book, we make the analogy precise.

1.3 Power Sets

Let A be a set. The reader is acquainted with the idea of a subset of A . We will take that idea to a new level by introducing *the set of all subsets of A* . This power set will be an important device in the later chapters of this book.

Definition 1.3.1 Let A be a set. The power set of A is the set of all subsets of A . We write

$$\mathcal{P}(A) = \{\text{sets } X \mid X \subset A\}.$$

Some examples will be useful. Since power sets grow quickly our examples of necessity will be small. It may be hard to accept at first that *the elements of $\mathcal{P}(A)$ are themselves sets*. If $A = \{w, x, y, z\}$ then $\mathcal{P}(A)$ contains *as elements* the sets $\emptyset, \{w, z\}, \{x, y, z\}$, and 11 more sets. See if you can find them. If we let

$$\mathbb{R}^+ = \{x \in \mathbb{R} \mid x > 0\}$$

denote the set of *positive real numbers* then the power set $\mathcal{P}(\mathbb{R}^+)$ of \mathbb{R}^+ contains *as an element* the set \mathbb{Q}^+ . $\mathcal{P}(\mathbb{R}^+)$ also contains the element $\{1, 2, 3\}$ and the element $\{\sqrt{2}, \sqrt{3}, \sqrt{5}, \sqrt{6}, \dots\}$. (What is the next element on the last implied list, reader?) Seeing a set as an element in a larger set can be done but it takes practice.

Since \emptyset and A are always subsets of A , $\mathcal{P}(A)$ contains \emptyset and A .

$$\emptyset, A \in \mathcal{P}(A) \quad \text{for any set } A.$$

For example,

$$\mathcal{P}(\{\bullet\}) = \{\emptyset, \{\bullet\}\}$$

and

$$\mathcal{P}(\{a, b\}) = \{\emptyset, \{a\}, \{b\}, \{a, b\}\}.$$

The reader can find the eight elements in $\mathcal{P}(A)$ when $A = \{a, b, c\}$. Be sure that you include $\emptyset, \{b\}, \{a, c\}$, and $\{a, b, c\}$. Because \emptyset is the only subset of \emptyset we have the equation

$$\mathcal{P}(\emptyset) = \{\emptyset\}.$$

The importance of the power set is that it allows us to construct a much larger set from a smaller one. In general, if A is a finite set with exactly n elements then

$$\mathcal{P}(A) \text{ contains exactly } 2^n \text{ elements.}$$

Thus $\mathcal{P}(\{a, b\})$ has $2^2 = 4$ elements while $\mathcal{P}(\{a, b, c\})$ has exactly $2^3 = 8$ elements. Try writing down the 8 sets that comprise $\mathcal{P}(\{a, b, c\})$.

In the case where A is an infinite set we can ask for a self-contained method for describing the elements of A . For example, $\mathcal{P}(\mathbb{N})$ contains *as elements* the set $\{1\}$, the set $\mathbb{E} = \{2, 4, 6, \dots\}$ of even numbers, the set \mathbb{P} of prime numbers, and \mathbb{N} . But when we try to make a list of the elements of $\mathcal{P}(\mathbb{N})$ we are met with the following deep result of mathematics.

We cannot make an implied list of the elements in $\mathcal{P}(\mathbb{N})$.

Any list that we make of $\mathcal{P}(\mathbb{N})$ will miss some element of $\mathcal{P}(\mathbb{N})$ and any correction we attempt to make will still result in an incomplete list of the elements of $\mathcal{P}(\mathbb{N})$. We will have much more to say about this mathematical mystery in later chapters.

1.4 Something From Nothing

Here is an application of sets that appears infrequently in the popular press. Although humans have used the natural numbers since they could herd sheep, most of us are unaware of the mathematically precise construction of, say, 1. We will use sets and subsets to construct the first few natural numbers. The process with which we start is easily extended to a construction of \mathbb{N} from nothing. For this we must initiate a discussion of *cardinality* that will be improved upon throughout the chapters of this book.

We give a superficial and *mathematically imprecise* definition of counting that can be understood without a significant mathematical

background. In the next chapter we will present the mathematical background needed to make a precise accounting of the counting process. Let A be a finite set. The *cardinality* of A is defined to be

$\text{card}(A) = \text{all the sets } X \text{ that can be matched element by element with } A.$

When treating $\text{card}(A)$ we must be careful because $\text{card}(A)$ is *not a set*. It is proper to call it something else like a collection or a class. But $\text{card}(A)$ is just too big to be a set. We will avoid the issue of what $\text{card}(A)$ is by carefully avoiding set theoretic issues surrounding sets, collections, and classes.

We will use $\text{card}(A)$ as a precise way of counting the number of elements in A . That is why we need a precise understanding of how to “match element by element” the elements of X and A . On the one hand, we have a word *cardinality* that you should interpret as meaning *the number of elements in* a set. On the other hand, we have a notation $\text{card}(A)$ that represents all sets that have the same number of elements as A . Some examples will help you visualize what we mean by this.

For the set

$$\{\bullet\}$$

the set $\text{card}(\{\bullet\})$ consists of all sets that can be matched element by element with $\{\bullet\}$. This is a simple thing to see, mostly because the set $\{\bullet\}$ is so small. The sets $\{a\}$, $\{*\}$, $\{1\}$ are in $\text{card}(\{\bullet\})$ because they can be matched element by element with $\{\bullet\}$. The matching between $\{a\}$ and $\{\bullet\}$ is obvious enough. This matching is

$$a \longrightarrow \bullet.$$

Similarly, $\text{card}(\{a, b\})$ consists of the sets that can be matched element by element with $\{a, b\}$. The sets $\{x, y\}$, $\{0, 1\}$, $\{H, W\}$ are in $\text{card}(\{a, b\})$. One matching between $\{0, 1\}$ and $\{a, b\}$ is

$$\begin{aligned} 0 &\longrightarrow a \\ 1 &\longrightarrow b. \end{aligned}$$

There are others and you are invited to write them down, but we only need one. At this point I hope that the pattern is clear to you. The set $\{a, b, c\}$ is in $\text{card}(\{1, 2, 3\})$ because there is a matching

$$\begin{aligned} a &\longmapsto 3 \\ b &\longmapsto 2 \\ c &\longmapsto 1. \end{aligned}$$

Notice we didn't need to pick the one you were thinking of. Any matching of elements will do. We just chose a different one.

Let's continue the discussion by looking at the set without elements, \emptyset . We have

$\text{card}(\emptyset) = \text{all sets that have no elements.}$

Inasmuch as the notation $\text{card}(\emptyset)$ is a bit cumbersome, we will use the symbol 0 to denote $\text{card}(\emptyset)$.

$0 = \text{card}(\emptyset).$

This notation is perfect. The identification of $\text{card}(\emptyset)$ with the traditional symbol 0 makes perfect sense. We want to choose a symbol for $\text{card}(\emptyset)$ that will not confuse your mathematical sensibilities. The cardinality of \emptyset is denoted by 0 because we say so and for no other reason. But the use of 0 is compelling. It makes you think about the right ideas, the right quantities, and the right magnitudes. We could just as easily have decided that $\text{card}(\emptyset)$ should be denoted by Z . But since the world has been using 0 for a similar purpose for centuries we will continue the tradition. We have thus *constructed the natural number 0 from nothing*.

Next we construct the number 1. Take the power set $\mathcal{P}(\emptyset)$ of \emptyset . $\mathcal{P}(\emptyset)$ is the set of all subsets of \emptyset . Thus $A \in \mathcal{P}(\emptyset)$ exactly when $A \subset \emptyset$. But since \emptyset has no elements, $A \subset \emptyset$ implies that $A = \emptyset$. Thus the only element of $\mathcal{P}(\emptyset)$ is \emptyset , or equivalently

$\mathcal{P}(\emptyset) = \{\emptyset\}.$

Observe that $\mathcal{P}(\emptyset)$ is not the empty set since it has an element, namely, \emptyset . Initially, you may not like thinking of \emptyset as an element in a set, so keep trying. Since it has been used in a similar way for ages, we will use the symbol 1 to denote the cardinality of $\mathcal{P}(\emptyset)$.

$$1 = \text{card}(\{\emptyset\}).$$

This is in perfect agreement with the ages old use for 1, isn't it? Before you read this book, if I asked you to count the number of elements in the set $\{\emptyset\}$ you would most certainly say "One." So this notation agrees with your intuition, with your experiences, and with your senses. It is the idea of *defining* 1 in this way that may unsettle your inner moral mathematical compass. But since this use is not in conflict with the rest of society, then why not make the identification. Nothing spiritual will be harmed. What we identify and what we believe are the same thing.

Okay, let's define the natural number 2 in such a way that it does not conflict with our education of what 2 means. I will respect your difficulty in seeing \emptyset as an element in a set and we will for the moment look at $\{\bullet\}$. We will see together that

$$\mathcal{P}(\{\bullet\}) = \{\emptyset, \{\bullet\}\}.$$

Write down a subset $A \subset \{\bullet\}$. Since A is a set either $\bullet \in A$ or $\bullet \notin A$, there can be no other cases to consider. Because \bullet is the only element of $\{\bullet\}$ either $A = \{\bullet\}$ or A is empty, $A = \emptyset$. Thus, as we predicted, $\mathcal{P}(\{\bullet\}) = \{\emptyset, \{\bullet\}\}$.

Now force yourself to see \emptyset as the lone element of $\{\emptyset\}$. Convince yourself that $\{\emptyset\}$ is the set whose only element is \emptyset . Applying the above argument to $\{\emptyset\}$ instead of $\{\bullet\}$ yields the power set of $\{\emptyset\}$.

$$\mathcal{P}(\{\emptyset\}) = \{\emptyset, \{\emptyset\}\}.$$

We have constructed 2 once we have defined

$$2 = \text{card}(\{\emptyset, \{\emptyset\}\}).$$

This definition of 2 agrees with everything you learned as a child. There are just enough elements to make perfect sense of this use of the *symbol* 2. You can think of 2 as $\text{card}(\{\emptyset\})$ or you can think of it as a number you learned long ago. It doesn't matter. They represent the same idea, don't they? That is the beauty of this discussion. We are constructing natural numbers in an entirely precise mathematical format without losing the traditional meaning of the symbols we are using.

Next, construct 4. We will be careful not to upset our mathematical traditions concerning the number or symbol 4. Use bullets for \emptyset if you need to, but see the set $\{\emptyset, \{\emptyset\}\}$ as a set with two elements. We will examine the subset structure of $\{x, y\}$ and then restrict our attention to $\{\emptyset, \{\emptyset\}\}$. This is the last time we will do this, so get comfy.

Let A be a subset of $\{x, y\}$. If A has no elements then $A = \emptyset$; if A has exactly one element then $A = \{x\}$ or $A = \{y\}$. If A has other than one element then $A = \{x, y\}$. There are no other choices since all of the elements of $\{x, y\}$ are exhausted at this point. Thus

$$\mathcal{P}(\{x, y\}) = \{\emptyset, \{x\}, \{y\}, \{x, y\}\}.$$

If we replace x by \emptyset and y by $\{\emptyset\}$ then we have written the power set of $\{\emptyset, \{\emptyset\}\}$.

$$\mathcal{P}(\{\emptyset, \{\emptyset\}\}) = \{\emptyset, \{\emptyset\}, \{\{\emptyset\}\}, \{\emptyset, \{\emptyset\}\}\}.$$

We define the number 4 to be the cardinality of $\mathcal{P}(\{\emptyset, \{\emptyset\}\})$. Thus

$$4 = \text{card}(\{\emptyset, \{\emptyset\}, \{\{\emptyset\}\}, \{\emptyset, \{\emptyset\}\}\}),$$

which agrees with our previous uses of the symbol 4. *We have thus constructed the natural numbers 0, 1, 2, 4.*

Wait, there is another way to see this. We observe that

$$\begin{aligned}
 \emptyset &= \{\} \\
 \mathcal{P}(\emptyset) &= \{\emptyset\} \\
 \mathcal{P}(\mathcal{P}(\emptyset)) &= \mathcal{P}(\{\emptyset\}) = \{\emptyset, \{\emptyset\}\} \\
 \mathcal{P}(\mathcal{P}(\mathcal{P}(\emptyset))) &= \mathcal{P}(\mathcal{P}(\{\emptyset\})) = \mathcal{P}(\{\emptyset, \{\emptyset\}\}) \\
 &= \{\emptyset, \{\emptyset\}, \{\{\emptyset\}\}, \{\emptyset, \{\emptyset\}\}\}
 \end{aligned}$$

From this chart we can see that

$$\begin{aligned}
 0 &= \text{card}(\emptyset) \\
 1 &= \text{card}(\mathcal{P}(\emptyset)) \\
 2 &= \text{card}(\mathcal{P}(\mathcal{P}(\emptyset))) \\
 4 &= \text{card}(\mathcal{P}(\mathcal{P}(\mathcal{P}(\emptyset))))
 \end{aligned}$$

This last list shows us that the numbers 0, 1, 2, 4 can be realized as an iterated application of the power set operation $\mathcal{P}(\cdot)$ by starting with \emptyset . So in a real sense we have constructed the numbers 0, 1, 2, 4 from nothing. I leave it to you to construct the natural number 3. Try to do so in the spirit of the above construction. However, 3 will not be $\mathcal{P}(A)$ for any set A . Try to think of 3 as $4 - 1$.

1.5 Indexed Families of Sets

To this point we have only considered \cup , \cap , and $(\cdot)'$ for sets A and B . These were used to make us comfortable with the logical properties of the words *and*, *or*, and *not*. In this section we will work with \cup , \cap , and $(\cdot)'$ for *families* of sets. In order to make the discussion as accessible as possible we will work only with *finite* or *countable families* of sets $\{A_1, A_2, A_3, \dots\}$. These countable families will prove to be enough of a mind expander for this book.

The smallest countable families of sets are the finite families. Thus $\{\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}\}$ is a finite family of sets as is $\{\emptyset, \{\bullet\}\}$. You know what a finite family of sets is. When you list the students in the classrooms at the local elementary school, you are making finitely

many finite sets. You make one finite set for each classroom. There may be 6 such classes. The lists of enrolled students in a classroom form a family of 6 sets as in the next list.

$$\text{List}_1, \text{List}_2, \text{List}_3, \text{List}_4, \text{List}_5, \text{List}_6$$

Each list List_k is the list of names of students in classroom k . The symbol k will be a natural number between 1 and 6.

Our examples of countable families of sets begin with

$$\begin{aligned} & \{1\}, \{2\}, \{3\}, \dots \\ & \{1\}, \{1, 2\}, \{1, 2, 3\}, \dots \end{aligned}$$

Oftentimes the implied list of sets will be listed vertically, as in the following family of sets.

$$\begin{aligned} A_1 &= \{0\} \\ A_2 &= \{0, 1\} \\ A_3 &= \{0, 1, 2\} \\ A_4 &= \{0, 1, 2, 3\} \\ &\vdots \end{aligned}$$

This list is infinite and it consists of sets. That is why we call such a list an *infinite family of sets*.

Here is another. For each $m \in \mathbb{N}$ let \mathbb{N}_m be the set of natural numbers divisible by m . Then the list

$$\begin{aligned} \mathbb{N}_2 &= \{2, 4, 6, \dots\} \\ \mathbb{N}_3 &= \{3, 6, 9, \dots\} \\ &\vdots \\ \mathbb{N}_m &= \{m, 2m, 3m, \dots\} \\ &\vdots \end{aligned}$$

is an infinite family of sets.

For each natural number m let

\mathbb{Q}_m^+ be the positive rational numbers whose numerator (the top part) is m .

We can make an implied list of this set since the denominators of a positive fraction can be only $1, 2, 3, 4, \dots$

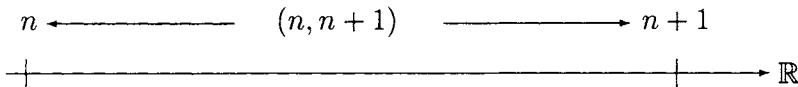
$$\mathbb{Q}_m^+ = \left\{ \frac{m}{1}, \frac{m}{2}, \frac{m}{3}, \frac{m}{4}, \dots \right\}.$$

Notice the denominator increases while the numerator remains m .

$$\begin{aligned}\mathbb{Q}_1^+ &= \left\{ \frac{1}{1}, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots \right\} \\ \mathbb{Q}_2^+ &= \left\{ \frac{2}{1}, \frac{2}{2}, \frac{2}{3}, \frac{2}{4}, \dots \right\} \\ \mathbb{Q}_3^+ &= \left\{ \frac{3}{1}, \frac{3}{2}, \frac{3}{3}, \frac{3}{4}, \dots \right\} \\ \mathbb{Q}_{24}^+ &= \left\{ \frac{24}{1}, \frac{24}{2}, \frac{24}{3}, \frac{24}{4}, \dots \right\} \\ &\vdots\end{aligned}$$

We will not have a set $\mathbb{Q}_{\sqrt{2}}^+$ since $\sqrt{2}$ is not a natural number.

Here is another example. Given an $n \in \mathbb{N}$ let $(n, n + 1) = \{x \in \mathbb{R} \mid n < x < n + 1\}$ be the set of real numbers strictly between n and $n + 1$. In picture form we would draw



Then

$$\{(n, n + 1) \mid n \in \mathbb{N}\} = \{(0, 1), (1, 2), (2, 3), \dots\}$$

is an infinite family of sets.

At this point we will define \cap and \cup for countable families of sets. The point behind doing this is that the notation corresponds to some important language. In particular, we will have the opportunity to use phrases like *for some*, or *at least*, or *for all*. The introduction of unions and intersections of infinite families of sets is meant to bring these phrases and their uses to your attention.

Definition 1.5.1 Let $\{A_0, A_1, A_2, \dots\}$ be an infinite family of sets.

1. $\bigcap\{A_n \mid n \in \mathbb{N}\} = \bigcap_{n \in \mathbb{N}} A_n = \{x \mid x \in A_n \text{ for all } n \in \mathbb{N}\}$. The phrase for all means that x is an element of every set in the family of sets.
2. $\bigcup\{A_n \mid n \in \mathbb{N}\} = \bigcup_{n \in \mathbb{N}} A_n = \{x \mid x \in A_n \text{ for some } n \in \mathbb{N}\}$. The phrase for some means at least one and at most all.

We certainly can use some examples here. Let

$$A_0, A_1, A_2, A_3, \dots$$

be an infinite family of sets. Then

$$\bigcup_{n \in \mathbb{N}} A_n = A_0 \cup A_1 \cup A_2 \cup \dots$$

Some examples will help you visualize what we are talking about.

Example 1.5.2 Let

$$\begin{aligned} A_1 &= \{0\} \\ A_2 &= \{0, 1\} \\ A_3 &= \{0, 1, 2\} \\ A_4 &= \{0, 1, 2, 3\} \\ &\vdots \end{aligned}$$

Then $1 \in A_2$, $2 \in A_3$, $3 \in A_4$, and continue. In general we see that

$$n \in A_{n+1} \text{ for all } n \in \mathbb{N}.$$

The phrase *for all* means that each and every set A_n in the list satisfies the predicate $n \in A_{n+1}$. Thus if we begin with a number k then

$$k \text{ is an element of } A_n \text{ for some } n \in \mathbb{N}.$$

We conclude that $k \in \bigcup_{n \in \mathbb{N}} A_n$. In fact since there was no extra condition on k we can say that *each* natural number k is in $\bigcup_{n \in \mathbb{N}} A_n$.

$$\mathbb{N} \subset \bigcup_{n \in \mathbb{N}} A_n.$$

Suppose we look at $\bigcap_{n \in \mathbb{N}} A_n$ now. Since $0 \in A_n$ for each $n \in \mathbb{N}$ we can see that

$$0 \in \bigcap_{n \in \mathbb{N}} A_n.$$

If $m \in \mathbb{N}$ is a natural number other than 0, then $m \notin A_m$ so that $m \notin A_n$ for some natural number n . In fact $m \notin A_n$ for many natural numbers n but we only need the one occurrence to conclude that

$$m \notin \bigcap_{n \in \mathbb{N}} A_n.$$

Therefore $\bigcap_{n \in \mathbb{N}} A_n = \{0\}$.

Example 1.5.3 Let us make another infinite family of sets. Examine the list

$$\begin{aligned} B_0 &= \{0, 1 \cdot 5\} \\ B_1 &= \{1 \cdot 5, 2 \cdot 5\} \\ B_2 &= \{2 \cdot 5, 3 \cdot 5\} \\ B_3 &= \{3 \cdot 5, 4 \cdot 5\} \\ &\vdots \end{aligned}$$

of sets of two natural numbers. Let

$$\mathbb{F} = \{5n \mid n \in \mathbb{N}\}$$

denote the set of *nonnegative multiples of 5*. We will prove that

$$\bigcup_{n \in \mathbb{N}} B_n = \mathbb{F}.$$

Proof: Let $x \in \bigcup_{n \in \mathbb{N}} B_n$. Then $x \in B_n$ for some $n \in \mathbb{N}$. By the definition of B_n , $x \in \{5n, 5(n+1)\}$ so that x is a multiple of 5. Hence $x \in \mathbb{F}$. As sets this means that

$$\bigcup_{n \in \mathbb{N}} B_n \subset \mathbb{F}.$$

On the other hand, the second inclusion begins with an element $x \in \mathbb{F}$. By the definition of \mathbb{F} there is an $n \in \mathbb{N}$ such that $x = n5$. Then $x \in \{5n, 5(n+1)\} = B_n$ for some $n \in \mathbb{N}$. We conclude that $x \in \bigcup_{n \in \mathbb{N}} B_n$ and so

$$\mathbb{F} \subset \bigcup_{n \in \mathbb{N}} B_n.$$

Therefore $\bigcup_{n \in \mathbb{N}} B_n = \mathbb{F}$ and the proof is complete.

Example 1.5.4 In this example we let $p \geq 2$ be a prime and we let \mathbb{N}_p denote the set of positive natural numbers divisible by p . Thus \mathbb{N}_2 is the set of positive even numbers $= \{2, 4, 6, \dots\}$, \mathbb{N}_3 is the set of positive numbers divisible by 3, and

$$\mathbb{N}_5 = \{5, 10, 15, 20, \dots\}.$$

We let $\mathbb{N}_1 = \{0, 1\}$ by default. This is just a notation we are using. It has nothing to do with the prime notation. Then we have an infinite family of sets

$$\begin{aligned}\mathbb{N}_1 &= \{0, 1\} \\ \mathbb{N}_2 &= \{2, 4, 6, \dots\} \\ \mathbb{N}_3 &= \{3, 6, 9, \dots\} \\ \mathbb{N}_5 &= \{5, 10, 15, \dots\} \\ &\vdots\end{aligned}$$

We have an infinite family

$$\mathbb{N}_1, \mathbb{N}_2, \mathbb{N}_3, \mathbb{N}_5, \dots$$

indexed by 1 and the set of *prime numbers* \mathbb{P} . Thus \mathbb{N}_7 is a member of the family but \mathbb{N}_9 is not included in the family. Since each natural number $n \geq 2$ is divisible by some prime number, I leave it to you to justify that

$$\begin{aligned}\bigcup_{n \in \mathbb{P}} \mathbb{N}_n &= \mathbb{N}_2 \cup \mathbb{N}_3 \cup \mathbb{N}_5 \cup \dots \\ &= \text{those natural numbers divisible by some prime} \\ &= \text{the set of natural numbers } n \geq 2.\end{aligned}$$

Remember 1 is not a prime number, no matter what you may believe right now. Thus 1 cannot be divisible by any prime. Moreover 1 is not in \mathbb{N}_n for any prime number n , so that $1 \notin \bigcup_{n \in \mathbb{P}} \mathbb{N}_n$.

Therefore

$$\begin{aligned}\mathbb{N}_1 \cup \bigcup_{n \in \mathbb{P}} \mathbb{N}_n &= \mathbb{N}_1 \cup \mathbb{N}_2 \cup \mathbb{N}_3 \cup \mathbb{N}_5 \cup \dots \\ &= \{0, 1, 2, 3, \dots\} \\ &= \mathbb{N}.\end{aligned}$$

Example 1.5.5 Using the notation in the previous example let us examine the set

$$X = \mathbb{N}_2 \cap \mathbb{N}_3 \cap \mathbb{N}_5 \cap \dots = \bigcap_{p \in \mathbb{P}} \mathbb{N}_p.$$

Suppose, for the purposes of a proof by contradiction, that we assume that there is an element $x \in X$. Then $x \in \mathbb{N}_p$ for each prime $p \in \mathbb{P}$ so that x is divisible by each prime $p \in \mathbb{P}$. What are the natural numbers divisible by all prime numbers? You are passingly familiar with the answer each time you factor a natural number. Natural numbers are divisible by only finitely many prime numbers. This contradiction (that x is divisible by all primes and that x can be divisible by only finitely many primes) shows us that our assumption is incorrect. Hence $X = \emptyset$. That is,

$$\mathbb{N}_2 \cap \mathbb{N}_3 \cap \mathbb{N}_5 \cap \dots = \emptyset$$

even though the intersection of any two of these sets is nonempty.

$pq \in \mathbb{N}_p \cap \mathbb{N}_q$ since pq is a multiple of p and a multiple of q

so that $\mathbb{N}_p \cap \mathbb{N}_q \neq \emptyset$.

Example 1.5.6 Let A_0, A_1, A_2, \dots be an infinite family of sets. The following negation of the definitions of \cap and \cup is a good exercise in using mathematical language properly.

$x \notin \bigcap_{n \in \mathbb{N}} A_n$ exactly when $x \notin A_n$ for some $n \in \mathbb{N}$.

Notice the change in phrasing. We went from *for all* to its logical negation *at least*. Thus the negation of *for all* is *at least* or *for some*.

The statement *all colors are impressive* is negated by stating *some colors are not impressive*. Try negating $x \in S$ for all sets

S before reading on. You would be correct if your negation reads $x \notin S$ for some set S .

The statement $n < n + 1$ for all $n \in \mathbb{N}$ is negated by writing $n \geq n + 1$ for some $n \in \mathbb{N}$. Said another way this negation reads

there is some natural number n such that $n \geq n + 1$.

This is obviously a false statement, which we should expect since we are negating the true statement $n < n + 1$ for all $n \in \mathbb{N}$.

The phrase *for some* means *at least one*, *at most all*. Its meaning will be reinforced with several examples.

The statement *some children have blue eyes* means that there is one child or perhaps several children who have blue eyes. A quick check of European children will verify that *many* children have blues eyes, so logically *some* children have blue eyes. The statement $x \in A_n$ for some $n \in \mathbb{N}$ means that there is one, and possibly more than that, natural number n such that $x \in A_n$. The statement $3x = 0$ for some $x \in \mathbb{R}$ means that there is one, and perhaps more, real number x such that $3x = 0$. Algebra shows us immediately that in this case there is exactly 1 number x such that $3x = 0$. But still it is proper to say that *there are some real numbers x such that $3x = 0$* .

If we have a predicate P and we write *P is satisfied by at least one natural number*, then we mean that there is a natural number n that satisfies P , and that there may be more natural numbers m that satisfy P . We do not rule out the possibility that every natural number n might satisfy X . For instance, let

$P =$ there are n musicians in a rock-and-roll band.

Then P is satisfied by at least the natural number 4. In fact, P is satisfied by 3, 5, and 7 as well. Rock-and-roll bands do not run into the thousands of musicians so there are some natural numbers that do not satisfy P . I leave it to you to name the bands that have that many people in them, keeping in mind that your author's narrow musical taste runs to sixties and seventies rock-and-roll.

Example 1.5.7 Let

$$\begin{aligned} A_0 &= \{0\} \\ A_1 &= \{0, 1\} \\ A_2 &= \{0, 1, 2\} \\ &\vdots \end{aligned}$$

be an infinite family of sets. Then $2 \notin A_1$, $3 \notin A_2$, and, in general, $n \notin A_{n-1}$. Hence

$$2, 3, \dots, n \notin \bigcap_{n \in \mathbb{N}} A_n$$

because n is not in some set A_m in this infinite family of sets. Since $0 \in A_n$ for all natural numbers n we see that

$$\{0\} = \bigcap_{n \in \mathbb{N}} A_n.$$

We say that $x \in \bigcup_{n \in \mathbb{N}} A_n$ if $x \in A_n$ for some $n \in \mathbb{N}$.

The negation of *for some* is *never*.

Thus the negation of the union looks like this.

$$x \notin \bigcup_{n \in \mathbb{N}} A_n$$

exactly when

$$x \notin A_n \text{ for any } n \in \mathbb{N}.$$

Thus

$$\{0, 1, 2, \dots\} = \bigcup_{n \in \mathbb{N}} A_n.$$

The following versions of DeMorgan's Laws are left as exercises for the reader. Let $\{A_n \mid n \in \mathbb{N}\}$ be an infinite family of sets contained in some universal set \mathcal{U} .

$$1. \left(\bigcap_{n \in \mathbb{N}} A_n \right)' = \bigcup_{n \in \mathbb{N}} A'_n.$$

$$2. \left(\bigcup_{n \in \mathbb{N}} A_n \right)' = \bigcap_{n \in \mathbb{N}} A'_n.$$

Chapter 2

Functions

Throughout this book we will want to compare sets. We do this by studying the *functions* that exist between them. A function from a set A into a set B is a rule that associates each element $x \in A$ with exactly one element $y \in B$. If we name the rule f then we will write

$$f(x) = y$$

to denote the fact that the element x is associated with y . We will also say that y is *the image of x under f* . Some of you are familiar with functions from that third year of high school math. In that case a function was usually a mapping or a general rule that acted on real numbers. For example,

$$f(x) = x^2$$

would be the function from \mathbb{R} into \mathbb{R} that associates each number x with its square x^2 . Others may have written this function as $x^2 = y$. In either case we are using some algebra to denote the image of x .

Our functions will only occasionally operate on real numbers. The reason for this is that our deliberations are on general sets and not exclusively real numbers. For example, we might define the function f that takes each person on Earth to his/her age in years, or we might define a function that associates each woman with exactly one man. You might even associate the elements of

$\{a, b, c\}$ with the elements from the set $\{x, y, z\}$ by defining

$$\begin{aligned} a &\longmapsto x \\ b &\longmapsto z \\ c &\longmapsto z. \end{aligned}$$

Notice that the function, which we will call f , misses y as an image and that two values b and c map to one value z . This type of abstract association is more closely related to the functions that we will encounter as our discussion evolves.

2.1 Functional Preliminaries

Suppose we are given sets $\{x, y, z\}$ and $\{0, 1, 2\}$. Some of the more important questions we will pursue in this book are the following.

1. What do we really mean by an *element by element matching between $\{x, y, z\}$ and $\{0, 1, 2\}$* ?
2. Can we make the notion of the *cardinality of $\{x, y, z\}$* a mathematically precise idea?
3. It seems clear that $\{a, b\}$ has fewer elements than $\{x, y, z\}$. Can we make this precise?

We will approach these questions by considering a general comparison of sets called a *function* that requires us to investigate the idea of a function on *abstract* sets.

The following takes our definition of function and gives it some mathematical style. It may read as a dry definition to you but it has the advantage of being mathematically precise. Perhaps you could try to define functions in such a way that it encompasses the examples we give in this section. There is something to be learned by trying to do that, so please do.

Definition 2.1.1 Let A and B be sets. A function from A to B is a rule f with the following two properties.

1. f associates each $x \in A$ with an element $f(x) \in B$.

2. There is exactly one value for $f(x)$.

A bit of very useful notation is the following. We will write

$$f : A \longrightarrow B$$

if f is a function from A to B . We call A *the domain of f* , and we call B *the codomain of f* . We call $f(x)$ *the image of x under f* just as we did in our algebra class in high school.

Another *description* of a function that you may find a bit easier to absorb is the following. A function f from A to B assigns to each element $x \in A$ an image $f(x) \in B$. There must be no ambiguity in the value of the image $f(x)$. This definition gives you the mental picture of a function that you had in high school, but it contains so much more. You may have read that

$$f(x) = y$$

and that is not wrong. We often let y denote the image of x . It is common to do so especially when graphing the function f as we will do presently. You may also have read that

$$f(x) = 2x + 1,$$

which is an example of a function on real numbers. This is not the only type of function, which is why we had to define functions in such a broad abstract way. There is a function $f : \{0, 1, 2\} \longrightarrow \{x, y, z\}$ defined by the rule

$$\begin{aligned} 0 &\longmapsto z \\ 1 &\longmapsto x \\ 2 &\longmapsto z. \end{aligned}$$

Under this function, y is not associated with any element in $\{0, 1, 2\}$.

Our next example of a function $f : \mathbb{R} \longrightarrow \mathbb{R}$ is defined by setting

$$f(x) = 7x + 6 \quad \text{for each } x \in \mathbb{R}.$$

We should all be familiar with the fact that the graph of this function is a line that is inclined upward when we read it from left to right. Notice that to each $x \in \mathbb{R}$ there is exactly one real number $y = 7x + 6$. Observe that the images of 0, 1, and -1 can be calculated using the formula $f(x) = 7x + 6$.

$$\begin{aligned}f(0) &= 7 \cdot 0 + 6 = 6 \\f(1) &= 7 \cdot 1 + 6 = 13 \\f(-1) &= 7 \cdot (-1) + 6 = -1.\end{aligned}$$

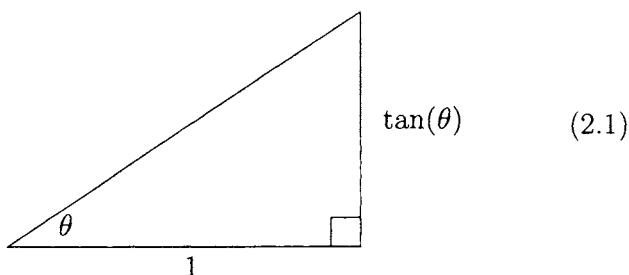
A familiar function $f : \mathbb{R} \longrightarrow \mathbb{R}$ is defined by $f(x) = x^2$. The value x^2 is a real number that is unique to x so f is indeed a function. We can find the images of 0, 1, and -1 easily enough.

$$\begin{aligned}f(0) &= 0^2 = 0 \\f(1) &= 1^2 = 1 \\f(-1) &= (-1)^2 = 1.\end{aligned}$$

If you had any trigonometry in your education then you may know that

$$f(\theta) = \tan(\theta)$$

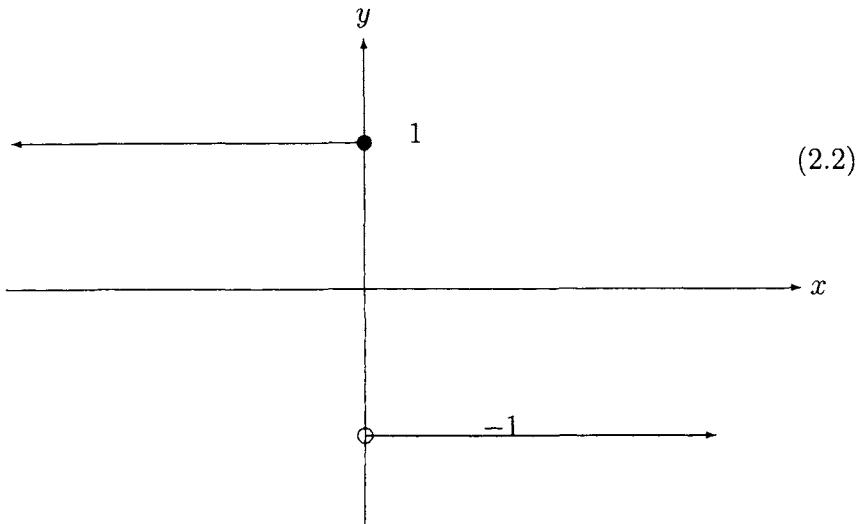
defines a function that associates with each angle θ its tangent, $\tan(\theta)$. We can construct the value $\tan(\theta)$ as follows. Draw a triangle whose base length is 1 and whose base angles are θ and a right angle, as in the diagram below. Then the side opposite θ has length $\tan(\theta)$. If you possessed a ruler that was mathematically accurate to all decimal places, you could read $\tan(\theta)$ from picture (2.1).



The functions $f(x) = 7x + 6$, $f(x) = x^2$, and $f(x) = \tan(x)$ will take real numbers x and associate them with or *send them to* a real

number $f(x)$. Most of the functions that you have experienced in your lifetime are like this. Now let's see something more abstract.

Consider the graph (2.2).



It has a naturally occurring feature that you may not have seen before. It has a gap and a hole in it. These are common features of graphs, even though they have been kept from you, and they occur more often than the smooth connected curves with which you are familiar. We will present some of these functions with broken graphs.

Example 2.1.2 Define a function $f : \mathbb{R} \rightarrow \{1, -1\}$ by setting

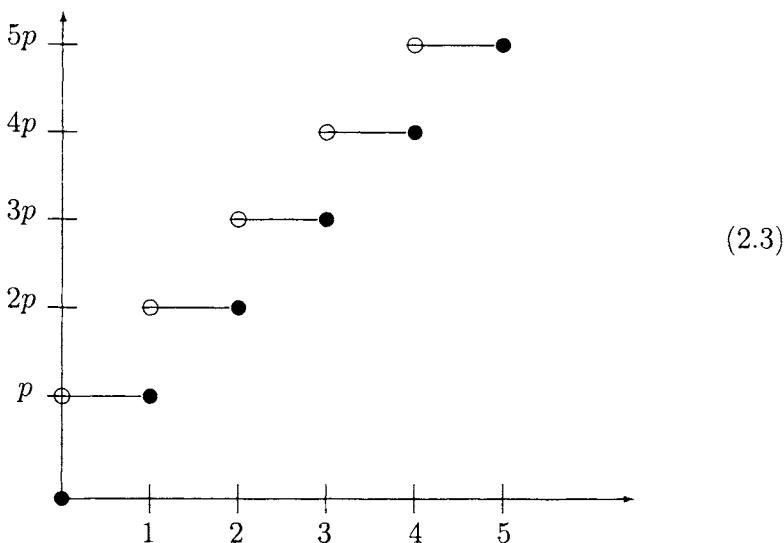
$$f(x) = \begin{cases} 1 & \text{if } x \leq 0 \\ -1 & \text{if } x > 0 \end{cases}.$$

We observe that given $x \in \mathbb{R}$ then $f(x) = 1$ or $f(x) = -1$ but not both. Thus f satisfies our definition of function. Its graph is the one given in (2.2).

Example 2.1.3 A function whose graph has many more steps is the *postage rate function*.

$$\begin{aligned} f(x) = & \text{ cost in pennies to send a letter that} \\ & \text{weighs } x \text{ ounces by snail mail.} \end{aligned}$$

Its graph is the graph (2.3) below.



If a letter weighs up to and including 1 ounce, the cost of postage is p . (At the time of this writing $p = 39$ cents.) If it weighs more than 1 ounce but at most (and including) 2 ounces, then the cost to mail the letter is $2p$. This continues on until you get tired of it. The graph of this function must reflect that jump in price at 1 ounce and at 2 ounces and at 3 ounces, and so on.

The fact that we have a hole \circ on a graph means that the price is not evaluated at that point. The line corresponding to p has a hole at the left end. This is the cost of mailing a letter weighing 0 ounces. There is no letter to mail in this case so the cost is 0 cents. The line corresponding to p has a solid dot \bullet at its right end, and the line corresponding to $2p$ has a hole at the left end. That means that the cost for a 1 ounce letter is to be evaluated from the lower line. Thus we pay p cents when mailing a 1 ounce letter.

The graph tells us that it will cost $2p$ cents to mail a letter that weighs more than 1 ounce and up to 2 ounces. The black dot \bullet on that line means that if the letter weighs more than 1 ounce then you must pay $2p$ to mail it. Similarly if the letter weighs more than 2 ounces or up to 3 ounces then the letter costs $3p$ to mail. You can guess what happens when the letter weighs more than 3 ounces but up to 4 ounces.

Example 2.1.4 Define a function $f : \mathbb{R} \longrightarrow \{0, 1\}$ by setting

$$f(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q} \\ 0 & \text{if } x \notin \mathbb{Q} \end{cases} \quad \text{That is, } x \text{ is a fraction}$$

This one cannot be graphed, but it is still a function. If we were to try to graph this function we would have to graph a horizontal line that is full of holes and full of dark spots. For instance,

$$f(0) = 1, f\left(\frac{1}{2}\right) = 1, \text{ and } f\left(\frac{4}{7}\right) = 1,$$

while

$$f\left(\frac{1}{\sqrt{2}}\right) = 0, f\left(\frac{1}{\pi}\right) = 0, \text{ and } f\left(\frac{2}{\sqrt[3]{3}}\right) = 0.$$

Furthermore, between every two real numbers x and y there is a rational number q

$$x < q < y,$$

so between every two numbers x and y at which $f(x) = f(y) = 0$ there is a number q such that $f(q) = 1$.

$$f(x) = 0, f(q) = 1, f(y) = 0.$$

Current technology does not permit us to graph such a line. However, since a given real number x is either a rational number or an irrational number but not both, the rule f defines a function $f : \mathbb{R} \longrightarrow \{0, 1\}$.

Next let us consider functions that are not so algebraic.

Example 2.1.5 Let \mathbf{P} be the set of people on the Earth, and define a function $a : \mathbf{P} \longrightarrow \mathbb{N}$ as the function that assigns to each person on Earth his/her age in years as of January 1, 2005. Thus $a(\text{Wendy}) = 15$, $a(\text{new born}) = 0$, and $a(\text{author}) > 25$. You cannot at this time determine $a(\text{author})$, but there is a natural number that the author calls his age. However, at the time of this writing there is no person x such that $a(x) = 200$.

Example 2.1.6 Let \mathbf{L} be the set of locations on the Earth, and define a function $t : \mathbf{L} \rightarrow \mathbb{N}$ to be the function that assigns to each location on Earth the temperature in degrees Celsius at that location at exactly 12 noon on January 1, 2005. So you might find that $t(\text{South Pole}) = -80^\circ\text{C}$, $t(\text{Guam}) = 30^\circ\text{C}$, and $t(\text{outside}) = 20^\circ\text{C}$. You cannot at this time determine $t(300 \text{ miles under New York City})$, but there is a natural number that we would all agree is that temperature. There is no place x on Earth such that $t(x) = \text{absolute zero}$.

A rule f must satisfy both of the conditions in our definition of function if it is to be a function. The following rules are not functions.

(Failure of Condition 1.) Let f be the association from \mathbb{R} to \mathbb{R} defined by $f(x) = \sqrt{x}$. Then f is not a function since $f(-1) = \sqrt{-1} \notin \mathbb{R}$.

(Failure of Condition 2.) Let \mathbb{R}^+ be the set of positive real numbers, and let f be the association from \mathbb{R}^+ to \mathbb{R} defined by $f(x) = \pm\sqrt{x}$. Then f is not a function since $f(1) = 1, -1$ is more than one value.

We think we know what an implied list is, but here is a mathematically precise way to think of it. Let A be a set. Any set will do. We define an implied list of elements in A to be a function

$$f : \mathbb{N} \longrightarrow A.$$

In this case we will write

$$f(0) = a_0, f(1) = a_1, f(2) = a_2, \dots$$

and in general

$$f(n) = a_n \text{ for each element } n \in \mathbb{N}.$$

For instance, we can define $f : \mathbb{N} \longrightarrow \mathbb{R}$ as

$$f(n) = n.$$

Then f is the implied list

$$0, 1, 2, 3, \dots$$

Another function $f : \mathbb{N} \rightarrow \mathbb{R}$ defined by

$$f(n) = \frac{1}{n+1} \text{ for all } n \in \mathbb{N}$$

is the implied list

$$\frac{1}{1}, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots$$

The following is a pretty little rule that is not a function unless we are very careful.

In this age of computers it should be no surprise that we can find for every real number $0 < x < 1$ a *binary series*

$$x = \frac{b_1}{2} + \frac{b_2}{2^2} + \frac{b_3}{2^3} + \dots,$$

where the digits $b_1, b_2, b_3 \dots$ take values in the set $\{0, 1\}$. We would then write x as a *binary expansion*.

$$x = .b_1 b_2 b_3 \dots \quad (2.4)$$

For example,

$$\frac{1}{2} = \frac{1}{2} + \frac{0}{2^2} + \frac{0}{2^3} + \dots$$

and

$$\frac{1}{3} = \frac{0}{2} + \frac{1}{2^2} + \frac{0}{2^3} + \frac{1}{2^4} + \dots,$$

where the coefficients for $\frac{1}{3}$ are 1 for the even exponents of $\frac{1}{2}$ and 0 elsewhere. Thus

$$\frac{1}{2} = .100\bar{0}$$

$$\frac{3}{4} = .1100\bar{0}$$

$$\frac{1}{3} = .0101\bar{0}\bar{1}.$$

In the general spirit of mathematics we denote the set of all binary sequences by \mathbf{S} .

$$\mathbf{S} = \text{the set of all binary sequences } b_1 b_2 b_3 \dots$$

That is, \mathbf{S} consists of all sequences of 0's and 1's. For instance, \mathbf{S} contains elements like $10\bar{0}\dots$ and $01\bar{0}\bar{1}\dots$, but it does not contain $.33\bar{3}\dots$

We define the *unit interval* $(0, 1)$ to be the real numbers that lie properly between 0 and 1 on the real number line.

$$(0, 1) = \{x \in \mathbb{R} \mid 0 < x < 1\}.$$

Let us define a rather obvious function

$$f : (0, 1) \longrightarrow \mathbf{S}$$

as follows. Given an $x \in (0, 1)$, write x as a binary expansion as in (2.4),

$$x = .b_1 b_2 b_3 \dots,$$

and then define

$$f(x) = b_1 b_2 b_3 \dots.$$

Thus $f(x)$ just drops the decimal point of the binary expansion for x . Evidently given $x \in (0, 1)$, $f(x)$ is a binary sequence, so that $f(x) \in \mathbf{S}$ for each $x \in (0, 1)$.

However, the value $f(x)$ has some ambiguity surrounding it. For example, let's show that $\frac{1}{2}$ has two binary expansions.

$$\begin{aligned} \frac{1}{2} &= \frac{1}{2} + \frac{0}{2^2} + \frac{0}{2^3} + \dots \\ &= \frac{0}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \dots \end{aligned}$$

This will require some trickery and guile that we will use several times in the future. Given

$$\frac{0}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \dots,$$

let's multiply by $\frac{1}{2} = 1 - \frac{1}{2}$ to realize the following equations. Use the distributive law or your friend foil to do the multiplication.

$$\begin{aligned}
 & \left(\frac{1}{2} \right) \left(\frac{1}{2^2} + \frac{1}{2^3} + \frac{1}{2^4} + \dots \right) \\
 &= \left(1 - \frac{1}{2} \right) \left(\frac{1}{2^2} + \frac{1}{2^3} + \frac{1}{2^4} + \dots \right) \\
 &= \left\{ \begin{array}{cccccc} \frac{1}{2^2} & + & \frac{1}{2^3} & + & \frac{1}{2^4} & + \dots \\ -\frac{1}{2} \left(\frac{1}{2^2} & + & \frac{1}{2^3} & + & \frac{1}{2^4} & + \dots \right) \end{array} \right\} \\
 &= \left\{ \begin{array}{cccccc} \frac{1}{2^2} & + & \frac{1}{2^3} & + & \frac{1}{2^4} & + \frac{1}{2^5} + \dots \\ -\frac{1}{2^3} & - & \frac{1}{2^4} & - & \frac{1}{2^5} & - \dots \end{array} \right. \\
 &= \frac{1}{2^2}.
 \end{aligned}$$

Thus

$$\left(\frac{1}{2} \right) \left(1\frac{1}{2^2} + 1\frac{1}{2^3} + 1\frac{1}{2^4} + \dots \right) = \frac{1}{2^2},$$

so that multiplication by 2 yields the equation

$$1\frac{1}{2^2} + 1\frac{1}{2^3} + 1\frac{1}{2^4} + \dots = \frac{1}{2}.$$

So, reader, which one is it? According to our definition of f , there are two choices for $f(x)$. Which choice do you make?

$$f \left(\frac{1}{2} \right) = \left\{ \begin{array}{l} 100\bar{0} \text{ or} \\ 011\bar{1} \end{array} \right..$$

The point here is that there should be no choice. The value $f(x)$ should be unique, unambiguous, singular to x . If two values present themselves, they should be equal. They might look different but they should be the same thing. This is not the case here. The value

for $f\left(\frac{1}{2}\right)$ has on the one hand only 1 entry of 1, while on the other hand it has infinitely many entries 1. As it is defined, the rule f does not satisfy Condition 2 of the definition of function, so it is not a function.

The only way to correct this deficiency is to take advantage of the number of 1's in these binary sequences. This, it turns out, is all we need to change the rule f into a function. Define a rule

$$F : (0, 1) \longrightarrow S$$

related to f by requiring that

$$\begin{aligned} F(x) &= b_1 b_2 b_3 \dots, \text{ where we choose the binary} \\ &\quad \text{expansion for } x = .b_1 b_2 b_3 \dots \text{ with the} \\ &\quad \text{smallest number of 1's.} \end{aligned} \tag{2.5}$$

Under this new rule F , the sequence $01\bar{1}\dots$ is ruled out as the image of $\frac{1}{2}$ since there is a “shorter” binary sequence available. Hence

$$F\left(\frac{1}{2}\right) = 100\bar{0}\dots$$

This definition for the new rule F avoids the ambiguity that the rule f possessed.

Let us examine an operation on functions. Functions in abstract sets cannot be added or multiplied, so we will not be dealing with the addition or multiplication of functions as we might have done in algebra. There is, however, an important operation on functions called *composition*.

Definition 2.1.7 Let $f : A \rightarrow B$ and $g : B \rightarrow C$ be functions. (Note that B serves as both the domain of g and the codomain of f .) There is a function

$$g \circ f : A \longrightarrow C$$

whose rule is

$$g \circ f(x) = g(f(x)) \text{ for each } x \in A.$$

For example, let $f : \mathbb{R} \rightarrow \mathbb{R}$ and $g : \mathbb{R} \rightarrow \mathbb{R}$ be defined by $f(x) = x + 1$ and $g(x^2)$. Then

$$g \circ f(x) = g(f(x)) = g(x+1) = (x+1)^2$$

and

$$f \circ f(x) = f(f(x)) = f(x+1) = (x+1)+1 = x+2.$$

Example 2.1.8 Let $f(x) = \frac{1}{x}$. Then

$$f \circ f(x) = f(f(x)) = f\left(\frac{1}{x}\right) = \frac{1}{1/x} = x.$$

That is, $f \circ f(x) = x$.

Example 2.1.9 Let $f(x) = \sqrt{x}$ and $g(x) = x^2$. Then

$$f \circ g(x) = f(g(x)) = f(x^2) = \sqrt{x^2} = \begin{cases} x & \text{if } x \geq 0 \\ -x & \text{if } x < 0 \end{cases}.$$

Experiment with this.

$$f \circ g(-1) = f(g(-1)) = f((-1)^2) = f(1) = \sqrt{1} = 1 = -(-1),$$

while

$$f \circ g(1) = f(g(1)) = f(1^2) = f(1) = \sqrt{1} = 1.$$

You can check this formula for $x = -2, 2$ to see that $f \circ g(x) \neq x$. How do you reconcile this fact with what you know today?

Example 2.1.10 Let $F : (0, 1) \rightarrow \mathbf{S}$ be the function defined in (2.5). Then $F(x)$ is the binary sequence $b_1 b_2 b_3 \dots$ that defines x . Define a function

$$G : \mathbf{S} \longrightarrow \{0, 1\}$$

by requiring that

$$G(b_1 b_2 b_3 \dots) = b_1 = \text{the first entry in the sequence } b_1 b_2 b_3 \dots$$

The composition $G \circ F$ has the rule

$$G \circ F(x) = G(F(x)) = G(b_1 b_2 b_3 \dots) = b_1 \in \{0, 1\}.$$

Thus $G \circ F$ is a function that assigns to each $x \in (0, 1)$ either a 0 or a 1, and it makes the assignment without ambiguity. We challenge the reader to find the set of numbers $x \in (0, 1)$ such that $G \circ F(x) = 1$. Hint: Look at the definition of F . The answer to this challenge is given at the end of this section.

We will need a notation for the set of all functions from A into B .

Definition 2.1.11 Let A and B be sets. We let

$$B^A = \text{the set of all functions } f : A \longrightarrow B.$$

Let $A = \{a, b\}$ and let $B = \{0, 1\}$. We will write down all of the functions in B^A .

$$\begin{array}{ll} f(a) = 0, f(b) = 0 & g(a) = 0, g(b) = 1 \\ h(a) = 1, h(b) = 0 & k(a) = 1, k(b) = 1. \end{array}$$

Since we have completely exhausted the possible images for a and b we have written a complete list of the elements of B^A . Another way of writing these functions presents itself.

Let us agree that each function f will be written as

$$f = [x, y] \text{ provided that } f(a) = x \text{ and } f(b) = y.$$

For instance,

$$f = [1, 0] \text{ means that } f(a) = 1 \text{ and } f(b) = 0.$$

Thus we can write down all of the functions in $\{0, 1\}^{\{a, b\}}$ by systematically writing down the possible pairs $[x, y]$. They are

$$\begin{array}{ll} [0, 0] & [0, 1] \\ [1, 0] & [1, 1]. \end{array}$$

We can then easily count the number of functions in $\{0, 1\}^{\{a, b\}}$. Recall that the *cardinality* of X is denoted by $\text{card}(X)$. Then

$$\text{card}(\{0, 1\}^{\{a, b\}}) = 4 = 2^2 = \text{card}(\{0, 1\})^{\text{card}(\{a, b\})},$$

which explains why we chose this notation.

Other examples are found by letting $A = \{a, b, c\}$ and $B = \{0, 1\}$. To count the number of functions in $\{0, 1\}^{\{a,b,c\}}$, let us agree to write the function

$$f : \{a, b, c\} \longrightarrow \{0, 1\}$$

in terms of its values at a, b , and c . Then f is denoted by

$$f = [x, y, z], \quad \text{where } f(a) = x, f(b) = y, \text{ and } f(c) = z.$$

For example, the function f such that $f(a) = 1, f(b) = 1, f(c) = 0$ is denoted by

$$f = [1, 1, 0].$$

The left entry in $[1, 1, 0]$ is 1 because $f(a) = 1$. The middle entry is 1 because $f(b) = 1$, and the right hand entry is 0 because $f(c) = 0$. Every function $f : \{a, b, c\} \longrightarrow \{0, 1\}$ can be represented in this way. This provides us with a very simple way to count the elements in $\{0, 1\}^{\{a,b,c\}}$. There are two possible values for each of the three entries of $f = [x, y, z]$ so that there are $2 \cdot 2 \cdot 2 = 8$ possible entries for f and therefore there are 8 possible functions in $\{0, 1\}^{\{a,b,c\}}$. That is,

$$\text{card}(B^A) = \text{card}(\{0, 1\}^{\{a,b,c\}}) = 8 = \text{card}(\{0, 1\})^{\text{card}(\{a,b,c\})}.$$

We leave it to the reader to write down the eight functions in B^A .

As a further exercise, the functions in $\{0, 1, 2\}^{\{a,b\}}$ can be written as

$$[x, y] \text{ where } f(a) = x \text{ and } f(b) = y,$$

where x, y are numbers in $\{0, 1, 2\}$. Since there are three numbers that can fill the two entries $[x, y]$, there are exactly $3 \cdot 3 = 9$ functions in $\{0, 1, 2\}^{\{a,b\}}$.

$$\text{card}(A^B) = \text{card}(\{0, 1, 2\}^{\{a,b\}}) = 3^2 = \text{card}(\{0, 1, 2\})^{\text{card}(\{a,b\})}.$$

Four of these functions are

$$[0, 0], [1, 0], [0, 2], [1, 2].$$

The first function takes each of a and b to 0. $[1, 2]$ is a function that takes a to 1 and that takes b to 2. The reader is invited to list all 9 functions and their rules using the $[x, y]$ notation that we described above.

This is the answer to the $G \circ F$ challenge. You can use the sum (2.4) to convince yourself that $G \circ F(x) = 1$ exactly when $\frac{1}{2} \leq x < 1$.

2.2 Images and Preimages

Consider the function $f : \mathbb{R} \longrightarrow \mathbb{R}$ defined by

$$f(x) = x^2.$$

A moment's thought reveals that

$$f(1) = f(-1) = 1.$$

Let's write that in a more useful notation.

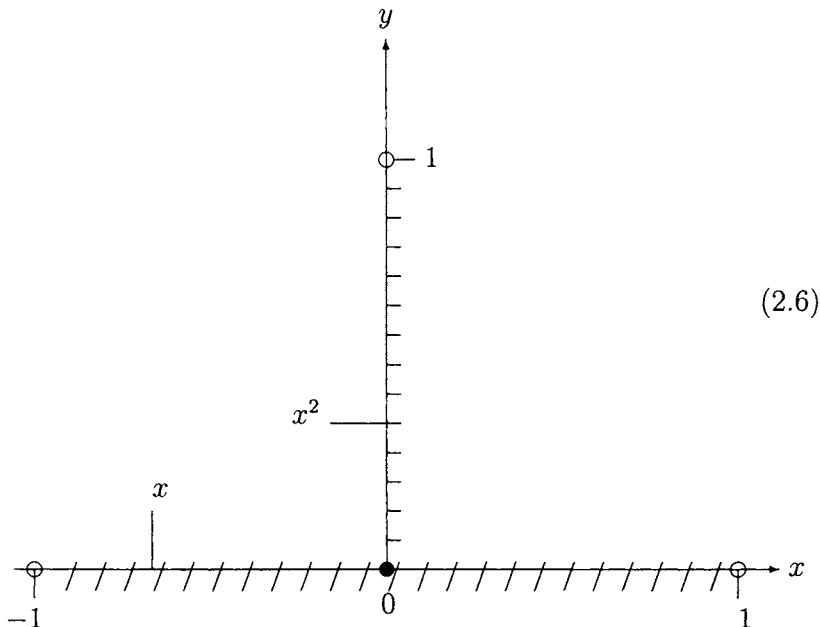
$$f(\{1, -1\}) = \{1\}.$$

Furthermore, if we recall that $(-1, 1) = \{x \in \mathbb{R} \mid -1 < x < 1\}$ then

$$f((-1, 1)) = \{f(x) \mid -1 < x < 1\} = \{y \mid 0 \leq y < 1\}.$$

The following picture may help you see this. The interval $(-1, 1)$ defined by the slashes “/” is mapped into the interval $[0, 1)$ defined by the dashes “—”. The open circles \circ mean that $x = -1, x = 1$,

and $y = 1$ are not included.



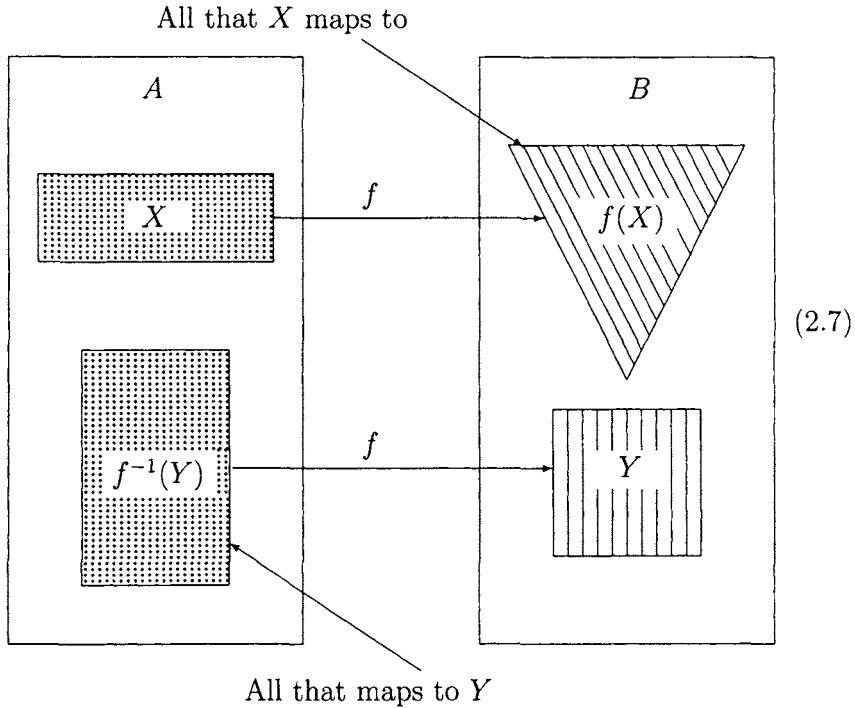
We will say that $[0, 1]$ is *the image of $(-1, 1)$* under the function $f(x) = x^2$. Since $(-1, 1)$ is everything that $f(x)$ sends into $[0, 1]$ we say that $(-1, 1)$ is *the preimage of $[0, 1]$* . Another example would be that $\{1\}$ is the image of $\{-1\}$ and since $\{-1, 1\}$ is all that f maps to $\{1\}$, $\{-1, 1\}$ is the preimage of $\{1\}$. In more general terms we have the following definition.

Definition 2.2.1 Let $f : A \rightarrow B$ be a function and let $X \subset A$ and $Y \subset B$ be subsets.

1. $f(X) = \{f(x) \mid x \in X\}$. $f(X)$ is called the image of X .
2. $f^{-1}(Y) = \{x \in X \mid f(x) \in Y\}$. $f^{-1}(Y)$ is called the preimage of Y .

If $x \in A$ and $y \in B$ then we will write $f(x)$ for $f(\{x\})$ and $f^{-1}(y)$ for $f^{-1}(\{y\})$. Thus $f(X)$ is the set of things that X is mapped to and $f^{-1}(Y)$ is the set of things that map into Y . Picture (2.7) is there to help you visualize images and preimages. Suppose we

think of A and B as boxes and suppose f is an arrow between them. Inside A is a box we call X and inside B is a box we call Y .



The arrows in picture (2.7) indicate that the box X maps to the triangle $f(X)$ and that the square Y comes from the rectangle $f^{-1}(Y)$. In fact, *everything* that maps into Y is in the rectangle $f^{-1}(Y)$.

Here are more examples to illustrate these ideas.

Example 2.2.2 Let $a : \text{People of Earth} \rightarrow \mathbb{N}$ be defined as

$$a(x) = \text{age in years of the person } x \text{ on January 1, 2005.}$$

Then $a(\text{your author}) = 50$,

$$\begin{aligned} a^{-1}(15) &= \{\text{people } x \mid a(x) = 15\} \\ &= \text{set of 15 year olds.} \end{aligned}$$

Suppose that x is some individual on Earth whose age is 50. Then $a(x) = 50$ and hence

$$\begin{aligned} a^{-1}(a(x)) &= a^{-1}(50) \\ &= \text{the set of all people having age 50}. \end{aligned}$$

Notice that $x \in a^{-1}(a(x))$ but that x is hardly everyone who is 50. We leave it to the reader to justify that $a^{-1}(200) = \emptyset$.

Example 2.2.3 In this example, let $t : \text{places on Earth} \rightarrow \mathbb{R}$ be the function that takes a place on Earth x and returns the temperature $t(x)$ in degrees Celsius of that place at 12 noon on January 1, 2005. Then $t(\text{outside}) = 20^\circ \text{ C}$,

$$\begin{aligned} t^{-1}(30^\circ \text{C}) &= \{\text{places } x \mid t(x) = 30^\circ \text{C}\} \\ &= \{\text{places at which the temperature is } 30^\circ \text{C}\} \end{aligned}$$

and

$$\begin{aligned} t^{-1}(100^\circ \text{C}) &= \{\text{places } x \mid t(x) = 100^\circ \text{C}\} \\ &= \{\text{places at which the temperature is the} \\ &\quad \text{boiling point of water at sea level}\}. \end{aligned}$$

Notice that $t^{-1}(100^\circ \text{C})$ contains the places where people at sea level are boiling water. It also contains those places, like volcanoes and geysers, at which the temperature is that of boiling water.

Example 2.2.4 An example from trigonometry is given by the function $f : \mathbb{R} \rightarrow \mathbb{R}$ defined by $f(\theta) = \tan(\theta)$, where \tan is the *tangent function*. We have already defined the tangent function for angles θ . (See page 40.) Then $f(0) = \tan(0) = 0$, while

$$\begin{aligned} f^{-1}(0) &= \text{the set of all angles whose tangent is } 0 \\ &= \{\dots, -2\pi, -\pi, 0, \pi, 2\pi, \dots\} \\ &= \{n\pi \mid n \in \mathbb{Z}\}. \end{aligned}$$

Here are some results on images and preimages that we are using as educational tools to prepare the reader for the more advanced arguments presented in the next chapter.

Theorem 2.2.5 *Let $f : A \rightarrow B$ be a function. If $Y \subset Y' \subset B$ then $f^{-1}(Y) \subset f^{-1}(Y')$.*

Proof: We must show that every element of $f^{-1}(Y)$ is in $f^{-1}(Y')$. Let $x \in f^{-1}(Y)$. By definition of preimage $f(x) \in Y$, and since $Y \subset Y'$, $f(x) \in Y'$. Then $x \in f^{-1}(Y')$ by definition of preimage. Hence $f^{-1}(Y) \subset f^{-1}(Y')$, which completes the proof.

Look at what we did in the above proof. We started with an element $x \in f^{-1}(Y)$ and we evaluated what that means to us. The definition of the preimage must be used. It tells us that x maps into Y . That is, $f(x) \in Y$. The subset hypothesis is now used. $f(x) \in Y \subset Y'$ implies that $f(x) \in Y'$. Since x maps into Y' then x is in the preimage, the set of all that maps into Y' : $x \in f^{-1}(Y')$. Hence $f^{-1}(Y) \subset f^{-1}(Y')$.

Notice the detail with which I am arguing. This detail is necessary for two reasons. First, we cannot know what the other is thinking so I must show all of the thoughts that I want you to think. Second, before we can skip steps without fear of error setting in, we must first pay our dues with this type of detail. I hope you are paying dues. The later material will more than make up for the present effort I am asking you to exert.

Theorem 2.2.6 *Let $f : A \rightarrow B$ be a function and let $X \subset A$. Then $X \subset f^{-1}(f(X))$.*

Picture (2.8) is an illustration of what we will show in the proof of this theorem.

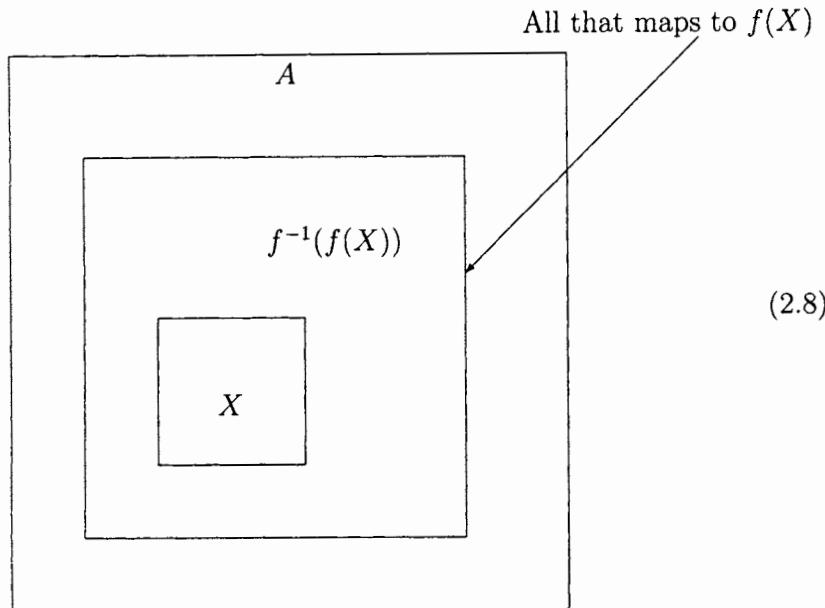
Proof: We must show that each element of X is an element of $f^{-1}(f(X))$. Let $x \in X$. By the definition of image $f(x) \in f(X)$. For the sake of clarity let $W = f(X)$. Then $f(x) \in W$. The definition of preimage and the definition of W show us that

$$x \in f^{-1}(W) = f^{-1}(f(X)).$$

Thus $X \subset f^{-1}(f(X))$, which completes the proof.

Once again we will detail and motivate the given proof. To prove that $X \subset f^{-1}(f(X))$ we must show that each given $x \in X$

is in $f^{-1}(f(X))$. Let $x \in X$. It must be clear to you by now that $f(x) \in f(X)$. Thus x is a part of all that maps into $f(X)$. Put symbolically, $x \in f^{-1}(f(X))$. We conclude that $X \subset f^{-1}(f(X))$.



Example 2.2.7 Here are two examples of how $f^{-1}(f(X))$ can be much larger than X .

1. Consider the function $f : \mathbb{R} \rightarrow \mathbb{R}$ defined by $f(x) = x^2$, and let $X = \{2\}$. A small calculation reveals that $f(2) = 4$. Thus $2 \in f^{-1}(4)$. The set $f^{-1}(4)$ is the set of *all* x such that $x^2 = 4$. A moment's thought will show you that $f(2) = f(-2) = 4$ and that $\{-2, 2\} = f^{-1}(4)$. Thus $X = \{2\} \neq \{-2, 2\} = f^{-1}(f(X))$.
2. Let $a : \text{People on Earth} \rightarrow \mathbb{N}$ be the function that assigns to each person x on Earth his/her age as of January 1, 2005. Then $a(\text{your author}) = 50$ while $a^{-1}(50)$ is the set of all 50 year olds on the planet. Hence $a^{-1}(a(\text{your author})) = a^{-1}(50)$ is quite a large set, not equal to just $\{\text{your author}\}$.

This is why the boxes in diagram (2.8) properly contain each other.

There is a relationship between subsets and the composition of functions. Notice the reversal of order in the functions f and g in the next result. This is also a test. The only question on this exam is: *Have you learned enough of the basics to understand a more terse mathematical proof?* Before you read on, remember that sets A and B are equal, $A = B$, exactly when $A \subset B$ and $B \subset A$. Nothing less than these two conditions may be proved if we wish to conclude that $A = B$. You will start with an $x \in A$ and prove that it is in B . Then you will start with an $x \in B$ and prove that it is in A . Notice that in the next several proofs this is exactly what we are practicing.

Theorem 2.2.8 *Let $f : A \rightarrow B$ and $g : B \rightarrow C$ be functions, and let $Z \subset C$. Then*

$$(g \circ f)^{-1}(Z) = f^{-1}(g^{-1}(Z)).$$

Proof: To prove the equality we must show that

$$(g \circ f)^{-1}(Z) \subset f^{-1}(g^{-1}(Z))$$

and that

$$f^{-1}(g^{-1}(Z)) \subset (g \circ f)^{-1}(Z).$$

We will prove the inclusion $(g \circ f)^{-1}(Z) \subset f^{-1}(g^{-1}(Z))$.

Let $x \in (g \circ f)^{-1}(Z)$. By definitions of composition and preimage,

$$g(f(x)) = (g \circ f)(x) \in Z.$$

Since $g(f(x)) \in Z$, $f(x) \in g^{-1}(Z)$, and then $x \in f^{-1}(g^{-1}(Z))$. Hence

$$(g \circ f)^{-1}(Z) \subset f^{-1}(g^{-1}(Z)).$$

Conversely, we will prove that $f^{-1}(g^{-1}(Z)) \subset (g \circ f)^{-1}(Z)$. Suppose that $x \in f^{-1}(g^{-1}(Z))$. We must show that $x \in (g \circ f)^{-1}(Z)$. Now $x \in f^{-1}(g^{-1}(Z))$ implies that $f(x) \in g^{-1}(Z)$ and hence that $g(f(x)) \in Z$. By definition of \circ we have

$$(g \circ f)(x) = g(f(x)) \in Z,$$

so that

$$x \in (g \circ f)^{-1}(Z).$$

Consequently,

$$f^{-1}(g^{-1}(Z)) \subset (g \circ f)^{-1}(Z),$$

and therefore $f^{-1}(g^{-1}(Z)) = (g \circ f)^{-1}(Z)$. This completes the proof.

Finally, as an exercise in the use of the quantifiers *for all* and *there exists*, we will prove the following theorems.

Theorem 2.2.9 *Let $f : A \rightarrow B$ be a function. If $\{Y_n \mid n \in \mathbb{N}\}$ is a family of subsets of B then*

$$\bigcup_{n \in \mathbb{N}} f^{-1}(Y_n) = f^{-1} \left(\bigcup_{n \in \mathbb{N}} Y_n \right).$$

Proof: In shorthand notation, we will first prove the inclusion

$$\bigcup_{n \in \mathbb{N}} f^{-1}(Y_n) \subset f^{-1} \left(\bigcup_{n \in \mathbb{N}} Y_n \right).$$

Let

$$x \in \bigcup_{n \in \mathbb{N}} f^{-1}(Y_n).$$

By the definition of \cup , $x \in f^{-1}(Y_m)$ for some $m \in \mathbb{N}$, so that $x \in f^{-1}(Y_m)$. By definition of preimage, $f(x) \in Y_m$, so that

$$f(x) \in \bigcup_{n \in \mathbb{N}} Y_n$$

by definition of \cup . The definition of preimage shows us that

$$x \in f^{-1} \left(\bigcup_{n \in \mathbb{N}} Y_n \right)$$

and hence that

$$\bigcup_{n \in \mathbb{N}} f^{-1}(Y_n) \subset f^{-1} \left(\bigcup_{n \in \mathbb{N}} Y_n \right).$$

Conversely, we will prove the reverse inclusion

$$f^{-1} \left(\bigcup_{n \in \mathbb{N}} Y_n \right) \subset \bigcup_{n \in \mathbb{N}} f^{-1}(Y_n).$$

Let

$$x \in f^{-1} \left(\bigcup_{n \in \mathbb{N}} Y_n \right).$$

The definition of preimage implies that

$$f(x) \in \bigcup_{n \in \mathbb{N}} Y_n$$

and the definition of \cup implies that $f(x) \in Y_k$ for some $k \in \mathbb{N}$. Then $x \in f^{-1}(Y_k)$ for that $k \in \mathbb{N}$, and so

$$x \in \bigcup_{n \in \mathbb{N}} f^{-1}(Y_n).$$

Consequently,

$$f^{-1} \left(\bigcup_{n \in \mathbb{N}} Y_n \right) \subset \bigcup_{n \in \mathbb{N}} f^{-1}(Y_n)$$

and therefore

$$f^{-1} \left(\bigcup_{n \in \mathbb{N}} Y_n \right) = \bigcup_{n \in \mathbb{N}} f^{-1}(Y_n).$$

This completes the proof.

Here are a few exercises that will build your mental mathematical muscles. The reader can sample the thrill of the creative and discovery processes by providing proofs of all these results. Rather than repeat the hypotheses over and over again, let us agree that, for each of these exercises, $f : A \rightarrow B$ is a function, that $X, X' \subset A$, and that $Y, Y' \subset B$.

1. If $X \subset X' \subset A$ then $f(X) \subset f(X')$.
2. $f(f^{-1}(Y)) \subset Y$.
3. If $X, X' \subset A$ then $f(X \cap X') \subset f(X) \cap f(X')$.

4. If $X, X' \subset A$ then $f(X \cup X') = f(X) \cup f(X')$.
5. If $Y, Y' \subset B$ then $f^{-1}(Y \cap Y') = f^{-1}(Y) \cap f^{-1}(Y')$.
6. If $Y, Y' \subset B$ then $f^{-1}(Y \cup Y') = f^{-1}(Y) \cup f^{-1}(Y')$.
7. If $\{X_n \mid n \in \mathbb{N}\}$ is a family of subsets of A then $f(\bigcup_{n \in \mathbb{N}} X_n) = \bigcup_{n \in \mathbb{N}} f(X_n)$.
8. If $\{Y_n \mid n \in \mathbb{N}\}$ is a family of subsets of B then $\bigcap_{n \in \mathbb{N}} f^{-1}(Y_n) = f^{-1}(\bigcap_{n \in \mathbb{N}} Y_n)$.

2.3 One-to-One and Onto Functions

Functions come in a variety of different colors. For instance, $f(x) = x^2$ defines a function on \mathbb{R} such that $f(-1) = f(1)$. That is, f takes two different numbers to the same number. The function $f(x) = x + 2$, on the other hand, takes different numbers $x \neq x'$ to different numbers $x + 2 \neq x' + 2$. When you stand in a particularly long line to see a movie and you count the number of people in front of you, you are forming a function f that associates a natural number n with a person in that line. Then $f(1)$ is the first person in line, $f(2)$ is the second person in line, and so on. In this function we see that if $n \neq m$ then the people $f(n)$ and $f(m)$ are different. One person will not hold two different places in line.

This type of *counting* function goes back to the first humans to herd animals. As the goats would enter the pasture land the ancient goatherd would drop a rock or stone in a pile. In this way he sets up a function from the goats to the pile of stones. As the goats leave the pasture he removes one stone for each goat that passes. If he has stones left he might conclude that the wolf has found one of his goats. If there are more goats than stones he might conclude that he has picked up another goat from somewhere. Different stones are associated with different goats. Thus when the goats and the stones match up element by element he concludes that he has just as many goats that evening as he had that morning.

The next definition identifies this property of functions.

Definition 2.3.1 Let $f : A \longrightarrow B$ be a function.

1. We say that f is a one-to-one function if

$$f(x) = f(x') \text{ implies that } x = x'$$

or equivalently if

$$x \neq x' \text{ implies that } f(x) \neq f(x').$$

Said in a colloquial manner, f is one-to-one if different elements of A map to different elements of B .

2. We say that f is an onto function if

given $y \in B$ then we can find an $x \in A$ such that $f(x) = y$.

In an informal way, f is onto if every element in B is the image of an element from A .

Consider what this definition represents. We have uncovered a pair of properties about functions that the reader probably has not encountered before. This discovery is the tip of an iceberg, or a general principle, that is called *abstraction*. Abstraction is one of the processes through which mathematics grows. For example, you can study lines on a sheet of paper and you will understand lines and linear processes. But once you abstract lines to linear approximation you have discovered the calculus. Now you can leave that sheet of paper and study nature and the universe around us. If you allow your definition of line to be abstracted you grow from points and lines on a plane to studying the geometry of points and lines on a sphere, or a globe. Lines here would be different because *they must curve along the sphere*. Quite a change, isn't it? Abstraction usually has this effect on mathematics.

Some examples will help you visualize one-to-one and onto functions.

Example 2.3.2 Start with a function $f : \{a, b\} \longrightarrow \{x, y, z\}$ defined by

$$f(a) = x \quad f(b) = y.$$

A glance at the definition shows us that $f(a) \neq f(b)$ so that f is a one-to-one function. f is not an onto function since $f(a), f(b) \neq z$. That is, z is not an image of an element from $\{a, b\}$.

Define a function $f : \{a, b, c\} \longrightarrow \{x, y\}$ by

$$f(a) = x \quad f(b) = x \quad f(c) = y.$$

Since $f(a) = x$ and $f(c) = y$, each element of the codomain $\{x, y\}$ is an image of an element from $\{a, b, c\}$. Thus f is an onto function. f is not a one-to-one function since the different elements $a \neq b$ map to the same element $f(a) = f(b) = x$.

You may recall that we examined sets that could be *matched element by element*. The matching is a limited way of describing a function that is both one-to-one and onto. These one-to-one and onto functions will thus replace the previous notion of an element by element matching of sets.

Example 2.3.3 Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be defined by $f(x) = 7x + 6$. We claim that f is both one-to-one and onto.

1. f is one-to-one: Suppose that $x, x' \in \mathbb{R}$ are such that $f(x) = f(x')$. Then $7x + 6 = 7x' + 6$ implies that $7x = 7x'$. We can divide by 7 to show that $x = x'$. As claimed f is one-to-one.

2. f is onto: Suppose that $y \in \mathbb{R}$. We must find a number $x \in \mathbb{R}$ such that $f(x) = y$. Let $x = \frac{y - 6}{7} \in \mathbb{R}$. A little algebra shows us that

$$f(x) = f\left(\frac{y - 6}{7}\right) = 7\left(\frac{y - 6}{7}\right) + 6 = y.$$

Thus $f(x) = y$, so that f is onto, as claimed.

In the above example we made use of the arithmetic of 7 and 6. We added and subtracted at will and we divided by 7 like there was no problem with it. (There isn't.) But now let us abstract this problem. We will need to acknowledge the fact that we could divide by 7 because $7 \neq 0$. We also need to see that $6 - 6 = 0$. Let's see how that is used in a more abstract setting.

Example 2.3.4 Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be defined by $f(x) = mx + b$ for some $m \neq 0$. We claim that f is both one-to-one and onto.

1. f is one-to-one: Suppose that $x, x' \in \mathbb{R}$ are such that $f(x) = f(x')$. Then

$$\begin{aligned} mx + b &= mx' + b \\ (mx + b) - b &= (mx' + b) - b \\ mx &= mx' \\ \frac{mx}{m} &= \frac{mx'}{m} \\ x &= x'. \end{aligned}$$

We could divide by m in the second to last step only because we chose $m \neq 0$. As claimed f is one-to-one.

2. f is onto: Suppose that $y \in \mathbb{R}$. We must find a number $x \in \mathbb{R}$ such that $f(x) = y$. Let $x = \frac{y - b}{m}$. Since $m \neq 0$, we can divide by m to form y , and so $x \in \mathbb{R}$. A little algebra shows that

$$f(x) = f\left(\frac{y - b}{m}\right) = m\left(\frac{y - b}{m}\right) + b = y.$$

Thus $f(x) = y$, so that f is onto, as claimed.

Our choice of x in the above example is typical of mathematical proof. Before setting down to write this little mathematical essay, several drafts of the essay were made. We found the value $x = \frac{x - b}{m}$ before we set pen to paper to write out our proof. Our choice of x may at first seem magical, coming from nowhere, without motivation as it did, but that is because you were not there while I worked out all of the details in the drafts of this proof. That is the sign of a well written well constructed argument. You have done all of your homework surrounding the proof so that when you attempt to type that final draft you can appeal to values as if they come naturally from the ether as opposed to the preparation. I will use this style of mathematics as little as possible, but I will use it.

Example 2.3.5 Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be defined by $f(x) = x^2$. We will show that f is neither one-to-one nor onto.

1. f is not one-to-one: We need only produce two different numbers that map to the same number. $f(-2) = 4 = f(2)$ shows

that f is not one-to-one. (Where did -2 and 2 come from, reader? If you write out a draft of this proof yourself you might find their origins.)

2. f is not onto: We need only produce one element in \mathbb{R} that cannot be written as $f(x)$. Since $x^2 \geq 0$ for all $x \in \mathbb{R}$, $f(x) = x^2 \neq -1$ for any $x \in \mathbb{R}$. (A little preparation gives us the value -1 as a good example of a number that is not an image under $f(x) = x^2$.)

A simple change of the domain or the codomain can significantly change the function. That is, by changing the domain and the codomain of a function that is neither one-to-one nor onto we change the function into one that is both one-to-one and onto. As we said above these one-to-one and onto functions are the element by element matchings that we encountered earlier as functions for counting sets. Such a special function deserves a name. We use the French term. A function is a *bijection* if it is both one-to-one and onto.

Definition 2.3.6 *The function $f : A \longrightarrow B$ is a bijection if f is a one-to-one and onto function.*

2.4 Bijections

Some functions on finite sets will further develop your intuition about bijections, or one-to-one and onto functions.

The function $f : \{a, b, c\} \longrightarrow \{x, y, z\}$ defined by

$$f(a) = x, \quad f(b) = y, \quad f(c) = z$$

or by a list of assignments

$$\begin{array}{rcl} a & \longmapsto & x \\ b & \longmapsto & y \\ c & \longmapsto & z \end{array}$$

is a bijection since a glance at the list shows us that different elements in the domain $\{a, b, c\}$ are mapped to different elements in

$\{x, y, z\}$, and each element in the codomain $\{x, y, z\}$ is the image of an element from $\{a, b, c\}$. We might have said as in Chapter 1 that there is an element by element matching of $\{a, b, c\}$ with $\{x, y, z\}$. Our use of a bijection makes this intuitive matching of elements mathematically concrete.

Example 2.4.1 Example 2.3.3 shows that if $f : \mathbb{R} \rightarrow \mathbb{R}$ is defined by $f(x) = 7x + 6$ then f is one-to-one and onto. Thus $f(x) = 7x + 6$ defines a bijection $f : \mathbb{R} \rightarrow \mathbb{R}$.

Example 2.4.2 Example 2.3.4 shows that if $m \neq 0$ is a real number and if $f : \mathbb{R} \rightarrow \mathbb{R}$ is defined by $f(x) = mx + b$ then f is one-to-one and onto. Thus $f(x) = mx + b$ defines a bijection $f : \mathbb{R} \rightarrow \mathbb{R}$.

Example 2.4.3 Let \mathbb{R}^+ be the set of positive real numbers, and let $f : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ be defined by $f(x) = \frac{1}{x}$. We will show that f is a bijection.

1. To show that f is one-to-one suppose we have $x, x' \in \mathbb{R}^+$ such that $f(x) = f(x')$. The definition of f shows us that then $\frac{1}{x} = \frac{1}{x'}$, and so a little algebra shows us that $x = x'$. Thus f is one-to-one.

2. To show that f is onto begin with a $y \in \mathbb{R}^+$. We must choose an $x \in \mathbb{R}^+$ such that $f(x) = y$. Since $y \in \mathbb{R}^+$, $y \neq 0$, so we can form the real number $x = \frac{1}{y} \in \mathbb{R}^+$. Observe that

$$f(x) = f\left(\frac{1}{y}\right) = \frac{1}{1/y} = y$$

to conclude that f is onto. Therefore f is a bijection.

Example 2.4.4 Recall that \mathbb{R}^+ is the set of positive real numbers. Let $g : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ be defined by $g(x) = x^2$. We will show that g is a bijection. That is, it is both one-to-one and onto.

1. g is one-to-one: Suppose that $x, x' > 0$ are such that $g(x) = g(x')$. Then $x^2 = (x')^2$. Since $x, x' > 0$ we can take square roots to see that

$$x = \sqrt{x^2} = \sqrt{(x')^2} = x'.$$

Thus g is one-to-one.

The above chain of equations holds only because $x, x' > 0$. To see this notice that $1^2 = (-1)^2$ while $1 \neq -1$. Only the *positive* numbers x satisfy $\sqrt{x^2} = x$.

2. g is onto: Let $y \in \mathbb{R}^+$. Since $y > 0$ we can let $x = \sqrt{y} \in \mathbb{R}^+$. Then

$$g(x) = (\sqrt{y})^2 = y,$$

which shows us that g is onto. (Where did that value x come from? Obviously, it is from the preparations we made before writing down the final draft of this proof.)

These examples concerning the rule x^2 should be compared. Even though the functions $f : \mathbb{R} \longrightarrow \mathbb{R}$ and $g : \mathbb{R}^+ \longrightarrow \mathbb{R}^+$ have the same rule, namely, $f(x) = g(x) = x^2$, their domains and their codomains are different. This difference makes g both one-to-one and onto, while f is neither one-to-one nor onto. These must be different functions because they behave differently. They enjoy different properties. Thus the sets used to define a function will have a significant effect on the properties possessed by the function.

Example 2.4.5 Example 2.3.2 shows us that the function $f : \{a, b\} \longrightarrow \{x, y, z\}$ defined by

$$f(a) = x, \quad f(b) = y$$

is one-to-one but not onto. Thus f is not a bijection.

However, the function $g : \{a, b\} \longrightarrow \{x, y\}$ defined by

$$g(a) = x, \quad g(b) = y$$

is a bijection. Notice that f and g have the same rule. The difference in their codomains makes them different functions.

Example 2.4.6 By Example 2.3.2 the function $f : \{a, b, c\} \longrightarrow \{x, y\}$ defined by

$$f(a) = x, \quad f(b) = x, \quad f(c) = y$$

is onto and not one-to-one. Thus f is not a bijection.

However, the function $g : \{b, c\} \longrightarrow \{x, y\}$ defined by

$$g(b) = x, \quad g(c) = y$$

is a bijection. Again we see that a change in domain changes a function that is not one-to-one into a function that is one-to-one.

2.5 Inverse Functions

The operation of composition $g \circ f$ looks very much like a multiplication of functions. It is natural to ask if we can divide functions as well. Unfortunately, *the division of functions is not defined in general*. Thus we must examine an analogous operation, namely the undoing of a function. We consider each function f as a rule that manipulates or that changes an input value x . The value $f(x)$ is seen then as some kind of variation on x . Our inverse functions will be this manipulation in reverse. If f squares then its inverse g takes square roots. If f adds 3 to x then g will subtract 3 from x . Do you see it? The inverse of f is the function that performs the opposite manipulation that f performs.

The best solution in our search for an inverse function is to find the largest class of functions that have inverse functions. Specifically, we will discuss functions for which the composition division exists. Within this collection a specific function f will be associated with a function g such that

$$\begin{aligned} g \circ f(x) &= g(f(x)) = x \text{ and} \\ f \circ g(x) &= f(g(x)) = x. \end{aligned}$$

In this case we will say that g is the *inverse of f* . This inverse is not found by dividing by f but by *undoing* whatever it is that f does. It happens that this kind of functional inverse exists only for bijections.

Definition 2.5.1 Let $f : A \rightarrow B$ and $g : B \rightarrow A$ be functions. Notice the change in the domain and codomain of f and g . Then g is called the *inverse of f* if

1. $g \circ f(x) = x$ for each $x \in A$, and
2. $f \circ g(y) = y$ for each $y \in B$.

In this case f is also the inverse of g . If f has an inverse then we say that f is *invertible*.

Observe that the compositions $g \circ f$ and $f \circ g$ are defined since their domains and codomains match up.

Example 2.5.2 Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be defined by

$$f(x) = mx + b$$

for some numbers $m \neq 0$ and b . Define a function $g : \mathbb{R} \rightarrow \mathbb{R}$ by

$$g(x) = \frac{x - b}{m}.$$

Since $m \neq 0$, $g : \mathbb{R} \rightarrow \mathbb{R}$ is a function. We will show that g is the inverse of f .

Let $x \in \mathbb{R}$. We must show that $g \circ f(x) = f \circ g(x) = x$. A little algebra shows us that

$$\begin{aligned} g \circ f(x) &= g(f(x)) \\ &= g(mx + b) \\ &= \frac{(mx + b) - b}{m} \\ &= \frac{mx}{m} \\ &= x \end{aligned}$$

and that

$$\begin{aligned} f \circ g(x) &= f(g(x)) \\ &= f\left(\frac{x - b}{m}\right) \\ &= m\left(\frac{x - b}{m}\right) - b \\ &= (x - b) - b \\ &= x. \end{aligned}$$

Thus g is the inverse of f .

Now where do you suppose the inverse g of f came from in the above example? In preparing for this example, I assumed that $f(x) = mx + b$ had an inverse $g(x)$. Then I calculated as follows. Since $f(g(x)) = x$ we have

$$m \cdot g(x) + b = x.$$

Now solve for $g(x)$.

$$\begin{aligned} m \cdot g(x) + b &= x \\ m \cdot g(x) &= x - b \\ g(x) &= \frac{x - b}{m}. \end{aligned}$$

The only reason that $g(x)$ can be found in this case is because $m \neq 0$. If $m = 0$ then f is not a bijection and its inverse g cannot be found.

The above theorem connects bijections with inverses. Try thinking of its proof as another examination of how far your mathematical thought processes have progressed over the last 10 pages or so. This approach is drier than the examples you have met in this chapter and the presentation will cause you to think about my line of reasoning. The idea of choosing values x or y in some previous draft is used throughout this argument. However, all of the details are there and if you can read it without too much difficulty then you are ready for the next chapter, where we begin our investigation of infinite sets.

Theorem 2.5.3 *Let $f : A \longrightarrow B$ be a function.*

1. *If f is invertible then f is a bijection.*
2. *If f is a bijection then f is invertible.*

Proof: 1. Suppose that f is invertible. We must show that f is one-to-one and onto. Since f is invertible it has an inverse g .

Suppose that $f(x) = f(x')$ for some $x, x' \in A$. Then

$$g(f(x)) = g(f(x')),$$

and since g is the inverse of f we have

$$x = g(f(x)) = g(f(x')) = x'.$$

Thus f is one-to-one.

Let $y \in B$, and let $x = g(y)$. Since g is the inverse of f we have

$$f(x) = f(g(y)) = y$$

so that f is onto.

Therefore f is a bijection.

2. Suppose that f is a bijection. We must *construct* a function g that will serve as the inverse of f . Define a rule g from B into A by

$$g(x) = y \text{ exactly when } x = f(y).$$

This is a reasonable definition for an inverse. We must show that g is a function.

The definition of function (page 38) asks us to define g on all of B . So let $x \in B$. Since f is onto there is a $y \in A$ such that $x = f(y)$. Then by definition of g , $g(x) = y$.

Also the definition of function asks us to show that $g(x)$ is exactly one value. To prove that there is exactly one of something you assume that you have two perhaps equal things, and then you prove that they are the same. So suppose there are two elements $y, y' \in A$ such that

$$g(x) = y \text{ and } g(x) = y'.$$

By the definition of g ,

$$x = f(y) \text{ and } x = f(y')$$

and because f is one-to-one, $y = y'$. Thus $g(x)$ is exactly one value.

Therefore $g : B \rightarrow A$ is a function.

To complete the proof we must show that $g(f(x)) = x$ and that $f(g(y)) = y$ for each $x \in A$ and $y \in B$.

Let $x \in A$ and write

$$f(x) = y.$$

Then, by the definition of g ,

$$x = g(y)$$

so that

$$g(f(x)) = g(y) = x.$$

On the other hand, if $y \in B$ is given then write x for $g(y)$. That is,

$$g(y) = x.$$

By the definition of g , $y = f(x)$, so that

$$f(g(y)) = f(x) = y.$$

Hence g is the inverse of f , which is what we had to prove.

The above theorem can be used to anticipate the existence of inverses without telling us what that inverse might be. For instance,

$$f(x) = mx + b \text{ where } m \neq 0$$

is known to be a bijection. Thus the above theorem tells us that f is invertible. It does not tell us what that inverse might be. We found (see Example 2.3.4) that

$$g(x) = \frac{x - b}{m}$$

is the inverse of f .

The function

$$h(x) : \mathbb{R}^+ \longrightarrow \mathbb{R}^+ \text{ such that } h(x) = x^2$$

is known to be a bijection. (See Example 2.4.4.) Thus h has an inverse. Its inverse can be found by emulating the proof of the above theorem. Solve the equation

$$x = y^2$$

for y . Of course, we just take a square root so that

$$\sqrt{x} = y.$$

Hence

$$s(x) = \sqrt{x}$$

is the inverse of h .

Now that we have plowed through the preliminaries, let me present a story.

When we learn to play a game like checkers or chess we play with a person learned in the ways of the game. As we play we learn the rules over several initial games. It might take three or four games to learn the rules. There is a deeper understanding that comes with playing more games. The only way to learn the game well is to lose a lot of games to a better player. If we keep coming back to challenge this master of the game we eventually will learn enough to challenge him or her. However, maybe you don't play games like checkers or chess. Maybe you like video games.

How do we learn to play video games? We start with a two dimensional character who moves across the screen prodded by our control box. Initially we play for 15 seconds and then we die. Do we give up? Do we go onto something else? We do not. We hit *reset* and begin again. This time we get a little further. Eventually, after resetting the game for what feels like an infinite number of times, we reach the end of the game. We defeat the Monster who was holding your Sister with the Golden Hair hostage, and we rejoice. We won the game. But it took a while. It took a lot of time.

I ask you to devote a fraction of that energy into the topics that we will cover in the next few chapters. The derived rewards will be far more than those enjoyed by checkers or video games, I assure you.

This Page Intentionally Left Blank

Chapter 3

Counting Infinite Sets

3.1 Finite Sets

The *natural numbers* are the numbers introduced by the Hindu and Arabic cultures. Today they are written as

$$\mathbb{N} = \{0, 1, 2, 3, \dots\}.$$

This is an implied list. The three dots, \dots , indicate that the pattern continues indefinitely, or without stopping. Although this set appears to be small it contains numbers that we do not encounter in our daily lives. Some examples of natural numbers that you have probably not met before are:

1. The number of people on the Earth, about 6,000,000,000 or 6×10^9 .

2. The Gross National Product of the USA is about 100 trillion or 10^{14} dollars. This is a 1 followed by 14 zeros,

$$100,000,000,000,000.$$

3. The distance across the observable universe, about 4×10^{10} light years. This is a 4 followed by 10 zeros. It takes light 4×10^{10} years to travel this distance. This number doesn't have a name, only a scientific designation.

4. The number of fish in the ocean on January 1, 2003. Even though we can't write down this number, it is still a number and we can refer to it by a symbol, say, α .

5. The number of grains of sand on the Earth. We will figure out how large this number is, but I think we can all agree that even though the number exists, it is impossible to write the exact number down.

Let us see how these numbers compare. To do this we will have to expand your imagination a bit. Once enlarged we will use the natural numbers $\{0, 1, 2, \dots\}$ to write down some really large numbers. Suppose we write down the largest number that you can imagine, say,

$$X.$$

We can get a larger number by adding 1:

$$X + 1.$$

How big was X ? Was it the number of people on Earth, 6 billion? Was it the speed of light, 186,000 miles/second? Maybe it was a *light year*, the distance light travels in a year, or about 2.6×10^9 meters. In miles that would be 1.6×10^9 . Now those were large numbers. But this is hardly big enough. We can produce a larger number by adding 1:

$$6,000,000,001.$$

Not really much larger, though, is it? Maybe we can form an exponential number:

$$(6 \times 10^9)^2 = 36 \times 10^{18}.$$

This is larger than a light year since it has 18 zeros following the 36. We can make a larger number:

$$36 \times 10^{18} + 1.$$

Ones are not what we need here. I want numbers that grow so large that we cannot print them with meaning. Let me explain that with examples.

Now that you have seen how big these numbers can be, let us write down implied lists of these numbers called *sequences*. One such sequence starts with 100:

$$100, 101, 102, 103, 104, \dots$$

These numbers do seem as large as they might have seemed to you, but they are getting larger. They just grow slowly. Adding 1 moves us only a little bit when compared to our first number.

So let us take multiples of 100:

$$100, 200, 300, 400, \dots$$

The second number is twice as large as the first number, and the third number is three times as large as the first, etc. Once you've done this 100 times you have reached a number that is 100 times larger than 100. This is the number

$$100 \times 100 = 10,000.$$

These numbers get larger more quickly than the numbers do in the sequence 100, 101, 102, 103, ..., but they are still quite small compared to the numbers we can form.

The sequence

$$100, 100^2, 100^3, 100^4, \dots \tag{3.1}$$

gets large relatively quickly. Each time we produce another entry in this sequence we add 2 zeros to the previous entry. The first number has 2 zeros, the second number has 4 zeros, the third has 6, the fourth number

$$100,000,000$$

has 8 zeros, and so on. The number 100^2 is already as large as the hundredth entry in the sequence 100, 200, 300, 400, The hundredth entry in the sequence 100, 100^2 , 100^3 , 100^4 , ... is 100^{100} , which is a 100 followed by a hundred zeros. These numbers are getting larger in *magnitude* with each entry in the sequence. So things are quickly getting large, but maybe the values are not large enough.

For the next demonstration we will write

$$100 = 10^2.$$

The reasoning here is that the exponential form is more easily manipulated. Consider the sequence

$$10^2, 10^{10^2}, 10^{10^{10^2}}, 10^{10^{10^{10^2}}}, \dots$$

of exponents of 10. These numbers get large so fast that the second one, the number

$$10^{100} = 1 \text{ followed by a hundred zeros}$$

is too large to comfortably print in this space. It is just a 1 followed by 100 zeros, but the usual way of writing it down as 1,000, . . . is utter folly. Try to write down 10^{100} as a 1 followed by 100 zeros just to get the feel for how useful this exponential notation is for this number. No one can read that many zeros with meaning. Did we read 99 zeros or did we read 100 zeros? This number 10^{100} is larger than the number of grains of sand you would need to fill a bag as big as the Earth. (More on that sand number later.) It is hard to understand just how big the other numbers are in this sequence. And yet each of them is finite.

Now let us compare the sizes of the numbers in the different sequences. We will all agree that

$$10^{10} + 1 < 2 \times 10^{10} < 10^{20} < 10^{100}.$$

One way to see that a number x is large is to see how small its reciprocal $\frac{1}{x}$ is. The number 10^{100} is so large that if you used your calculator to find a decimal equivalent for the fraction

$$\frac{1}{10^{100}}$$

your display would probably read 0. And yet this number is not 0, no matter what the display says.

Now that we have some large numbers we can examine how large the relatively small numbers 10^{100} and $10^{10^{100}}$ are. We will use the sand that makes up the Earth to give us all a mental picture of these numbers.

To write down the number of grains of sand on the Earth we will have to use *scientific notation*. Scientific notation is the use of powers of 10 to abbreviate numbers that have a daunting length. Certainly we can write down

This number is large if you happen to be 9 years old. We can also write

$$10^2 = 100.$$

One hundred is a large number if you earned a 75 on your last mathematics examination, but it may not require the scientific notation 10^2 . Larger numbers look like this:

$$10^{15} = 1,000,000,000,000,000.$$

You see that 10^{15} is a more compact way of writing 1 followed by 15 zeros. As we indicated before,

$$10^{100}$$

is a more practical method for writing down that number. It is an example of a number that can only be written in a meaningful way as a power of 10.

Now that is one large number. And that is where we draw the line. One hundred zeros is beyond the patience of most people. Scientific notation prevails.

Small numbers can also be written in scientific notation. For instance,

$$10^{-1} = .1$$

$$10^{-2} = .01 \text{ notice the 1 zero.}$$

$$10^{-10} = .0000000001 \text{ notice the 9 zeros.}$$

$$10^{-100} = \text{a decimal point “.” followed by 99 zeros and then a 1.}$$

The only useful way to write down 10^{-100} is as a power of 10. In other words, there are some numbers that can only be realized in scientific notation.

The reader may ask: "Why examine this? Surely 10^{100} and 10^{-100} are unique enough that they do not apply to anything in the real world." This is not the case. For example, when banks transfer sums of money between themselves they need to encode their transaction. They need to hide this little bit of business from the rest of the world. The numbers used to encode have more than 100 digits. That is, in order to transfer information between banks, a 100 digit number, a number larger than 10^{100} is required. Just

how this encoding takes place is too far afield for our elementary intuitive discussion.

Another large number is found in chemistry. There are about 6×10^{23} atoms per *mole of matter*. All we need to know about a *mole of matter* is that it is a fundamental unit of measurement in chemistry. There are about 10^{28} atoms in your body and about $10^{80} = 10^{8 \times 10}$ protons, neutrons, and electrons exist in the space of the observable universe. This is the fortieth entry in the implied list $100, 100^2, 100^3, \dots$ above, or a 1 followed by 80 zeros. The point is, we have discovered something new. Numbers with up to 100 digits have a very practical purpose.

The story goes that two mathematicians and a child were sitting around the kitchen table one afternoon when one adult asks the other just how many grains of sand there are on the Earth. To make this problem something we can see in our mind's eye, suppose we have a typical grain of sand. The typical grain of sand has the following dimensions.

1. weight of one grain of sand = 10^{-2} grams.
2. diameter of one grain of sand = 10^{-4} meters.

We duplicate this grain many times over until we have a bag of sand that, in our minds eye, is the size of the Earth. By size I mean that the bag of sand and the Earth have the same size and shape and mass. The question we are asking now becomes: "How many grains of sand do we need to fill a bag that large?"

The grain of sand is light, but not too light. It seems that 100 copies of this grain of sand weigh

$$100 \times 10^{-2} = 10^2 \cdot 10^{-2} = 1 \text{ gram.}$$

So 100 grains of sand would not make a large or heavy bag. Suppose we have 1 trillion or 10^{12} copies of this grain of sand. Then the mass of that bag of sand is

$$1 \text{ trillion} \cdot 10^{-2} = 10^{12} \cdot 10^{-2} = 10^{12-2} = 10^{10}$$

or 10 billion grams. Now that is a large number of grains of sand and yet modern computer hard drives have many more than 10 billion bytes (10 gigabytes) of available storage. So this number is

still something you might have in front of you right now. Suppose that we have 10^{102} copies of that grain of sand. Then the mass of that bag of sand is

$$10^{102} \cdot 10^{-2} = 10^{102-2} = 10^{100} \text{ grams},$$

which it turns out is pretty close to the mass of the Earth, so we can say that there are about 10^{102} grains of sand on the Earth.

The mathematicians I mentioned above were so impressed by this number that they asked a nine month old child to name the number. The child said *googol* and that name stuck.

The number 10^{100} is called a *googol*.

Even though we have answered our question of just how many grains of sand make up the Earth, that is not the end of the story. As mathematicians will do, they asked for some kind of concrete realization of the number *googolplex*.

The number $10^{10^{100}}$ is called a *googolplex*.

Let us demonstrate just how large a googolplex is.

For this demonstration we will need a common object from around the house, toilet paper. But this roll of toilet paper has a special printing on it. On the first square of paper is a 1 and on each following square there is a 0. The roll contains enough squares so that the number 1 followed by all of those zeros is a googolplex. Just how large is that roll? It is so large that you would never run out of this household necessity. This roll could not be kept in your house, it is too large. You might try keeping it somewhere on your block. No. The roll is much larger than that. Perhaps the mayor of your city would allow you to use the city as a storage for this roll. This also is too small. The roll expands beyond the boundaries of any city. How about the country you live in? The roll is larger than that. Perhaps you would keep it on the North American continent. Nope. Too small for this roll. Let us jump and ask if the

Earth will hold it. No. The Earth is too small to contain such a roll. Be patient, we have a long journey to make here. Perhaps the roll would fit into the solar system as encompassed by the orbit of Pluto? This roll is too big to be kept in that space. Perhaps the Galaxy is large enough to encompass this roll. No. The roll will spill out into intergalactic space. How much larger should we go? Is this roll large enough to fit into the observable universe? Now that would be one huge roll of toilet paper. And yet this roll is too large to fit into the observable universe. We are faced with the following thought: *This roll is larger than our physical universe can hold.* That googolplex is one large number. It would seem that a googolplex is a number for which there is no physical significance. However, it is still a small value when compared with numbers like the exponential

$$10^{\text{googolplex}}.$$

I can't even begin to hint at how large that monster might be. These numbers are large but they are still finite. The numbers that we will encounter in the next section, called *infinite cardinals*, outstrip these natural numbers for size.

The numbers googol and googolplex seem to have struck a nerve with modern American culture. The numbers turn up in cartoons, movies, and a number of novels published in the United States. This is the first example I know of an obscure mathematical idea that appears in a children's cartoon as well as adult entertainment.

3.2 Hilbert's Infinite Hotel

In Chapter 2 we started a process that would eventually define each natural number n from nothing. Thus we will assume that we have constructed the set

$$\mathbb{N} = \{0, 1, 2, 3, \dots\}$$

of natural numbers. To decide what an infinite set is, we first decide what a finite set is.

Definition 3.2.1 Let $n \in \mathbb{N}$. We say that the set A has cardinality n and we write

$$\text{card}(A) = n$$

if there is a bijection $f : A \longrightarrow \{1, \dots, n\}$. The set A is finite if $\text{card}(A) = n$ for some $n \in \mathbb{N}$.

Example 3.2.2 $\text{card}(\{\bullet\}) = 1$ since every function $f : \{\bullet\} \longrightarrow \{1\}$ is a bijection.

The set $\{\bullet, :\}$ has cardinality 2 since the map

$$f : \{\bullet, :\} \longrightarrow \{1, 2\}$$

defined by

$$f(\bullet) = 2 \quad \text{and} \quad f(:) = 1$$

is a bijection.

Example 3.2.3 We claim that \mathbb{N} is not a finite set. Even though this is an intuitively obvious fact, mathematicians require some kind of justification or proof. We will show that there are no bijections from $f : \mathbb{N} \longrightarrow \{1, \dots, n\}$. To do this we will show that no onto function $f : \mathbb{N} \longrightarrow \{1, \dots, n\}$ is a one-to-one function. (In this way f cannot be both one-to-one and onto.) To prove that f is not one-to-one we will find two *different* numbers x and x' that map to the same value. That is, we must produce $x \neq x' \in \mathbb{N}$ such that $f(x) = f(x')$.

Given an integer $n > 0$, suppose we are given an onto function $f : \mathbb{N} \longrightarrow \{1, \dots, n\}$. There are $x_1, \dots, x_n \in \mathbb{N}$ such that $f(x_1) = 1, \dots, f(n) = n$. Since $\{x_1, \dots, x_n\}$ is a finite set of numbers it has a largest number. We will assume that this number is x_1 . Since x_1 is the largest element and since $x_1 + 1 > x_1$, $x_1 + 1 \notin \{x_1, \dots, x_n\}$. But because $x_1 + 1 \in \mathbb{N}$, $f(x_1 + 1) \in \{1, \dots, n\}$. Hence $f(x_1 + 1) = k = f(x_k)$ for some $k \in \{1, \dots, n\}$. Since $x_1 + 1$ is not in $\{x_1, \dots, x_n\}$,

$x_1 + 1 \neq x_k$, but we also have $f(x_1 + 1) = f(x_k)$. This is what we wanted to prove. Therefore f is not one-to-one.

Subsequently, there can be no bijections $f : \mathbb{N} \longrightarrow \{1, \dots, n\}$. Since n was arbitrarily chosen (i.e., there is no special property about n), \mathbb{N} is not a finite set.

Let us look at that proof again. A specific n is called for. Suppose that $f : \mathbb{N} \longrightarrow \{1, \dots, 101\}$ is a bijection. Then there are natural numbers $\{x_1, \dots, x_{101}\} \subset \mathbb{N}$ such that $f(x_k) = k$. There is a largest number $x \in \{x_1, \dots, x_{101}\}$. “What made you do that?” you ask. “Writing and rewriting this example enables me to do that,” I respond. It, x , need not be x_{101} . The number x is just the largest number in $\{x_1, \dots, x_{101}\}$. “Why look there, why look for x ? What inspired you to do that?” you ask. “My preparation of this proof is what inspired that choice,” I say. That and a number of years of experience. Because $x < x + 1$ we conclude that $x + 1 \notin \{x_1, \dots, x_{101}\}$ (x is the largest thing in that set). Hence $x + 1 \neq x_k$ for any of the $x_k \in \{x_1, \dots, x_{101}\}$. Meanwhile, however,

$$f(x + 1) \in \{1, \dots, 101\} = \{f(x_1), \dots, f(x_{101})\}$$

so that $f(x + 1) = f(x_k)$ for some $x_k \in \{x_1, \dots, x_{101}\}$. We have shown that there is an x_k such that $x + 1 \neq x_k$ but such that $f(x + 1) = f(x_k)$. These $x + 1$ and x_k are two different numbers in \mathbb{N} that map under f to the same number, and hence f is not one-to-one.

Congratulations! You have just learned *how* mathematicians count and *what* mathematicians count. We count elements in sets via bijections. If the object is not in a set then we cannot count it. Moreover, you have seen how a mathematician does his work. He counts on his experiences, he counts on his mathematical instincts, and he counts on his mathematical inspirations. Do not discount the wonderful consequences of mathematical inspiration.

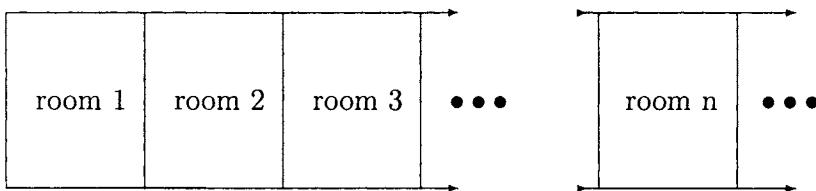
With this, we can define what mathematics means by infinity or by an infinite set.

Definition 3.2.4 A set A is said to be infinite if A is not a finite set.

Of course \mathbb{N} is an infinite set. Since $\mathbb{N} \subset \mathbb{Q}$, and since \mathbb{N} is infinite then \mathbb{Q} is infinite. Since $\mathbb{Q} \subset \mathbb{R}$, \mathbb{R} is infinite.

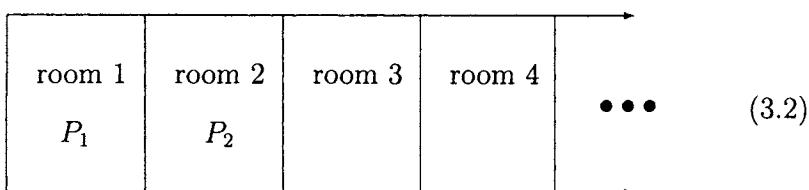
We will attempt to expand your horizons in preparation for a more mathematical treatment of infinite sets. The thought experiment that we will use is called *Hilbert's Infinite Hotel*. Do not try to think of this Hotel as a hammer and nail construction. It exists only in the fertile minds of human beings. If you require your mathematics to be practical, you are about to get your mind significantly stretched.

Hilbert owns a unique hotel. It has just one floor but it has *infinitely many rooms* in it.



The three dots $\bullet \bullet \bullet$ indicate that the room numbers continue in the same manner. Obviously Hilbert's Hotel does not exist in our universe, but in some other universe of imagined mathematical objects.

Initially the Hotel is empty. A set $\{P_1, P_2\}$ of people (people travel in sets in this universe) visits the Hotel and they want different rooms. So Hilbert assigns them rooms as follows.



Hilbert observes that infinitely many rooms $3, 4, 5, \dots$ are vacant.

Next, a set $\{N_1, N_2, N_3, \dots\}$ of Naturalists visits the Hotel. They are in town for a convention protesting the overuse of paradoxes involving infinity in modern mathematics. They all want private rooms. Hilbert hits upon the following scheme for assigning them rooms. Since rooms 1 and 2 are occupied Hilbert looks at picture (3.2) and assigns rooms as follows.

room 1	room 2	room 3	room 4	room 5	$\bullet \bullet \bullet$
P_1	P_2	N_1	N_2	N_3	(3.3)

Notice that in general the n th Naturalist N_n is put up in room $n+2$ in picture (3.3). In this way everyone has a single room and every room is occupied.

Next to arrive at Hilbert's Hotel is a lone stranger L who wants a private room. Hilbert studies picture (3.3) for a while and comes up with the following scheme. Every occupant will move down one room.

$$\begin{array}{ccc} \text{room } 1 & \longleftrightarrow & \text{room } 2 \\ \text{room } 2 & \longleftrightarrow & \text{room } 3 \\ \text{room } 3 & \longleftrightarrow & \text{room } 4 \\ & & \vdots \end{array}$$

The result is summarized by the picture (3.4).

room 1	room 2	room 3	room 4	room 5	$\bullet \bullet \bullet$
	P_1	P_1	N_1	N_2	(3.4)

In this way everyone has a private room. As he anticipated, Hilbert sees that room 1 in picture (3.4) is unoccupied. (There is no room 0 to send to room 1.) Thus Hilbert assigns the lone stranger L to room 1 as in picture (3.5).

room 1	room 2	room 3	room 4	room 5	$\bullet \bullet \bullet$	(3.5)
L_1	P_1	P_1	N_1	N_2		

Everyone has a single room and every room is occupied as in picture (3.5). There are no vacancies at Hilbert's Infinite Hotel.

A troop of 1023 Boy Scouts arrives at the Hotel. After extracting a promise from them that they will not practice the art of making fire from two sticks, Hilbert assigns them rooms in his filled Hotel as follows. First he moves his guests.

$$\begin{aligned} \text{room 1} &\longmapsto \text{room } 1023 + 1 = 1024 \\ \text{room 2} &\longmapsto \text{room } 1023 + 2 = 1025 \\ \text{room 3} &\longmapsto \text{room } 1023 + 3 = 1026 \\ &\vdots \end{aligned}$$

Draw this picture for yourself, reader. This empties rooms 1 through 1023 so he assigns each of the 1023 Scouts a single room. The Hotel is filled to capacity once again. I leave it to the reader to draw the pictures that relate to this assignment of rooms.

The next day everyone checks out. The Hotel is not empty for long, though, as an infinite set $\{S_1, S_2, S_3, \dots\}$ of college students arrive for spring break. Hilbert assigns student S_n to room n . The Hotel is filled as in picture (3.6).

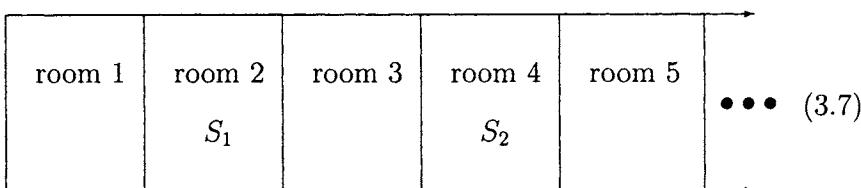
room 1	room 2	room 3	room 4	room 5	$\bullet \bullet \bullet$	(3.6)
S_1	S_2	S_3	S_4	S_5		

Just as Hilbert thinks that the worse of this continual stream of guests is over, an infinite set of Psychologists $\{X_1, X_2, X_3, \dots\}$ arrives at Hilbert's Hotel. They are here for The Conference on the Theory of Group Therapy for Infinite Groups being held in the Hotel's Infinite Conference Room. Hilbert tells his staff to set up infinitely many chairs and a podium for the lectures, so now he has to assign these new guests rooms. Of course, everyone wants a single room. But the Hotel is filled. How can he do this? He decides to make room by making every other room vacant. Try to figure this one out yourself before reading on.

Here's how he does it. The student S_n in room n is sent to room $2n$ as follows.

$$\begin{aligned} S_1 &\longmapsto \text{room } 2 \\ S_2 &\longmapsto \text{room } 4 \\ S_3 &\longmapsto \text{room } 6 \\ &\vdots \\ S_n &\longmapsto \text{room } 2n \\ &\vdots \end{aligned}$$

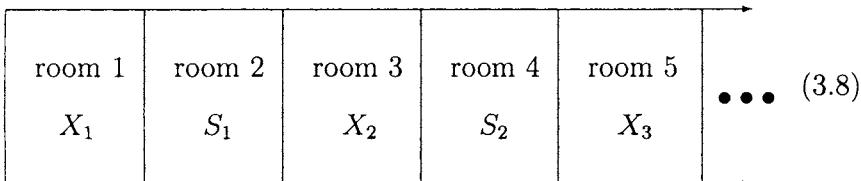
The room assignments in the Hotel look like picture (3.7).



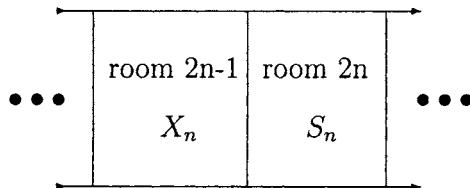
Each of the odd numbered rooms is empty so Hilbert assigns the Psychologists rooms as follows.

$$\begin{aligned} X_1 &\longmapsto \text{the first odd numbered room } 1 \\ X_2 &\longmapsto \text{the second odd numbered room } 3 \\ X_3 &\longmapsto \text{the third odd numbered room } 5 \\ &\vdots \\ X_n &\longmapsto \text{the } n\text{th odd numbered room } 2n - 1 \\ &\vdots \end{aligned}$$

This gives each student a single room and each odd numbered room is occupied as in the following picture.



In general, we are assigning rooms as follows. The rooms $2n - 1$ and $2n$ receive people X_n and S_n as follows.



Notice that the even numbered room $2n$ is occupied by student S_n who had occupied room n . The odd numbered room $2n - 1$ is occupied by Psychologist X_n . Thus each room is occupied by exactly one person. The Hotel is filled once again.

Here are a few exercises on Hilbert's Infinite Hotel.

1. Suppose that the Hotel is filled with infinitely many students S_1, S_2, S_3, \dots . Two infinite groups $\{Y_1, Y_2, Y_3, \dots\}$ and $\{Z_1, Z_2, Z_3, \dots\}$ show up looking for rooms for the night. Show how to accommodate these new guests with single rooms.
2. The Hotel is empty when a group $\{U_1, U_4, U_9, \dots, U_{n^2}, \dots\}$ shows up. Assign them rooms so that the Hotel is filled.

There is an interesting thing going on in these room assignments. Hilbert assigns rooms so that when he is done his Hotel is filled to capacity. That is right. He fills the hotel so that each person has a

room and each room is filled. So far he has been able to do that for any group that arrives, but shortly we will see that there is a large collection of Realists that cannot be housed in Hilbert's Hotel. I will not give details, only the punch line. Hilbert has filled his Hotel with everyone who has arrived so far, but he will be unable to accommodate all of the Realists.

The next day all the guests at Hilbert's Infinite Hotel check out. Suddenly, infinitely many groups of Rationalists show up and require single rooms.

$$\begin{aligned}\mathbf{L}_1 &= \{A_{11}, A_{12}, A_{13}, \dots\} \\ \mathbf{L}_2 &= \{A_{21}, A_{22}, A_{23}, \dots\} \\ \mathbf{L}_3 &= \{A_{31}, A_{32}, A_{33}, \dots\} \\ &\vdots\end{aligned}$$

Notice that

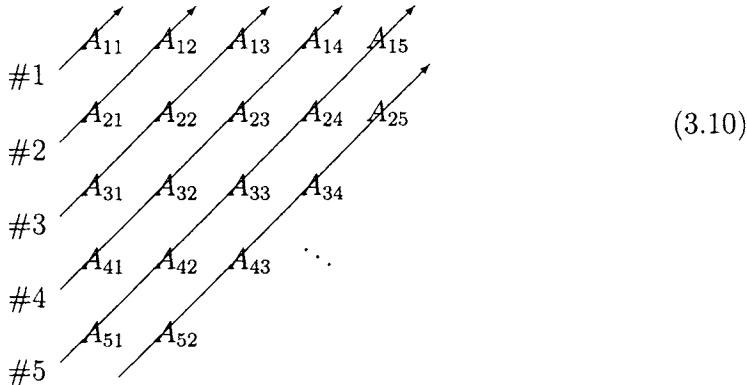
A_{nm}

is Rationalist m in implied list \mathbf{L}_n . Thus A_{12} is Rationalist 2 in the implied list \mathbf{L}_1 , A_{47} is Rationalist 7 in the implied list \mathbf{L}_7 , and $A_{101\text{googol}}$ is Rationalist number googol in the implied list \mathbf{L}_{101} . We will demonstrate how to accommodate these guests so that each room is occupied and each Rationalist has a single room.

Write the names in each implied list \mathbf{L}_n as row n in the infinite rectangular array (3.9).

$$\begin{array}{ccccccc} A_{11} & A_{12} & A_{13} & A_{14} & \dots \\ A_{21} & A_{22} & A_{23} & A_{24} & \dots \\ A_{31} & A_{32} & A_{33} & A_{34} & \dots \\ A_{41} & A_{42} & A_{43} & A_{44} & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{array} \tag{3.9}$$

Draw parallel arrows starting at the elements A_{k1} and ending at the elements A_{1k} as in array (3.10).



These arrows can be used to define a bijection

$$f : \mathbb{N} \longrightarrow A$$

as follows. These arrows define a means of reading the array and assigning rooms to the Rationalists. Hold the page at an angle if you must, but start reading the array along the arrows as though there were words to be read. The first thing you read is A_{11} . The next two things you read are A_{21} and A_{12} . The next things you would read would be A_{31}, A_{22}, A_{13} . The process continues indefinitely. In the more precise language of functions we write

$$\begin{aligned} f(0) &= A_{11} \\ f(1) &= A_{21} \quad f(2) = A_{12} \\ f(3) &= A_{31} \quad f(4) = A_{22} \quad f(5) = A_{13} \\ f(6) &= A_{41} \quad f(7) = A_{32} \quad f(8) = A_{23} \quad f(9) = A_{14} \\ &\dots \end{aligned}$$

Trace these values for f through the first five arrows in the above array. We will show you that this rule f is the promised bijection. Try to plow through this argument. However, if you feel overwhelmed then try to understand that in drawing these arrows we have intersected each of the entries of the array in such a way that each arrow

intersects finitely many entries in the array and each entry is met by some arrow.

We take it as clear that f is one-to-one. Let $A_{nm} \in \mathbf{L}_n$. By the above scheme we see that the subscripts lying along the first arrow add up to $1 + 1 = 2$, the subscripts lying along the second arrow add up to 3, the subscripts lying along arrow #3 add up to 4, and in general the subscripts lying along arrow $\#k$ add up to $k + 1$. Thus A_{nm} lies along arrow $\#n + m - 1$. Now consider the number that we will assign A_{nm} with f .

On arrow #1 there is 1 entry. On arrow #2 there are 2 entries, and in general on arrow $\#n$ there are n entries. Thus, when we have finished counting the entries from A_{11} to A_{1n} , we will have counted a total of

$$\text{from arrow } \#1 \quad \#2 \quad \dots \quad \#\ell \\ 1 + 2 + \dots + \ell = \frac{\ell(\ell + 1)}{2}$$

entries. We will see why this sum is $\frac{\ell(\ell + 1)}{2}$ in a little bit. For now just accept it knowing that there is the promise of a proof. Meanwhile back at the argument, we see that A_{nm} is on arrow $\#n + m - 1$ so that

$$\text{there is some } k \leq \frac{(n + m - 1)(n + m)}{2} \text{ such that } f(k) = A_{nm}.$$

This proves that $f(k) = A_{mn}$ for some number k , and this in turn proves that f is onto. Since we have already stated that f is one-to-one, $f : \mathbb{N} \rightarrow A$ is a bijection. We would then send Rationalist A_{nm} to room k . The assignment of rooms is complete.

Let us take a break at this time and examine the sum $1 + 2 + \dots + \ell$. A German school teacher was feeling a bit hungry so he prepared to go out for coffee and a danish. To keep his pupils busy he instructed them to add up the numbers from 1 to 100. We will do this problem for an arbitrary positive integer ℓ . A child, Carl F. Gauss, decided that there must be a faster way. In a flash of genius the 11 year old Gauss discovered the following identity. Let us call the sum S .

$$S = 1 + 2 + \dots + \ell.$$

Find the sum in two different ways. Start at 1 and end at ℓ , and in the other start at ℓ and end at 1.

$$\begin{aligned} S &= 1 + 2 + \dots + \ell - 1 + \ell \\ S &= \ell + \ell - 1 + \dots + 2 + 1 \end{aligned}$$

Notice that in each column we have a sum

$$\ell + 1.$$

Add these two sums together to find that

$$\begin{array}{rccccccccc} S & = & 1 & + & 2 & + & \dots & + & \ell - 1 & + & \ell \\ +S & = & \ell & + & \ell - 1 & + & \dots & + & 2 & + & 1 \\ \hline 2S & = & (\ell + 1) & + & (\ell + 1) & + & \dots & + & (\ell + 1) & + & (\ell + 1) \end{array}$$

There are ℓ terms $\ell + 1$ so we have $2S = \ell(\ell + 1)$ or equivalently

$$S = 1 + 2 + \dots + \ell = \frac{\ell(\ell + 1)}{2}.$$

Quite a pretty little argument for a child, eh? Before the school teacher could button his coat, Gauss turned his slate over and folded his hands. The teacher, thinking something was amiss, examined the slate and saw the correct sum written there. History does not record whether or not the school teacher got his coffee that day or if he suffered for a lack of it.

The purpose behind the above stories is to develop an intuition for the infinite. The story about Hilbert's Infinite Hotel shows us that in an infinite set there is always *plenty of room*. It would seem that we can make room for any number of guests. Or can we?

The conventions and conferences are over so that the Hotel is empty. Hilbert breathes a sigh of relief and counts his revenue. Accommodating all of those guests has taxed him. There is a knock at the door and Hilbert sees that an infinite collection of Realists

$$\{R_x \mid x \in \mathbb{R} \text{ and } 0 \leq x \leq 1\}$$

is in town for a conference on complex issues. Hilbert sets out to assign each of them a private room, but his first attempt at

assigning rooms leaves infinitely many Realists without a room. He tries again and still he cannot assign every Realist a single room. For some reason he does not yet fathom, no matter how he assigns rooms to the Realists, some Realists do not get a single room. The mathematics behind this conundrum is a jolt to our *common sense* and will have to wait until Theorem 4.3.1. The problem of assigning rooms to Realists shows us that our *intuition* about finite sets is useless when we consider matters about infinite sets. We will have to develop entirely new insights if we are to understand the infinite. Hilbert does not yet have the mathematical experience to address this problem. He sends most of these potential guests packing.

Let us discover what Hilbert did not fathom. Let us see why there is no way to accommodate these realists with single rooms in the Infinite Hotel.

Suppose, for the sake of contradiction, that we can make an implied list of these guests and their single rooms. That is, we assign Realists to single rooms and then we make an implied list of the rooms n and the associated occupants R_{x_n} .

room	occupant	decimal						
1	R_{x_1}	$x_1 = .d_{11} d_{12} d_{13} d_{14} \dots$	d_{11}	d_{12}	d_{13}	d_{14}	\dots	
2	R_{x_2}	$x_2 = .d_{21} d_{22} d_{23} d_{24} \dots$	d_{21}	d_{22}	d_{23}	d_{24}	\dots	
3	R_{x_3}	$x_3 = .d_{31} d_{32} d_{33} d_{34} \dots$	d_{31}	d_{32}	d_{33}	d_{34}	\dots	
4	R_{x_4}	$x_4 = .d_{41} d_{42} d_{43} d_{44} \dots$	d_{41}	d_{42}	d_{43}	d_{44}	\dots	
:	:							

We have also given the decimal expansion of x_n . In this implied list, each d_{ij} is a digit, a number in the set $\{0, 1, \dots, 9\}$, and we choose the shortest decimal expansion for x_n . That is, if the decimal expansion for x_n ends with $99\bar{9}$ then we replace it with the number that ends in 0's. For instance, the following numbers are equal.

$$\begin{aligned} 1.00\bar{0} &= .99\bar{9} \\ .100\bar{0} &= .099\bar{9} \\ .123\bar{0} &= .122999. \end{aligned}$$

In these cases we choose the shorter finite decimal expansion, *except for* 1. The number 1 is written as a decimal $.99\bar{9}$.

Furthermore, d_{11} represents the first digit in the decimal form of x_1 , d_{23} represents digit number 3 in the decimal form of x_2 , and in general d_{nm} represents digit number m in the decimal form of x_n . We can now produce a Realist R_x who does not get a room.

Let us define a digit d_1 as follows.

$$d_1 = \begin{cases} d_{11} + 1 & \text{if } d_{11} \neq 9 \\ 0 & \text{if } d_{11} = 9 \end{cases}.$$

Notice that we have defined d_1 so that $d_1 \neq d_{11}$. For example, if $d_{11} = 5$ then $d_1 = 6$ and if $d_{11} = 9$ then $d_1 = 0$.

We do the same thing for d_{22} . Let

$$d_2 = \begin{cases} d_{22} + 1 & \text{if } d_{22} \neq 9 \\ 0 & \text{if } d_{22} = 9 \end{cases}.$$

Notice that we have defined d_2 so that $d_2 \neq d_{22}$. For example, if $d_{22} = 0$ then $d_2 = 1$ and if $d_{22} = 7$ then $d_2 = 8$.

In general, let us define

$$d_n = \begin{cases} d_{nn} + 1 & \text{if } d_{nn} \neq 9 \\ 0 & \text{if } d_{nn} = 9 \end{cases}.$$

Then $d_n \neq d_{nn}$.

The roomless Realist R_x is given by writing down the decimal expansion

$$x = .d_1d_2d_3d_4\dots$$

If x ends in $99\bar{9}$ then we replace it with the shorter decimal representation that ends in $00\bar{0}$. Then $x \neq x_1$ because by our choice of d_1 , x and x_1 differ in the first decimal place, $d_1 \neq d_{11}$. Also, $x \neq x_2$ because by our choice of d_2 , x and x_2 differ in decimal place number 2, $d_2 \neq d_{22}$. In general, $x \neq x_n$ because by our choice of d_n , x and x_n differ in decimal place number n , $d_n \neq d_{nn}$. Thus R_x is not on the implied list of Realists who got a room. This is the person we have sought. This is why Hilbert could not give each Realist a single room. There are simply too many Realists! No matter what rooms

he assigned there would always be a Realist R_x who does not get a room. There are more Realists than there are rooms.

How can that be, reader? There are infinitely many rooms and infinitely many Realists. How can there be more Realists than rooms? Read on.

Let's examine the above choices for d_n and the construction of x on a concrete example. Suppose that we are given specific real numbers x_1, x_2, x_3, \dots that form the following implied list.

$x_1 = .$	$\boxed{1}$	2	3	4	5	\dots
$x_2 = .$	0	$\boxed{9}$	0	1	0	\dots
$x_3 = .$	1	4	$\boxed{9}$	5	7	\dots
$x_4 = .$	4	1	4	$\boxed{2}$	4	\dots
$x_5 = .$	0	1	0	1	$\boxed{3}$	\dots
						\vdots

Then

$$x = .2 \ 0 \ 0 \ 3 \ 4 \ \dots$$

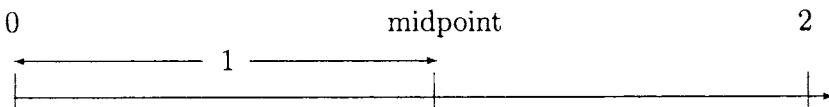
The first decimal of x is 2 because the first decimal of x_1 is 1.
 The second decimal of x is 0 because the second decimal of x_2 is 9.
 The second decimal of x is 0 because the third decimal of x_3 is 9.
 The fourth decimal of x is 3 because the fourth decimal of x_4 is 2.
 The fifth decimal of x is 4 because the fifth decimal of x_5 is 3.

The comparisons of x to the numbers on our implied list are obvious.

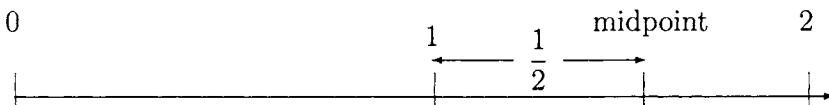
$x \neq x_1$ because they differ in decimal place 1,
 $x \neq x_2$ because they differ in decimal place 2,
 $x \neq x_3$ because they differ in decimal place 3.

These comparisons continue indefinitely. In this way we show that x is not on the implied list x_1, x_2, x_3, \dots .

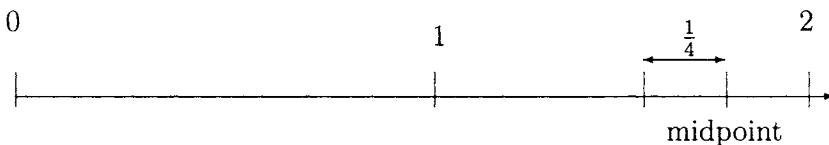
There is a person at the Hotel who haunts the Lobby. Each day he tries to walk a 2 meter long straight line across the Lobby floor. He starts at one end of the line and wants to get to the other end, but he has a peculiar way of walking. His next step is always half the length of his previous step. He starts with a step of one meter. One meter remains to be crossed.



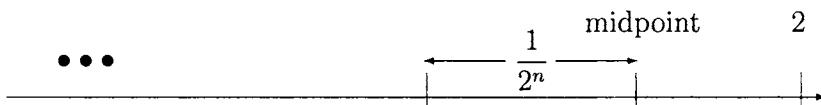
He follows this with a $\frac{1}{2}$ meter step. Half a meter remains.



Then he steps off a $\frac{1}{4} = \frac{1}{2^2}$ meter step. A $\frac{1}{4}$ meter remains.



Continuing on, on his n th step he moves $\frac{1}{2^n}$ meters, and that leaves $\frac{1}{2^n}$ meters left to be covered. The scale is magnified.



He continues on indefinitely. Upon completion there are 0 meters left to cover and so he has crossed

$$2 = 1 + \frac{1}{2} + \frac{1}{4} + \dots + \frac{1}{2^n} + \dots$$

meters. He has walked the entire 2 meter length.

Each of the rooms at Hilbert's Infinite Hotel has to be cleaned and there is one and only one cleaning lady to do the job. Her name is Mary. (Hilbert's mother needed a job.) Her routine is to

clean the rooms in order. She begins in the morning with room 1, followed by room 2, and then room 3, and on inductively. She is an industrious woman who picks up momentum as her day progresses so she cleans the first room in 1 hour, the second room in $\frac{1}{2}$ hour,

the next in $\frac{1}{2^2}$ hour, and so on, cleaning room n in $\frac{1}{2^n}$ hours. When she has cleaned each room she is done. Thus she works

$$1 + \frac{1}{2} + \frac{1}{4} + \dots + \frac{1}{2^n} + \dots = 2$$

hours a day. She certainly has a good job. She only has a 2 hour work day. But something else is wrong here. Where is she at the end of her day? There is no last room for her to clean because the Hotel has *infinitely many rooms*. However, we can be assured that each of the rooms is spotless. We will learn the fate of the cleaning lady when we finish our discussion of ordinals on page 184.

3.3 Equivalent Sets and Cardinality

In this section we will define the equivalence of cardinalities.

Definition 3.3.1 Sets A and B are said to be equivalent if there is a bijection $f : A \longrightarrow B$.

Theorem 3.3.2 A and B are equivalent exactly when $\text{card}(A) = \text{card}(B)$.

If we look at the different steps in Hilbert's Infinite Hotel we see that in making space for the lone stranger Hilbert used the fact that the function

$$f : \{1, 2, 3, \dots\} \longrightarrow \{2, 3, 4, \dots\}$$

defined by

$$f(n) = n + 1$$

defines a bijection. Thus $\{1, 2, 3, \dots\}$ is equivalent to the proper subset $\{2, 3, 4, \dots\}$.

In moving the students $\{S_1, S_2, S_3, \dots\}$ into new rooms to accommodate all of the Psychologists $\{X_1, X_2, X_3, \dots\}$, we are making use of the bijection

$$f : \{1, 2, 3, \dots\} \longrightarrow \{2, 4, 6, \dots\}$$

given by

$$f(n) = 2n.$$

Thus $\{1, 2, 3, \dots\}$ is equivalent to the proper subset $\{2, 4, 6, \dots\}$.

One of the exercises asks you to show a result due to Galileo Galilei. Galileo observed that *there are no more and no fewer perfect squares than there are natural numbers*. In our language Galileo was observing that there is a bijection

$$f : \{1, 2, 3, \dots\} \longrightarrow \{1, 4, 9, \dots\}$$

given by

$$f(n) = n^2.$$

Thus $\{1, 2, 3, \dots\}$ is equivalent to its proper subset $\{1, 4, 9, \dots\}$ of perfect squares. This property turns out to be a characteristic of infinite sets that can be used to define infinite sets without referring to finite sets.

Theorem 3.3.3 *The set A is infinite exactly when A is equivalent to a proper subset of itself.*

For example, by the functions f that we discussed above, \mathbb{N} is an infinite set.

Example 3.3.4 Let us show that \mathbb{Z} is equivalent to \mathbb{N} . This may seem strange. \mathbb{N} is unbounded to the right while \mathbb{Z} is unbounded in both directions. A bijection between them exists nonetheless.

Define a rule $f : \mathbb{Z} \longrightarrow \mathbb{N}$ by writing

$$f(z) = \begin{cases} 2z & \text{if } z \geq 0 \\ -1 - 2z & \text{if } z < 0 \end{cases}$$

for all $z \in \mathbb{Z}$. (“Now where did that come from?” you ask. “From a great deal of preparation,” I respond.) Let us get a feel for the rule given here. For $z \geq 0$ we have

$$f(0) = 0, f(1) = 2, f(2) = 4,$$

and in general for $n \in \mathbb{N}$ the n th positive number n is sent to the n th even number:

$$f(n) = 2n.$$

The negative z are a different matter.

$$f(-1) = -1 - 2(-1) = 1$$

$$f(-2) = -1 - 2(-2) = 3$$

$$f(-3) = -1 - 2(-3) = 5$$

and in general for $n \in \mathbb{N}$ the n th negative number $-n$ is sent to the n th odd number:

$$f(-n) = 2n - 1.$$

We casually observe that f is onto since a natural number is either even or odd. Furthermore, f is one-to-one. This will require us to look at a few cases. Let $z \neq z' \in \mathbb{Z}$.

1. If $z \neq z' \geq 0$ then $f(z) = 2z \neq 2z' = f(z')$.
2. If $z \neq z' < 0$ then $f(z) = -1 - 2z \neq -1 - 2z' = f(z')$.
3. In the last case, we can assume without loss of generality that $z < 0 < z'$. Then $f(z)$ is odd and $f(z')$ is even so that $f(z) \neq f(z')$.

In any case, $f(z) \neq f(z')$ so that f is one-to-one. Hence f is a bijection and consequently \mathbb{Z} is equivalent to \mathbb{N} . Subsequently, since \mathbb{Z} is equivalent to its proper subset \mathbb{N} , \mathbb{Z} is infinite.

Example 3.3.5 We will show that \mathbb{R} is equivalent to the proper subset \mathbb{R}^+ of positive real numbers, thus proving that \mathbb{R} is infinite. This may seem obvious, but remember that every statement in mathematics must be accompanied by a justification or proof. Besides, the exercise prepares us for more complex arguments later.

The bijection $f : \mathbb{R} \longrightarrow \mathbb{R}^+$ is given by

$$f(x) = 2^x.$$

The argument requires a little high school algebra involving the logarithm $\log_2(x)$. All you have to know about $\log_2(x)$ is that

$$\log_2(2^x) = x = 2^{\log_2(x)}.$$

To see that f is one-to-one, suppose that $f(x) = f(x')$. Then

$$2^x = 2^{x'}.$$

Apply \log_2 to each side of this equation to find that

$$\log_2(2^x) = \log_2(2^{x'}),$$

so that $x = x'$. Thus $f : \mathbb{R} \longrightarrow \mathbb{R}^+$ is one-to-one.

To see that f is onto let $y \in \mathbb{R}^+$. Then $x = \log_2(y)$ is a number such that

$$f(x) = 2^x = 2^{\log_2(y)} = y.$$

Thus $f : \mathbb{R} \longrightarrow \mathbb{R}^+$ is onto, hence f is a bijection, whence \mathbb{R} is equivalent to \mathbb{R}^+ .

Example 3.3.6 We will show in Theorem 4.3.1 that the function

$$\tan(\theta) : \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \longrightarrow \mathbb{R}$$

is a bijection. Then \mathbb{R} is equivalent to its proper subset $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$, hence \mathbb{R} is infinite, and therefore $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ is infinite. The algebraically educated reader should try this as an exercise.

We present two classic examples due to G. Cantor, which will demonstrate that our intuition about infinity is inaccurate. These examples will present ideas that are contrary to what we learned as children, if indeed we learned anything about infinity as children. We begin with an alternative approach to cardinality.

Let A be an infinite set. The cardinality of A can be defined as follows.

$$\text{card}(A) = \text{the sets } B \text{ that are equivalent to } A.$$

In this case we call $\text{card}(A)$ a *cardinal number* or a *cardinal*. Thus x is a cardinal exactly when there is a set A such that $x = \text{card}(A)$. Although this appears to be different from the original definition of cardinal number, it defines the same thing.

For example, $\text{card}(\emptyset)$ is the collection of sets that have no elements. Naturally \emptyset is the only such set, so we will use *the symbol* 0 to denote $\text{card}(\emptyset)$.

$$0 = \text{card}(\emptyset).$$

Since any bijection

$$f : \{\bullet\} \longrightarrow X$$

is onto, we can write

$$X = \{f(\bullet)\}$$

or equivalently

$$X = \{x\} \text{ for some element } x.$$

Thus it is natural to use *the symbol* 1 to denote $\text{card}\{\bullet\}$.

$$1 = \text{card}\{\bullet\}.$$

A tradition started by Georg Cantor [1] is to use the first Hebrew letter \aleph to denote *small infinite cardinals*. Thus we let

$$\aleph_0 = \text{card}(\mathbb{N}).$$

Chapter 4

Infinite Cardinals

As we defined in the previous chapter the cardinality of a set A , $\text{card}(A)$, is the class of all sets B that are equivalent to A . For a finite set $A = \{1, \dots, n\}$, $\text{card}(A)$ is the collection of sets that contain exactly n elements. Thus $\text{card}(A)$ is a good way to describe the size of A , *even when A is infinite*. We might naively say that if A and B are infinite sets then $\text{card}(A) = \text{card}(B)$. However, we showed in Chapter 3 that Hilbert's Infinite Hotel, whose rooms are numbered $1, 2, 3, \dots$ cannot fit the Realists whose group is indexed by real numbers $(0, 1) = \{x \in \mathbb{R} \mid 0 < x < 1\}$. In another light, there is no bijection from the set $(0, 1)$ onto the set of natural numbers \mathbb{N} . If such a bijection does not exist then we must conclude intuitively and accurately that

$\text{card}(\mathbb{N})$ and $\text{card}(0, 1)$ are different cardinalities.

This should come as a shock to your common sense. Both cardinals are infinite so how can they be different values? The inequality $\text{card}(\mathbb{N}) \neq \text{card}((0, 1))$ is true enough, so we have learned that our common sense is wrong when we deal with infinity. This is an important point. We do not have any common sense when it comes to infinite sets. We have to develop a new sense for size by relying on the mathematical properties of infinite sets. Once this is done, we can find some new intuitive feelings about infinity.

We will find in this chapter that there are two kinds of infinite sets: the *countable* ones and the *uncountable* ones. Any set equiv-

alent to \mathbb{N} is said to be *countable* (hence the name countable) so we say that $\text{card}(\mathbb{N})$ is a countable cardinal. If S is an infinite set and if $\text{card}(\mathbb{N}) \neq \text{card}(S)$ then S is said to be *uncountable*. We will show that $\text{card}(\mathbb{N}) \neq \text{card}((0, 1))$ so the set $(0, 1)$ is uncountable. Other uncountable sets include $\mathcal{P}(\mathbb{N})$ and $\text{card}(\{0, 1\}^{\mathbb{N}})$ so there are several uncountable sets. We end this chapter with the amazing fact that there is *an infinite chain of infinite cardinals*. We might colloquially rephrase this by saying that *there are infinitely many infinities*.

4.1 Countable Sets

A set A is said to be *countable* if A is finite or if $\text{card}(A) = \text{card}(\mathbb{N})$. Equivalently, A is countable if A is equivalent to some subset of \mathbb{N} . Countable sets are appropriately named since these are the sets A that we can list as the possibly finite sets:

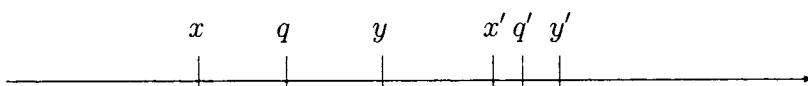
$$A = \{a_1, a_2, a_3, \dots\}.$$

Thus we can use subsets of \mathbb{N} to list the elements of A . The sets \emptyset , $\{1, 2, 3\}$, and $\{2, 4, 6, \dots\}$ are countable. There are plenty of countable sets and we will give several examples of them in this section.

Let

$$\mathbb{Q}^+ = \left\{ \frac{n}{m} \mid n, m > 0 \text{ and } n, m \in \mathbb{Z} \right\}$$

be the set of *positive rational numbers*. It is a fact (not proved here) that between any two positive real numbers $x < y$ there is a positive rational number q . If you pick two different real numbers $x' < y'$, then there is another rational number q' between them. Let's draw that picture.



Thus between 0 and $\frac{1}{\pi^2}$ there is a positive rational number. You can check that one such number is $\frac{1}{4^2}$. Compare this to the fact that the distance between two different *natural* numbers is at least 1. We say that \mathbb{Q}^+ is a *dense subset* of \mathbb{R}^+ while \mathbb{N} is a *discrete subset* of \mathbb{R}^+ . Another way to think of this property of \mathbb{Q}^+ is that each real number can be approximated by a rational number to any degree of accuracy. We are familiar with the notion of the density of the rationals since each calculator readout demonstrates that each real number x can be approximated by a number with a finite decimal expansion. Finite decimal numbers are rational numbers. For example, the readout on your calculator for $\sqrt{2}$ is not completely accurate even though it may give 64 decimal places of $\sqrt{2}$. The calculator value for $\sqrt{2}$ is accurate only to those few decimal places. The reason for this is that the number $\sqrt{2}$ is neither a finite decimal nor a repeating decimal. An infinite number of decimal places is necessary to write down $\sqrt{2}$ with complete accuracy, but we can approximate $\sqrt{2}$ with rational numbers.

We might be led to believe that \mathbb{N} and \mathbb{Q}^+ are not equivalent since \mathbb{N} is more thinly distributed in \mathbb{R}^+ than is \mathbb{Q}^+ . That is, we might suspect that $\text{card}(\mathbb{N}) \neq \text{card}(\mathbb{Q}^+)$. However, the next theorem, due to Georg Cantor, shows us that \mathbb{Q}^+ is a *countable* set. Thus the distribution or density of numbers on the real line does not necessarily reflect cardinality.

Theorem 4.1.1 \mathbb{Q}^+ is a countable set.

Proof: We must produce a bijection $f : \mathbb{N} \longrightarrow \mathbb{Q}^+$. Begin by enumerating \mathbb{Q}^+ as the infinite rectangular array (4.1).

$\frac{1}{1}$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{4}$	\dots
$\frac{1}{2}$	$\frac{2}{2}$	$\frac{2}{3}$	$\frac{2}{4}$	\dots
$\frac{2}{1}$	$\frac{2}{2}$	$\frac{3}{3}$	$\frac{4}{4}$	\dots
$\frac{1}{3}$	$\frac{3}{2}$	$\frac{3}{3}$	$\frac{3}{4}$	\dots
$\frac{3}{1}$	$\frac{2}{2}$	$\frac{3}{3}$	$\frac{4}{4}$	\dots
$\frac{4}{1}$	$\frac{4}{2}$	$\frac{4}{3}$	$\frac{4}{4}$	\dots
\vdots	\vdots	\vdots	\vdots	

(4.1)

Notice that the numbers in Row 1 have numerator 1, the numbers in Row 2 have numerator 2, and in general the numbers in Row n have numerator n . Similarly, the denominator of the fractions in Column 1 is 1, the denominator of the fractions in Column 2 is 2, and in general the fractions in Column m have denominator m . Since each number in \mathbb{Q}^+ is a fraction $\frac{n}{m}$, the array (4.1) contains all of the numbers in \mathbb{Q}^+ . (Which Row and Column contain that last fraction?)

Delete from this array any fractions that are not in reduced form. For example, we will delete the fractions

$$\frac{3}{3}, \frac{4}{2}, \frac{9}{27}, \frac{212}{424}$$

as well as any other fraction that can be reduced. The remaining fractions cannot be reduced. Some of those fractions include

$$\frac{1}{1}, \frac{1}{2}, \frac{19}{27}, \frac{101}{53}.$$

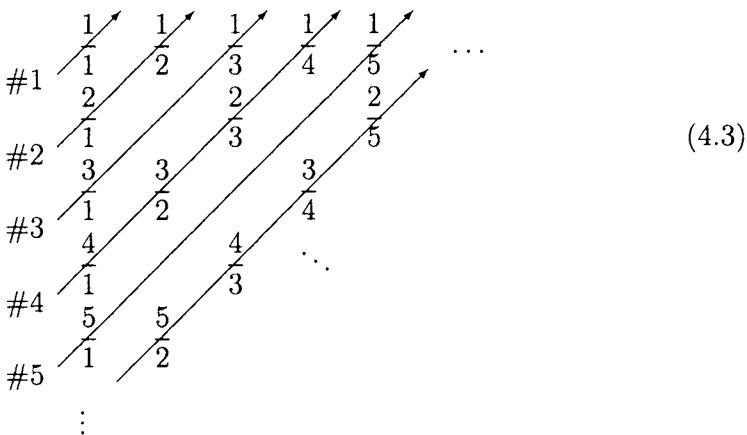
A portion of the resulting array is

$$\begin{array}{ccccccc}
 \frac{1}{1} & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \cdots & & \\
 \frac{2}{1} & & \frac{2}{3} & & \cdots & & \\
 \frac{1}{1} & & & \frac{3}{4} & & \cdots & \\
 \frac{3}{1} & \frac{3}{2} & & \frac{3}{4} & & \cdots & \\
 \frac{1}{1} & & & & & & \\
 \frac{4}{1} & & \frac{4}{3} & & \cdots & & \\
 \frac{1}{1} & & & & & & \\
 \vdots & \vdots & \vdots & \vdots & & &
 \end{array} \tag{4.2}$$

At this point we introduce arrows into array (4.2). The arrow #1 passes through $\frac{1}{1}$ only. The arrow #2 starts at $\frac{2}{1}$ and extends to $\frac{1}{2}$.

In general, arrow $\#n$ starts at $\frac{n}{1}$ on the left edge of the array and extends to $\frac{1}{n}$ at the top edge of the array. The array (4.3) will help

you envision this process.



These arrows give us a means of writing down a bijection $f : \mathbb{N} \longrightarrow \mathbb{Q}^+$ as follows. We will define f by reading the fractions as they appear along the arrows.

$$\begin{aligned}
 \text{LIST: } f(0) &= \frac{1}{1}, \\
 f(1) &= \frac{2}{1}, \quad f(2) = \frac{1}{2}, \\
 f(3) &= \frac{3}{1}, \quad f(4) = \frac{1}{3}, \\
 f(5) &= \frac{4}{1}, \quad f(6) = \frac{3}{2}, \quad f(7) = \frac{2}{3}, \quad f(8) = \frac{1}{4}, \\
 &\vdots
 \end{aligned}$$

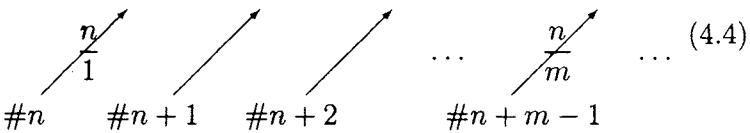
Before reading on, try to use arrows #5 and #6 to fill in the next two rows of this rule.

You can peek now. The next three arrows will produce the values

$$\begin{aligned}
 f(9) &= \frac{5}{1}, \quad f(10) = \frac{1}{5}, \\
 f(11) &= \frac{6}{1}, \quad f(12) = \frac{5}{2}, \quad f(13) = \frac{4}{3}, \quad f(14) = \frac{3}{4}, \quad f(15) = \frac{2}{5}, \quad f(16) = \frac{1}{6}, \\
 f(17) &= \frac{7}{1}, \quad f(18) = \frac{5}{3}, \quad f(19) = \frac{3}{5}, \quad f(20) = \frac{1}{7}
 \end{aligned}$$

for the function f . Now that we have some intuition about the rule f , our next task is to show that f is a bijection.

To see that f is onto let $\frac{n}{m}$ be on the array. Then $\frac{n}{m}$ is in Row n and Column m and the number $\frac{n}{1}$ begins the n th arrow in our array as in the picture (4.4).



By shifting to the right one place, we shift to the next arrow. It is clear from the diagram that $\frac{n}{m}$ is on arrow $\#n + m$. How many numbers have we counted when we finish arrow $\#n + m$? On arrow #1 we counted 1 number, on arrow #2 we counted 2 more, and in general on arrow $\#\ell$ we counted *at most* ℓ fractions. Thus when we have finished counting along arrow $\#n + m$ we have counted *at most* a total of

$$1 + 2 + \dots + (n + m - 1) = \frac{(n + m - 1)(n + m)}{2}$$

fractions. You may remember this sum as the identity found by an eighteenth century elementary school student Carl F. Gauss. See page 92. Since sometime before we reach the end of arrow $\#n+m-1$ we will reach $\frac{n}{m}$, there will be some number

$$k \leq \frac{(n + m - 1)(n + m)}{2}$$

such that

$$f(k) = \frac{n}{m}.$$

For example, before we reach the end of arrow #7, we will count at most

$$1 + 2 + \dots + 7 = \frac{7(7 + 1)}{2} = 28$$

fractions. Thus there is some number $k \leq 28$ such that $f(k) = \frac{3}{5}$.

The reader will find k by reading Row 7 of the above **LIST** that defines f that $k = 19$. That is, $f(19) = \frac{3}{5}$. Thus f is onto.

It is clear that different natural numbers k and k' are sent to different fractions $f(k)$ and $f(k')$ on the **LIST**, so f is one-to-one. Hence f is a bijection, and as required to complete our task, \mathbb{N} is equivalent to \mathbb{Q}^+ .

The inclusions

$$\{2, 4, 6, \dots\} \subset \mathbb{N} \subset \mathbb{Q}^+$$

can be used to intuitively deduce that

$$\text{card}(\{2, 4, 6, \dots\}) \leq \text{card}(\mathbb{N}) \leq \text{card}(\mathbb{Q}^+).$$

But the compelling work that we have done to date allows us to deduce the stronger thought

$$\text{card}(\{2, 4, 6, \dots\}) = \text{card}(\mathbb{N}) = \text{card}(\mathbb{Q}^+).$$

Some might try to say that because \mathbb{N} and \mathbb{Q}^+ are infinite they must be equivalent. This represents old common sense so it has to be rejected as an argument. Cardinalities are equal only when we can produce a bijection between two of their elements. (Remember that $\text{card}(A)$ is the set of all sets B that are equivalent to A .) Some of our subsequent work will show why that kind of reasoning is not a part of mathematics.

The next result shows us that *the countable union of countable sets is countable*. That is, we will start with a family of sets

$$A_1, A_2, A_3, \dots$$

such that each A_k is itself a countable set. Thus we can take set A_1 and list it as

$$A_1 = \{a_{11}, a_{12}, a_{13}, \dots\},$$

where a_{12} represents element 2 in set A_1 , and where $a_{1,11}$ is the 11th element in set A_1 . We will also list A_2 as

$$A_2 = \{a_{21}, a_{22}, a_{23}, \dots\}.$$

Notice that the first subscript gives the set A_2 that the element is in, and the second subscript gives the place that the element takes in the implied list A_2 . Thus $a_{2,15}$ is element 15 in the set A_2 . In general, we will list A_n as

$$A_n = \{a_{n1}, a_{n2}, a_{n3}, \dots\}.$$

For instance, in the implied list A_{121} , the 200th element on the implied list is $a_{121,200}$. The next result shows that if we have a countable implied list A_1, A_2, A_3, \dots of countable sets, then when we put them together using the union operation we will have a countable set.

Theorem 4.1.2 *Let A_1, A_2, A_3, \dots be an implied list of countable sets. Then $\bigcup_{n \in \mathbb{N}} A_n$ is a countable set.*

Proof: We have used this argument several times before.

Since each set A_n is a countable set we can make an implied list of their elements as we did in the discussion preceding this theorem.

$$\begin{aligned} A_1 &= \{a_{11}, a_{12}, a_{13}, a_{14}, \dots\}, \\ A_2 &= \{a_{21}, a_{22}, a_{23}, a_{24}, \dots\}, \\ A_3 &= \{a_{31}, a_{32}, a_{33}, a_{34}, \dots\}, \\ A_4 &= \{a_{41}, a_{42}, a_{43}, a_{44}, \dots\}, \\ &\vdots \end{aligned}$$

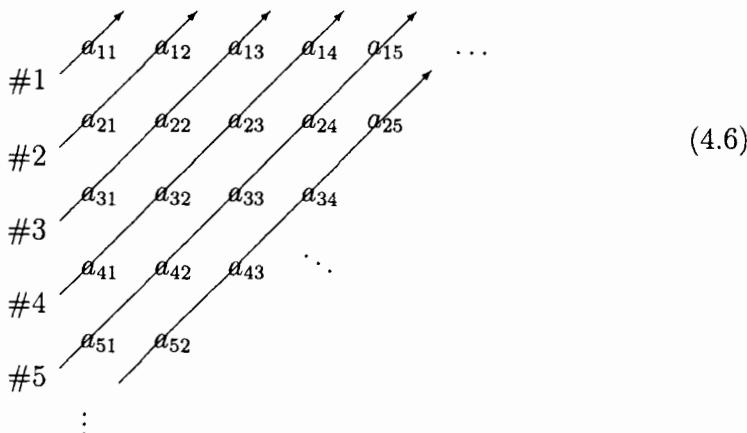
To make the notation easier, let $A = \bigcup_{n \in \mathbb{N}} A_n$ and list the elements

of A as in array (4.5).

$$\begin{array}{ccccccc}
 a_{11} & a_{12} & a_{13} & a_{14} & \dots \\
 a_{21} & a_{22} & a_{23} & a_{24} & \dots \\
 a_{31} & a_{32} & a_{33} & a_{34} & \dots \\
 a_{41} & a_{42} & a_{43} & a_{44} & \dots \\
 \vdots & \vdots & \vdots & \vdots & \vdots
 \end{array} \tag{4.5}$$

We are going to use arrows to count the elements in the array (4.5). Your instincts might say that we should count left to right, the way we read. But that approach leads us to exhaust the counting numbers \mathbb{N} before we have left the first row. No, the best way is to use these diagonal arrows because *each of them is finite* and each a_{ij} lies on some arrow. Thus we are assured that each element will be counted.

Draw parallel line segments starting at the elements a_{k1} and ending at the elements a_{1k} as in the array (4.6).



Use these arrows to define a bijection

$$f : \mathbb{N} \longrightarrow A$$

whose images $f(n)$ satisfy the following rule:

$$\begin{aligned} f(0) &= a_{11}, \\ f(1) &= a_{21}, \quad f(2) = a_{12}, \\ f(3) &= a_{31}, \quad f(4) = a_{22}, \quad f(5) = a_{13}, \\ f(6) &= a_{41}, \quad f(7) = a_{32}, \quad f(8) = a_{23}, \quad f(9) = a_{14}, \\ f(10) &= a_{51}, \quad \dots \end{aligned}$$

It is clear that f is one-to-one. Let $a_{mn} \in A$. By the above scheme we see that the subscripts lying along arrow #1 add up to $1+1=2$, the subscripts lying along arrow #2 add up to 3, the subscripts lying along arrow #3 add up to 4, and in general the subscripts lying along arrow $\#l$ add up to $\ell+1$. Thus a_{mn} lies along arrow $\#m+n-1$, which proves that $f(k) = a_{mn}$ for some number k . In fact, using the argument used to prove Theorem 4.1.1, we could show that

$$k \leq 1 + 2 + \dots + (n+m-1) = \frac{(n+m-1)(n+m)}{2}.$$

Try to do that, reader. This proves that f is onto and hence that $f : \mathbb{N} \longrightarrow A$ is a bijection. Consequently, A is equivalent to \mathbb{N} and therefore A is a countable set. This completes the proof.

The next result shows that if we increase \mathbb{N} by finitely many elements then we have not increased the cardinality of \mathbb{N} .

Theorem 4.1.3 *Let $n \in \mathbb{N}$ and suppose that $\{a_1, \dots, a_n\}$ is a set. Then $\text{card}(\mathbb{N} \cup \{a_1, \dots, a_n\})$ is countable.*

Proof: The sets \mathbb{N} and $\{a_1, \dots, a_n\}$ are countable, so by Theorem 4.1.2 their union $\mathbb{N} \cup \{a_1, \dots, a_n\}$ is also countable. This completes the proof.

Theorem 4.1.2 can be used to give a quick proof that $\text{card}(\mathbb{N}) = \text{card}(\mathbb{Q})$.

Theorem 4.1.4 \mathbb{Q} is countable.

Proof: Observe that $\mathbb{Q} = \mathbb{Q}^+ \cup \{0\} \cup \mathbb{Q}^-$, where \mathbb{Q}^- is the set of negative rational numbers. We will prove that \mathbb{Q}^- is countable.

Define a function $f : \mathbb{Q}^+ \longrightarrow \mathbb{Q}^-$ as

$$f\left(\frac{n}{m}\right) = -\frac{n}{m}.$$

That is, the function $f(x)$ sends a rational number x to its additive inverse $-x$. Thus $f(1) = -1$ and $f(9) = -9$. The function itself is not complex but we will find it to be very useful. We will show that this f is a bijection.

To see that f is one-to-one let $x \neq x'$ be different numbers in \mathbb{Q}^+ . It is then clear that $-x = -x'$ so that $f(x) \neq f(x')$. Hence different fractions x and x' map to different fractions $f(x)$ and $f(x')$, whence f is one-to-one.

Prove that f is onto as follows. Let $y \in \mathbb{Q}^-$ be a negative fraction. Then $x = -y$ is a positive fraction, now isn't it? By forming $-y$ you are simply removing the negative sign from y . Try it with $-\frac{1}{2}$ and $-\frac{5}{3}$. Then $x \in \mathbb{Q}^+$ and

$$f(x) = -x = -(-y) = y.$$

Hence f is an onto function.

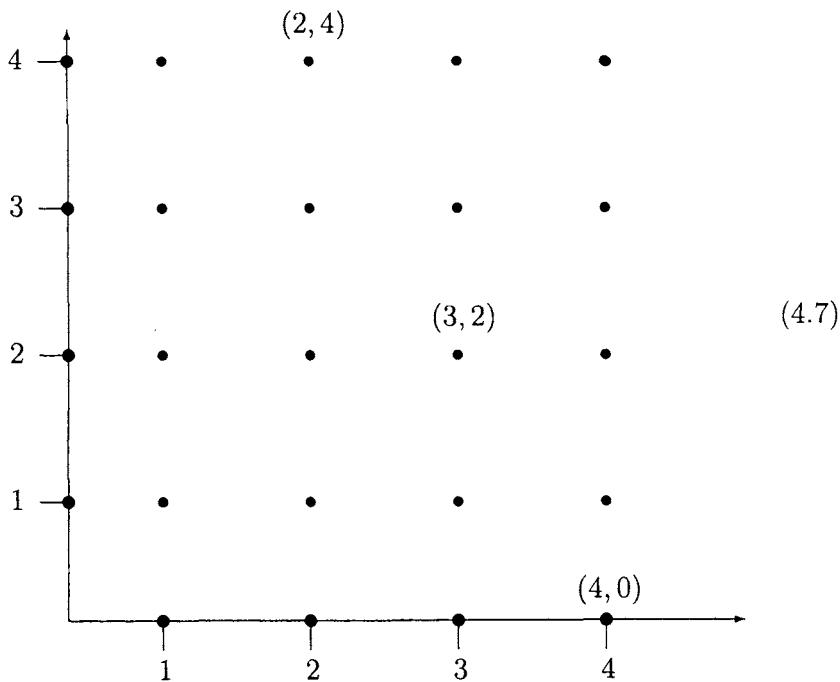
As we claimed at the onset, $f : \mathbb{Q}^+ \longrightarrow \mathbb{Q}^-$ is a bijection.

By Theorem 4.1.1, \mathbb{Q}^+ is countable, so \mathbb{Q}^- is countable. . Thus $\mathbb{Q}^+, \{0\}$, and \mathbb{Q}^- are countable sets, and hence \mathbb{Q} is a union of three countable sets. Theorem 4.1.2 then implies that \mathbb{Q} is countable. This completes the proof.

Summarizing the above proof we showed that \mathbb{Q}^+ and \mathbb{Q}^- have the same cardinality because there is a bijection $f : \mathbb{Q}^+ \longrightarrow \mathbb{Q}^-$. The bijection f is defined by $f(x) = -x$ for each $x \in \mathbb{Q}^+$. That is there is a way to count off the elements of \mathbb{Q}^+ and \mathbb{Q}^- so that when we are done with the elements of \mathbb{Q}^+ we are also done with the elements of \mathbb{Q}^- . One element of \mathbb{Q}^+ corresponds to one element of \mathbb{Q}^- and one element of \mathbb{Q}^- corresponds to one and only one element of \mathbb{Q}^+ . This is what makes a bijection. This is why $\text{card}(\mathbb{Q}^+) = \text{card}(\mathbb{Q}^-)$.

The next result shows us how to use Theorem 4.1.2 to show that a set is countable. Consider the set of pairs $\mathbb{N} \times \mathbb{N} = \{(n, m) \mid n, m \in \mathbb{N}\}$. Since $\mathbb{N} \subset \mathbb{R}$ we can graph $\mathbb{N} \times \mathbb{N}$ as points in the plane \mathbb{R}^2 as in figure (4.7).

The pairs $(2, 4)$ and $(3, 2)$ label a couple of randomly selected points. This pleasant geometric object of evenly spaced points in the plane is called a *lattice*. The lattice extends indefinitely in all directions. Given the theme of this chapter it is natural to ask for the cardinality of the lattice.



Theorem 4.1.5 $\mathbb{N} \times \mathbb{N}$ is a countable set.

Proof: Arrange $\mathbb{N} \times \mathbb{N}$ as an infinite rectangular array (4.8).

$$\begin{array}{ccccccc}
 (0, 0) & (0, 1) & (0, 2) & (0, 3) & \dots \\
 (1, 0) & (1, 1) & (1, 2) & (1, 3) & \dots \\
 (2, 0) & (2, 1) & (2, 2) & (2, 3) & \dots \\
 (3, 0) & (3, 1) & (3, 2) & (3, 3) & \dots \\
 \vdots & \vdots & \vdots & \vdots & \vdots
 \end{array} \tag{4.8}$$

We will avoid the arrows used in Theorems 4.1.1 and 4.1.2. The rows of the array (4.8) are actually countable sets A_0, A_1, A_2, \dots , and there are countably many of them.

$$\begin{aligned}
 A_0 &= \{(0,0), (0,1), (0,2), (0,3), \dots\} \\
 A_1 &= \{(1,0), (1,1), (1,2), (1,3), \dots\} \\
 A_2 &= \{(2,0), (2,1), (2,2), (2,3), \dots\} \\
 A_3 &= \{(3,0), (3,1), (3,2), (3,3), \dots\} \\
 &\vdots
 \end{aligned} \tag{4.9}$$

In general,

$$A_n = \{(n,0), (n,1), (n,2), (n,3), \dots\}$$

so that (n,m) is the m th element on the implied list A_n . This holds for each $n, m \in \mathbb{N}$ so that

$$\mathbb{N} \times \mathbb{N} = A_0 \cup A_1 \cup A_2 \cup \dots$$

Hence $\mathbb{N} \times \mathbb{N}$ is a countable union of countable sets, whence $\mathbb{N} \times \mathbb{N}$ is countable by Theorem 4.1.2. This completes the proof.

The next result shows us that we can increase the dimension of a set and still preserve its countability.

Theorem 4.1.6 *Let A and B be countable sets. Then $A \times B$ is a countable set.*

Proof: Since A and B are countable we can list them. Let

$$\begin{aligned}
 A &= \{a_1, a_2, a_3, \dots\}, \\
 B &= \{b_1, b_2, b_3, \dots\}.
 \end{aligned}$$

Then we can list the elements of $A \times B$ as an infinite array or implied list (4.10).

$$\begin{aligned}
 & (a_1, b_1) \ (a_1, b_2) \ (a_1, b_3) \ (a_1, b_4) \dots \\
 & (a_2, b_1) \ (a_2, b_2) \ (a_2, b_3) \ (a_2, b_4) \dots \\
 & (a_3, b_1) \ (a_3, b_2) \ (a_3, b_3) \ (a_3, b_4) \dots \\
 & (a_4, b_1) \ (a_4, b_2) \ (a_4, b_3) \ (a_4, b_4) \\
 & \vdots \qquad \vdots \qquad \vdots \qquad \vdots
 \end{aligned} \tag{4.10}$$

Name the rows of array (4.10) with symbols X_1, X_2, \dots as in array (4.11).

$$\begin{aligned}
 X_1 &= (a_1, b_1) \ (a_1, b_2) \ (a_1, b_3) \ (a_1, b_4) \dots \\
 X_2 &= (a_2, b_1) \ (a_2, b_2) \ (a_2, b_3) \ (a_2, b_4) \dots \\
 X_3 &= (a_3, b_1) \ (a_3, b_2) \ (a_3, b_3) \ (a_3, b_4) \dots \\
 X_4 &= (a_4, b_1) \ (a_4, b_2) \ (a_4, b_3) \ (a_4, b_4) \dots \\
 &\vdots \qquad \vdots \qquad \vdots \qquad \vdots
 \end{aligned} \tag{4.11}$$

Each X_k is a countable set and there are countably many of them, so by Theorem 4.1.2, the union

$$A \times B = X_1 \cup X_2 \cup X_3 \cup \dots$$

is a countable set.

Since \mathbb{Q} is countable we have proved the following theorem.

Theorem 4.1.7 $\mathbb{Q} \times \mathbb{Q}$ is a countable set.

Using the same kind of argument the reader should try to prove the next result.

Theorem 4.1.8 $\mathbb{N} \times \mathbb{N} \times \mathbb{N} = \{(n, m, \ell) \mid n, m, \ell \in \mathbb{N}\}$ is countable.

4.2 Uncountable Sets

The next stage of our work requires us to compare the size of cardinals.

Definition 4.2.1 Let A and B be sets. We write

$$\text{card}(A) \leq \text{card}(B)$$

if there is a one-to-one function $f : A \rightarrow B$.

This is quite a reasonable definition of inequality for cardinals. For example, $3 < 4$ because $\{1, 2, 3\} \subset \{1, 2, 3, 4\}$, and in general

$m < n$ exactly when $\{1, \dots, m\} \subset \{1, \dots, n\}$.

For larger sets we see that

$$\text{card}(\mathbb{N}) \leq \text{card}(\mathbb{Z}) \leq \text{card}(\mathbb{Q}) \leq \text{card}(\mathbb{R})$$

because $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}$. One property that real numbers share with cardinals is the following property of natural numbers.

$\text{If } n \leq m \text{ and } m \leq n \text{ then } n = m.$

What this says is that we can deduce the equality of numbers if we can show that they are mutually related by \leq . It is a remarkable fact that this property also holds for *infinite cardinals*. Its proof is too far afield for us to write here. We will just accept its truth in this book.

$\text{If } \text{card}(A) \leq \text{card}(B) \text{ and } \text{card}(B) \leq \text{card}(A)$
 $\text{then } \text{card}(A) = \text{card}(B).$

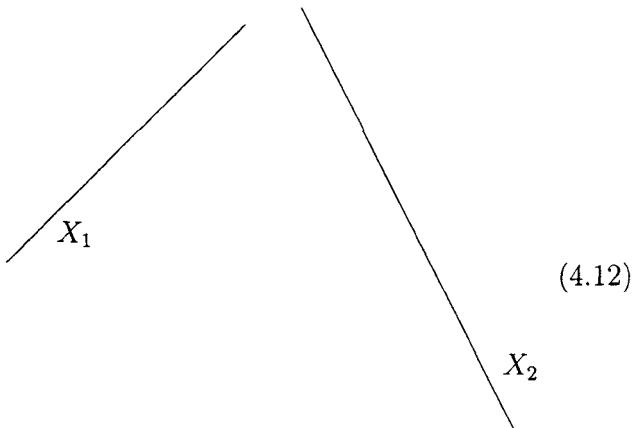
What we need now are some good examples. The set A is said to be *uncountable* if $\text{card}(\mathbb{N}) \neq \text{card}(A)$. The question is, are there

any uncountable sets at all, or is every set countable? We will show the remarkable fact that there are indeed uncountable sets. This must seem strange if you think about it for a second. Why should there be two types of infinite set? And yet that is where we are headed.

Let $(0, 1)$ denote the set of real numbers properly between 0 and 1:

$$(0, 1) = \{x \in \mathbb{R} \mid 0 < x < 1\}.$$

The next series of results provides us with a collection of uncountable sets by showing that many of the geometric objects you are familiar with are themselves uncountable. For example, every line segment is uncountable. Thus the cardinality of a set is independent of its length. The first result shows us that all line segments of *finite length* have the same cardinality.

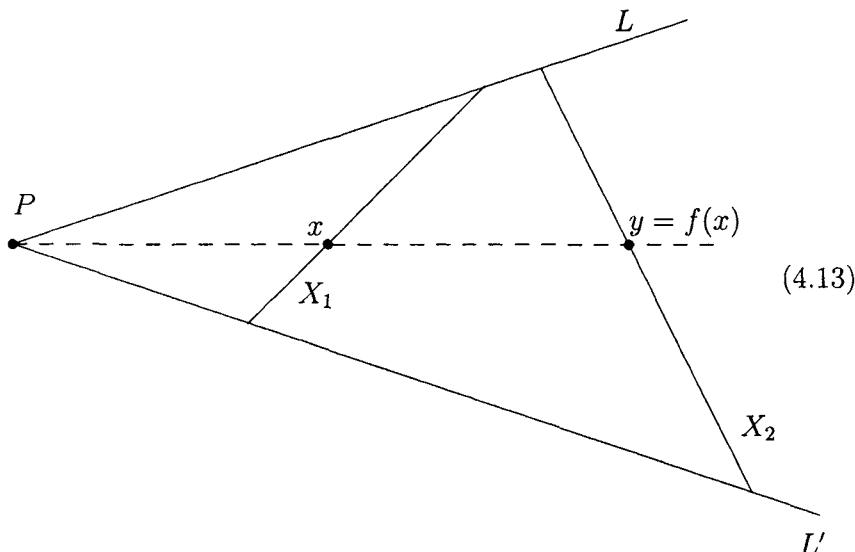


Theorem 4.2.2 *If X_1 and X_2 are finite line segments then $\text{card}(X_1) = \text{card}(X_2)$. Equivalently, any two line segments are equivalent.*

Proof: Begin with two line segments as in picture (4.12). The segments X_1 and X_2 in picture (4.12) are representative of all segments. We will use only their most general properties in this discussion. In picture (4.13) the end points of these segments are contained in lines L and L' . Lines L and L' intersect at a point P .

We define a function $f : X_1 \rightarrow X_2$ as follows. Given any point $x \in X_1$ draw a line (the dotted one in picture (4.13)) from P through x such that it intersects the segment X_2 at the point y . We define

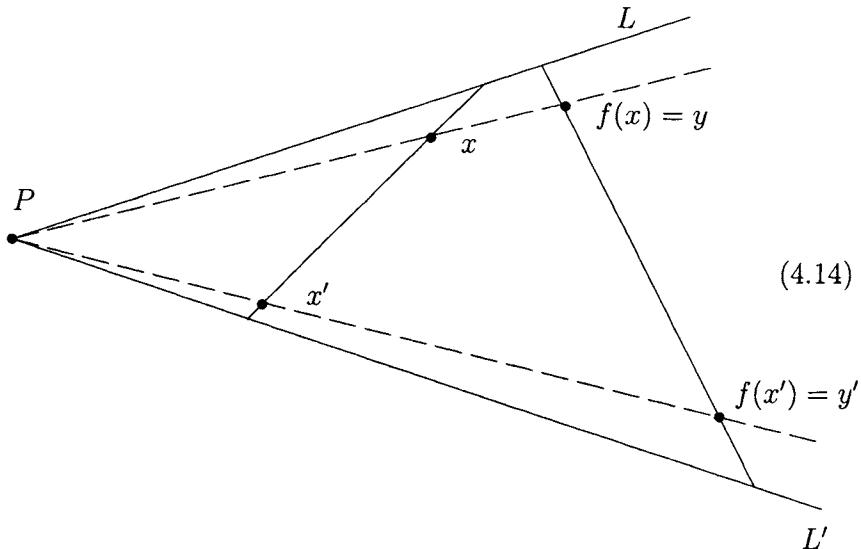
$$f(x) = y.$$



To see that f is onto, refer to picture (4.13). Let $y \in X_2$ be a point. Extend a dotted line from P to y . This line intersects X_1 at some point x . Then by the definition of f , $f(x) = y$, and hence f is onto.

To see that f is one-to-one suppose that $x \neq x'$. Then as in picture (4.14), x and x' determine different (dotted) lines through P . Let y and y' be the points where the dotted lines intersect X_2 . The intersections of these dotted lines with X_2 are different points y and y' . (If otherwise, the intersection is one point y , then y and the point P yield *exactly one* dotted line, not two.) Then $f(x) \neq f(x')$,

which implies that f is one-to-one.



Thus f is a bijection and therefore X_1 is equivalent to X_2 . This ends the proof.

A colloquial way of stating the above result is that *any two line segments have the same number of elements*. However, our later work will show that to say “the same number of elements” is a terribly inadequate way to describe the cardinality of a set.

For instance,

$$\text{card}(0, 1) = \text{card}(0, \text{googolplex}) = \text{card}(-2, -1) = \text{card}\left(-\frac{\pi}{2}, \frac{\pi}{2}\right).$$

Consequently,

$$\text{card}(0, 1) = \text{card}(a, b) \text{ for any numbers } a < b.$$

That is, any two segments have the same cardinality, namely, that of $(0, 1)$.

The next example shows that all segments have the cardinality of \mathbb{R} .

Theorem 4.2.3 $\text{card}(0, 1) = \text{card}(\mathbb{R})$.

Proof: By the comment preceding this theorem

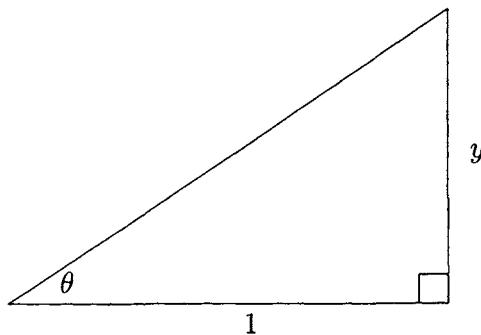
$$\text{card}(0, 1) = \text{card}\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$$

so, without losing generality, we can prove that $\text{card}\left(-\frac{\pi}{2}, \frac{\pi}{2}\right) = \text{card}(\mathbb{R})$.

Consider the function

$$\tan(\theta) : \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \longrightarrow \mathbb{R}.$$

Given a real number y draw a right triangle labelled as in the diagram below.



Elementary trigonometry shows that

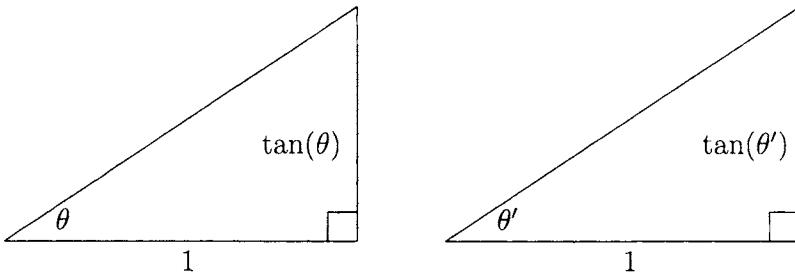
$$\tan(\theta) = \frac{y}{1} = y.$$

Thus $\tan(\theta)$ is onto.

To see that $\tan(\theta)$ is one-to-one, suppose that θ and θ' are given such that

$$\tan(\theta) = \tan(\theta').$$

Draw two triangles:



Certainly the bases of these triangles are equal as are the indicated right angles. By hypothesis the right hand legs are equal in length. Then by a theorem you learned in high school geometry the two triangles are congruent. Thus the measures θ and θ' of the corresponding base angles are equal:

$$\theta = \theta',$$

so that $\tan(\theta)$ is one-to-one.

Then $\tan(\theta) : \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \longrightarrow \mathbb{R}$ is a bijection and therefore

$$\text{card}\left(-\frac{\pi}{2}, \frac{\pi}{2}\right) = \text{card}(\mathbb{R}).$$

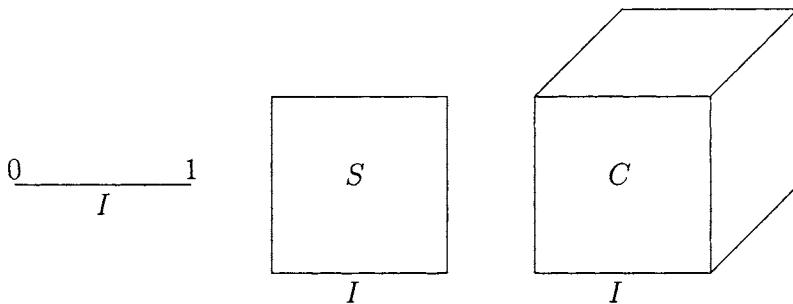
This completes the proof of the theorem.

That may come as a surprise to you. The cardinality of a finite segment and the cardinality of the unbounded real number line are the same. This implies that greater length does not produce more points in a line. In simple but inaccurate terms, the number of points on a finite segment is equal to the number of points in all of the real number line.

Let

$$I = \{x \in \mathbb{R} \mid 0 \leq x \leq 1\}.$$

This is the set whose elements are 0 and 1 and *all* of the numbers between 0 and 1. Let S be the *unit square and its interior*, and let C be the *unit cube and its interior*. In pictures we have



Let us be clear about which figures we mean. I includes the end points and the interior of the interval $(0, 1)$, S is the interior and the boundary of the square whose sides have length 1, and C is the solid cube whose sides have length 1. The following example will show us that the cardinality of a set is independent of its geometry and dimension. If you draw a line segment on a piece of paper, that segment has the same number of points as the piece of paper, which has the same number of points as the room in which you are reading this. Thus even though these sets have different geometric dimension they have the same cardinality.

Theorem 4.2.4 $\text{card}(I) = \text{card}(S) = \text{card}(C)$.

Proof: It seems clear enough that

$$I \subset S \subset C$$

so that

$$\text{card}(I) \leq \text{card}(S) \leq \text{card}(C). \quad (4.15)$$

To complete the proof we will produce a one-to-one function $f : C \rightarrow I$. This will allow us to conclude that

$$\text{card}(C) \leq \text{card}(I)$$

and hence that $\text{card}(I) = \text{card}(C)$. Play with that one for a while. Suppose that a, b are numbers such that $a \leq b$ and $b \leq a$. The only way for this to happen is for $a = b$. We are simply using the fact that this same property holds for cardinals. If you find the following argument difficult, just skip forward. There is plenty of time and no shame in coming back to these deliberations.

From analytic geometry we recall that the unit cube C is a set of ordered triples,

$$C = \{(x, y, z) \mid x, y, z \in I\}.$$

The one-to-one function $f : C \longrightarrow I$ that we need is defined as follows. Given an ordered triple $(x, y, z) \in C$, write x, y, z as decimals

$$\begin{array}{ccccccc} x & = & x_0. & x_1 & x_2 & x_3 & \dots \\ y & = & y_0. & y_1 & y_2 & y_3 & \dots \\ z & = & z_0. & z_1 & z_2 & z_3 & \dots \end{array}$$

where $x_0 = 0$ or 1 , and where x_1, x_2, x_3, \dots are digits in the set $\{0, 1, \dots, 9\}$. Similar statements apply to the decimal expansions of y and z . Furthermore, to avoid that nasty case where numbers end in $99\bar{9}$, we assume that the shorter decimal expansion that ends in $00\bar{0}$ is used for x, y , and z . Then we define

$$f(x, y, z) = .x_0y_0z_0\ x_1y_1z_1\ x_2y_2z_2\ x_3y_3z_3\ \dots$$

Since the numbers x_k, y_k, z_k are in the set $\{0, 1, \dots, 9\}$, $f(x, y, z)$ is a decimal between 0 and 1 . That is, $f(x, y, z) \in I$ and hence

$$f : C \longrightarrow I$$

is a function. Since the numbers x, y , and z do not have an infinite string of 9 's in their decimal expansion, the number $f(x, y, z)$ does not end in $99\bar{9}$.

For example, to find

$$f(1.00\bar{0}, 0.41415\dots, 0.12\bar{2})$$

we write the three numbers as

$$\begin{aligned}x &= 1. \quad 0 \quad 0 \quad 0 \quad \dots, \\y &= 0. \quad 4 \quad 1 \quad 4 \quad \dots, . \\z &= 0. \quad 1 \quad 2 \quad 1 \quad \dots.\end{aligned}$$

Now shuffle them.

$$f(1.0\bar{0}, 0.41415\dots, 0.12\bar{2}) = .100 \ 041 \ 012 \ 041 \ \dots$$

We will show that f is one-to-one. Suppose that

$$w = f(x', y', z') = f(x, y, z) = .x_0y_0z_0 \ x_1y_1z_1 \ x_2y_2z_2 \ x_3y_3z_3 \ \dots$$

for some ordered triples $(x, y, z), (x', y', z') \in C$. This is the unique short way of writing w since by design w does not end in $99\bar{9}$. Then by the definition of f (applied in reverse) we have

$$\begin{aligned}x &= x_0. \quad x_1 \quad x_2 \quad x_3 \quad \dots = x' \\y &= y_0. \quad y_1 \quad y_2 \quad y_3 \quad \dots = y' \\z &= z_0. \quad z_1 \quad z_2 \quad z_3 \quad \dots = z'.\end{aligned}$$

Then $(x, y, z) = (x', y', z')$, which implies that f is a one-to-one function. By our definition of \leq for cardinals,

$$\text{card}(C) \leq \text{card}(I).$$

Combining this with the inequalities (4.15), we see that

$$\text{card}(I) = \text{card}(S) = \text{card}(C)$$

as required to complete the proof.

The above example should strike you as contrary to the conservation of mass or some such concept from physics. How could a line and a cube have the same anything? The issue here is that we are counting the points in a mathematical object, and points do not have any mass and they do not have any other physical properties. Physics and physical properties of lines and cubes never enter the picture.

The surprises continue in the next result where we state, but do not prove, that infinite space, a child's toy block, and a 1 inch long segment have the same cardinality.

Theorem 4.2.5 Let \mathbb{R} denote the one dimensional real line, let \mathbb{R}^2 denote the two dimensional entire plane, and let \mathbb{R}^3 denote all of three dimensional real space. Then

$$\text{card}(0, 1) = \text{card}(\mathbb{R}) = \text{card}(\mathbb{R}^2) = \text{card}(\mathbb{R}^3).$$

That is, cardinality is independent of spatial dimension.

This example should also upset your common sense. If we identify $(0, 1)$ with an inch long line segment, then you might believe that certainly an inch has fewer points than the entire real line, which has fewer points than the plane or all of three dimensional space. Once again we are shown that our intuition regarding infinity does not agree with the surrounding mathematical facts. The theorems in the next section will show you why we are being so careful in our discussion.

4.3 Two Infinities

Let us summarize what we have learned to this point. The cardinality of a set X is a way of measuring in precise mathematical terms the number of elements in X . We saw that

$$\text{card}(\mathbb{N}) = \text{card}\{1^2, 2^2, 3^2, \dots\} = \text{card}(\mathbb{Q}).$$

These sets are called countable. We saw that

$$\text{card}(0, 1) = \text{card}(\text{unit cube}) = \text{card}(\mathbb{R}) = \text{card}(\text{space}).$$

These sets are called *uncountable* and I will explain why shortly.

We also saw that

$$\text{card}(\mathbb{N}) \leq \text{card}(\mathbb{R}).$$

We ask the natural question: How are $\text{card}(\mathbb{N})$ and $\text{card}(\mathbb{R})$ otherwise related? Are the uncountable sets and the countable sets two different types of sets? The uninitiated reader might argue that since both sets are infinite, \mathbb{N} and \mathbb{R} have the same cardinality. The next result shows that this is not the case. A colloquial and accurate interpretation of this theorem is the philosophical and intuitive shocker that *there is more than one infinity*.

Theorem 4.3.1 $\text{card}(\mathbb{N}) < \text{card}(\mathbb{R})$.

Proof: Evidently $\mathbb{N} \subset \mathbb{R}$, so by the definition of inequality of cardinals,

$$\text{card}(\mathbb{N}) \leq \text{card}(\mathbb{R}).$$

What we have to prove now is that $\text{card}(\mathbb{N}) \neq \text{card}(\mathbb{R})$. To do this we borrow from the Realists who visited Hilbert's Infinite Hotel on page 94.

Suppose, for the sake of contradiction, that we have a one-to-one function

$$f : \mathbb{N} \longrightarrow [0, 1].$$

Make a list of f and its values in $[0, 1]$.

$$\begin{array}{lllllll} f(1) = x_1 = & . & \boxed{d_{11}} & d_{12} & d_{13} & d_{14} & \dots \\ f(2) = x_2 = & . & d_{21} & \boxed{d_{22}} & d_{23} & d_{24} & \dots \\ f(3) = x_3 = & . & d_{31} & d_{32} & \boxed{d_{33}} & d_{34} & \dots \\ f(4) = x_4 = & . & d_{41} & d_{42} & d_{43} & \boxed{d_{44}} & \dots \\ \vdots & & \vdots & & \vdots & & \end{array}$$

In the above implied list we have given the shorter decimal expansion of $f(n) = x_n$ for each $n \in \mathbb{N}$, if the shorter decimal expansion exists. In this implied list, each d_{ij} is a digit, a number in the set $\{0, 1, \dots, 9\}$, and $f(n)$ does not end in 999. Furthermore, d_{11} represents the first digit in the decimal form of x_1 , d_{23} represents digit number 3 in the decimal form of x_2 , and in general d_{nm} represents digit number m in the decimal form of x_n . We can now produce a real number that is not on this implied list.

Define the digit d_1 as

$$d_1 = \begin{cases} d_{11} + 1 & \text{if } d_{11} \neq 9 \\ 0 & \text{if } d_{11} = 9 \end{cases}.$$

Notice that we have defined d_1 so that $d_1 \neq d_{11}$.

Let

$$d_2 = \begin{cases} d_{22} + 1 & \text{if } d_{22} \neq 9 \\ 0 & \text{if } d_{22} = 9 \end{cases}$$

Notice that we have defined d_2 so that $d_2 \neq d_{22}$.

In general, define

$$d_n = \begin{cases} d_{nn} + 1 & \text{if } d_{nn} \neq 9 \\ 0 & \text{if } d_{nn} = 9 \end{cases}.$$

Then $d_n \neq d_{nn}$.

The extra number x is given by the following decimal expansion:

$$x = .d_1 d_2 d_3 d_4 \dots$$

If the number x ends in $99\bar{9}$, replace it with the shorter decimal expansion that ends in $00\bar{0}$.

Then $x \neq x_1$ because by our choice of d_1 , x and x_1 differ in the first decimal place, $d_1 \neq d_{11}$. Also, $x \neq x_2$ because by our choice of d_2 , x and x_2 differ in decimal place number 2, $d_2 \neq d_{22}$. In general, $x \neq x_n$ because by our choice of d_n , x and x_n differ in decimal place number n , $d_n \neq d_{nn}$. Thus x is not on the implied list x_1, x_2, x_3, \dots of real numbers in $[0, 1]$. This is what we wanted. Hence f is not a bijection.

We have thus shown that there are no bijections $f : \mathbb{N} \longrightarrow [0, 1]$, which proves that

$$\text{card}(\mathbb{N}) \neq \text{card}[0, 1].$$

Since Theorem 4.2.3 states that $\text{card}[0, 1] = \text{card}(\mathbb{R})$, we have shown that

$$\text{card}(\mathbb{N}) \neq \text{card}(\mathbb{R})$$

and therefore that

$$\text{card}(\mathbb{N}) < \text{card}(\mathbb{R}).$$

This completes the proof.

Subsequently, there cannot be a list of the real numbers \mathbb{R} . Furthermore, we conclude that the terms *countable* and *uncountable* describe two different types of infinite sets. Those sets equivalent to \mathbb{N} are not equivalent to \mathbb{R} . Since \mathbb{Q} is a dense subset of \mathbb{R} the following result shows us that cardinality does not depend on the density of the set in the real line.

Theorem 4.3.2 $\text{card}(\mathbb{Q}) < \text{card}(\mathbb{R})$.

Proof: Theorem 4.1.4 shows us that $\text{card}(\mathbb{N}) = \text{card}(\mathbb{Q})$ and Theorem 4.3.1 shows us that $\text{card}(\mathbb{N}) < \text{card}(\mathbb{R})$. Then $\text{card}(\mathbb{Q}) < \text{card}(\mathbb{R})$. This is what we wanted to prove.

Summarizing some of our results to this point, we see that

$$\text{card}(\mathbb{N}) = \text{card}(\mathbb{Q}) < \text{card}(\mathbb{R}) = \text{card}(\text{space}).$$

How does that make you feel? There it is in plain mathematical language. There are several infinities. We will see later that there are more than just two. That should jolt you. Infinity is infinity and that is that, you might have said. The answer to that argument is that yes, all infinite sets are not finite, but infinity is not a number. Infinity is not a cardinal. Infinity is not the last word on the size of a set. It is only the beginning, like saying that something is large. There will always be larger sets. Saying that a set is infinite is like saying that a car has color. Yes, it has color, but which one? Is it tan or silver? So when we encounter cardinals we will ask, “Is it finite or infinite?” And if it is infinite we will ask “Is it countable or uncountable?” This will describe somewhat which cardinal we have but it will not be the last word on how large that cardinal is. Keep an open mind as you read this section.

Because $\text{card}(\mathbb{Q}) < \text{card}(\mathbb{R})$ there must be real numbers that are not fractions, not repeating decimals. A number x is said to be *irrational* if it is not rational; that is, if $x \notin \mathbb{Q}$. We let

$$\mathbb{Q}' = \text{the set of irrational numbers.}$$

Unlike \mathbb{Q} , there can be no implied list of irrational numbers. Theorem 4.3.2 and a little intuition will convince us that \mathbb{Q}' and \mathbb{R} are equivalent sets.

Theorem 4.3.3 $\text{card}(\mathbb{Q}') = \text{card}(\mathbb{R})$.

It is interesting to observe that this simple counting $\text{card}(\mathbb{Q}') = \text{card}(\mathbb{R})$ shows us that *most numbers are not rational*. This is contrary to your experience since every number you have ever seen has

been a rational number. That changes nothing. There are many more irrational numbers than there are rational ones. However professional mathematicians consider it to be very hard to prove that any given number is irrational. While it is true that π is irrational, the proof is quite hard. We will show shortly that $\sqrt{2}$ is an irrational number, but we cannot show that $\pi^{\sqrt{2}}$ is irrational. Currently, no one knows its rationality. If you have some facility with logarithms, you might find it challenging to try to prove that

$$\log_2(3) \text{ is irrational.}$$

(Hint: Read the proof of Theorem 4.3.4 first.) A simple question that currently remains unanswered is this. *If x and y are irrational numbers, under what conditions is $x + y$ irrational?* For example, it is known that $\sqrt{3} + \sqrt{2}$ is irrational. But what can be said of more mysterious numbers like $2^{\sqrt{2}}$?

Let's settle back and catch our breath now. The earliest example of an irrational number occurs around 600 BC in ancient Greece, where the Pythagoreans showed that $\sqrt{2}$ is an irrational number. They found this fact to be so disturbing that they threatened people with a removal of their hands if they betrayed the great secret that $\sqrt{2}$ is not a fraction of integers. We laugh at this today or we gape in awe that rational people would react so extremely to a mathematical fact. Today this fact is considered to be one of the highlights of early mathematics. The proof is just as compelling today as it was 2600 years ago. The proof that $\sqrt{2}$ is irrational is a pivotal moment in mathematics. Prior to this, people who thought about the universe thought that all of it could be described by fractions, rational numbers. Thus the following result is a moment of transformation in mathematics, physics, and science in general. Do your best to follow it.

Theorem 4.3.4 $\sqrt{2}$ is an irrational number.

Proof: Before we begin we need to establish:

$$\text{If } n^2 \text{ is divisible by 2 then } n \text{ is divisible by 2.} \quad (4.16)$$

Equivalently, we will prove:

$$\text{If } n \text{ is not divisible by 2 then } n^2 \text{ is not divisible by 2.}$$

The proof is a calculation. Suppose that n is not divisible by 2. Then $n = 2k + 1$ for some $k \in \mathbb{N}$ so that

$$n^2 = (2k + 1)^2 = 4k^2 + 4k + 1 = 2(2k^2 + 2k) + 1.$$

Thus $n^2 = 2K + 1$ where K is the integer $2k^2 + 2k$, so that n^2 is an odd number. This proves statement (4.16). Armed with this fact, we can attack the irrationality of $\sqrt{2}$.

Assume for the sake of contradiction that $\sqrt{2}$ is rational. Then

$$\sqrt{2} = \frac{n}{m}$$

for some integers $n, m > 0$. We assume without losing our generality that the fraction $\frac{n}{m}$ is in reduced form. Specifically,

$$\begin{aligned} &\text{at least one of the numbers } n \text{ and } m \\ &\text{is not divisible by 2.} \end{aligned} \tag{4.17}$$

Squaring both sides yields

$$2 = \frac{n^2}{m^2}.$$

Thus $2m^2 = n^2$ and so n^2 is divisible by 2. Statement (4.16) implies that n is divisible by 2 and hence we can write

$$n = 2k$$

for some integer k . Squaring both sides we find that

$$2m^2 = n^2 = 4k^2,$$

so division by 2 yields

$$m^2 = 2k^2.$$

Then m^2 is divisible by 2 and hence m is divisible by 2 by statement (4.16). Thus both n and m are divisible by 2, contrary to the statement (4.17). Hence our original assumption is in error, and we conclude that $\sqrt{2}$ is irrational. This completes the proof.

The reader can use the same argument to show that $\sqrt{3}$ is irrational and that in general \sqrt{p} is irrational for each prime p . This

reinforces the above work that there are infinitely (actually uncountably) many irrational numbers. The present work has the advantage that we have implied the existence of a **List** of some irrational numbers.

$$\sqrt{2}, \sqrt{3}, \sqrt{5}, \sqrt{7}, \dots \text{ are irrational numbers.}$$

Compare this to Theorem 4.3.3, which states that there are infinitely many irrational numbers but which does not list them.

There are at least two types of cardinals, the countable and the uncountable. We will see in the next section that uncountable cardinals also come in many different sizes.

4.4 Power Sets

Let A be a set. Recall from Definition 1.3.1 that *the power set of A* , $\mathcal{P}(A)$, is the set of all subsets of A . We will use the power set to show that there are many infinite cardinals.

From Theorem 4.3.1 we know that $\text{card}(\mathbb{N})$ and $\text{card}(\mathbb{R})$ are different infinite cardinals. The next result shows us that \mathbb{R} and $\mathcal{P}(\mathbb{N})$ have the same cardinality.

Theorem 4.4.1 $\text{card}(\mathbb{R}) = \text{card}(\mathcal{P}(\mathbb{N}))$.

Proof: We will produce a one-to-one function

$$f : [0, 1] \longrightarrow \mathcal{P}(\mathbb{N}),$$

thus proving that $\text{card}[0, 1] \leq \text{card}(\mathcal{P}(\mathbb{N}))$. Then by Theorem 4.2.3, we will have shown that $\text{card}(\mathbb{R}) = \text{card}[0, 1] \leq \text{card}(\mathcal{P}(\mathbb{N}))$. If you find this proof is stretching you a bit too far, skip to the end. There is no examination at the end of this discussion.

In this age of computers it should come as no surprise that each real number can be written as a *binary decimal*. That is, to each real number $x \in [0, 1]$ there is a string

$$x = .b_1 b_2 b_3 \dots$$

of bits $b_1, b_2, b_3, \dots \in \{0, 1\}$. Specifically, we can write down an equation $x = .b_1 b_2 b_3 \dots$ exactly when

$$x = b_1 \frac{1}{2} + b_2 \frac{1}{2^2} + b_3 \frac{1}{2^3} + \dots$$

For example,

$$\frac{1}{2} = 1 \frac{1}{2} + 0 \frac{1}{2^2} + 0 \frac{1}{2^3} + \dots,$$

where the coefficients of the remaining fractions $\frac{1}{2^n}$ are 0. Thus we will write

$$\frac{1}{2} = .10\bar{0}.$$

Another example is

$$\frac{5}{8} = 1 \frac{1}{2} + 0 \frac{1}{2^2} + 1 \frac{1}{2^3} + 0 \frac{1}{2^4} + 0 \frac{1}{2^5} + \dots,$$

where the coefficients of the remaining fractions $\frac{1}{2^n}$ are 0. Then

$$\frac{5}{8} = .1010\bar{0}.$$

Furthermore, we can write

$$\frac{1}{3} = 0 \frac{1}{2} + 1 \frac{1}{2^2} + 0 \frac{1}{2^3} + 1 \frac{1}{2^3} + 0 \frac{1}{2^3} + \dots$$

where the pattern of coefficients 0, 1, 0, 1, 0, 1, ... continues indefinitely. Thus

$$\frac{1}{3} = .0101\bar{01}.$$

Now, write the elements of $[0, 1]$ in the usual way. Given an $x \in [0, 1]$, write x as a binary number as above

$$x = .b_1 b_2 b_3 \dots$$

and pair the decimal bits b_1, b_2, b_3, \dots with \mathbb{N} as follows.

$$\begin{array}{ccccccc} 1 & 2 & 3 & 4 & \dots \\ b_1 & b_2 & b_3 & b_4 & \dots \end{array}$$

Construct a subset $f(x) = U_x \subset \mathbb{N}$ as follows.

$$\left\{ \begin{array}{ll} n \in U_x \text{ exactly when} & b_n = 1 \\ n \notin U_x \text{ exactly when} & b_n = 0 \end{array} \right..$$

That this defines a subset of \mathbb{N} is clear since the bits b_n can only be 0 or 1. For example, let's see what set U_x corresponds to $\frac{1}{3}$.

Because $\frac{1}{3} = .0101\overline{01}$, the pairing for $\frac{1}{3}$ looks like this.

$$\begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & \dots \\ 0 & 1 & 0 & 1 & 0 & 1 & \dots \end{array}$$

Then the elements of U_x are those paired with a 1. That is,

$$f\left(\frac{1}{3}\right) = \{2, 4, 6, \dots\}.$$

The reader can show that the pairing for

$$x = .101001000100001000001\dots$$

results in the set

$$U_x = \{1, 3, 6, 10, 15, 21, \dots\}.$$

Search for the pattern in the decimal expansion for x .

There is a possible point of confusion that we should discuss. There are two very different ways to write $\frac{1}{2}$ as a sum of powers of $\frac{1}{2}$. They are

$$\frac{1}{2} = 1\frac{1}{2} + 0\frac{1}{2^2} + 0\frac{1}{2^3} + 0\frac{1}{2^4} + \dots$$

$$\frac{1}{2} = 0\frac{1}{2} + 1\frac{1}{2^2} + 1\frac{1}{2^3} + 1\frac{1}{2^4} + \dots,$$

or in other words

$$\frac{1}{2} = .1\bar{0} = .0\bar{1}.$$

Which of these sums should we use in defining $f(\frac{1}{2})$? Let us agree that, in case there are two or more ways to write a fraction x as a binary decimal expansion, we will take the one with fewer 1's. Thus we take $\frac{1}{2} = .1\bar{0}$ and so

$$f\left(\frac{1}{2}\right) = \{1\}.$$

Our goal now is to show that f is one-to-one. This is done by showing that

$$x \neq y \text{ implies that } f(x) \neq f(y).$$

So suppose that $x \neq y$ and write each as a binary decimal.

$$\begin{aligned} x &= .b_1 \ b_2 \ b_3 \ b_4 \ \dots \\ y &= .c_1 \ c_2 \ c_3 \ c_4 \ \dots \end{aligned}$$

Since $x \neq y$, they must differ in some position, say, the n th position. That is,

$$b_n \neq c_n.$$

We may assume without loss of generality that $b_n = 1$ and that $c_n = 0$. Then by the definition of f

$$\begin{aligned} n \in f(x) &= U \text{ and} \\ n \notin f(y) &= V, \end{aligned}$$

so that U and V do not contain the same elements. It follows that

$$f(x) = U \neq V = f(y)$$

and so f is a one-to-one function.

Thus $f : [0, 1] \rightarrow \mathcal{P}(\mathbb{N})$ is a one-to-one function, and therefore

$$\text{card}[0, 1] \leq \text{card}(\mathcal{P}(\mathbb{N})). \quad (4.18)$$

Conversely, the following careful inspection of $\mathcal{P}(\mathbb{N})$ will reveal that

$$\text{card}(\mathcal{P}(\mathbb{N})) \leq \text{card}[0, 1]. \quad (4.19)$$

Combining the inequality (4.19) with the inequality (4.18) will then yield the equation

$$\text{card}(\mathbb{R}) = \text{card}[0, 1] = \text{card}(\mathcal{P}(\mathbb{N})).$$

When we are done with the inspection, the proof will be complete.

What we are going to do is to define a one-to-one function

$$f : \mathcal{P}(\mathbb{N}) \longrightarrow [0, 1].$$

First, write each subset of \mathbb{N} as a binary sequence. Given a set $U \subset \mathbb{N}$, define a binary bit b_n as

$$b_n = \begin{cases} 1 & \text{if } n \in U \\ 0 & \text{if } n \notin U \end{cases}$$

and then

$$U \text{ corresponds to } b_0 b_1 b_2 \dots$$

This correspondence is so detailed that we can identify U with $b_0 b_1 b_2 \dots$. They both contain the same information, namely, which numbers are in U and which are not.

For instance to construct the binary sequence that corresponds to the set $U = \{0, 2, 4, \dots\}$ of even numbers, we find each binary bit, beginning with 0.

$$\begin{aligned} b_0 &= 1 \text{ since } 0 \in U \\ b_1 &= 0 \text{ since } 1 \notin U \\ b_2 &= 1 \text{ since } 2 \in U \\ b_3 &= 0 \text{ since } 0 \notin U \\ &\vdots \end{aligned}$$

(Yes, 0 is an even number.) Thus the set of even numbers corresponds to

$$10\overline{10}.$$

The set $U = \{5\}$ corresponds to $000001\bar{0}$. If U corresponds to $0101\bar{0}\bar{1}$, then $U = \{1, 3, 5, \dots\}$ = the set of odd numbers. If U corresponds to $000011\bar{1}$, then U contains all the natural numbers except $0, 1, 2, 3$. The set U of prime numbers corresponds to the binary sequence that begins with

$$00110101000101000101.$$

You can find the next five terms in that sequence if you wish. Thus we will treat each $U \in \mathcal{P}(\mathbb{N})$ as a binary sequence $b_0b_1b_2\dots$

Consider the function

$$f : \mathcal{P}(\mathbb{N}) \longrightarrow [0, 1]$$

defined as follows. Write each $U \in \mathcal{P}(\mathbb{N})$ as a binary sequence

$$U = b_0b_1b_2\dots$$

as we did above. Then define $f(U)$ to be the real number

$$f(x) = .b_0b_1b_2\dots$$

written as a *decimal number*. For example,

$$\begin{aligned} f(01\bar{0}) &= .01 = \frac{1}{100}, \\ f(11\bar{1}) &= .11\bar{1} = \frac{1}{9}, \\ f(10\bar{1}\bar{0}) &= .10\bar{1}\bar{0}. \end{aligned}$$

Evidently, $f(U)$ just places a decimal point in front of the binary sequence $b_0b_1b_2\dots$ corresponding to U , making the sequence a real number $.b_0b_1b_2\dots$

We will show that f is one-to-one. Suppose that

$$U \neq V \text{ for some sets } U, V \in \mathcal{P}(\mathbb{N}).$$

We must show that $f(U) \neq f(V)$. As we did above, write the sets

$$U = b_0b_1b_2\dots \text{ and } V = c_0c_1c_2\dots$$

as *binary sequences*. Since $U \neq V$ there are binary bits, say, b_n, c_n , such that

$$b_n \neq c_n.$$

But then the *decimal numbers* $f(U)$ and $f(V)$ differ in the n th decimal places. That is, these are different numbers.

$$f(U) = .b_0 b_1 b_2 \dots \neq .c_0 c_1 c_2 \dots = f(V).$$

Hence f is a one-to-one function.

Consequently, $\text{card}(\mathcal{P}(\mathbb{N})) \leq \text{card}[0, 1]$. Since we have already proved that $\text{card}[0, 1] \leq \text{card}(\mathcal{P}(\mathbb{N}))$, we conclude that

$$\text{card}[0, 1] = \text{card}(\mathcal{P}(\mathbb{N})),$$

which concludes the proof.

In Theorem 4.3.1 we showed that

$$\text{card}(\mathbb{N}) < \text{card}[0, 1] = \text{card}(\mathbb{R}),$$

thus showing that there are at least two different infinite cardinals. This presents the problem of determining how many infinite cardinals there are. It seems unlikely that there are just two so we look for more.

The following beautiful result shows us that no matter which set A we choose, $\mathcal{P}(A)$ is always a larger set than A . That is, the cardinality of $\mathcal{P}(A)$ is always more than the cardinality of A . This fact is certainly true of *finite* cardinals as we have proved that

$$\text{if } \text{card}(A) = n \text{ is finite then } \text{card}(\mathcal{P}(A)) = 2^n.$$

We see immediately that

$$\text{card}(A) = n < 2^n = \text{card}(\mathcal{P}(A))$$

for any $n \in \mathbb{N}$, thus proving that $\mathcal{P}(A)$ has more elements than the finite set A . If $A = \{a_1, \dots, a_n\}$ has exactly n elements and if we let

$$\begin{aligned} \mathcal{P}^2(A) &= \mathcal{P}(\mathcal{P}(A)), \\ \mathcal{P}^3(A) &= \mathcal{P}(\mathcal{P}(\mathcal{P}(A))) &= \mathcal{P}(\mathcal{P}^2(A)), \\ \mathcal{P}^4(A) &= \mathcal{P}(\mathcal{P}(\mathcal{P}(\mathcal{P}(A)))) &= \mathcal{P}(\mathcal{P}^3(A)), \\ &\vdots \end{aligned}$$

(note the use of 4 \mathcal{P} 's in the third line), then we see that

$$\begin{aligned}\text{card}(\mathcal{P}(A)) &= 2^n, \\ \text{card}(\mathcal{P}^2(A)) &= \text{card}(\mathcal{P}(\mathcal{P}(A))) = 2^{2^n} \quad \text{since } \text{card}(\mathcal{P}(A)) = 2^n, \\ \text{card}(\mathcal{P}^3(A)) &= \text{card}(\mathcal{P}(\mathcal{P}^2(A))) = 2^{2^{2^n}} \quad \text{since } \text{card}(\mathcal{P}^2(A)) = 2^n, \\ \text{card}(\mathcal{P}^4(A)) &= \text{card}(\mathcal{P}(\mathcal{P}^3(A))) = 2^{2^{2^n}} \quad \text{since } \text{card}(\mathcal{P}^3(A)) = 2^{2^n}, \\ &\vdots\end{aligned}$$

This leads us to the infinite sequence of *finite cardinals*

$$n < 2^n < 2^{2^n} < 2^{2^{2^n}} < \dots$$

Our next result shows that there is a similar sequence of *infinite* cardinals. We challenge the reader to find the similarities between the argument used below and the argument used to prove Theorem 4.3.1.

Theorem 4.4.2 *If A is a set then $\text{card}(A) < \text{card}(\mathcal{P}(A))$.*

Proof: First we establish that $\text{card}(A) \leq \text{card}(\mathcal{P}(A))$. The symbol λ is the Greek letter called *lambda*. The one-to-one function we need is

$$\lambda : A \longrightarrow \mathcal{P}(A)$$

defined by

$$\lambda(x) = \{x\}.$$

Certainly $\lambda(x) = \lambda(x')$ implies that $\{x\} = \{x'\}$ or equivalently that $x = x'$. Thus λ is a one-to-one function and hence $\text{card}(A) \leq \text{card}(\mathcal{P}(A))$.

To prove that $\text{card}(A) < \text{card}(\mathcal{P}(A))$ we need to show that no function

$$F : A \longrightarrow \mathcal{P}(A)$$

is onto, thus showing that there are no bijections between A and $\mathcal{P}(A)$. It will then follow that $\text{card}(A) \neq \text{card}(\mathcal{P}(A))$ and hence that $\text{card}(A) < \text{card}(\mathcal{P}(A))$.

Given a function $F : A \longrightarrow \mathcal{P}(A)$ and an $x \in A$, observe that $F(x) \in \mathcal{P}(A)$ so that $F(x)$ is a subset of A , $F(x) \subset A$. We can ask, *Is $x \in F(x)$ or is $x \notin F(x)$?* Define a special set $X \subset A$ by

$$X = \{x \in A \mid x \notin F(x)\}.$$

The set X is a remarkable observation made by Bertrand Russell and has no motivation. Let's accept the genius that uncovered this idea, that it works well, and then see how it is used in our argument. We claim that there does not exist an $x \in A$ such that $F(x) = X$.

Suppose to the contrary that there is an $x \in A$ such that $F(x) = X$. We seek a contradiction. Since X is a set we ought to be able to decide whether or not $x \in F(x) = X$ or $x \notin F(x) = X$. Suppose that $x \in F(x) = X$. Then x satisfies the predicate defining X , which is $x \notin F(x)$. This is not possible since $F(x) = X$ is a set. So it must be that $x \notin F(x) = X$. In this case x satisfies the predicate for X , so that $x \in X$. But this is also impossible ($x \notin X$ and $x \in X$) since X is a set. Thus we cannot decide if $x \in F(x) = X$ or if $x \notin F(x) = X$. This contradiction to the definition of set shows us that our initial assumption that $F(x) = X$ for some x is a falsehood. Thus there is no $x \in A$ such that $F(x) = X$. This proves our claim, thus $F : A \longrightarrow \mathcal{P}(A)$ is not onto, and therefore $\text{card}(A) \neq \text{card}(\mathcal{P}(A))$. This completes the proof.

Theorem 4.4.3 $\text{card}(\mathbb{N}) < \text{card}(\mathcal{P}(\mathbb{N}))$.

Theorem 4.4.2 allows us to prove the following counterintuitive result. In a language devoid of mathematics it implies that *there are infinitely many infinities*.

Theorem 4.4.4 *There is an infinite sequence $\alpha_0 < \alpha_1 < \alpha_2 < \dots$ of infinite cardinals.*

Proof: If we let $\alpha_0 = \text{card}(\mathbb{N})$ and let $\alpha_1 = \text{card}(\mathcal{P}(\mathbb{N}))$, then Theorem 4.4.2 shows us that

$$\alpha_0 < \alpha_1.$$

Next, consider the sequence

$$\mathcal{P}(\mathbb{N}), \quad \mathcal{P}^2(\mathbb{N}) = \mathcal{P}(\mathcal{P}(\mathbb{N})), \quad \mathcal{P}^3(\mathbb{N}) = \mathcal{P}(\mathcal{P}^2(\mathbb{N})), \quad \dots$$

and let

$$\alpha_1 = \text{card}(\mathcal{P}(\mathbb{N})), \quad \alpha_2 = \text{card}(\mathcal{P}^2(\mathbb{N})), \quad \alpha_3 = \text{card}(\mathcal{P}^3(\mathbb{N})), \quad \dots$$

In other words, if we let $\mathcal{P}(\mathbb{N}) = A$ then $\mathcal{P}^2(\mathbb{N}) = \mathcal{P}(A)$ is the power set of $A = \mathcal{P}(\mathbb{N})$. Theorem 4.4.2 shows us that

$$\alpha_0 = \text{card}(\mathcal{P}(\mathbb{N})) = \text{card}(A) < \text{card}(\mathcal{P}(A)) = \text{card}(\mathcal{P}^2(\mathbb{N})) = \alpha_2.$$

That is, $\alpha_0 < \alpha_1 < \alpha_2$.

In general, let $A = \mathcal{P}^n(\mathbb{N})$. Because

$$\mathcal{P}^{n+1}(\mathbb{N}) = \mathcal{P}(\mathcal{P}^n(\mathbb{N})) = \mathcal{P}(A),$$

Theorem 4.4.2 shows us that

$$\text{card}(\mathcal{P}^n(\mathbb{N})) = \text{card}(A) < \text{card}(\mathcal{P}(A)) = \text{card}(\mathcal{P}(\mathcal{P}^n(\mathbb{N}))) = \text{card}(\mathcal{P}^{n+1}(\mathbb{N})).$$

By setting

$$\alpha_n = \text{card}(\mathcal{P}^n(\mathbb{N})) \quad \text{for each } n \in \mathbb{N}$$

we have constructed *an infinite chain of infinite cardinals*

$$\alpha_0 < \alpha_1 < \dots < \alpha_n < \alpha_{n+1} < \dots$$

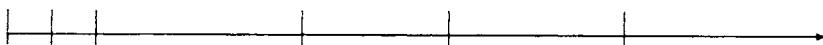
Specifically,

$$\text{card}(\mathbb{N}) < \text{card}(\mathcal{P}(\mathbb{N})) < \text{card}(\mathcal{P}^2(\mathbb{N})) < \text{card}(\mathcal{P}^3(\mathbb{N})) < \dots$$

This ends the proof.

In pictures, the above chain can be realized as in the following line. This is not the real line. It is just a visual means of describing a chain of cardinals.

$$0 \quad 1 \quad 2 \quad \bullet \bullet \bullet \quad \text{card}(\mathbb{N}) \quad \text{card}(\mathcal{P}(\mathbb{N})) \quad \text{card}(\mathcal{P}^2(\mathbb{N})) \quad \bullet \bullet \bullet$$



Let us examine another method for constructing cardinals and infinities. Recall that if A is a set then $\{0, 1\}^A$ is the set of functions $f : A \rightarrow \{0, 1\}$.

Theorem 4.4.5 Let A be a set. Then

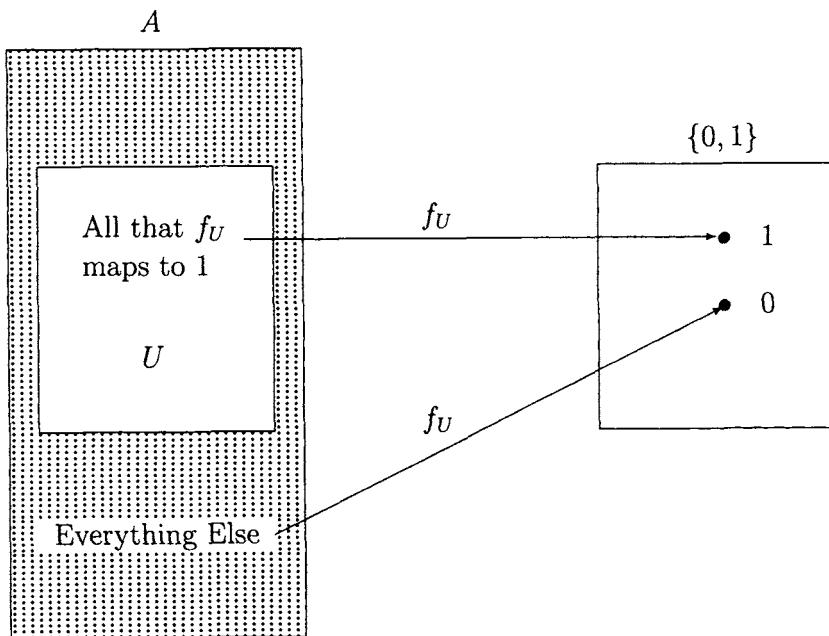
$$\text{card}(\mathcal{P}(A)) = \text{card}(\{0, 1\}^A).$$

Proof: We will produce a bijection

$$F : \mathcal{P}(A) \longrightarrow \{0, 1\}^A$$

from which we will conclude that

$$\text{card}(\mathcal{P}(A)) = \text{card}(\{0, 1\}^A).$$



Given a subset $U \subset A$, define a function

$$f_U : A \longrightarrow \{0, 1\}$$

as in the above diagram.

$$f_U(x) = \begin{cases} 1 & \text{if } x \in U \\ 0 & \text{if } x \notin U \end{cases}.$$

Thus $f_U(x) = 1$ precisely when $x \in U$. Otherwise it is 0. Clearly

$$f_U \in \{0, 1\}^A.$$

so that the rule

$$F(U) = f_U$$

defines a function $F : \mathcal{P}(A) \rightarrow \{0, 1\}^A$. We will show that F is a bijection.

To see that F is one-to-one, let

$$U \neq V$$

be subsets of A . That is, take distinct elements of $\mathcal{P}(A)$. By the definition of equal sets some element x is in one set and not the other. We assume without loss of generality that there is an element $x \in U$ that is not in V . Let f_V be the function that takes V to 1 and all else to 0. Then

$$x \in U \quad \text{and} \quad x \notin V,$$

so that

$$f_U(x) = 1 \quad \text{while} \quad f_V(x) = 0$$

by the definitions of the functions f_U and f_V . Then f_U and f_V have different rules

$$F(U) = f_U \neq f_V = F(V),$$

so that $F : \mathcal{P}(A) \rightarrow \{0, 1\}^A$ is a one-to-one function.

To see that $F : \mathcal{P}(A) \rightarrow \{0, 1\}^A$ is an onto function, let $f \in \{0, 1\}^A$. Define a subset of A ,

$$U = \{x \in A \mid f(x) = 1\}.$$

That is, $U = f^{-1}(1)$, or in other words, U is the set of all $x \in A$ that are sent by f to 1. (See the above diagram.) Then

$$\begin{aligned} f(x) &= 1 &= f_U(x) &\quad \text{for all } x \in U \text{ and} \\ f(x) &= 0 &= f_U(x) &\quad \text{for all } x \notin U. \end{aligned}$$

Since f and f_U have the same rules, $f = f_U$, and since $F(U) = f_U$ we have

$$f = f_U = F(U).$$

It follows that $F : \mathcal{P}(A) \longrightarrow \{0, 1\}^A$ is onto.

Hence F is a bijection and therefore

$$\text{card}(\mathcal{P}(A)) = \text{card}(\{0, 1\}^A).$$

This completes the proof.

Theorem 4.4.6 1. $\text{card}(\mathbb{N}) < \text{card}(\{0, 1\}^{\mathbb{N}})$.

2. $\text{card}(\mathbb{R}) < \text{card}(\{0, 1\}^{\mathbb{R}})$.

Proof: 1. By Theorem 4.4.2, $\text{card}(\mathbb{N}) < \text{card}(\mathcal{P}(\mathbb{N}))$ and by Theorem 4.4.5, $\text{card}(\mathcal{P}(\mathbb{N})) = \text{card}(\{0, 1\}^{\mathbb{N}})$. Thus $\text{card}(\mathbb{N})$ is strictly smaller than $\text{card}(\{0, 1\}^{\mathbb{N}})$. Part 2 is handled in a similar fashion. This completes the proof.

The notation we will use for the exponents is also an exponent. Given a set A let

$$\text{card}(A) = \aleph$$

(read as *aleph*) and then let

$$\text{card}(\{0, 1\}^A) = 2^\aleph.$$

This does not imply that we are taking a power of 2 to some infinite cardinal. It is just a convenient method for writing some cardinalities. There is something more to see here though.

Theorem 4.4.7 *Let \aleph be a cardinal. That is, $\aleph = \text{card}(A)$ for some set A . Then*

$$\aleph < 2^\aleph.$$

Proof: From Theorems 4.4.2 and 4.4.5 we see that

$$\aleph = \text{card}(A) < \text{card}(\mathcal{P}(A)) = \text{card}(\{0, 1\}^A) = 2^{\text{card}(A)} = 2^\aleph.$$

This completes the proof.

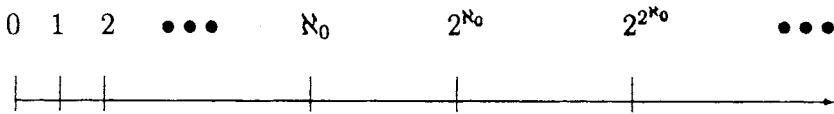
It is a tradition going back over 100 years to let

$$\text{card}(\mathbb{N}) = \aleph_0,$$

where \aleph_0 is read as *aleph nought*. Then with Theorem 4.4.5 and our new notation we can write

$$\begin{aligned}\aleph_0 &= \text{card}(\mathbb{N}), \\ 2^{\aleph_0} &= \text{card}(\{0, 1\}^{\aleph_0}) = \text{card}(\mathcal{P}(\mathbb{N})), \\ 2^{2^{\aleph_0}} &= \text{card}(\{0, 1\}^{\{0, 1\}^{\aleph_0}}) = \text{card}(\mathcal{P}(\mathcal{P}(\mathbb{N}))), \\ &\vdots\end{aligned}$$

which by Theorem 4.4.5 leads us to an infinite sequence of infinite cardinals



The next question we will explore is, can we find some inequality between $\text{card}(\{0, 1\}^{\mathbb{N}})$ and $\text{card}(\mathbb{R})$?

4.5 The Arithmetic of Cardinals

We will use the Greek letters α (alpha) and β (beta) and the Hebrew letter \aleph (aleph) to denote general cardinals. We will also use the traditional notation of

$$\aleph_0 = \text{card}(\mathbb{N})$$

(pronounced *aleph nought*). It was Georg Cantor in his ground breaking paper on cardinal numbers [1] who first used the notation \aleph_0 for cardinals.

What you are about to read should be interesting. We are about to define the addition and multiplication of natural numbers. This is the arithmetic that you learned in elementary school. We will define $1 + 1$ and show that it is 2. We will define $2 \cdot 3$ and show that it is 6. We will eventually use this arithmetic to define operations like addition and multiplication on infinite cardinals. And surprise! $\aleph_0 + \aleph_0 = \aleph_0$. Now isn't that what you expected of infinity? What would you expect of the following sum?

$$\aleph_0 + 2^{\aleph_0} = ?$$

Let us define *the addition of natural numbers*. Sets E and F are said to be *disjoint* if $E \cap F = \emptyset$.

Definition 4.5.1 Let $n, m \in \mathbb{N}$. Choose disjoint finite sets E and F such that $\text{card}(E) = n$ and $\text{card}(F) = m$. Then we define

$$n + m = \text{card}(E \cup F).$$

As our first example of how to use this definition, we will prove an elementary arithmetic fact. Recall that 1 is the set of all sets that are equivalent to the set $\{\bullet\}$, and similarly 2 denotes the set of all sets that are equivalent to $\{\bullet, \square\}$. Then

$$\text{card}(\{\bullet\}) = \text{card}(\{\square\}) = 1$$

so that

$$\begin{aligned} 1 + 1 &= \text{card}(\{\bullet\} \cup \{\square\}) \\ &= \text{card}(\{\bullet, \square\}) \\ &= 2 \end{aligned}$$

We have just proved that $1 + 1 = 2$. We did not state this numerical identity, we did not ask for measurements, and we did not accept the teacher's word for it. We proved it with the same certainty that a proof about congruent triangles would possess in a geometry

course. Thus set theory is a point of view that allows us to examine all of the mathematical fundamentals. That is its power.

This idea of considering the addition of natural numbers as the union of finite sets is what we will use to define *the addition of infinite cardinals*.

Definition 4.5.2 Suppose that α and β are infinite cardinals. Choose any sets A and B such that $\text{card}(A) = \alpha$ and $\text{card}(B) = \beta$. Then we define

$$\alpha + \beta = \text{card}(A \cup B).$$

The power here is that to add the infinite cardinals α and β we may choose any of the sets A and B for which $\text{card}(A) = \alpha$ and $\text{card}(B) = \beta$. Unlike the addition of finite cardinals, the sets A and B do not have to be disjoint when adding $\text{card}(A)$ and $\text{card}(B)$. This statement is actually a theorem. Professionals use a weaker statement as a definition. But for our purposes, this definition is a good working definition. Thus $\alpha + \beta$ is independent of our choice of sets A and B . The professional would say that the definition of the addition of infinite cardinals is *well defined*.

The addition of infinite cardinals has some of the properties associated with the addition of real numbers, and as we will see the addition of infinite cardinals can be quite unexpected. One similarity is the *commutative property of addition*. Given any cardinals α and β , then

$$\alpha + \beta = \beta + \alpha.$$

To see the truth of this matter choose sets A and B such that $\alpha = \text{card}(A)$ and $\beta = \text{card}(B)$. Then

$$\begin{aligned} \alpha + \beta &= \text{card}(A \cup B) \\ &= \text{card}(B \cup A) \\ &= \beta + \alpha, \end{aligned}$$

which is what we wanted to prove. We have just *proved* that addition is commutative. This is a fact that has been handed down to you through educational tradition but no one (I'm betting) in your past proved that this property was true of *all* cardinals or numbers.

For example, if we use the traditional symbol

$$\mathfrak{c} = \text{card}(\mathbb{R}),$$

then

$$\begin{aligned}\aleph_0 + \mathfrak{c} &= \mathfrak{c} + \aleph_0 \text{ and} \\ 2^{\aleph_0} + \mathfrak{c} &= \mathfrak{c} + 2^{\aleph_0}.\end{aligned}$$

The symbol \mathfrak{c} stands for *the continuum*, an old term used to describe the real line. This shows us that the addition is commutative but it does not show us what the indicated sums are.

As children we might have heard an adage that *infinity plus 1 is infinity* or that *infinity plus infinity is infinity*. This is mildly true since any number of elements added to an infinite set still yields an infinite set. However, as we saw in the previous sections, the term *infinite* is just too coarse a measurement for the size of a set. It is like saying, *the house has color*. Sure it has color, but which ones? Saying that something is infinite simply says that it is not finite. It does not tell us a thing about the cardinality of the set. Is the cardinality \aleph_0 , or \mathfrak{c} , or some other value?

As an example of how to add infinite cardinals, consider the following result. One might read this as the old adage *Infinity plus one is infinity*. Our work to date brings mathematical truth to this adage.

Theorem 4.5.3 *Let \aleph be any infinite cardinal.*

1. $0 + \aleph = \aleph$.

2. *If $n > 0$ is a natural number then $n + \aleph = \aleph$.*

Proof: 1. By definition there is a set A such that $\aleph = \text{card}(A)$, and by definition

$$0 = \text{card}(\emptyset).$$

Then we have

$$\begin{aligned} 0 + \aleph &= \text{card}(\emptyset \cup A) \\ &= \text{card}(A) \\ &= \aleph. \end{aligned}$$

This proves part 1.

2. Since A is infinite and n is finite, there is a set $\{a_1, \dots, a_n\} \subset A$ of n elements contained in A . Then

$$n = \text{card}\{a_1, \dots, a_n\}$$

and so

$$\begin{aligned} n + \aleph &= \text{card}(\{a_1, \dots, a_n\} \cup A) \\ &= \text{card}(A) \\ &= \aleph. \end{aligned}$$

This proves part 2, which completes the proof.

For example, part 2 above shows us that

$$\aleph_0 + 1 = \aleph_0$$

and that

$$\aleph_0 + 10^{10^{10^{10^{10}}}} = \aleph_0$$

even though $10^{10^{10^{10^{10}}}}$ is a huge finite natural number.

The next result might be read as *infinity plus infinity is infinity*. This is a true statement, but we can say with certainty which infinite cardinal the sum is.

Theorem 4.5.4 *Let \aleph be any infinite cardinal. Then*

$$\aleph + \aleph = \aleph.$$

Proof: We know that $\aleph = \text{card}(A)$ for some set A . Hence

$$\begin{aligned}\aleph + \aleph &= \text{card}(A \cup A) \\ &= \text{card}(A) \\ &= \aleph.\end{aligned}$$

This ends the proof.

Then, for example,

$$\aleph_0 + \aleph_0 = \aleph_0$$

and

$$\mathfrak{c} + \mathfrak{c} = \mathfrak{c}.$$

and for any infinite cardinal \aleph ,

$$2^\aleph + 2^\aleph = 2^\aleph.$$

You see? An infinite cardinal plus the same infinite cardinal is the same infinite cardinal. Or in other words, any infinite cardinal added to itself is itself.

For that reason, we will not subtract cardinals because *subtraction makes no sense on infinite cardinals*. To see the truth of this matter, consider the following example.

Example 4.5.5 We will show the following.

There is no way to define $\aleph_0 - \aleph_0$ as a cardinal.

Assume to the contrary that we could define the subtraction $\aleph_0 - \aleph_0$ as the opposite of addition. The implied value of the subtraction $\aleph_0 - \aleph_0$ is

$$\aleph_0 - \aleph_0 = 0.$$

Now, by part 2 of the above theorem,

$$1 + \aleph_0 = \aleph_0.$$

Using the assumed subtraction, we would then subtract \aleph_0 from both sides of the equation, thus revealing that

$$\begin{aligned} 1 + (\aleph_0 - \aleph_0) &= \aleph_0 - \aleph_0 \\ 1 + 0 &= 0 \\ 1 &= 0, \end{aligned}$$

which is ridiculous. This mathematical mistake shows us that our initial assumption is wrong, and thus $\aleph_0 - \aleph_0$ cannot be defined. This concludes the proof.

In fact, given *any* cardinal \aleph we cannot define $\aleph - \aleph$. If we cannot define $\aleph - \aleph$ then *a general subtraction of cardinals cannot be defined*.

So let us decide what the addition of two cardinals is. That is, what is $\alpha + \beta$? The next result gives further mathematical certainty to the old chestnut *infinity plus infinity is infinity*.

Theorem 4.5.6 *Let $\alpha < \beta$ be infinite cardinals. Then*

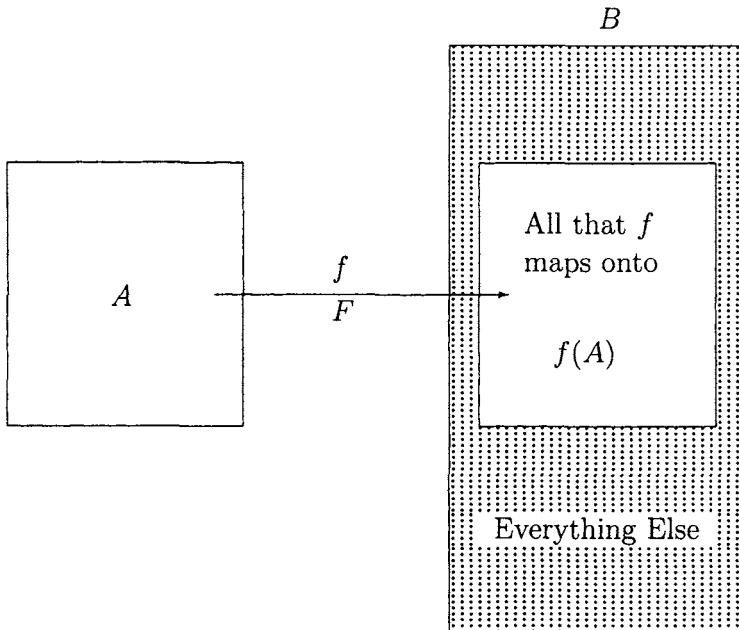
$$\alpha + \beta = \beta = \text{the larger of two cardinals.}$$

Proof: To prove this result we will need sets A and B such that $\alpha = \text{card}(A)$ and $\beta = \text{card}(B)$. By the definition of inequality of cardinals (see page 117), there is a one-to-one function

$$f : A \longrightarrow B.$$

We will show that A and $f(A)$ are equivalent sets. The function $f : A \longrightarrow B$ restricts to one $F : A \longrightarrow f(A)$. Thus f and F have the same rule but they map into different sets. The following

picture will help you see what we are doing with f , F , and $f(A)$.



Specifically, f may not be onto, but we will show that F is a bijection. The function F is one-to-one because that is how we chose f . To see that F is onto, let $y \in f(A)$. By definition of $f(A)$ there is an $x \in A$ such that

$$F(x) = f(x) = y.$$

Then F is onto, hence F is a bijection, whence A and $f(A)$ are equivalent.

Since $f(A) \subset B$, $f(A) \cup B = B$, so that

$$\begin{aligned} \alpha + \beta &= \text{card}(A) + \text{card}(B) \\ &= \text{card}(f(A)) + \text{card}(B) \\ &= \text{card}(f(A) \cup B) \\ &= \text{card}(B) \\ &= \beta. \end{aligned}$$

This completes the proof.

For more familiar cardinals there are the following examples.

$$\aleph_0 + \mathfrak{c} = \mathfrak{c}$$

because by Theorem 4.3.1

$$\text{card}(\mathbb{N}) = \aleph_0 < \mathfrak{c} = \text{card}(\mathbb{R}).$$

Furthermore,

$$\aleph_0 + 2^{\aleph_0} = 2^{\aleph_0},$$

because by Theorem 4.4.7

$$\aleph_0 < 2^{\aleph_0}.$$

In general, for infinite cardinals \aleph we will apply Theorem 4.4.7,

$$\aleph < 2^\aleph,$$

to see that

$$\aleph + 2^\aleph = 2^\aleph.$$

A curious consequence of this addition is that \aleph_0 behaves like 0 on the infinite cardinals. You see

$$\aleph_0 \leq \aleph$$

for each infinite cardinal \aleph , so by Theorem 4.5.6

$$\aleph_0 + \aleph = \aleph \text{ for infinite cardinals } \aleph.$$

Don't try to cancel that \aleph from the equation. In general, the cancellation of infinite cardinals does not work. Again, our finite intuition is worthless in the infinite setting.

Since we can add cardinals to themselves we can define *multiples of cardinals*. Let \aleph be an infinite cardinal and let $n \in \mathbb{N}$. Then

$$n \cdot \aleph = \underbrace{\aleph + \dots + \aleph}_{n \text{ summands}} = \aleph.$$

Specifically,

$$\begin{aligned} 2 \cdot \aleph_0 &= \aleph_0 \text{ and} \\ n \cdot c &= c \text{ for each } n \in \mathbb{N}. \end{aligned}$$

It is important for us to see that

this multiplication is not associated with a division of cardinals.

Here's why. Assume to the contrary that we can divide by \aleph_0 . One of the fundamental identities implied by the existence of a division is that

$$\aleph_0 \cdot \frac{1}{\aleph_0} = 1.$$

By applying the assumed division we would then have the following series of equations:

$$\begin{aligned} 2 \cdot \aleph_0 &= \aleph_0 \\ 2 \cdot \aleph_0 \cdot \frac{1}{\aleph_0} &= \aleph_0 \cdot \frac{1}{\aleph_0} \\ 2 \cdot 1 &= 1 \\ 2 &= 1, \end{aligned}$$

another clear contradiction. Therefore our initial assumption must be tossed out and we conclude that *we cannot divide by \aleph_0* . This may be your first experience with a multiplication that is not associated with a division. It might seem strange at first but the laws of mathematics demand it.

Moving on to the *multiplication of infinite cardinals*, observe that

$$\{1, 2\} \times \{1, 2, 3\}$$

consists of the following 6 pairs.

$$\begin{array}{lll} (1, 1) & (1, 2) & (1, 3) \\ (2, 1) & (2, 2) & (2, 3) \end{array}$$

We have just proved the arithmetic fact

$$2 \times 3 = 6.$$

You should be sitting straight up in your chair when I tell you something like this. Here is a fundamental concept of human intellect and I am showing you how to prove it from more basic facts.

The reader can write down the 15 pairs that make up $\{1, 2, 3\} \times \{1, 2, 3, 4, 5\}$. The point behind these examples is that *if E and F are finite sets*, then we have

$$\text{card}(E \times F) = \text{card}(E)\text{card}(F).$$

To see the truth of the matter suppose that $E = \{1, \dots, m\}$ and that $F = \{1, \dots, n\}$. Arrange $E \times F$ as a rectangular array of pairs (x, y) .

$$\begin{array}{llll} (1, 1) & (1, 2) & \dots & (1, n) \\ (2, 1) & (2, 2) & \dots & (2, n) \\ \vdots & & & \vdots \\ (m, 1) & (m, 2) & \dots & (m, n) \end{array}$$

There are $m = \text{card}(E)$ columns and $n = \text{card}(F)$ rows in this array so there must be $m \cdot n = \text{card}(E)\text{card}(F)$ pairs in this rectangular array. This is what we wanted to prove.

Thus motivated, it is reasonable to define the product of infinite cardinals as follows.

Definition 4.5.7 Let α and β be infinite cardinals. Choose sets A and B such that $\text{card}(A) = \alpha$ and $\text{card}(B) = \beta$. We define

$$\alpha \cdot \beta = \text{card}(A \times B).$$

As before, we want to verify that *the multiplication of cardinals is commutative*. That is,

$$\alpha \cdot \beta = \beta \cdot \alpha.$$

To see this, choose sets A and B such that $\alpha = \text{card}(A)$ and $\beta = \text{card}(B)$. Then the reader is invited to show that the function

$$f : A \times B \longrightarrow B \times A$$

defined by switching components

$$f(a, b) = (b, a)$$

is a bijection. Then $\text{card}(A \times B) = \text{card}(B \times A)$ and hence

$$\begin{aligned} \alpha \cdot \beta &= \text{card}(A \times B) \\ &= \text{card}(B \times A) \\ &= \beta \cdot \alpha. \end{aligned}$$

This is what we wanted to prove.

Thus, like the multiplication of numbers, the multiplication of infinite cardinals is commutative. This may be the totality of the similarity between the multiplication of natural numbers, and the multiplication of infinite cardinals.

Let us examine the product $\aleph_0 \cdot \aleph_0$ under this multiplication. Since $\aleph_0 = \text{card}(\mathbb{N})$ we can write

$$\begin{aligned} \aleph_0 \cdot \aleph_0 &= \text{card}(\mathbb{N} \times \mathbb{N}) \\ &= \text{card}(\mathbb{N}) \quad \text{by Theorem 4.1.5} \\ &= \aleph_0. \end{aligned}$$

That is,

$$\aleph_0 \cdot \aleph_0 = \aleph_0.$$

This multiplication must strike you as strange. The only real numbers x for which $x \cdot x = x$ are 0 and 1. And \aleph_0 is not special in this regard. In general,

$$\aleph \cdot \aleph = \aleph \quad \text{for any infinite cardinal } \aleph.$$

Furthermore, since $\aleph \cdot \aleph = \aleph^2$ we are led to consider square roots of cardinals. Things are not what our old common sense might have said to us.

$$\sqrt{\aleph} = \aleph \quad \text{for any infinite cardinal } \aleph.$$

This is another difference between natural numbers and infinite cardinals. The square root of most natural numbers is not a natural number. However, the square root of *any* infinite cardinal is an infinite cardinal. Here is another peculiarity. Question: How many natural numbers do you know of that are their own square roots? Answer: Only two, 0 and 1. And yet *every* infinite cardinal is its own square root. The land of infinite cardinals is indeed strange.

Here is a property of multiplication that may not agree with your calculus lectures. (Skip this discussion if you have never had calculus.) Limits of the form $\frac{0}{0}$ require an application of L'Hopital's Rule. That is, if $f(x)$ and $g(x)$ are functions such that $f(a) = g(a) = 0$, then the limit

$$\lim_{x \rightarrow a} \frac{f(x)}{g(x)}$$

can be almost any real number. The value of the limit depends on the functions $f(x)$ and $g(x)$. For example,

$$\lim_{x \rightarrow 0} \frac{\sin(x)}{x} = 1$$

and

$$\lim_{x \rightarrow 1} \frac{2x^2 - 1}{x^3 + 2x + 1} = 2.$$

Specifically, the limits of the form $0 \cdot \infty$ can have any value. In set theory no such rule is necessary.

Theorem 4.5.8 *Let \aleph be an infinite cardinal. Then*

$$0 \cdot \aleph = 0.$$

Proof: This one is tricky. There is a set A such that $\text{card}(A) = \aleph$. See Example 1.2.2 to see that $\emptyset \times A = \emptyset$ for any set A . Hence $0 \cdot \aleph = \text{card}(\emptyset \times A) = 0$. This completes the proof.

The following theorem shows us that like addition, the multiplication of cardinals is in some respects different from the multiplication of real numbers. One major difference between the multiplication of numbers and infinite cardinals is the following rule of evaluation.

Theorem 4.5.9 *Let $\alpha \leq \beta$ be infinite cardinals. Then*

$$\alpha \cdot \beta = \beta = \text{the larger of the two cardinals.}$$

We will not prove this one, but we offer the following examples.

$$\aleph_0 \cdot \aleph_0 = \aleph_0,$$

$$\mathfrak{c} \cdot \mathfrak{c} = \mathfrak{c},$$

$$\aleph_0 \cdot \mathfrak{c} = \mathfrak{c}, \text{ by Theorem 4.3.1,}$$

$$\aleph_0 \cdot \aleph = \aleph \text{ for any infinite cardinal } \aleph,$$

$$\aleph \cdot \aleph \cdot \aleph = \aleph \text{ for any infinite cardinal } \aleph.$$

The results in this section show us that the arithmetic of infinite cardinals will require some thought before we can apply it or manipulate with it. Read the above arithmetic facts while reminding yourself that these equations do not hold for real numbers other than 0 and 1.

Inasmuch as we have defined

$$\aleph^2 = \aleph \cdot \aleph,$$

it is important that we define powers of cardinals. Recall that given sets A and B

$$B^A \text{ is the set of all functions } f : A \longrightarrow B.$$

Let us write down the rules of all of the functions in $\{1, 2\}^{\{1, 2, 3\}}$. A function $f : \{1, 2, 3\} \longrightarrow \{1, 2\}$ is given by its images of 1, 2, and 3. Thus we can write down the rule for f by just listing how it behaves on 1, 2, and 3. A very effective method is to list

$$f = [a, b, c],$$

where $a, b, c \in \{1, 2\}$ and where

$$f(1) = a, f(2) = b, f(3) = c.$$

For instance, suppose that we have a function

$$f : \{1, 2, 3\} \longrightarrow \{1, 2\}$$

whose rule is given by

$$f(1) = 1, f(2) = 1, f(3) = 1.$$

Then the very effective method for writing down f is

$$f = [1, 1, 1].$$

The function

$$g : \{1, 2, 3\} \longrightarrow \{1, 2\}$$

whose rule is given by

$$g(1) = 2, g(2) = 1, g(3) = 2$$

is written down as

$$g = [2, 1, 2].$$

Furthermore, given a triple

$$[1, 2, 2]$$

we have defined a function

$$h : \{1, 2, 3\} \longrightarrow \{1, 2\}$$

whose rule is given by

$$h(1) = 1, h(2) = 2, h(3) = 2.$$

Let us count the number of functions in $\{1, 2\}^{\{1, 2, 3\}}$. Since $[a, b, c]$ is a random function and since there are 2 choices for a , 2 choices for b , and 2 choices for c we see that there are 2^3 functions f in $\{1, 2\}^{\{1, 2, 3\}}$. That is,

$$\text{card}(\{1, 2\}^{\{1, 2, 3\}}) = 2^3 = \text{card}(\{1, 2\})^{\text{card}(\{1, 2, 3\})}.$$

In general, if $E = \{1, \dots, n\}$ and $F = \{1, \dots, m\}$ are finite sets then each function in F^E can be written as a finite sequence

$$[y_1, y_2, \dots, y_n],$$

where $y_1, \dots, y_n \in F$ and where

$$f(1) = y_1, \dots, f(n) = y_n.$$

For each k there are exactly m choices y_k for the image $f(k)$ under f . Since there are m choices made for each of the n values y_1, \dots, y_n , there are exactly m^n functions in F^E . In other words, if E and F are finite sets then

$$\boxed{\text{card}(F^E) = \text{card}(E)^{\text{card}(F)}}.$$

This motivates the following natural definition of exponential values of infinite cardinals.

Definition 4.5.10 Let α and β be cardinals. Choose any sets A and B such that $\alpha = \text{card}(A)$ and $\beta = \text{card}(B)$. Then

$$\boxed{\beta^\alpha = \text{card}(B^A)}.$$

In particular,

$$2^{\aleph_0} = \text{card}(\{0, 1\})^{\text{card}(\mathbb{N})} = \text{card}(\{0, 1\}^{\mathbb{N}}).$$

Theorems 4.4.2 and 4.4.5 show us that

$$\text{card}(\mathbb{N}) < \text{card}(\mathcal{P}(\mathbb{N})) = \text{card}(\{0, 1\}^{\mathbb{N}})$$

for the *infinite set* \mathbb{N} so that by Theorem 4.3.1

$$\aleph_0 < 2^{\aleph_0}.$$

Consider this inequality in a more general context. Let \aleph be an infinite cardinal and let A be a set such that $\aleph = \text{card}(A)$. Then

$$2^\aleph = \text{card}(\{0, 1\}^A)$$

and hence Theorem 4.4.5 shows us that

$$\aleph < 2^\aleph.$$

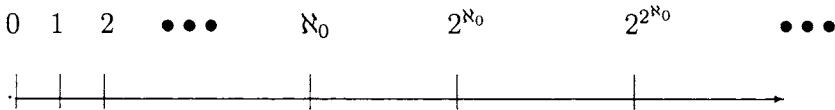
Moreover, since 2^\aleph is also a cardinal, we can form the cardinal

$$2^{2^\aleph},$$

which is properly larger than 2^\aleph . Iterating this process beginning with \aleph_0 produces *an infinite chain of infinite cardinals*

$$\aleph_0 < 2^{\aleph_0} < 2^{2^{\aleph_0}} < 2^{2^{2^{\aleph_0}}} < \dots$$

The diagram below illustrates these inequalities. *This is not the real line.* It is just a visual means of describing a chain of cardinals.



We leave this section with a question and a list of small infinite cardinals that summarizes our research to date.

$$\aleph_0 = \text{card}(\mathbb{N}) = \text{card}(\mathbb{Q}) = \text{card}(\mathbb{N} \times \mathbb{N})$$

$$\text{card}(0, 1) = \text{card}(\mathbb{R}) = \text{card}(\mathbb{R}^3) = \text{card}(\mathcal{P}(\mathbb{N})) = 2^{\aleph_0}$$

$$\aleph_0 < 2^{\aleph_0} < 2^{2^{\aleph_0}} < 2^{2^{2^{\aleph_0}}} < \dots$$

Question: Is there a cardinal between $\text{card}(\mathbb{N})$ and $\text{card}(\mathbb{R})$? Specifically, is there a cardinal \aleph_1 such that

$$\text{card}(\mathbb{N}) < \aleph_1 < \text{card}(\mathbb{R})?$$

Chapter 5

Well Ordered Sets

5.1 Successors of Elements

The set of natural numbers \mathbb{N} has an interesting property.

The Well Ordered Property: Given any element $n \in \mathbb{N}$ there is a *unique* next element or *successor* element n^+ .

Thus there is exactly one successor element n^+ . You know n^+ as $n + 1$, for example, $0^+ = 1$, $12^+ = 13$, and $121^+ = 122$. There is no element $x \in \mathbb{N}$ such that $n^+ = 0$. Other common sets have this property and we tend to think that *every* collection has the Well Ordered Property. For example, suppose that $\mathbf{P} = \{P_1, P_2, P_3\}$ is a set of three people. Our culture automatically endows \mathbf{P} with several well orderings. We give \mathbf{P} a well ordering by saying that the smallest element will be the shortest person, say, P_1 , the next element will be the next tallest person, say, P_2 , and then the tallest person, P_3 . We would write

$$P_1 < P_2 < P_3.$$

Good. This is a well ordering of \mathbf{P} by saying that the symbol $<$ points to the shorter person. Furthermore,

$$P_1^+ = P_2, \quad P_2^+ = P_3.$$

There is no element x in \mathbf{P} such that $x^+ = P_1$ or $P_3^+ = x$.

Height is not the only way to well order \mathbf{P} . Suppose that P_1 is older than P_3 and that P_3 is older than P_2 . Then we might well order \mathbf{P} by age by writing

$$P_2 < P_3 < P_1.$$

In this case $<$ points to the younger person. Thus

$$P_2^+ = P_3, P_3^+ = P_1$$

under the age well ordering. There is no element x in \mathbf{P} such that $x^+ = P_2$ or $P_1^+ = x$.

You might also well order \mathbf{P} alphabetically on the last names of the people in \mathbf{P} . Thus if P_1 is named Himmel, if P_2 is named Frugal, and if P_3 is named Briar, then under the alphabet ordering

$$P_3 < P_2 < P_1.$$

Thus $P_3^+ = P_2$, $P_2^+ = P_1$, and there is no element x in \mathbf{P} such that $P_1^+ = x$.

We have ordered \mathbf{P} in different ways by considering different numbers associated with people. We are constantly well ordering people and their lives when we say “I love you more” or “You’re so much better than him.” These sentences imply that given any two people we can assign an inequality to them:

$$P_1 < P_2 \text{ if } P_1 \text{ is “better than” } P_2.$$

In this example $<$ points to the better person. Most of us well order people on a continual basis even when an ordering is not called for. After all, can a parent declare which of their children they most love? I don’t think so.

The *power set* $\mathcal{P}(A)$ of a set A is a naturally occurring example of a set that is not linearly ordered in the natural way. We define an order $<$ on $\mathcal{P}(A)$ that is not a linear order as follows.

Given $X, Y \in \mathcal{P}(A)$ then $X < Y$ if $X \subset Y$.

Take, for example, $\mathcal{P}(\{1, 2, 3\})$. The elements $\{2\}$, $\{1, 2\}$, and $\{2, 3\} \in \mathcal{P}(\{1, 2, 3\})$ satisfy

$$\{2\} \subset \{1, 2\} \quad \text{and} \quad \{2\} \subset \{2, 3\}.$$

We might consider the two elements to be successors of $\{2\}$. Since $\{2\}$ does not have a *unique* successor element, $\mathcal{P}(\{1, 2, 3\})$ is not a well ordered set. A similar argument shows that \emptyset^+ does not exist because it cannot be defined *uniquely* under the subset ordering $<$.

$$\emptyset \subset \{1\}, \quad \emptyset \subset \{2\}, \quad \emptyset \subset \{3\}.$$

However, we can define an *abstract* well ordering \leq on $\mathcal{P}(\{1, 2, 3\})$ by picking an ordering at random on the 8 elements of $\mathcal{P}(\{1, 2, 3\})$ as follows.

$$\{1\} \leq \emptyset \leq \{1, 2, 3\} \leq \{2\} \leq \{2, 3\} \leq \{1, 3\} \leq \{1, 2\} \leq \{3\}.$$

Notice that this ordering does not have much to do with the subset inclusion \subset . It is simply an ordering that your author chose to make a point. Abstract well orderings are still well orderings, and they will be treated as legitimate well orderings on sets.

The set \mathbb{R} of *real numbers* is an example of a set that possesses a linear order $<$ but that is not well ordered. For example, $\sqrt{2}$ is a real number but there is no *next real number*, $\sqrt{2}^+$. That is, there is no real number that is next on the real number line after $\sqrt{2}^+$. $\sqrt{2} + 1$ will not do since

$$\sqrt{2} < \sqrt{2} + \frac{1}{2} < \sqrt{2} + 1.$$

There should be nothing between a number and its successor.

It is also apparent that \mathbb{Q} is not well ordered. 0^+ does not exist as a *rational number*. We cannot use 1 as 0^+ in \mathbb{Q} since 1 is not the next *rational* number larger than 0. This is clear once we write

$$0 < \frac{1}{2} < 1.$$

So well ordered sets are going to be thin sets. Given an element x , there is a next value or successor value x^+ .

There is another interesting example of a set that is not well ordered, but is still quite thin. Let

$$W = \left\{ 0, 1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots \right\}$$

and let $<$ be the usual ordering of real numbers. Then in the set W , $\frac{1}{2}^+ = 1$, $\frac{1}{4}^+ = \frac{1}{3}$, but 0^+ is not to be found in W . Let's argue about that. Suppose Q^+ exists in W . Then

$$0^+ = \frac{1}{n} \text{ for some } n \in \mathbb{N}.$$

But then $\frac{1}{n}$ should be the next value in the set that is larger than 0. This is not the case since

$$0 < \frac{1}{n+1} < \frac{1}{n}.$$

Hence, in W , 0^+ does not exist. Therefore W is not well ordered.

I leave it to the reader as a challenge problem to justify to yourself that the set

$$U = \left\{ 1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots \right\}$$

is well ordered. Just try to find x^+ for each $x \in U$.

The well ordered set of natural numbers \mathbb{N} satisfies some interesting properties that we will use later in a more general setting. \mathbb{N} satisfies the following property.

The Trichotomy Property: Given $n, m \in \mathbb{N}$ then *exactly one* of the following relationships holds.

$$n < m \text{ or } m < n \text{ or } n = m.$$

For example, the solution to $2 = x + 1$ satisfies exactly one of the comparisons $x < 0$, $0 < x$, or $x = 0$.

Another example is the age n_o of the youngest mathematics professor in Canada. This number n_o satisfies exactly one of the following options. Either

- $n_o < 22$ and we say that the mathematics professor is younger than 22 ,
- or $n_o > 22$ and we say that the mathematics professor is older than 22 ,
- or $n_o = 22$ and we say that the mathematics professor is 22 years old.

Even though we do not know the age n_o , we can still state that it satisfies the Trichotomy Property.

A property closely aligned with the Trichotomy Property is the following.

The Minimum Property: If A is a well ordered set then each subset of A has a unique minimum element.

Each nonempty set in \mathbb{N} contains a unique least element. That is, given $X \subset \mathbb{N}$ there is an element $x_0 \in X$ such that $x_0 \leq x$ for each $x \in X$. For example, $\{7, 8, 9, 10, 11\} \subset \mathbb{N}$ has least element 7, while

$$\{n \mid n \text{ is the age in years of a Canadian mathematician}\}$$

has a unique least age n_o that we cannot specifically identify. We would all agree, though, that this age n_o is somewhat more than 10 and somewhat less than 100.

A hard to verify minimum natural number is the minimum naturally occurring temperature in degrees Celsius on the surface of the Earth. For our purposes here this temperature will not have a decimal value, but only a whole number value. While we can safely say that this minimum temperature is smaller than 100°C and more than -100°C , we cannot say what that minimum temperature is without first measuring temperatures everywhere on the Earth's surface. However, the Minimum Property tells us that such a minimum temperature exists, whether or not we have made the measurements.

If you are given a set of people measured only as closely as inches (i.e., their heights are in inches but not in fractions of inches), then these heights have a unique minimum height. It might be that the heights are smaller than 120 inches and that they are greater than 1

inch, but there is a unique minimum height h_o and there is a person whose height is h_o .

The Well Ordered Property and the Minimum Property combine as follows. If $n \in \mathbb{N}$ then

n^+ is the smallest of all the elements that are larger than n .

In other words,

if $n < m \in \mathbb{N}$ then $n^+ \leq m$.

For example, $1 < 5$ so that $1^+ = 2 \leq 5$; $12 < 14$ so that $12^+ \leq 14$; and $121 < 122$ so that $121^+ \leq 122$.

Another example is as follows. We will argue in this chapter that the set of cardinals is a well ordered set. A complete justification is well beyond the scope of this book. The framed inequality on page 161 shows us that

$$\aleph_0 < 2^{\aleph_0}.$$

So where is the successor \aleph_0^+ to \aleph_0 ? The best we can do at this time is to say that

$$\aleph_0 < \aleph_0^+ \leq 2^{\aleph_0}.$$

Furthermore, $\aleph_0^+ \neq \aleph_0 + 1 = \aleph_0$ since, as must be apparent, $\aleph_0^+ \neq \aleph_0 = \aleph_0 + 1$.

With these properties in mind let us abstract the notion of a well ordered set.

Definition 5.1.1 Let A be any set. We say that A is a well ordered set if it satisfies the following two properties.

1. A satisfies the Trichotomy Property. That is, given $x, y \in A$ then x and y satisfy exactly one of the following options.

$$x < y, y < x, \text{ or } x = y.$$

2. *A satisfies the Minimum Property. That is, each subset of A contains a unique least element. Equivalently, to each element $x \in A$ there is a unique element $x^+ \in A$ such that*

$$\text{if } y \in A \text{ and if } x < y \text{ then } x^+ \leq y.$$

We call x^+ the successor of x .

Notice that the element $x^+ \neq x + 1$ since we do not know if 1 is in A or if there is a way to add the elements of A .

Under this definition \mathbb{N} is a well ordered set. We have already discussed that \mathbb{N} satisfies the Trichotomy Property, and that each element $n \in \mathbb{N}$ possesses a successor element $n^+ = n + 1$.

We can also show that \mathbb{N} satisfies the Minimum Property. Each subset of \mathbb{N} has a least element found as follows. Let $X \subset \mathbb{N}$ and assume that $X \neq \emptyset$. If $0 \in X$ then we are done. 0 is the smallest element of \mathbb{N} . Otherwise, test $1 \in X$. If it is then it is the smallest element of X . Continue in this way, taking 2 and then 3, and on inductively. By choosing the natural numbers in this way $0, 1, 2, 3, \dots$ we will eventually find an element of X (since X is nonempty) and this will be the smallest element of X . Anything smaller has been ruled out by the way we chose $0, 1, 2, \dots$. Thus \mathbb{N} satisfies the Minimum Property.

Given a set $A = \{x_0, x_1, \dots, x_n\}$ we can write down a well ordering on the elements of A by requiring that

$$x_0 < x_1 < \dots < x_n.$$

Notice that every element $x \neq x_n \in A$ has a successor $x^+ \in A$. For example,

$$x_0^+ = x_1, \quad x_1^+ = x_2.$$

The successor x_n^+ to x_n does not exist since

$$\{x \in A \mid x_n < x\}$$

is empty. Moreover, it is improper to say that $x_n^+ = x_{n+1}$ in A since x_{n+1} is not an element of A .

We can make $\{x_1, x_2, x_3\}$ a well ordered set if we define

$$x_2 < x_3 < x_1.$$

Notice the peculiar order of the subscripts when compared with the ordering of the x 's. The order of the subscripts does not follow the ordering of the elements. It is the order of the elements that matters, not how we name them. Observe that x_1^+ does not exist.

Partially order the pages of this book by the rule

$$x \leq y \text{ if page } y \text{ of the book follows page } x.$$

You should try to convince yourself that this ordering makes the pages in this book a well ordered set. If x is the last page in the book then x^+ does not exist since there is no page following the last page of the book.

Convince yourself that the subset ordering \subset makes the set

$$A = \{\emptyset, \{a\}, \{a, b\}, \{a, b, c\}\}$$

a well ordered set.

Order the set

$$A = \{\{1, 2\}, \{2, 3\}\}$$

using the subset order \subset . The elements of A are *incomparable*. That is, $\{1, 2\} \not\subset \{2, 3\}$ and $\{2, 3\} \not\subset \{1, 2\}$. (Read $\not\subset$ as *not a subset of*.) Because the set A does not satisfy the Trichotomy Property, it is not well ordered. Write this one down for yourself.

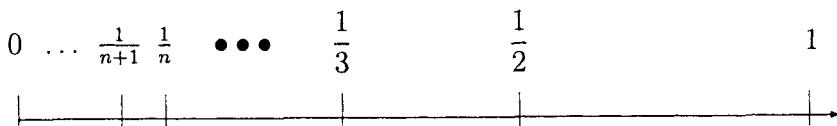
If we use the ordering $<$ of the real numbers then

$$W = \left\{ 0, 1, \frac{1}{2}, \frac{1}{3}, \dots \right\}$$

is *not* a well ordered set since the set

$$U = \left\{ 1, \frac{1}{2}, \frac{1}{3}, \dots \right\}$$

does not contain a minimum element. The following diagram will help show us why. The line below is the line \mathbb{R}^+ of positive real numbers. It is not to scale.



This picture shows us that in A we have $\frac{1^+}{2} = 1$, $\frac{1^+}{3} = \frac{1}{2}$, and in general

$$\frac{1^+}{n+1} = \frac{1}{n}.$$

If we try to choose an element $\frac{1}{n}$ in X then we have a smaller one $\frac{1}{n+1}$. Thus X cannot contain a minimum element, something in X that is smaller than any other element of X . Thus A does not have the Minimum Property.

Moreover, 0 is without a successor 0^+ in A . To see this, suppose that 0^+ exists in A . Then 0^+ must satisfy

$$0^+ = \frac{1}{n}$$

for some $n \in \mathbb{N}$. In that case $0 < \frac{1}{n+1}$ so that $0^+ \leq \frac{1}{n+1}$ by the definition of the successor element. But then

$$\frac{1}{n+1} < \frac{1}{n} = 0^+ \leq \frac{1}{n+1},$$

a clear contradiction. Thus 0^+ does not exist in A .

We can construct new well ordered sets from old ones in the following way. For reasons that will become clear later, let us assume that

ω_0

is a *symbol* not in \mathbb{N} . Any element not in \mathbb{N} will do. You might choose \bullet , you might choose \square , or you might even choose the *symbol* \mathbb{N} . We reserve the symbol ∞ for another purpose. It is traditional to choose ω_0 .

We then turn the set

$$\mathbb{N} \cup \{\omega_0\}$$

into a well ordered set by requiring that

$$n < \omega_0 \text{ for each } n \in \mathbb{N}.$$

We can list the elements in this well ordered set as follows.

$$\mathbb{N} \cup \{\omega_0\} = 0 < 1 < 2 < 3 < \dots < \omega_0.$$

A picture of $\mathbb{N} \cup \{\omega_0\}$ will help.



This is not a picture of the real line. The three dots in the picture indicate that the values 0, 1, 2 continue on indefinitely in the same manner. The placement of ω_0 indicates that ω_0 is larger than 0, 1, 2, and every natural number n : for example,

$$10^{10^{10^{10^{10}}}} < \omega_0$$

Even the very large natural numbers are smaller than ω_0 .

Reader Be Warned: We are dealing with an abstract mathematical construction. We treat \mathbb{N} as a featureless set. Recall how we constructed the natural numbers. \mathbb{N} is the collection of all sets that are equivalent to $\{x, y\}$. \mathbb{N} is not a distance under this construction. $1 < 2$ does not mean that 1 and 2 are a distance of 1 unit apart. We have not introduced a means of defining distance on \mathbb{N} . You should not see ω_0 as being infinitely far away. ω_0 is not a number. It is simply another *element* in a larger set.

We have added ω_0 in such a way that ω_0 is the *unique largest element in $\mathbb{N} \cup \{\omega_0\}$* . Furthermore,

there is no element $x \in \mathbb{N} \cup \{\omega_0\}$ such that $x^+ = \omega_0$.

This is not hard to show. Let $n \in \mathbb{N}$ be any element. Then by elementary arithmetic $n^+ = n + 1 \in \mathbb{N}$. Thus $n^+ \neq \omega_0$.

For this reason, we say that

ω_0 is the first infinite ordinal

or *the first countably infinite ordinal*. It is natural to think of ω_0 as a countable infinity since it is larger than each $n \in \mathbb{N}$. We also refer to ω_0 as a *limit ordinal*.

Some of you may be having trouble thinking of this element ω_0 . You are so accustomed to seeing \mathbb{N} as a set whose elements are 1 unit of distance apart that the idea of \mathbb{N} as a set without distance makes you uncomfortable. You think of $<$ as a measure of the distance between elements. This is not the case.

Think of \mathbb{N} this way. Envision a mathematical bag that contains each of the elements $n \in \mathbb{N}$. There is no distance and there is no unit of measurement. There is just a bag that contains all of the natural numbers n . You can do this, you know. It has been done on a smaller scale. Pick up a large dictionary and look at the back of the book. There is a page of measurements in various units of distance. Thus we have, on one page, included numbers like 12 inches in 1 foot. Here we squeezed 12 numbers into one place. We also have 1 mile is 5280 feet. We put several thousands of natural numbers into one page. You might also find 186,000 miles per second if you looked up the speed of light. These are large numbers that are not far away. They are within our grasp. Think of our bag of natural numbers in the same way. They exist as numbers in some kind of set. Nothing more.

5.2 The Arithmetic of Ordinals

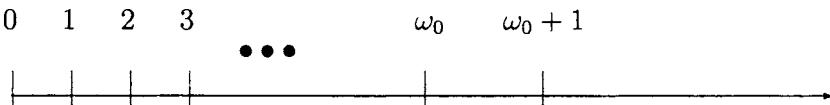
We now have a set $\mathbb{N} \cup \{\omega_0\}$ that contains a symbol ω_0 that can be identified with an infinite value. All we had to do was to introduce a new symbol ω_0 that we *defined* to be larger than each of the elements in \mathbb{N} . No distance was implied since infinite distance is undefined. The symbol ω_0 is just that, a symbol. Let us continue

using this idea of adding a symbol, and thus extend the well ordered set $\mathbb{N} \cup \{\omega_0\}$ to a larger well ordered set.

We will choose a symbol to serve as ω_0^+ in a well ordered set. We choose for convenience

$$\omega_0 + 1$$

and we place $\omega_0 + 1$ as the maximal element in a new well ordered set as in the following diagram.



The symbol does not really mean the sum of ω_0 and 1. We have simply chosen a symbol that we can recognize as the successor ω_0^+ of ω_0 . However, we will abuse the notation and read $\omega_0 + 1$ as *omega nought plus one*. In this way we have begun an induction process with which we will construct a new set of ordinals. We have already introduced two different infinite ordinals, $\omega_0, \omega_0 + 1$, and there is the promise of many more. Notice that this construction is in conflict with the adage that we were taught as children: *infinity plus one is infinity*. While this adage applies to cardinals, it does not anticipate the existence of ordinals. We might try bending the adage somewhat and say that *infinity plus one is infinite*.

Extend the above construction of ordinals by choosing different symbols

$$\omega_0 + 2, \omega_0 + 3, \dots$$

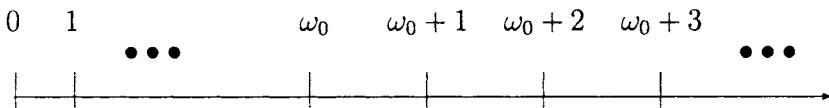
that we will call *ordinals*. We make a new well ordered set by defining a quite natural ordering

$$0 < 1 < 2 \dots < \omega_0 < \omega_0 + 1 < \omega_0 + 2 < \omega_0 + 3 < \dots$$

of the new set

$$\mathbb{N} \cup \{\omega_0, \omega_0 + 1, \omega_0 + 2, \omega_0 + 3, \dots\}.$$

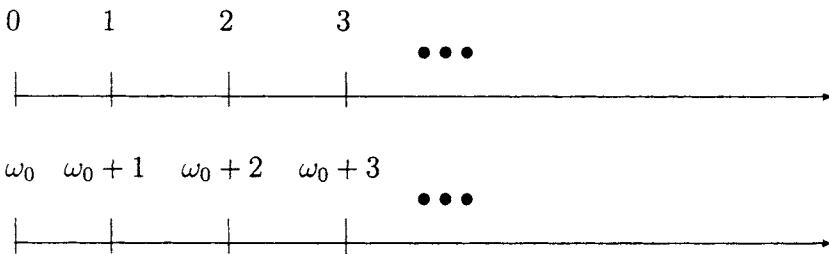
A picture will help the reader visualize the ordering.



In effect we have constructed a sequence of ordinals or infinities. If we ignore the addition and multiplication of \mathbb{N} then the set

$$\{\omega_0, \omega_0 + 1, \omega_0 + 2, \omega_0 + 3, \dots\}$$

appears to be no different from \mathbb{N} . Another picture will help.



For all intents and purposes, our new well ordered set is formed by attaching a copy of \mathbb{N} to the right of \mathbb{N} .

There is another interesting omission here. Where do we find the symbol $1 + \omega_0$ in this picture? The answer is hard to explain but easy to read.

$$1 + \omega_0 = \omega_0 \neq \omega_0 + 1.$$

This contradicts all of our intuition on addition of numbers. Why isn't $\omega_0 + 1 = 1 + \omega_0$? Common sense tells us that addition is commutative. The reality of it shows us that *our common sense about numbers cannot be applied meaningfully to infinite ordinals*. An explanation that might satisfy our curiosity and our common sense lies deep inside the study of well ordered sets. We will not go

that far. Suffice it to say that this startling unexpected equation is a property of ordinals and not of numbers.

We need a new element that will act as an element that is larger than $\omega_0 + n$ for each $n \in \mathbb{N}$. Since our new well ordered set is basically formed by putting one copy of \mathbb{N} behind another copy of \mathbb{N} , our next choice of a largest element might be chosen as $\omega_0 + \omega_0$. This *symbol* is not in our set. Certainly $2\omega_0$ is another symbol not in our set so we will make the identification

$$2\omega_0 = \omega_0 + \omega_0$$

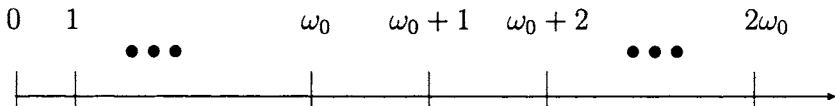
and then form the well ordered set

$$\mathbb{N} \cup \{\omega_0, \omega_0 + 1, \omega_0 + 2, \dots\} \cup \{2\omega_0\}$$

by requiring that

$$\omega_0 + n < 2\omega_0 \text{ for all } n \in \mathbb{N}.$$

In our new well ordered set the largest element is $2\omega_0$ as in the accompanying picture.



We have constructed another well ordered set. *This set is different from the previous well ordered sets* in that it has two limit ordinals while the previous well ordered sets have just had the one limit ordinal ω_0 .

$$\omega_0 < 2\omega_0.$$

Thus there is no element x such that $x^+ = \omega_0$ or $x^+ = 2\omega_0$. Furthermore, to each element x except the largest element there is an

element x^+ . Given the way that we have defined $x^+ \in \mathbb{N} \cup \{\omega_0\}$ as the symbol $x + 1$, it is natural and mathematically proper to write

$$x^- = \text{predecessor of } x$$

for ordinals x . That is, x^- is *the unique largest element that is smaller than x* . Hence

$$x^- = \text{the unique element } y \text{ such that } y + 1 = x.$$

Under this convenient notation we have

$$\omega_0 = (\omega_0 + 1)^-, \quad \omega_0 + 1 = (\omega_0 + 2)^-, \quad \omega_0 + 2 = (\omega_0 + 3)^-.$$

We must be aware of the fact that *not every element x has a predecessor x^-* . For example, ω_0^- and 0^- do not exist. There is no element y such that $y + 1 = 0$ or ω_0 . Thus there is no way to define $\omega_0 - 1$ since $\omega_0 - 1$ would be the ordinal x such that $x^+ = \omega_0$, and such an ordinal does not exist. In a similar way $2\omega_0 - 1$ cannot be defined since there is no ordinal x such that $x^+ = 2\omega_0$. Therefore, in general, even though we can add some ordinals and subtract some ordinals, *the subtraction of ordinals cannot be defined on all ordinals*. Here is an abstract addition whose associated subtraction is not defined everywhere.

We have defined the multiplication of at least two ordinals when we formed $2\omega_0$. We will show that *division is not defined on all ordinals*. To see this, assume to the contrary that there is some ordinal

$$\frac{1}{2}\omega_0 = x.$$

Then

$$\omega_0 = 2x$$

so that $\omega_0 > x$. Consequently, $x \in \mathbb{N}$ since the only ordinals less than ω_0 are in \mathbb{N} . This implies that

$$2x \in \mathbb{N},$$

while we chose ω_0 such that

$$2x = \omega_0 \notin \mathbb{N}.$$

This contradiction shows us that $\frac{1}{2}\omega_0$ cannot be defined.

If we place a copy of the well ordered set

$$\mathbb{N} \cup \{\omega_0, \omega_0 + 1, \omega_0 + 2, \dots\}$$

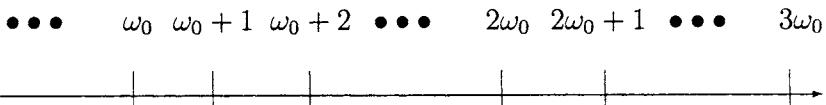
at its end, we have constructed a new well ordered set that is different from what we started with. That is, we can choose symbols $2\omega_0 + n$ for $n \in \mathbb{N}$ and we define a new well ordered set.

$$\mathbb{N} \cup \{\omega_0, \omega_0 + 1, \omega_0 + 2, \dots\} \cup \{2\omega_0, 2\omega_0 + 1, 2\omega_0 + 2, \dots\}. \quad (5.1)$$

Pictorially we have the following set.



Can you see it? Can you see how we will proceed next? We will insert a new element $3\omega_0$ that is larger than all elements in (5.1) thus creating a new well ordered set whose picture might look like this.



The above picture suggests a chain of limit ordinals that are integer multiples of ω_0 .

$$\omega_0 < 2\omega_0 < 3\omega_0 < \dots$$

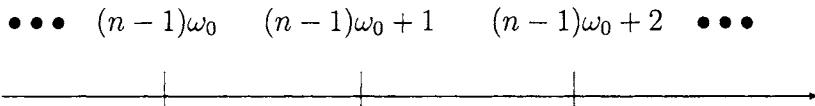
Given an $n \in \mathbb{N}$ we would define an ordinal

$$n\omega_0$$

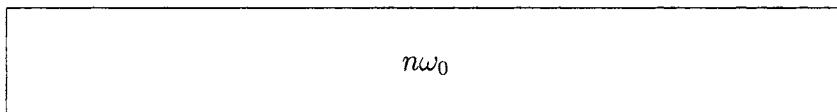
as follows. Provided that we know $(n - 1)\omega_0$, we can define the ordinals

$$(n - 1)\omega_0^+ = (n - 1)\omega_0 + 1, \quad (n - 1)\omega_0 + 1^+ = (n - 1)\omega_0 + 2, \quad \dots$$

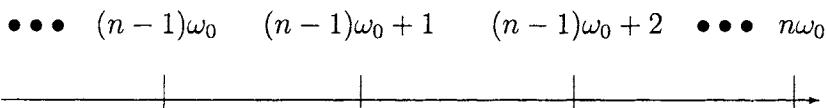
as we did above with ω_0 . A picture is given as follows.



Then we choose a symbol



and define it to be the unique element that is larger than all of the elements $(n - 1)\omega_0, (n - 1)\omega_0 + 1, (n - 1)\omega_0 + 2, \dots$. The well ordered set that we have just defined can be described by the following picture.



Then we have indeed defined an unending chain of limit ordinals:

$$\omega_0 < 2\omega_0 < 3\omega_0 < \dots < n\omega_0 < \dots$$

Thus there is a limit ordinal $100\omega_0$ and one $186,000\omega_0$. Why there is even a limit ordinal

$$10^{10^{10^{10^{10}}}}\omega_0$$

but this is so far out there that we would have trouble imagining just how large that ordinal is.

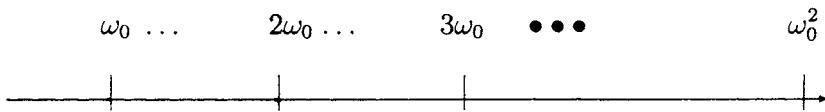
We might be tempted to say “and so on” when we read \dots but what does that mean now? Indefinitely? Infinitely? What do these

words mean. Do we mean that we should repeat the process once for each $n \in \mathbb{N}$, or once for each ordinal $< 3\omega_0$, or perhaps once for each ordinal less than $10^{10^2}\omega_0$? We need a more precise way of saying “and so on,” and we will see it in the next section.

As long as we have a chain of limit ordinals $\omega_0, 2\omega_0, 3\omega_0, \dots$ we can define a symbol

$$\omega_0^2 = \omega_0 \omega_0$$

that is larger than each of the limit ordinals $n\omega_0$ with $n \in \mathbb{N}$. Pictorially we have defined the new well ordered set whose limit ordinals look like this.



We have defined ω_0^2 by placing ω_0 copies of \mathbb{N} one right after the other. The ordinal ω_0^2 is larger than all limit ordinals of the form $n\omega_0$. Even the incredibly large limit ordinal

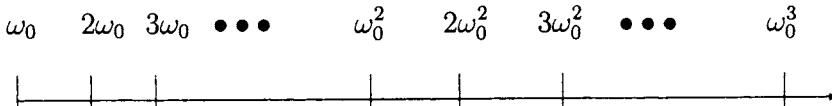
$$10^{10^{10^{10^{10}}}} \omega_0$$

is smaller than ω_0^2 .

Just for the fun of it let's try to define ω_0^3 . Let us define

$$\omega_0^3 = \omega_0^2 \cdot \omega_0.$$

Beginning with ω_0^2 we will define the ordinals $\omega_0^2 + n$ and $n\omega_0^2$ for each $n \in \mathbb{N}$. The diagram shows us what we have done. It is not to scale.



By repeating the same process we will define ordinal powers of ω_0 as the symbol

$$\omega_0^n = \omega_0^{n-1} \cdot \omega_0 \text{ for each } n \in \mathbb{N}.$$

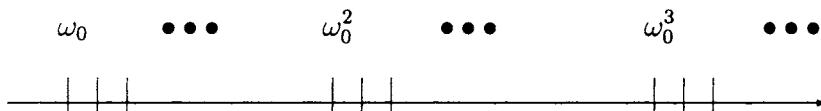
If we start with $\omega_0 = \omega_0^1$ then we can define $\omega_0^2 = \omega_0 \cdot \omega_0$, $\omega_0^3 = \omega_0^2 \cdot \omega_0$, and then $\omega_0^4 = \omega_0^3 \cdot \omega_0$. The process continues, defining all of the powers ω_0^n . We have then expanded our collection of limit ordinals by taking *powers of* ω_0 and all of their combinations

$$\omega_0 < \omega_0^2 < \omega_0^3 < \dots$$

and

$$10\omega_0 < 121\omega_0^2 < 10^{10^2}\omega_0^{33} + 101 < 2\omega_0^{81} + \omega_0 < \dots$$

Pictorially we have



If we continue this chain $\omega_0, \omega_0^2, \omega_0^3, \dots$ of limit ordinals we arrive at the much larger limit ordinal

$\omega_0^{\omega_0}$ = the smallest ordinal that is larger than all of the ordinals ω_0^n .

This limit ordinal $\omega_0^{\omega_0}$ can be pictured as follows.



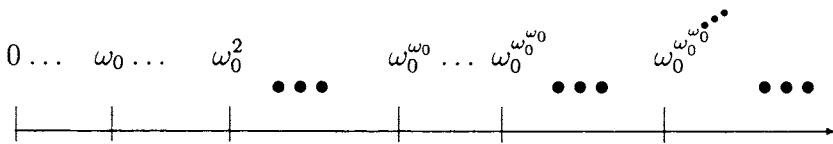
In fact, there is no end to the exponents we can form. We can even define an ordinal that has infinitely many exponents. Let us agree that

$$\omega_0^{\omega_0}$$

is the smallest ordinal that is larger than each of the composite powers

$$\omega_0 < \omega_0^{\omega_0} < \omega_0^{\omega_0^{\omega_0}} < \omega_0^{\omega_0^{\omega_0^{\omega_0}}} < \dots$$

Thus \mathbb{N} is part of a larger well ordered set, some of whose picture we give below.



This picture is not complete since we can extend this well ordered set by defining

$$\omega_0^{\omega_0^{\omega_0^{\omega_0}}} < \omega_0^{\omega_0^{\omega_0^{\omega_0}}} + 1 < \omega_0^{\omega_0^{\omega_0^{\omega_0}}} + 2 < \dots$$

It is difficult to explain which well ordered set these ordinals represent. They are so large that they are larger than every ordinal that we have discussed so far. Let us just agree that these are large ordinals whose existence we had not expected. Keep in mind, reader, that like any other graph, there are many unlabelled ordinals between the labelled points in this picture. That is what ... and ••• should say to you.

And yet, as large as this ordinal is, there is a larger ordinal. To find it we will have to construct an ordinal in a different way.

Let X be the set of real numbers. It is best if we use a different symbol like X as it will enhance the discussion. In Theorem 4.3.1 we showed that

$$\text{card}(\mathbb{N}) < \text{card}(\mathbb{R}) = \text{card}(X).$$

One of the accepted axioms of mathematics is that X can be turned into a well ordered set. The set X is so large that it contains a copy of every one of the ordinals we have discussed so far. For example,

$$1, \omega_0, \omega_0^{\omega_0}, \dots$$

are all (essentially) in X . But none of these ordinals is the largest element of X .

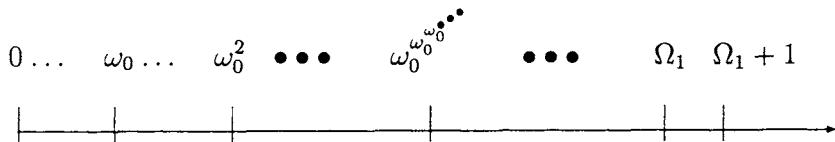
Let Ω_1 be a symbol not in X and define a new well ordered set $X^+ = X \cup \{\Omega_1\}$ by requiring that

$$x < \Omega_1 \text{ for each } x \in X.$$

Then Ω_1 becomes the largest element in X^+ . We say that

$$\Omega_1 = \text{the first uncountable ordinal.}$$

And yet Ω_1 is not the largest ordinal.



Indeed, if we seek the largest ordinal then our search is fruitless. There is no such object. Ordinals just go on without stopping. If we think that we have exhausted our chain of ordinals at some value a , we can always define a new ordinal $a < a + 1$ and kickstart our chain. If we have a set of ordinals

$$a_1 < a_2 < \dots$$

and we think that this is the end of the ordinal process, we can always choose a new symbol z not on the chain, place it at the end of the set

$$a_1 < a_2 < \dots < z,$$

and start the process of constructing ordinals again.

Let us end this section by reintroducing an old friend of ours. Do you remember the cleaning lady, Mary, in Hilbert's Infinite Hotel from page 98? She's the one who cleaned all of the rooms immaculately and was not heard from again. Well, as it happens, she was found resting quietly in a room outside the Infinite Hotel numbered ω_0 . Naturally, this is the room for Hotel Staff.

5.3 Cardinals as Ordinals

It is possible to order a set *linearly* but not well order the set. For example, the usual ordering of the real numbers \mathbb{R} makes \mathbb{R} a linearly ordered set, but not a well ordered one. *Linear order* implies that the Trichotomy Property is true of \mathbb{R} ; however, it does not imply that \mathbb{R} is well ordered. For example, we cannot define a successor for $\frac{1}{2}$ in \mathbb{R} that respects the usual ordering of \mathbb{R} . The reason would be that if we have $\frac{1}{2}^+$ then there should be no real number x such that

$$\frac{1}{2} < x < \frac{1}{2}^+.$$

Thus *there can be no next real number*.

Recall that $\text{card}(A) \leq \text{card}(B)$ if there is a one-to-one function $f : A \rightarrow B$. This ordering \leq of cardinals is linear because the given cardinals satisfy the Trichotomy Property. That is,

given any two cardinals α and β either
 $\alpha < \beta$, $\beta < \alpha$, or $\alpha = \beta$.

This allows us to make the following conclusion.

Given two cardinals α and β then
 $\alpha \leq \beta$ and $\beta \leq \alpha \implies \alpha = \beta$.

It is natural to ask if this *linear* ordering of cardinals is a *well* ordering of cardinals. For example, if we want to know the comparative sizes of $\text{card}(\mathcal{P}^3(\mathbb{N}))$ and $2^{\text{card}(\mathcal{P}(\mathbb{N}))}$ then we can be assured that exactly one of the three conditions

$$\text{card}(\mathcal{P}^3(\mathbb{N})) < 2^{\text{card}(\mathcal{P}(\mathbb{N}))}, \quad 2^{\text{card}(\mathcal{P}(\mathbb{N}))} < \text{card}(\mathcal{P}^3(\mathbb{N})), \\ \text{or} \quad \text{card}(\mathcal{P}^3(\mathbb{N})) = 2^{\text{card}(\mathcal{P}(\mathbb{N}))}$$

applies to the two cardinals. Now that we know that the cardinals are linearly ordered, we make the next leap of the mathematical imagination. The reason that we have taken such care with the ordinals and their arithmetic is the following deep result from mathematics that links ordinals with cardinals.

Theorem 5.3.1 *Given a cardinal \aleph there is a successor cardinal \aleph^+ . That is, the set of cardinals is well ordered.*

The theorem is deep and its proof is far too complex for us to present in this book, so let us just accept the truth of it.

Thus there is a cardinal \aleph_0^+ and one $(2^{\aleph_0})^+$. Since

$$\aleph_0 < 2^{\aleph_0}$$

and since the successor element x^+ is the smallest element that is larger than x , we can say that

$$\aleph_0^+ \leq 2^{\aleph_0},$$

but we do not know if this inequality is proper or an equation.

A more familiar inequality comes from Theorem 4.3.1:

$$\aleph_0 < \text{card}(\mathbb{R}).$$

Since the successor element \aleph_0^+ is smaller than every element that is larger than \aleph_0 , we can write that

$$\aleph_0^+ \leq \text{card}(\mathbb{R}).$$

Questions about this inequality turn out to be some of the deepest questions in mathematics. We will address this predicament presently.

The following notation will be used for the most often referred to cardinals:

$$\aleph_0 = \text{card}(\mathbb{N}) \quad (\text{read as } aleph \text{ nought}),$$

$$\aleph_1 = \aleph_0^+ \text{ is the successor of } \aleph_0 \quad (\text{read as } aleph \text{ one}),$$

$$\aleph_2 = \aleph_1^+ \text{ is the successor of } \aleph_1,$$

and so on.

Inasmuch as

$$\aleph_0 = \text{card}(\mathbb{N}) < \text{card}(\mathbb{R}),$$

we can write that

$$\aleph_1 \leq \text{card}(\mathbb{R}).$$

The reasoning here is that the successor $\aleph_1 = \aleph_0^+$ is *the smallest element that is larger than \aleph_0* . We have thus arrived at our first mysterious property of cardinals. Given that $\aleph_1 \leq \text{card}(\mathbb{R})$ it is natural to ask if $\aleph_1 = \text{card}(\mathbb{R})$. This question is so subtle a matter that the mathematician who answered it, Paul J. Cohen, was honored with the highest award in mathematics: the Field's Medal.

Theorem 5.3.2 [Paul J. Cohen] *We cannot prove and we cannot disprove that $\aleph_1 = \text{card}(\mathbb{R})$.*

We say that the condition $\aleph_1 = \text{card}(\mathbb{R})$ is *independent of the ZFC Axioms of Set Theory*. That is, it is unattainable as a theorem in regular mathematics. That statement needs some explanation. I suppose you have had experience in plane geometry, so we start there. The geometry we studied in high school was based on ten elementary truths, or axioms and propositions, given by a Greek mathematician Euclid (circa 300 BC). Euclid started with these ten statements and then proceeded to develop all of geometry from them. The truth of most of those statements is quite transparent. For example, Euclid starts by assuming that *the whole is the sum of its parts*. Obviously, this is true. He also assumes that *all right angles have equal measure*. Again, obviously true. There are eight others that we will not state because they do not impact on our

discussion. The point is, from ten clearly true statements Euclid could derive all of the mathematics in the geometry you studied in school. Thus, by starting with true statements about points, lines, and triangles and then extending his thoughts using *Aristotelian Logic*, (the logic of true and false statements), Euclid was able to develop a mathematics that was free of any mistakes or misconceptions. Euclid's method, now called the *axiomatic method*, was so compelling that mathematicians in every century over the last 2300 years have made use of it. We all begin our work with a few basic statements and proceed from there to demonstrate universal truths. These truths are not personal opinion. They are statements that would be true no matter who was reading them. These are truths that can be shared with advanced civilizations anywhere on Earth and, should the need arise, with extraplanetary civilizations. Assuming that a group of people is sufficiently advanced, assuming they know enough mathematics, we would share with them the same theorems of plane geometry. That is the power and the nature of the axiomatic method. Try to make a similiar statement concerning the literature, art, or religion of two different civilizations on Earth. That would not be possible. Mathematics took on the axiomatic method with vigor in the early part of the twentieth century. In this time every area of mathematics from algebra to calculus tried to find those few elementary truths that could be used as a springboard with Aristotelian Logic to produce universal truths about that area of mathematics. Sometime in the 1920s a mathematician came up with a few statements that all mathematicians could agree on as obviously true, and that most agreed were simple statements. These axioms today are called the *ZFC Axioms*, or *Zermelo-Franklin-Choice Axioms*. These are axioms about sets and they are the basis for most modern mathematics. For example, one of the statements is that there do not exist sets A and B that are elements of each other: that is,

There do not exist sets A and B such that $A \in B$ and $B \in A$.

A moment's thought might convince you that this statement is true, and that its negation is:

There are two sets A and B such that $A \in B$ and $B \in A$.

The point behind this is that by starting with elementary statements about sets we are able to *prove* advanced ideas in such areas as linear algebra, calculus, and trigonometry. All of mathematics assumes set theory so all of mathematics depends upon axioms.

What P. J. Cohen proved was that there is at least one statement about mathematics that cannot be proved or disproved with the ZFC Axioms. No amount of logic and mathematics will allow us to start with the axioms of set theory and then end up with $\aleph_1 = \text{card}(\mathbb{R})$. Mathematics, as large as it might seem right now, is too small to include $\aleph_1 = \text{card}(\mathbb{R})$ as a theorem. Now there is a statement. Why would you ever think that mathematics is small? And yet here we have an example of its limitations.

This is probably the first time you have seen such a statement. Think of this as a Catch-22 in your algebra course. Your teacher asks you to show that $x = y$ for some numbers x and y , and you tell your professor, “I cannot prove that $x = y$ but I cannot prove that $x \neq y$.” Your professor would certainly give you a high mark if $x = \aleph_1$ and $y = \text{card}(\mathbb{R})$.

Let’s examine what it means to be independent of the axioms of set theory. There are five axioms that define what set theory is. From these axioms almost all of modern mathematics can be proved as theorems. This occurs in exactly the same way you learned it in your high school geometry course. You begin with five axioms or postulates (these from Euclid) from which you can prove all of the theorems in elementary geometry. Mathematics is handled in the same way. Most of the mathematics we meet in elementary courses can be proved from the axioms of set theory. Thus the ZFC Axioms of Set Theory become the foundation of modern mathematics. With such a beginning, mathematicians are sure that modern mathematics is on the same firm foundation that Euclid’s plane geometry enjoys.

Before Cohen’s Theorem 5.3.2, mathematicians did not know of a *simply stated idea* that was independent of the ZFC Axioms of Set Theory. Indeed, these Axioms were relatively new discoveries when Cohen proved his theorem. It struck the mathematicians of the day as quite strange. It is still thought to be curious that some statement about mathematics is beyond mathematical proof or disproof. Before Cohen’s Theorem, the powerful mathematician David

Hilbert (remember him of Infinite Hotel fame) had suggested that mathematicians should start looking for a program or computer of sorts that would produce all of mathematics if it was allowed to run long enough. Cohen's Theorem 5.3.2 proved that Hilbert's suggestion could not be realized. There can be no program or computer that produces all of mathematics even if it was allowed to run in some thought experiment. Some statements of mathematics are simply beyond the reach of mathematical proof. They can be stated or *assumed* but not proved.

The cold fact is that the axioms of set theory are simply not strong enough to prove every statement about mathematics. Furthermore, no collection of mathematical axioms will allow us to prove every statement about mathematics. The only way to use a statement like $\aleph_1 = \text{card}(\mathbb{R})$ is to *include it as an axiom*. However, once that inclusion is made we *create a larger and different mathematics that properly contains the one we grew up with*. This new mathematics will contain all of the work you did in your high school mathematics class, including algebra and geometry, and it will contain a new and perhaps strange mathematical statement. Before the end of this discussion we will encounter a simple statement that is true in the mathematics that assumes that $\aleph_1 = \text{card}(\mathbb{R})$ but not in a smaller mathematics.

Georg Cantor did not know of Cohen's Theorem (since Cantor lived 80 years before Cohen's Theorem was published), so he accepted $\aleph_1 = \text{card}(\mathbb{R})$ as an axiom or hypothesis.

The Continuum Hypothesis: Assume that $\aleph_1 = \text{card}(\mathbb{R})$.

Because the Continuum Hypothesis is independent of our usual mathematics, it can be used as an axiom with which we can construct an entirely new mathematics.

CH

CH is the *usual mathematics together with the Continuum Hypothesis* used as an axiom. Many beautiful theorems are proved in **CH**,

or by assuming the Continuum Hypothesis. The mathematics you learned in high school is not strong enough to prove all of the theorems in **CH**.

On the other hand, because we cannot disprove the Continuum Hypothesis, we can include its *logical negation*, $\aleph_1 < \text{card}(\mathbb{R})$, as an axiom in mathematics, thus constructing a new mathematics:

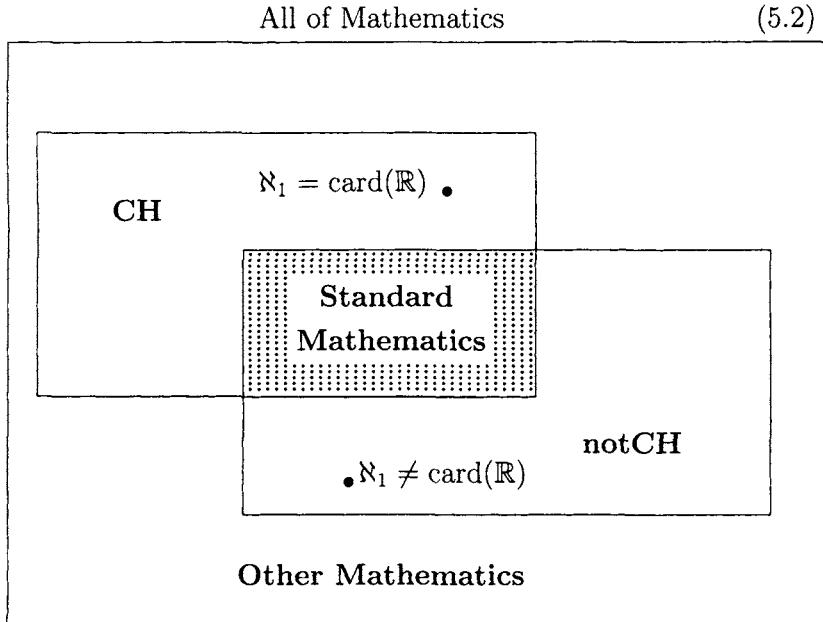
notCH.

While many theorems are proved in **notCH** these theorems cannot be compared to the theorems in **CH**. Theorems in **CH** may seem to contradict some theorems in **notCH**, but this is an inaccurate comparison because

CH and **notCH** exist as two different mathematical universes.

We cannot compare the theorems from **CH** and **notCH** anymore than we could compare recipes for bread and cake and claim that one is right and one is wrong. Each has its own uses and each has its own role on the dinner table. In the same way the theorems of **CH** have their place in mathematics, the theorems in **notCH** have their place in mathematics, and the two places cannot be compared.

This must seem odd. We were all led to believe that there was only one mathematics. This reaction is not unusual. The mathematics we learned in high school is contained in both **CH** and **notCH**. No one except the experts ever has the opportunity to distinguish between the mathematics in **CH** and the mathematics in **notCH**. In picture (5.2), **CH** and **notCH** are the regions that contain the shaded region called Standard Mathematics. From this picture we can imply certain facts about Mathematics. Here is what this means to you.



1. **CH** and **notCH** overlap in the Standard Mathematics.
 2. **CH** and **notCH** are different.
 3. Neither **CH** nor **notCH** equals the Standard Mathematics.
 4. Together **CH** and **notCH** do not encompass All of Mathematics.

Moreover, the diagram suggests that there are *other* regions of mathematics out there to be discovered. This is indeed the truth of the matter.

In the following discussion I will attempt to show you a mathematical fact that changes when viewed as a theorem in **CH** and in **notCH**.

Recall that $\aleph_1 = \aleph_0^+$ is the successor to \aleph_0 , and that \aleph_2 denotes the successor of \aleph_1 .

$$\aleph_2 = \aleph_1^+.$$

The next question we will consider is: *How do \aleph_0 , \aleph_1 , $\text{card}(\mathbb{R})$, and \aleph_2 compare?* It turns out that these relationships depend on the mathematics we choose to work in, **CH** or **notCH**. We have shown that

$$\aleph_1 \leq \text{card}(\mathbb{R}).$$

Using this with the already established Theorems 4.3.1 and 4.4.2, we find that there is a chain of cardinals

$$\aleph_0 < \aleph_1 \leq \text{card}(\mathbb{R}).$$

To fine tune these inequalities we would use the Continuum Hypothesis or its logical negation. The next two boxes show us how these inequalities change in **CH** and in **notCH**. Note the different placement of \aleph_1 and \aleph_2 .

In **CH**,

$$\aleph_0 < \boxed{\aleph_1 = \text{card}(\mathbb{R})} < \aleph_2,$$

while in **notCH**,

$$\aleph_0 < \aleph_1 < \boxed{\aleph_2 \leq \text{card}(\mathbb{R})}.$$

This is real evidence that **CH** and **notCH** are different mathematical systems.

Let us examine that thought. The cardinality of \mathbb{R} , essentially the number of elements in \mathbb{R} , has different values in **CH** and **notCH**. This infinite number $\text{card}(\mathbb{R})$ has two different values, depending upon which mathematics you choose to work in. The difference between **CH** and **notCH** could not be more plain.

$$\begin{aligned}\text{card}(\mathbb{R}) &= \aleph_1 \text{ in } \mathbf{CH}, \\ \text{card}(\mathbb{R}) &\geq \aleph_2 \text{ in } \mathbf{notCH}.\end{aligned}$$

How is that possible? Does the number of elements in \mathbb{R} change as it passes between **CH** and **notCH**? Are there fewer real numbers in **CH** than there are in **notCH**? Of course, the set \mathbb{R} does not change when we work in **CH** or **notCH**. If x is a real number in the mathematics **CH** then x is a real number in the mathematics **notCH**. The reason for this is that \mathbb{R} exists in the mathematics you learned in high school. The real numbers will be the same no matter which mathematical structure you work in. So then how does $\text{card}(\mathbb{R})$ change values between **CH** and **notCH**? The answer is that in **notCH** we add a cardinal between \aleph_0 and $\text{card}(\mathbb{R})$. When we form **CH** we are not allowing for any other cardinals between \aleph_0 and $\text{card}(\mathbb{R})$. **CH** tolerates the inequalities

$$\aleph_0 < \aleph_1 = \text{card}(\mathbb{R}) < \aleph_2.$$

When we list the cardinals in **notCH** we allow for the existence of a cardinal \aleph_1 such that

$$\aleph_0 < \aleph_1 < \aleph_2 \leq \text{card}(\mathbb{R}).$$

Since $\aleph_2 = \aleph_1^+$ there will not be cardinals strictly between \aleph_1 and \aleph_2 . Thus in **notCH** we have the chain of inequalities

$$\aleph_0 < \aleph_1 < \aleph_2 \leq \text{card}(\mathbb{R}).$$

It follows that the ways we *count* in **CH** and **notCH** are different. The cardinals (numbers) that we use change values when we pass from the mathematical system **CH** to **notCH**.

Now consider the cardinal 2^{\aleph_0} . Under the ZFC Axioms of Set Theory we can write down that

$$\aleph_0 < \aleph_1 \leq 2^{\aleph_0}.$$

Thus in **CH** we have

$$\aleph_0 < \aleph_1 = 2^{\aleph_0} < \aleph_2$$

while in **notCH** we have

$$\aleph_0 < \aleph_1 < \aleph_2 \leq 2^{\aleph_0}.$$

Once again we see that the size of a cardinal depends upon the mathematical system in which we are working.

Let me impart one final item. A more general version of the Continuum Hypothesis is

The Generalized Continuum Hypothesis (GCH):

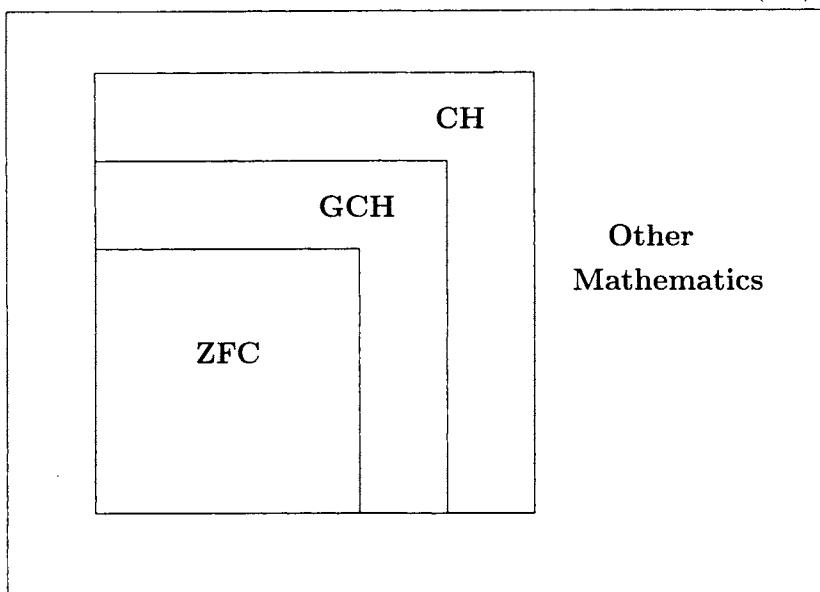
Assume that $\aleph^+ = 2^\aleph$ for each cardinal \aleph .

That is, given a cardinal \aleph assume that 2^\aleph is the next cardinal. The statement GCH assures us that all cardinals come in essentially one form, that of a power of 2.

Picture (5.3) illustrates the containing relationship between the mathematical system **GCH** that results from the addition of the Generalized Continuum Hypothesis into the Standard Mathematics **ZFC**. Picture (5.3) suggests that **ZFC**, the Standard Mathematics, is *properly* contained in the mathematical system **GCH** formed by adding GCH to the ZFC Axioms of Set Theory, and that **CH** properly contains **GCH**. Thus we have found a mathematics properly between **CH** and **ZFC**. There is every reason to expect that there is an *infinite chain* of mathematical systems containing **ZFC**. Indeed,

All of Mathematics

(5.3)



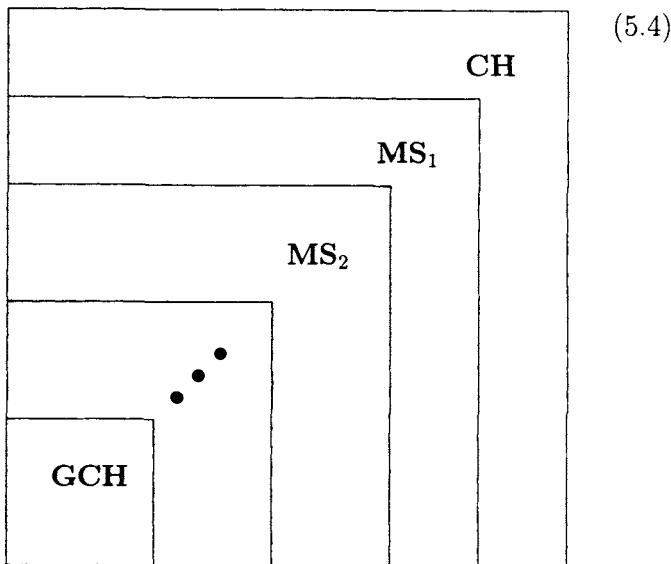
we leave it to the reader to find an infinite chain of mathematical systems

$$\mathbf{ZFC} \subset \dots \subset \mathbf{MS}_2 \subset \mathbf{MS}_1 \subset \mathbf{CH}$$

between **ZFC** and **CH**. Think of the Continuum Hypothesis and remember that we do not know if $\aleph^+ = 2^\aleph$ for cardinals \aleph . Just find these new mathematical systems. There is no reason to show that these chains behave precisely like the picture or your intuition suggests.

As we stated earlier, according to the Generalized Continuum Hypothesis each cardinal except \aleph_0 is a power of 2. In this way cardinals are not like numbers at all. After all, not every number is a power of 2. Thus under the Generalized Continuum Hypothesis each of the cardinals $\aleph_0, \aleph_1, \aleph_2, \dots$ is a power of 2.

$$\aleph_0, \quad \aleph_1 = 2^{\aleph_0}, \quad \aleph_2 = 2^{2^{\aleph_0}}, \quad \aleph_3 = 2^{2^{2^{\aleph_0}}}, \quad \dots$$



There is one easy question that I will leave the reader with in this chapter. We know that the Continuum Hypothesis cannot be proved using the ZFC Axioms of Set Theory. There are also mathematical statements that cannot be proved within **CH**. The Generalized Continuum Hypothesis is one such statement. Moreover, there is a true statement that cannot be proved within **GCH**. It exists but we cannot state it here. As we indicated in picture (5.4) there is a chain

$$\mathbf{GCH} \subset \dots \subset \mathbf{MS}_2 \subset \mathbf{MS}_1 \subset \mathbf{CH}$$

of mathematical systems. Each system **MS_n** is associated with a statement *S_n* that cannot be proved within **MS_n**. The question is this. Can we find one statement, say *S*, that cannot be proved in any of these systems **MS_n**? That is, can we find one statement, say *S*, that cannot be proved by **MS₁**, that cannot be proved by **MS₂**, that cannot be proved by **MS₃**, and so on? I will not answer this because the question is more philosophical than mathematical. Test your answer when you find it. Feel free to get emotional about it. Do you really think that your statement is beyond the proof of all of these mathematical systems?

5.4 Magnitude versus Cardinality

As with every discussion about the infinite, these ideas take place in a mathematical thought experiment or a Platonic Universe. So if these ideas do not agree with what you see around you, good! There is nothing infinite about this world around us. Let us expand on that for a little bit.

When you were introduced to the infinite as a child (i.e., before you read this book), you thought that infinity was an infinite distance from us. Somehow, infinity was a place that we could reach if we could move infinitely far. This is not possible within our physical universe. In our universe we have only a finite distance that we can travel before we come upon our starting point again. There are two infinities in that idea that have to be addressed, namely, infinite distance and infinite time.

The universe, as far as is known at the time of this printing, is a lumpy sphere having four dimensions. Its beginning can be pinned down to within such a small fraction 10^{-33} of a second that for our purposes we can assume that scientists have glimpsed the origin of the universe. Of course, this is not really the universe's origin, and perhaps that origin is beyond the scope of scientific observation, but it gives us a beginning, a starting point to talk about age *in years*. The beginning of our universe seems to have occurred some 20 billion years ago. It might be older or younger than that but we can assume with scientific certainty that the universe is no more than 30 billion years old. Moreover, outside this framework of time, it is relatively certain that time as we know it does not exist. Time seems to be peculiar to our universe. Thus there can be no infinite time prior to us. Time has not existed forever.

The expansion of the universe probably occurred in fits and starts: that is, the expansion of the universe was not smooth and not of uniform speed. But it has a maximum value, the speed of light. Given the age of the universe and a finite rate of expansion, it is clear that the universe is a bounded place, say, smaller than $3 \times 10^{10} \times c$, c being the speed of light. This gives us a bounded universe, so that distances in the universe do not appear to be infinite. Before anyone starts to suggest that we can reach the edge of the universe, let me bring you back to reality. There is no edge

to the universe. If you start out in a direction and hope to travel in a straight line you are thwarted by gravitational considerations since gravity warps space and time. Furthermore, you are on a four dimensional lumpy spheroid, not a three dimensional object. Thus your travels are necessarily distorted by the fourth dimension. For instance, suppose you are a point living on a sphere. You live in three dimensions but you only experience two. As you travel around the bounded spherical universe of yours, you are impressed with the fact that you can never reach the edge of the universe. And yet you know the universe is bounded. Try it on an orange and see if that helps you visualize the point moving about and seeking the edge of its universe.

Thus we do not find our universal edge. It bends away from us in the fourth dimension. In fact, cosmologists, people who study the universe, suspect that our universe exists as an 11 dimensional object. It would then be impossible for us to know today where that bending or curvature in the universe might be.

But just because the point's universe has no edge does that mean that the distances in the universe extend to the infinite? Are there two points in our universe that are infinitely far apart? Of course not. The distances might be large, perhaps even larger than our imagined point ever imagined, but the distances on the sphere are quite finite. In the same way but in a higher number of dimensions our universe does not possess two points that have an infinite distance between them. Infinite distances it seems must somehow extend beyond the universe. What then do we mean by infinite distances? This is the difference between cardinals and the infinite distances you may have become familiar with. Infinite distances do not exist while our notion of a cardinal does exist.

A cardinal is in a very loose sense a count of the number of elements in a set. As we saw there are many cardinals including the first infinite cardinal $\text{card}(\mathbb{N})$ and the cardinality of the reals $\text{card}(\mathbb{R})$. (If this notation makes little or no sense to you, then please read the previous chapters in this book.) We can continually add elements to these sets to form sets with larger cardinality. For example, if we let $\mathcal{P}(X)$ denote the *power set* of X , or equivalently $\mathcal{P}(X)$ is the set of subsets of the set X , then

$$\mathbb{N}, \mathcal{P}(\mathbb{N}), \mathcal{P}(\mathcal{P}(\mathbb{N})), \dots$$

is a list of sets with *different cardinalities*. Since we can always take the power set of a set, this list is infinite. Thus if we count the number of elements in a set carefully, we find that there are *infinitely many infinite cardinals*. These cardinals do not measure distance. They do not measure how far it is across a set. These cardinalities simply measure *how many elements are in the set* using one-to-one and onto functions.

The infinite ideas that you have seen in life are more closely related to distance because their size is determined relative to the natural numbers, \mathbb{N} . Mathematics does not recognize distance as an infinite number. That symbol you may have seen, the lazy eight ∞ , is not defined as a number. That's right. No matter what you may have heard ∞ is *not a number*. The division $\frac{1}{0}$ is not infinity, either. It is *undefined*, meaning that it holds no meaning at all. The symbol $\frac{1}{0}$ makes no mathematical sense at all. We do not even give it a special symbol or name. In fact, except for books like this one, you are not likely to see a professional mathematician write down $\frac{1}{0}$. That's enough of that. My fingers hurt when I type $\frac{1}{0}$. Ouch, I did it again.

Now, to speak about infinite quantities mathematicians have found that *the limit* is the only notion that allows us to describe and manipulate infinite or unbounded quantities in a consistent and mathematically precise manner. Let $f(x)$ be a function that takes real numbers x to real numbers $f(x)$. This is where that high school education comes in handy. Suppose we want to know about the size of $f(x)$ as x gets closer to some number, say, a . We write

$$\lim_{x \rightarrow a} f(x) = L$$

provided that $f(x)$ gets closer to L as x gets closer to a . Don't let the notation scare you. This is the smallest bit of notation we can use to get the idea across to the reader. It has been in use for 150 years now and shows no sign of being replaced by something else. What do we mean by x gets closer to a ? What we mean is that x is allowed to take on values that are physically closer to a . That is, *the difference between them* (i.e., the larger minus the

smaller), is allowed to become very small or close to 0. For example, a calculator will show you that $2x + 3$ gets close to 5 as x gets close to 1. I won't bore you with the table of numbers. In our notation

$$\lim_{x \rightarrow 1} 2x + 3 = 5.$$

We are more interested in limits where L is not finite. We say that $f(x)$ has an infinite limit at a if for each natural number N there is a small neighborhood of a , call it U , such that

$$f(x) > N \text{ for all } x \in U,$$

and we write

$$\lim_{x \rightarrow a} f(x) = \infty$$

in this case. For example, if we consider the function $f(x) = \frac{1}{x}$ at 0 then as we said above $f(0)$ is undefined. (Notice how I managed to say that without writing $\frac{1}{0}$. Ouch!) But we can take the limit as x gets close to 0 to see that

$$\lim_{x \rightarrow 0} \frac{1}{x} = \infty.$$

The table shows us why. We take values of x that are closer and closer to 0, but do not equal 0. The sequence of resulting values $f(x)$ gets larger.

x	$f(x) = \frac{1}{x}$
.1	10
.01	100
.001	1000
:	:

If we are given a small value for x , a number close to 0, then the number $\frac{1}{x}$ seems to get very large. If we are given a natural number N we can find a number $\frac{1}{N+1} = x$ very close to 0 such that $f(x) > N$. Thus it would seem that $\lim_{x \rightarrow 0} \frac{1}{x} = \infty$. Observe how the limit replaces the undefined term $\frac{1}{0}$. (Thought I'd say ouch, didn't you?)

This is an infinity associated with magnitude. This is an infinity that requires us to think of how tall or how long or how much of something there is. For this reason it would seem that ∞ has little to do with the infinite cardinals \aleph_0 , \aleph_1 , 2^{\aleph_0} .

The calculus also presents us with some other paradoxes associated with the infinite. Consider the fact that if we add three numbers together then we can push them around as follows.

$$\begin{aligned}x + (y + z) &= (x + y) + z \\x + y &= y + x.\end{aligned}$$

Thus

$$1 - 1 + 1 = (1 - 1) + 1 = 0 + 1 = 1$$

and

$$1 - 1 + 1 = 1 + 1 - 1 = (1 + 1) - 1 = 2 - 1 = 1.$$

We arrive at 1 in two different ways. If the number of terms is infinite, problems arise. Consider the sum

$$1 - 1 + 1 - 1 + 1 - \dots$$

Such infinite sums have to be treated with care. Just because we have written it down does not mean it equals a number! For example, if this sum exists then you might say that the sum should support any use of parentheses. However, , when we try the following two groupings of the sum *we get different values*. By grouping using parentheses we see that

$$\begin{aligned}1 - 1 + 1 - 1 + 1 - 1 + \dots &= (1 - 1) + (1 - 1) + (1 - 1) + \dots \\&= 0 + 0 + 0 + \dots \\&= 0,\end{aligned}$$

and by changing the order of terms and then grouping we get

$$\begin{aligned}1 - 1 + 1 - 1 + 1 - 1 + \dots &= 1 + 1 - 1 + 1 - 1 + 1 - 1 + \dots \\&= 1 + (1 - 1) + (1 - 1) + (1 - 1) \dots \\&= 1 + 0 + 0 + \dots \\&= 1.\end{aligned}$$

So which is the answer, 0 or 1? One august mathematician in the eighteenth century suggested that since 0 and 1 were answers that we could also include their average $\frac{1}{2}$. (Now that fraction didn't hurt a bit.) Today we say that the symbols

$$1 - 1 + 1 - 1 + 1 - 1 + 1 - \dots$$

do not converge to a number. The sum is just as undefined as the unmentionable fraction $\frac{1}{0}$. (I think I need an aspirin for this pain.)

Another sum that has problems with the infinite is a *geometric sum*. The reader may have heard that

$$1 + x + x^2 + \dots = \frac{1}{1-x} \quad (5.5)$$

for numbers x between -1 and 1 . Thus, if we let $x = \frac{1}{2}$ in (5.5) then

$$1 + \frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \dots = \frac{1}{1 - \frac{1}{2}} = 2$$

as we will prove in Chapter 6. However, as soon as we try the boundary value $x = -1$ in (5.5), we arrive at a familiar set of symbols that do not mean anything.

$$1 - 1 + 1 - 1 + \dots = \frac{1}{1 - (-1)} = \frac{1}{2}.$$

Maybe that august individual was not so far from the mark.

The sum (5.5) gives an absurd value when we try to use values beyond the boundaries set by mathematics. With $x = 2$ in (5.5) we have our mathematically sensitive mind injured by

$$1 + 2 + 2^2 + 2^3 + \dots = \frac{1}{1 - 2} = -1.$$

That must have hurt you as much as me.

So what is the message here? The message is that we cannot talk about infinite quantities except in the presence of limits. If we are using infinite sums then we must be very careful to work within what is called the *radius of convergence*. This is the neighborhood

in which the sum makes sense. If we take values x outside this neighborhood, the result will be mathematical nonsense. For that reason we say that

$$1 + x + x^2 + x^3 + \dots$$

is only defined for x 's between -1 and 1 , and nowhere else.

This Page Intentionally Left Blank

Chapter 6

Inductions and Numbers

Let us consider the well ordered set \mathbb{N} from a new perspective. We used two properties of \mathbb{N} to define well ordered sets. They were the *Trichotomy Property* and the *Minimum Property*. It is accepted by mathematicians that these properties together with the *Principle of Mathematical Induction* will give us all a common intellectual picture of the natural numbers. At this stage of intellectual development, the professionals agree that everyone who reads these statements will envision the same set of natural numbers that are right this minute dancing in your head. Furthermore, these principles can be extended to well ordered sets to give us a new and powerful tool or argument about well ordered sets. That new tool is called *Transfinite Induction*. Transfinite Induction is to well ordered sets what Mathematical Induction is to the natural numbers.

6.1 Mathematical Induction

We begin with a discussion of the *Principle of Mathematical Induction* or more briefly Mathematical Induction. Intuitively, Mathematical Induction allows us to make a statement for all of the natural numbers by knowing only that the statement holds at 0 and then the statement holds at $n + 1$ if it holds at n . For instance, suppose you are playing a board game (a thought experiment) whose object is to move through all of the playing spaces numbered 0, 1, 2, 3, You move from space to space by rolling a die whose every face has a single black spot \bullet on it. You are on space 0. Next, to get from

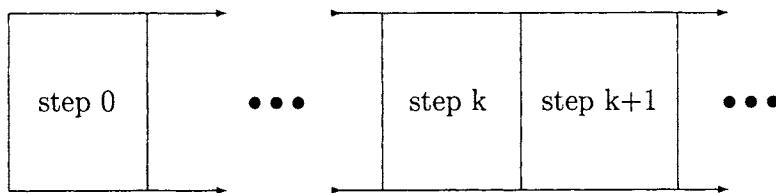
space 100 to 101 you roll the die and move one space. In general, to get from the space numbered n to the space numbered $n+1$ you roll the die and move one space. This is how Mathematical Induction works. You have a starting place and you are told how to get from place to place or, more precisely, from natural number to natural number. A more mathematical discussion follows.

The Principle of Mathematical Induction: Suppose we are given a subset $X \subset \mathbb{N}$. Then $X = \mathbb{N}$ if X satisfies the property

$$0 \in X \text{ and given } k \in X \text{ then } k + 1 \in X.$$

You may have encountered this principle in your Algebra III course in high school, although you may not have given it a name. The Principle of Mathematical Induction tells us that if we have a process P that can be done in the first place, (we say that $P(0)$ is true), and if we can show how to proceed from some statement $P(k)$ to the statement position $P(k+1)$, then we can infer that $P(n)$ is true for each $n \in \mathbb{N}$. That is one efficient way to prove something for infinitely many natural numbers. Please allow me to use ladders with steps instead of rungs.

Here is an illustration of how Mathematical Induction works. Suppose that we have an infinite ladder whose first step is numbered with 0.



We are on the first step of the ladder, and we have a set of instructions that shows us how to get from one step on the ladder to the next in a general way. These directions will not be of the form

Move from step 0 to step 1 and then to step 2.

Such directions will work for a finite ladder, one having, say, 5 steps, but it fails to help us climb a ladder with an infinite number of steps. Rather, the directions should read like this.

If you are on some step then here is how to get to the next step.

This set of directions ensures us that we will be able to traverse every step of this infinite ladder. In elementary mathematics, these instructions are usually a set of equations. In computer programming, this process is the set of instructions in a loop without a stop command. The only way to find some kind of intuition surrounding Mathematical Induction is to read several examples.

Example 6.1.1 The following algorithm writes all of the natural numbers $n > 0$.

1. Let $k = 1$.
2. Write k .
3. Increase k to $k + 1$.
4. Go to step 2.

We will prove that the claim of this algorithm is true using Mathematical Induction.

The *Initial Step* is to check that 1 is written on the first pass. One pass through this algorithm, that is, one reading of lines 1 through 4, confirms that step 2 will write 1 in step 2. Then k is increased from 1 to 2.

The *Induction Hypothesis*: Assume that k is written on the k -th pass through the algorithm.

The next step is the *Induction Step*. Assuming the Induction Hypothesis we must show that $k + 1$ is written on the $(k + 1)$ st pass. So assume that the Induction Hypothesis is true. Assume that k is written on the k th pass. Step 3 of the algorithm increases k to $k + 1$ so that on the next pass (i.e., on pass number $k + 1$) step 2 of the algorithm writes $k + 1$. This is what we had to prove.

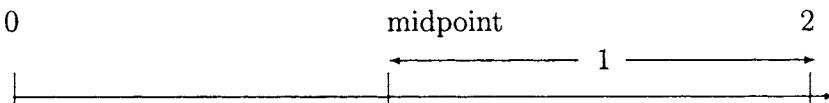
Then by *Mathematical Induction* we conclude that for each natural number $n > 0$, the algorithm writes n on its n th pass.

Notice what we did. We predicted the behavior of the algorithm through *all* of its passes knowing only how it behaves *at some point*

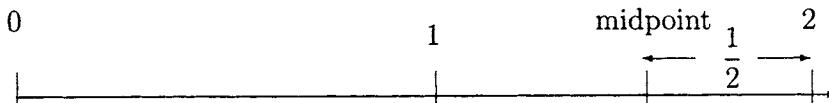
k. While physical restrictions limit a computer to print only finitely many integers using this algorithm, Mathematical Induction allows us to know the behavior of this algorithm through *any number* of passes. Thus no matter how fast or powerful computers get we will know precisely what this algorithm does on any pass.

Here is an example of how Mathematical Induction can replace hand waving in a mathematical discussion. Recall that on page 96 we watched a gentleman take a curious walk across the lobby in Hilbert's Infinite Hotel. He moved first one meter and then $\frac{1}{2}$ a meter and in general on his n th step he moved $\frac{1}{2^{n-1}}$ meters. We will use Mathematical Induction to show that when he has completed his n th step then he is $\frac{1}{2^{n-1}}$ meters from the end of his walk.

Example 6.1.2 Begin with the open interval $(0, 2)$ of numbers properly between 0 and 2. Mark the midpoint of the open interval. The segment yet to be crossed (i.e., the one bordering on 2) has length 1 as in the picture.



Take the interval bordering on 2 and find its midpoint. Continue this process *inductively*. That is, if we have an open interval bordering on 2, then find its midpoint and find the half that borders on 2. For example, in the second step we will find the midpoint of the chosen interval and choose the half bordering on 2 as in the picture.



The segment yet to be crossed has length $\frac{1}{2}$.

We will prove using Mathematical Induction that:

In step $n + 1$ the length of the chosen segment is $\frac{1}{2^n}$.

Initial Step: We are starting with $1 = 0 + 1$. On the 1st step the chosen segment has length $1 = 0 + 1 = \frac{1}{2^0}$. Thus our formula for length works for $n = 0$.

Induction Hypothesis: We will assume that in the $(k + 1)$ st step of our process the segment chosen has length $\frac{1}{2^k}$.

Induction Step: By the Induction Hypothesis, in the $(k + 1)$ st step of the man's walk he chooses a segment of length $\frac{1}{2^k}$. In the next step, the $(k + 2)$ nd step, we take half of the given segment. The result is a segment of length $\frac{1}{2} \cdot \frac{1}{2^k} = \frac{1}{2^{k+1}}$. Thus the segment in step $k + 2 = (k + 1) + 1$ has length $\frac{1}{2^{k+1}}$. This is what we had to prove.

We conclude by Mathematical Induction that the segment in step $n + 1$ has length $\frac{1}{2^n}$ for each $n \in \mathbb{N}$.

Example 6.1.3 Another identity using $\frac{1}{2}$ is the subject of the next induction argument. We will show that

$$2 - \frac{1}{2^n} = 1 + \frac{1}{2} + \cdots + \frac{1}{2^n}$$

is true for each $n \in \mathbb{N}$.

Proof: *Initial Step:* In step 0 check that

$$2 - \frac{1}{2^0} = 1 = \frac{1}{2^0}.$$

Induction Hypothesis: Assume that we have shown that

$$2 - \frac{1}{2^k} = 1 + \frac{1}{2} + \cdots + \frac{1}{2^k}$$

for some $k \in \mathbb{N}$. (This is true since we have proved it true for at least the value $k = 0$.)

Induction Step: We must show that

$$2 - \frac{1}{2^{(k+1)}} = 1 + \frac{1}{2} + \cdots + \frac{1}{2^k} + \frac{1}{2^{(k+1)}}.$$

We begin by grouping the first k terms in the sum.

$$1 + \frac{1}{2} + \cdots + \frac{1}{2^k} + \frac{1}{2^{(k+1)}} = \left(1 + \frac{1}{2} + \cdots + \frac{1}{2^k} \right) + \frac{1}{2^{(k+1)}}.$$

By the Induction Hypothesis and a small bit of algebra we can write

$$\begin{aligned} \left(1 + \frac{1}{2} + \cdots + \frac{1}{2^k} \right) + \frac{1}{2^{(k+1)}} &= \left(2 - \frac{1}{2^k} \right) + \frac{1}{2^{(k+1)}} \\ &= \left(2 - \frac{2}{2^{(k+1)}} \right) + \frac{1}{2^{(k+1)}} \\ &= 2 - \frac{1}{2^{(k+1)}} \end{aligned}$$

after a little bit of arithmetic. This shows us that

$$2 - \frac{1}{2^{(k+1)}} = 1 + \frac{1}{2} + \cdots + \frac{1}{2^{(k+1)}},$$

which completes the Inductive Step.

We conclude by Mathematical Induction that

$$2 - \frac{1}{2^n} = 1 + \frac{1}{2} + \cdots + \frac{1}{2^n}$$

for each $n \in \mathbb{N}$. This completes the proof.

We can use the above example to prove that the curious man in the lobby of Hilbert's Infinite Hotel covers 2 meters when he has finished walking all of the steps possible. On the man's $(n + 1)$ st step he has travelled

$$1 + \frac{1}{2} + \cdots + \frac{1}{2^n}$$

meters. The distance between 2 and this intermediate position is

$$2 - \left(1 + \frac{1}{2} + \cdots + \frac{1}{2^n} \right).$$

By the above example this can be simplified to

$$2 - \left(2 - \frac{1}{2^n} \right) = \frac{1}{2^n}.$$

Thus his final position x and 2 must be separated by less than $\frac{1}{2^n}$ for each n . That is,

$$2 - \frac{1}{2^n} < x \leq 2.$$

Then

$$\lim_{n \rightarrow \infty} 2 - \frac{1}{2^n} \leq x \leq 2.$$

However,

$$\lim_{n \rightarrow \infty} 2 - \frac{1}{2^n} = 2 - 0 = 2$$

so that

$$2 = \lim_{n \rightarrow \infty} 2 - \frac{1}{2^n} \leq x \leq 2.$$

Thus his final position x must be equal to 2.

An application of Mathematical Induction will show that the cleaning lady who cleaned all of the rooms in Hilbert's Infinite Hotel does indeed clean each of the rooms in a total of 2 hours.

Example 6.1.4 Recall that there is a cleaning lady Mary who cleans room 1 in 1 hour, and cleans out each successive room in Hilbert's Infinite Hotel in half the time it takes to clean the previous room. That is, if Mary takes x hours to clean room n then she takes $\frac{1}{2}x$ hours to clean room $n+1$. We will show that Mary cleans all of the rooms in Hilbert's Infinite Hotel and that she cleans room n in $\frac{1}{2^n}$ hours.

Initial Step: Mary cleans room 1. I know because I stayed there. The room was spotless. By design she takes only 1 hour to clean room 1.

Induction Hypothesis: Assume that Mary has cleaned a room, say, room k , in $\frac{1}{2^{k-1}}$ hours.

Induction Step: Mary has cleaned room k by the induction process. She moves to room $k + 1$ and proceeds to clean it. Thus room $k + 1$ is cleaned. She cleans room $k + 1$ in half the time it took to clean the previous room k . Then it takes $\frac{1}{2} \frac{1}{2^{k-1}} = \frac{1}{2^k}$ hours. This is what we had to prove.

We conclude by Mathematical Induction that Mary cleans all of the rooms at Hilbert's Infinite Hotel, and that she cleans room n in $\frac{1}{2^{n-1}}$ hours. This ends the Mathematical Induction proof.

Let us find the total time it takes Mary to complete her work. If we reread Example 6.1.4 we see that Mary takes $\frac{1}{2^{n-1}}$ hours to completely clean room n . We conclude then that the time taken for Mary to clean each of the first $n + 1$ rooms is

$$1 + \frac{1}{2} + \frac{1}{2^2} + \cdots + \frac{1}{2^n}.$$

According to the previous example

$$1 + \frac{1}{2} + \frac{1}{2^2} + \cdots + \frac{1}{2^n} = 2 - \frac{1}{2^n}.$$

It follows that as Mary cleans more rooms the difference between 2 hours and Mary's working time is

$$2 - \left(2 - \frac{1}{2^n}\right) = \frac{1}{2^n}.$$

Thus, in a mathematically precise way, we see that when Mary has completed her rounds (i.e. when she has cleaned room $n + 1$ for all natural numbers $n \in \mathbb{N}$) the difference between 2 hours and her working time is smaller than $\frac{1}{2^n}$ for all $n > 0$. But 0 is the only nonnegative number smaller than the fractions $\frac{1}{2^n}$ for all $n > 0$. We conclude that after Mary has completed her cleaning duties, she has worked exactly 2 hours. That is, Mary requires exactly 2 hours to clean all of these rooms.

Here is another example that uses the number 2 in a clever way.

Example 6.1.5 We will show that

$$1 + 2 + 2^2 + \cdots + 2^n = 2^{n+1} - 1 \text{ for each } n \in \mathbb{N}.$$

Initial Step: $1 = 2^0$ so that

$$2^0 = 2^{0+1} - 1 = 1,$$

providing us with the Initial Step.

Induction Hypothesis: We assume that we have proved that

$$1 + 2 + 2^2 + \cdots + 2^k = 2^{k+1} - 1 \text{ for some } k \in \mathbb{N}.$$

(Again, we did so for $k = 0$.)

Inductive Step: We must show that

$$1 + 2 + 2^2 + \cdots + 2^{k+1} = 2^{k+2} - 1,$$

where k is the integer chosen in the *Induction Hypothesis*.

A little algebra and the Induction Hypothesis shows us that

$$\begin{aligned} 1 + 2 + 2^2 + \cdots + 2^k + 2^{k+1} &= (1 + 2 + 2^2 + \cdots + 2^k) + 2^{k+1} \\ &= 2^{k+1} + 2^{k+1} - 1 \\ &= 2 \cdot 2^{k+1} - 1 \\ &= 2^{k+1+1} - 1 \\ &= 2^{k+2} - 1. \end{aligned}$$

This is what we had to prove.

Therefore, by Mathematical Induction, we have proved that

$$1 + 2 + 2^2 + \cdots + 2^n = 2^{n+1} - 1 \text{ for each } n \in \mathbb{N}.$$

Now here is another mathematically imprecise way to justify the identity given in the above example. Fix $n \in \mathbb{N}$ and use your old friend foil to find the product

$$(2 - 1)(1 + 2 + \cdots + 2^{n-1} + 2^n).$$

Do not simplify this expression as that will make our calculations useless. Then

$$\begin{aligned}
 1 + 2 + \cdots + 2^n &= (2 - 1)(1 + 2 + \cdots + 2^n) \\
 &= \left\{ \begin{array}{l} 2(1 + 2 + \cdots + 2^n) \\ -1(1 + 2 + \cdots + 2^n) \end{array} \right. \\
 &= \left\{ \begin{array}{l} 2 + \cdots + 2^n + 2^{n+1} \\ -1 - 2 - \cdots - 2^n \end{array} \right. \\
 &= -1 + (2 - 2) + \cdots + (2^n - 2^n) + 2^{n+1} \\
 &= -1 + 2^{n+1}.
 \end{aligned}$$

Therefore

$$1 + 2 + \cdots + 2^n = 2^{n+1} - 1,$$

and we have proved the identity in the above example. The reader might question why this argument is mathematically imprecise. The answer lies in the way we stacked the powers of 2. In the above equations we lined up the powers of 2 as follows.

$$\begin{array}{ccccccccc}
 2 & + & 2^2 & + & \cdots & + & 2^n & + & 2^{n+1} \\
 -1 & - & 2 & - & 2^2 & - & \cdots & - & 2^n
 \end{array}$$

We then implicitly declared that *since this stacking of powers of 2 works for the first 2 powers of 2 it works for all powers of 2*. This is the weakness in the argument. There is no mathematical law that says that a pattern that occurs twice will occur in all places. There is no way to imply from the first two places that the stacking will take place in all places. No way, that is, unless we use the Principle of Mathematical Induction.

You might be suggesting right now that the pattern is obvious. We should be able to imply a mathematical truth from the obvious nature of the pattern. I offer the following numerical pattern and ask the reader to fill in the value x in the pattern.

$$1, 2, 3, x.$$

Most readers will choose $x = 4$ as the next obvious value. Some might choose $x = 5$ since the (obvious) pattern could be that the next number is the sum of the previous two.

Another example is the sequence

$$2, 4, 8, x$$

with which we ask for the next number x . Some would say that the pattern is

$$2^1, 2^2, 2^3, 2^4$$

so that $x = 2^4 = 16$. Others might claim that the pattern is more recursive and conclude that the next number is the product of the previous two. Hence

$$2, 4, 2 \cdot 4, 4 \cdot 8 = 32$$

is the most obvious pattern, and so $x = 32$ is the next value. *So which pattern is the most obvious?*

The answer is that *neither is more obvious*. In fact it has been proved that the three numbers 1, 2, 3 and 2, 4, 8 satisfy *infinitely many patterns*, and so no pattern is the obvious pattern. We only choose the *most familiar* patterns. The most obvious pattern cannot be found because there is no way to judge *obviousness* for a pattern from the infinitely many patterns available. Thus we should not look at three numbers and then quickly point out some pattern as the only obvious correct one.

Here is an open question for the reader. Suppose I give you three natural numbers

$$2, 3, 5$$

and ask you for the fourth one x . What pattern did I have in mind? Remember that there are many values for x because there are many patterns whose first three values are 2, 3, 5. I had $x = 8$ in mind. Why is that? Another equally legitimate number would be $x = 7$. Why did I choose these numbers, reader? I leave the first to your imagination. The value $x = 7$ was chosen because it is the next prime number. 2, 3, 5, 7, ... could have been the implied list of prime numbers.

Let us consider an old chestnut as an example of Mathematical Induction in mathematics. We previously examined this example on page 92.

Example 6.1.6 Let $n \in \mathbb{N}$. Then

$$0 + 1 + \dots + n = \frac{n(n+1)}{2}.$$

Proof: As the *Initial Step* of this proof we observe that

$$0 = \frac{0(0+1)}{2}.$$

Thus the formula is true for 0.

Induction Hypothesis: Assume that you have proved the result for *some* natural number $k \geq 0$, and in fact you have. You have proved it for the natural number 0. In this example our *Induction Hypothesis* is

$$0 + \dots + k = \frac{k(k+1)}{2}.$$

We can begin the *Induction Step*. We must give a general argument that shows how the formula for $k+1$ can be derived from the *Induction Hypothesis*. That is, we must prove that

$$0 + \dots + k + (k+1) = \frac{(k+1)((k+1)+1)}{2} = \frac{(k+1)(k+2)}{2}.$$

The sum $0 + \dots + k + (k+1)$ can be regrouped as

$$(0 + \dots + k) + (k+1),$$

so by the Induction Hypothesis

$$(0 + \dots + k) + (k+1) = \frac{k(k+1)}{2} + (k+1).$$

Next, a little algebra shows us that

$$\frac{k(k+1)}{2} + (k+1) = \frac{k(k+1) + 2(k+1)}{2} = \frac{(k+2)(k+1)}{2},$$

or equivalently that

$$0 + \dots + k + (k+1) = \frac{(k+1)(k+2)}{2}.$$

This is what we had to show.

The most important step in the proof is this last one. By Mathematical Induction we conclude that

$$0 + \dots + n = \frac{n(n+1)}{2}$$

is true for each $n \in \mathbb{N}$.

You might ask why we are interested in Mathematical Induction. After all, you might say, there is nothing wrong with the argument we gave on page 92. There is a subtlety that can give us trouble in more general settings. The idea is this. We have written the argument so that the numbers on page 92 match up perfectly as in the following expression.

$$\begin{aligned} S &= 1 + 2 + \dots + \ell - 1 + \ell \\ S &= \ell + \ell - 1 + \dots + 2 + 1 \end{aligned}$$

We know that the sum of the first and last column entries in this expression is $\ell + 1$. *We assumed* that the sum on any column in this expression is $(k) + (\ell - k + 1) = \ell + 1$. But how do we know this? There must be some concrete mathematical reason why the columns in this expression match up as they do, and there is no hint of that reason in the argument we gave on page 92. The argument above using Mathematical Induction provides us with a bridge for the gap in our earlier argument. Mathematical Induction gives us a neat way of jumping from the fact that the first column adds up to $\ell + 1$ to the truth that *each* column adds up to $\ell + 1$. This is exactly the type of mathematical fact on which Mathematical Induction is designed to work.

Here is another algebraic application of Mathematical Induction. Recall that a natural number n is an *odd number* if $n = 2k - 1$ for some $k = 1, 2, 3, \dots$. In general, $2k - 1$ with $k = 1, 2, 3, \dots$ is the k th odd number. Thus $2(1) - 1 = 1$ is the first odd number, $2(10) - 1 = 19$ is the 10th odd number, and $2(101) - 1 = 201$ is the 101st odd number. Galileo (circa 1580) noticed that the sum of the first n odd numbers is the n th perfect square. Specifically, he noticed that if he wrote down the first 2 odd prime numbers that they added up to 2^2 , if he wrote down the first 3 odd prime numbers

that they added up to 3^2 , and if he wrote down the first 4 odd prime numbers that they added up to 4^2 .

$$\begin{aligned} 2^2 &= 1 + 3, \\ 3^2 &= 1 + 3 + 5, \\ 4^2 &= 1 + 3 + 5 + 7. \end{aligned}$$

He then concluded that

$$n^2 = 1 + 2 + 3 + \dots + (2n - 1)$$

but without proof. We will use Mathematical Induction to show that this identity is the case for all natural numbers $n > 0$.

Example 6.1.7 n^2 is the sum of the first n odd numbers.

Proof: *Initial Step:* By replacing n with 1 we see that

$$1^2 = 1 = 2 \cdot 1 - 1.$$

Thus the Initial Step has been established.

Induction Hypothesis: Assume that $k^2 = 1 + \dots + (2k - 1)$ for some natural number $k > 0$.

Induction Step: Show that $(k + 1)^2 = 1 + \dots + (2(k + 1) - 1)$.

Begin by grouping the sum of $k + 1$ terms.

$$\begin{aligned} &1 + \dots + (2k - 1) + (2(k + 1) - 1) \\ &= [1 + \dots + (2k - 1)] + (2k + 1). \end{aligned}$$

The Induction Hypothesis and a little algebra will show that

$$\begin{aligned} [1 + \dots + (2k - 1)] + (2k + 1) &= k^2 + 2k + 1 \\ &= (k + 1)^2. \end{aligned}$$

Thus

$$(k + 1)^2 = [1 + \dots + (2k - 1)] + (2(k + 1) - 1),$$

which is what we had to prove.

Therefore, by Mathematical Induction,

$$n^2 = 1 + 3 + 5 + \dots + (2n - 1)$$

for all natural numbers $n \geq 1$. This completes the proof.

Here is another story in mathematics that shows us that patterns are not always as obvious as they first appear. The equation

$$ax + b = 0 \text{ and } a \neq 0$$

can be solved by even the youngest of mathematical students. It is important to note that the exponent of x is 1 in this equation. Next, the high school student knows that the *quadratic formula*

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \quad (6.1)$$

will solve the quadratic polynomial equation

$$ax^2 + bx + c = 0 \text{ and } a \neq 0.$$

The solution features a square root (a radical) \sqrt{y} and the coefficients a, b, c of the polynomial equation. Notice that the highest power of x that occurs is 2. We say that the expression (6.1) is a *solution by radicals of the polynomial equation of degree 2*.

In about 1535 AD the Italian mathematician Tartaglia showed that the equation

$$ax^3 + bx^2 + cx + d = 0 \text{ and } a \neq 0$$

can be solved by radicals. His solution featured cube roots $\sqrt[3]{y}$, square roots, and the coefficients on the polynomial equation. Since this polynomial equation has degree 3, Tartaglia's solution is said to be a *solution by radicals of the polynomial equation of degree 3*.

Tartaglia's solution was published in 1545 in Cardano's book *Ars Magna*. The *Ars Magna* also contained the first publication of the solution to the polynomial equation

$$ax^4 + bx^3 + dx^2 + ex + f = 0 \text{ and } a \neq 0.$$

The solution to this polynomial equation of degree 4 also featured fourth roots $\sqrt[4]{y}$, cube roots, square roots, and the coefficients of the polynomial equation. Thus we have solutions by radicals for polynomial equations of degrees 1, 2, 3, and 4. Is there a pattern here?

The above history may seem to indicate a pattern. If polynomial equations of degree 1, 2, 3, 4 can be solved by radicals, we might

then suggest that it is obvious that the fifth degree polynomial equation

$$ax^5 + bx^4 + dx^3 + ex^2 + fx + g = 0 \text{ and } a \neq 0$$

has a solution by radicals. Unfortunately, the mathematical gods are playing games with us. The supposed pattern is misleading. In 1821 a 21 year old Norwegian mathematician, Niels Abel, showed that there is no solution *by radicals* to the polynomial equation of degree 5. In fact, an 18 year old French mathematician, Evariste Galois, showed that except for these first four degrees, the general polynomial equation is not solvable by radicals like the first four are. So what happened to our obvious pattern? It must be that there was never a pattern to find. The pattern evaporated. It disappeared. *It was never there.* This is an important thought. Mathematicians must be careful in deciding when a pattern exists and when it does not exist. This is not as easy as it might seem since most people seem to see patterns everywhere. The best defense against seeing phantom patterns is to use a good solid argument that includes Mathematical Induction.

The next story concerns a misuse of Mathematical Induction. This example elegantly demonstrates the importance of the *Initial Step* in a proof using Mathematical Induction. The story is called **The Unexpected Termination**.

Example 6.1.8 A company Boss wants to fire a Supervising Barber who works in the Hair Styling Branch of the company. The firing will take place at 9:00 AM sometime between Sunday and Saturday of next week. However, since the Boss does not want the Barber to anxiously await his last day, he decides not to tell the man which day will be his last. The Barber catches wind of this compassion and decides to use it to his advantage.

The Barber decides that he will not be terminated. His argument is as follows. He reasons that if he has not heard about his termination by 9 AM Friday then he most certainly will not be terminated the next day, Saturday. This is because the Boss must fire this Barber at 9 AM Saturday. He would then be upset all day Friday while he waited for the axe to fall on Saturday. The compassionate Boss does not want the Barber to suffer in that way, so the

Barber concludes that he will not be terminated Saturday. He reasons that he can use this fact as the initial step in a Mathematical Induction.

Suppose that the Barber has argued that he will not be terminated on day $k - 1$. (Think of Saturday as day 1, Friday as day 2, and so on to Sunday which is Day 7.) He reasons about Day k as follows. If he has not heard about his termination by 9 AM of Day $k + 1$ then he most certainly will not be terminated the next day, Day k , because in that case the Barber would be upset all day long Day $k + 1$ contrary to the compassionate Boss's wishes. Thus the Barber concludes that he will not be terminated on Day k . The Barber concludes by Mathematical Induction that he will not be terminated on any day n that week. Unfortunately, the Boss terminates the Barber at 9 AM on Wednesday. The Barber's argument must have been flawed, but where is that flaw?

Let us examine that argument in detail. The Barber will not be terminated Saturday for the reasons given in the Induction Step. So maybe Friday is the termination day. If so then Thursday morning at 10 AM he will know that he is not to be terminated Thursday, and so a Friday termination is the only thing possible. He would then fret, contrary to the compassionate Boss' wishes. He concludes that Friday is not the day. Will Thursday be his termination day? No. The same argument applies. Try it yourself. We then use the familiar dots, . . . , to conclude that the Barber will not be terminated that week.

So where is the error in that argument? I leave it to you to think about The Unexpected Termination for a while. Try not to peek at the answer before you work with the problem a little.

As we said earlier the answer concerns the Initial Step. His initial assumption is that he has a job on Friday. This renders the rest of the argument mute because if he has a job Friday then *he has not been terminated at all*. Thus he has assumed that which he wants to prove, and there is his mistake. If you want to prove that you will not be terminated then you cannot make an assumption that assumes that you will not be terminated.

We hope that these examples have helped you to understand the process called Mathematical Induction. Recapping, we begin with a process P that is defined at each $n \in \mathbb{N}$. Induction begins by

showing that the process is true for some initial value, say, 0. In the *Induction Hypothesis* we assume that $P(k)$ is true for some $k \in \mathbb{N}$, and in the *Induction Step* we show how $P(k)$ leads us to $P(k + 1)$. We then conclude by Mathematical Induction that $P(n)$ is true for all $n \in \mathbb{N}$.

Use Mathematical Induction to show the following.

1. $1^2 + 2^2 + \dots + n^2 = \frac{1}{6}n(n + 1)(2n + 1)$ for each $n \in \mathbb{N}$.
2. $1^3 + 2^3 + \dots + n^3 = \frac{1}{4}n^2(n + 1)^2$ for each $n \in \mathbb{N}$.
3. $(1 - x)(1 + x + x^2 + \dots + x^n) = 1 - x^{n+1}$ for each $n \in \mathbb{N}$.

6.2 Transfinite Induction

The power of Mathematical Induction comes from the fact that it works for *any* process defined for each $n \in \mathbb{N}$. The limits of Mathematical Induction come from the fact that it only works for processes defined for each $n \in \mathbb{N}$. There are some very important processes that are defined for *every ordinal* α . Thus if P is a process defined for *ordinals*, then Mathematical Induction will show that $P(n)$ is true for each $n \in \mathbb{N}$ but it will miss the truth of $P(\omega_0)$. For example, we used Mathematical Induction to show that the cleaning lady in Hilbert's Infinite Hotel on page 98 will eventually clean all of the rooms labelled with $n \in \mathbb{N}$. However, our discussion missed the room for Hotel Staff labelled with ω_0 . We found her in that room on page 184. In this way Mathematical Induction can be applied to processes defined for each natural number $n < \omega_0$ but it fails for processes defined for the ordinals $\alpha \geq \omega_0$. In this section we will present an induction that can be applied to infinite ordinals as well as \mathbb{N} .

Let P be a process that is defined for all ordinals α . In other words, $P(\alpha)$ is *defined* or makes mathematical sense for each ordinal α . A *proof by Transfinite Induction* will proceed as follows.

1. *Initial Step:* Prove that $P(0)$ is true.
2. *Induction Hypothesis:* Assume that there is an ordinal α such that $P(\beta)$ is true for each ordinal $\beta < \alpha$.

3. *Induction Step:* Show that $P(\alpha)$ is true.

4. Conclude that $P(\gamma)$ is true for all ordinals γ .

We will use Transfinite Induction to replace the phrases *continue indefinitely* and *continue in the same manner* that accompanied some vague arguments that we encountered in some earlier proofs. Some examples of arguments using Transfinite Induction will help us see what all of this formality is about.

Example 6.2.1 The first example is the simplest possible. We will construct a well ordered set.

Initial Step: We choose a smallest element which we denote by

$$P(0) = x_0.$$

Induction Hypothesis: We assume that for some ordinal α we have constructed a set

$$P(\alpha) = \{x_\beta \mid \beta < \alpha\}$$

with an ordering given by

$$x_0 < x_1 < \dots < x_\beta < \dots \text{ for each ordinal } \beta < \alpha.$$

Hopefully you see the pattern.

The order on $P(\alpha)$ is not a well ordering since it is possible that $x_{\alpha-1}$ is in $P(\alpha)$ but by definition of $P(\alpha)$ no successor element for $x_{\alpha-1}$ exists in $P(\alpha)$. For example, given $\alpha = 5$ our set is

$$P(5) = \{x_0, x_1, x_2, x_3, x_4\} = \{x_\beta \mid \beta < 5\}$$

and the ordering is

$$x_0 < x_1 < x_2 < x_3 < x_4.$$

The element x_4 does not have a successor element.

For $\alpha = \omega_0$ our set is

$$P(\omega_0) = \{x_0, x_1, x_2, \dots\} = \{x_\beta \mid \beta < \omega_0\}.$$

Unlike the previous set, this ordering is without end.

$$x_0 < x_1 < x_2 < \dots$$

Before reading on, the reader should write down the set corresponding to x_{ω_0+1} . It is

$$P(\omega_0 + 1) = \{x_0, x_1, x_2, \dots, x_{\omega_0}\} = \{x_\beta \mid \beta < \omega_0 + 1\}.$$

We make no special hypothesis about the nature of the symbols x_n . We *define* the order in the set $P(\omega_0 + 1)$ as

$$x_0 < x_1 < x_2 < \dots < x_{\omega_0}.$$

Notice that x_{ω_0} is the unique largest element in this set. It has no successor in $P(\omega_0 + 1)$.

Induction Step: The object here is to prove that there is a set

$$P(\alpha + 1) = \{x_\beta \mid \beta < \alpha + 1\}$$

whose ordering is given by

$$x_0 < x_1 < \dots < x_\beta < \dots < x_\alpha \text{ for each ordinal } \beta < \alpha.$$

Choose a symbol x_α that is not in $P(\alpha)$ and form the set

$$P(\alpha + 1) = \{x_\beta \mid \beta < \alpha + 1\} = \{x_\beta \mid \beta < \alpha\} \cup \{x_\alpha\}$$

by simply requiring that

$$x_\beta < x_\alpha$$

for each ordinal $\beta < \alpha$.

We conclude, by Transfinite Induction, that there is a well ordered set

$$P = \{x_\gamma \mid \gamma \text{ is an ordinal}\}.$$

This set is well ordered because for each element x_α there is a successor element $x_{\alpha+1}$. The element $x_{\alpha+1}$ exists by the Induction Step above. Actually, the limit P is not a set. It is a collection. It turns out that P is larger than any set so it cannot be a set. We will have more to say about this kind of anomaly later.

Notice in the above example that we did not have to prove anything for $\alpha + 1$ as we did in Mathematical Induction. We use the Induction Hypothesis to prove something about the construction process as it exists prior to α . Try to keep this in mind as these examples continue.

Example 6.2.2 Here is another example of Transfinite Induction. Suppose that there is some universal set \mathcal{U} given, and suppose that there is a *transfinite chain of subsets of \mathcal{U}*

$$U_0 \subset U_1 \subset \dots \subset U_\alpha \subset \dots \subset \mathcal{U},$$

where α ranges over all of the ordinals. Further suppose that there is some special element $u \in \mathcal{U}$ that is not in U_α for any ordinal α :

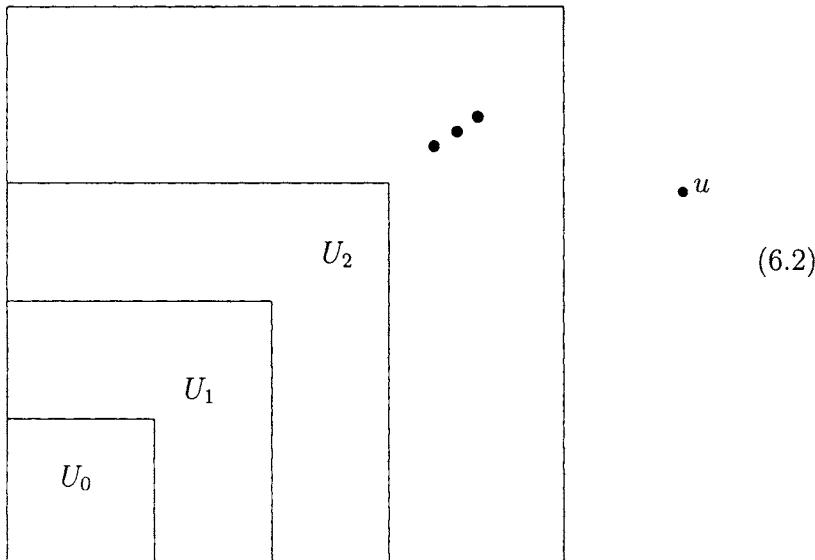
$$u \notin U_\alpha \text{ for each ordinal } \alpha.$$

We will show, using Transfinite Induction, that

$$u \notin \bigcup_{\text{all ordinals } \alpha} U_\alpha.$$

Picture (6.2) will help you to visualize what we are doing. The set U_2 includes U_1 and the set U_3 includes U_2 . The set \mathcal{U} is the union of all of the sets U_α , α an ordinal. Notice that the element u is not in any of the sets U_α .

$$\mathcal{U} = \bigcup_{\text{all ordinals } \alpha} U_\alpha$$



Initial Step: It is clear that $u \notin U_0$ so that

$$u \notin \bigcup_{\beta < 1} U_\beta = U_0.$$

This begins the induction process.

Induction Hypothesis: Let us assume that

$$u \notin \bigcup_{\beta < \alpha} U_\beta$$

for some ordinal α .

Induction Step: In the Induction Step we must show that

$$u \notin \bigcup_{\beta < \alpha+1} U_\beta.$$

From the Induction Hypothesis we know that

$$u \notin \bigcup_{\beta < \alpha} U_\beta,$$

and by the hypothesis that starts this problem $u \notin U_\alpha$. Then by the definition of the set operation \cup we have

$$u \notin U_\alpha \cup \left(\bigcup_{\beta < \alpha} U_\beta \right) = \bigcup_{\beta \leq \alpha} U_\beta = \bigcup_{\beta < \alpha+1} U_\beta.$$

This is what we had to prove.

We conclude, by Transfinite Induction, that

$$u \notin \bigcup_{\text{all ordinals } \beta} U_\beta.$$

What we have just shown is that if the U_α do not contain a point u then the union of the U_α fails to contain that point u . We accomplished this by showing that the union

$$U_0 \cup U_1 \cup U_2 \cup \dots \cup U_\beta \text{ for } \beta < \alpha$$

does not contain the point u and then concluded by Transfinite Induction that the total union does not contain the point u .

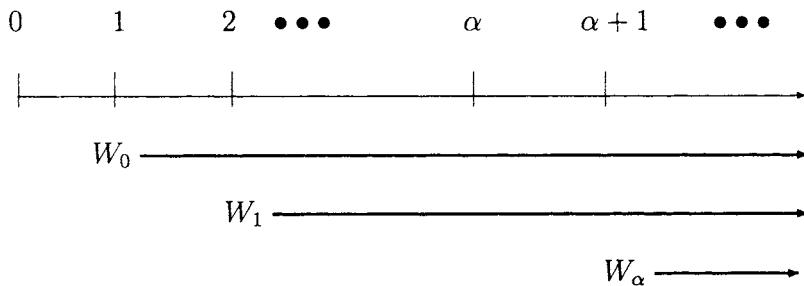
Example 6.2.3 Given an ordinal α let

$$W_\alpha = \{\text{ordinals } \gamma \mid \alpha < \gamma\}.$$

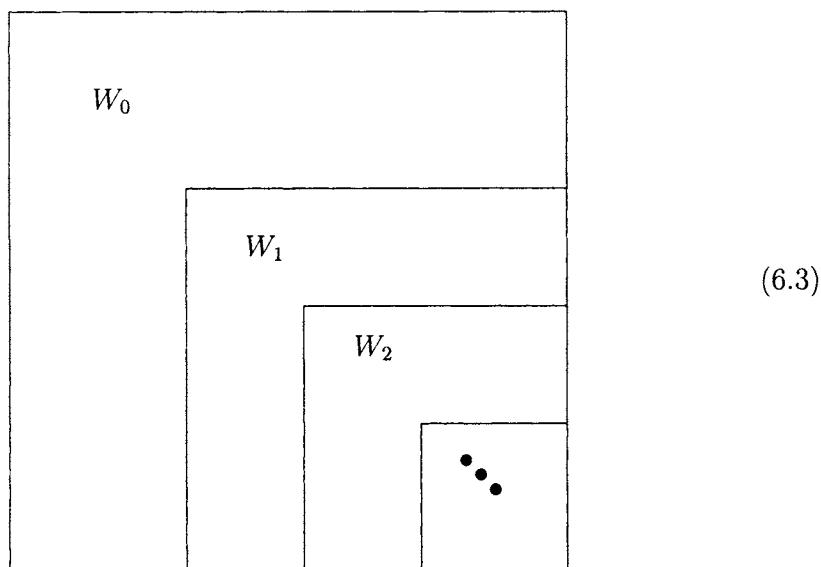
The set W_α is the chain of ordinals that lie above α on the chain of ordinals. Thus W_α can be thought of as a line of ordinals beginning at $\alpha + 1$ and continuing on indefinitely. So

$$\alpha + 1, \alpha + 2, \alpha + 3, \dots \in W_\alpha.$$

In terms of a picture we have



We might also envision the sets W_α as follows.



Using Transfinite Induction we will show that

$$\bigcap_{\text{all ordinals } \alpha} W_\alpha = \emptyset.$$

Initial Step: $1 \notin W_1$ since W_1 is the set $\{\gamma \mid 1 < \gamma\}$ of ordinals γ that are strictly larger than 1. Since

$$\bigcap_{\text{all ordinals } \alpha} W_\alpha \subset W_1,$$

we conclude that

$$1 \notin \bigcap_{\text{all ordinals } \alpha} W_\alpha.$$

Induction Hypothesis: Assume that for some ordinal β we have

$$\beta \notin \bigcap_{\text{all ordinals } \alpha} W_\alpha.$$

Induction Step: Since $W_{\beta+1} = \{\text{ordinals } \gamma \mid \beta + 1 < \gamma\}$, it follows that

$$\beta + 1 \notin W_{\beta+1}.$$

Since

$$\bigcap_{\text{all ordinals } \alpha} W_\alpha \subset W_{\beta+1},$$

we find that

$$\beta + 1 \notin \bigcap_{\text{all ordinals } \alpha} W_\alpha.$$

Therefore, by Transfinite Induction,

$$\gamma \notin \bigcap_{\text{all ordinals } \alpha} W_\alpha$$

for any ordinal γ . Thus $\bigcap_{\text{all ordinals } \alpha} W_\alpha$ does not contain any ordinals, which implies that

$$\bigcap_{\text{all ordinals } \alpha} W_\alpha = \emptyset.$$

Example 6.2.4 (With an acknowledgment to Douglas R. Hofstadter [5].) This is a story that illustrates how Transfinite Induction takes over where Mathematical Induction ends.

It seems that the magic lamp was passed down from generation to generation of Aladdin's family. The current heir to the lamp, Al, is a mathematician with a droll sense of humor. When he rubbed the lamp the genie appeared and said, "You have three wishes." Using his droll humor he said, "I wish for more wishes." "That is a *wish about wishes* and I cannot grant such a thing," the genie responded. "It would be like the Mayor of New York City acting like the President of the United States. Before I can grant your wish I must ask for my Boss' permission. His name is

M_1 .

M_1 grants my wishes."

So the genie contacted M_1 and said, "I wish to grant a wish about wishes." M_1 thought for a minute and said, "You have made a wish about a wish about wishes. Before I can grant your wish I must ask for my Boss' permission. His name is M_2 . M_2 grants my wishes."

Once M_2 is contacted M_1 asks, "I wish to grant a wish about a wish about wishes." Of course, M_2 saw through that wish right away and said, "That is a wish about a wish about a wish about wishes, and I cannot grant it without the permission of my boss, M_3 . M_3 grants my wishes."

The process of asking the boss M_n continued once for each $n \in \mathbb{N}$, where

M_{n+1} grants the wishes to M_n .

(The observant reader might observe that we did not define M_0 . We will assume that M_0 is Al. Thus M_1 grants M_0 's wishes.) The problem was that no one could give the permission that would eventually grant Al's wish. No matter how far up the chain the

request travelled no boss M_{n+1} could grant permission to the boss M_n . The chain of bosses $\{M_n \mid n \in \mathbb{N}\}$ has no maximal element.

The chain of bosses $\{M_n \mid n \in \mathbb{N}\}$ is only a part of a well ordered chain of bosses. Let me introduce you to the boss of all of the chain $\{M_n \mid n \in \mathbb{N}\}$. His name is

$$M_{\omega_0}.$$

Because he is a good administrator, he realized the predicament his people were in so he thought he would just give all of the M_n permission to grant those wishes. But you see, being an administrator M_{ω_0} must ask permission from his boss M_{ω_0+1} before he can grant wishes about wishes for each M_n with $n \in \mathbb{N}$.

The transfinite process here is now quite clear. Suppose that we are given an ordinal α such that

$$\text{boss } M_\beta \text{ exists for each ordinal } \beta < \alpha$$

and such that

$$M_{\beta+1} \text{ grants } M_\beta \text{'s wishes for each ordinal } \beta < \alpha.$$

Define boss M_α who is the boss of all M_β with $\beta < \alpha$. Then we have extended our discussion to include all bosses M_β such that $\beta \leq \alpha < \alpha + 1$. By Transfinite Induction, for each ordinal k , each boss M_{k+1} grants wishes for boss M_k .

The problem is that there is no boss M at the top of this chain to give everyone permission to grant wishes. The story has a happy ending though, as The Supreme One saw what was going on and said to each boss M_γ , “Grant Al’s wishes and let’s get on with the work of the day.” And so it was done.

Before continuing on to the next section let me present a process called *Student Induction*. It seems that a certain University cancels

classes the Wednesday prior to Thanksgiving Thursday. This gives the students an initial day off. They then rationalize that since there are no classes Wednesday they can leave for home Tuesday night. This is not too hard to understand. The anticipation of the holiday being what it is they decide to leave earlier that day, Tuesday afternoon. No, make it Tuesday morning because knowing that they are leaving for home at noon, they lose their concentration in the morning classes. But then, why not leave Monday. Well nobody is going to stay at college for one day of classes so they argue that they can leave Sunday night. Get real. No one will stay the weekend to leave Sunday night, so they leave late Friday. No. Not late. No one has enthusiasm for Friday classes so they decide to leave Thursday night. With this kind of *Student Induction* taking place the student is soon leaving on Halloween (October 31) in anticipation of the Thanksgiving holiday (the last Thursday in November). The same process is at work in the Spring Semester when the students prepare for Spring Break. They are soon leaving in January for a vacation that does not take place until March. We leave the details to be filled in by the reader.

6.3 Mathematical Recursion

Another process that is designed to take the ellipsis ... out of mathematical definitions is call *recursion*. Recursion is a mathematical device that defines one object in terms of another smaller object. This smaller object is similar to the larger one but not at quite the same size. The first recursion usually seen by the student is n factorial. This n factorial is defined as

$$\begin{aligned} 0! &= 1 \text{ and} \\ n! &= n(n-1)! \text{ for integers } n \geq 1. \end{aligned}$$

Do you see that $n!$ is defined in terms of the smaller number $(n-1)!$? For example,

$$\begin{aligned} 1! &= 1 \cdot 0! = 1 \cdot 1 = 1, \\ 3! &= 3 \cdot (2!) = 3 \cdot (2 \cdot 1!) = 3 \cdot (2 \cdot 1) = 6, \\ 10! &= 10 \cdot (9!) = 10 \cdot 9 \cdot 8! = 10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1. \end{aligned}$$

We do not define $n!$ for negative numbers and we do not define $n!$ for fractions. The number $n!$ in this book is defined only for whole nonnegative numbers n .

The definition of $n!$ is recursive because it is defined in terms of a factorial at a smaller value. We can define other numbers in this way. We let

$$\begin{aligned}T_0 &= 0 \text{ and} \\ T_n &= n + T_{n-1} \text{ for integers } n \geq 1.\end{aligned}$$

T_n is called the n th *triangular number* and it is the number of dots needed to form a triangular array. For instance, since one dot forms a (trivial) triangular array, $T_1 = 1$ is a triangular number.

•

The formula for triangular numbers shows us that

$$T_2 = 2 + T_1 = 2 + 1 = 3.$$

Thus 3 is the number of dots needed to form the next larger triangular array, as in the illustration below.



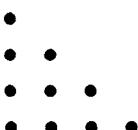
For larger triangular numbers proceed recursively.

$$T_3 = 3 + T_2 = 3 + 3 = 6$$

since we know that $T_2 = 3$. The triangular array for T_3 is formed by adding a row of three dots into the previous triangular array.



Notice that T_3 is found by counting the number of dots in a triangular array of dots whose base has 3 dots. The same is said for $T_4 = 4 + T_3 = 4 + 6 = 10$.



Actually, we were introduced to triangular numbers when we examined the sum

$$1 + 2 + 3 + \dots + n.$$

A little induction will show that

$$T_n = 1 + 2 + 3 + \dots + n \text{ for } n > 0.$$

Proof: Evidently $T_1 = 1$ so the Initial Step is defined.

Suppose that $T_k = 1 + 2 + \dots + k$ for some integer $k > 0$. This is the Induction Hypothesis.

We must show that $T_{k+1} = 1 + 2 + \dots + (k + 1)$. By definition and the Induction Hypothesis,

$$\begin{aligned} T_{k+1} &= (k + 1) + T_k \\ T_{k+1} &= (k + 1) + (1 + 2 + \dots + k) \\ T_{k+1} &= (1 + 2 + \dots + k) + (k + 1). \end{aligned}$$

Therefore, by Mathematical Induction,

$$T_n = 1 + 2 + \dots + n$$

for each integer $n > 0$.

Subsequently, by using the formula found by little Gauss on page 92, we see that

$$T_n = \frac{n(n + 1)}{2} \text{ for each integer } n > 0. \quad (6.4)$$

Our recursive formula for $T_n = n + T_{n-1}$ is then replaced by a closed formula for $T_n = \frac{n(n + 1)}{2}$.

Closed formulas have their advantages. The most immediate advantage is that we can calculate T_{10} without calculating T_n for all $0 < n < 10$.

$$T_{10} = \frac{10(10 + 1)}{2} = 55.$$

Larger triangular numbers can be calculated without the use of smaller triangular numbers or of triangular arrays.

$$T_{100} = \frac{100(100 + 1)}{2} = 5050.$$

This is the advantage of closed formulas.

There is an interesting formula involving triangular numbers and perfect squares. We will show that

$$T_n + T_{n-1} = n^2 \text{ for each integer } n > 0.$$

In words, consecutive triangular numbers add up to a perfect square. That should sound satisfying to you. Two triangular numbers add up to a square number. How many times have you broken a square into two triangles using a pencil and a ruler? Now you have a numerical identity that does the same thing to numbers. A square number is the sum of two triangular numbers.

Proof: By (6.4) we have

$$\begin{aligned} T_n + T_{n-1} &= \frac{n(n+1)}{2} + \frac{(n-1)n}{2} \\ &= \frac{n}{2}((n+1) + (n-1)) \\ &= \frac{n}{2}(2n) \\ &= n^2. \end{aligned}$$

This is what we had to prove. Therefore $T_n + T_{n-1} = n^2$ for each integer $n \geq 1$.

Another sum that adds up to a perfect square is due to Galileo (circa 1600 AD). Observe that

$$1 + 3 = 4$$

and that

$$1 + 3 + 5 = 9.$$

Thus the sum of the first two odd numbers is the second perfect square, and the sum of the first three odd numbers is the third perfect square. Care to guess about the sum

$$1 + 3 + \dots + 29$$

of the first 15 odd numbers? Of course the pattern is set. The sum is 15^2 , the 15th perfect square. Care to prove this pattern holds for all integers $n \geq 1$, reader? Let us do just that.

$$1 + 3 + \dots + (2n - 1) = n^2.$$

The sum of the first n odd numbers is the n th perfect square.

Proof: We use induction. Since $1 = 1^2$ we have established the initial step.

Assume that $1 + 3 + \dots + (2k - 1) = k^2$. This is the Induction Hypothesis. We must show that

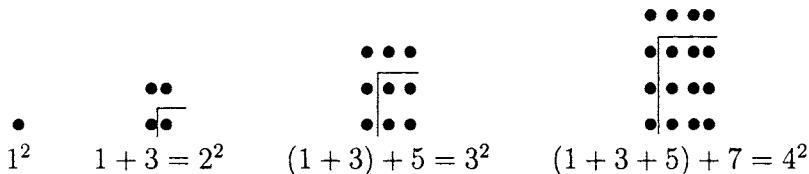
$$1 + 3 + \dots + (2(k + 1) - 1) = (k + 1)^2.$$

By the induction hypothesis

$$\begin{aligned} 1 + \dots + (2(k + 1) - 1) &= (1 + \dots + (2k - 1)) + (2(k + 1) - 1) \\ &= k^2 + (2k + 1) \\ &= (k + 1)^2. \end{aligned}$$

Therefore, by Mathematical Induction, the sum $1 + \dots + (2n - 1)$ of the first n odd numbers is the n th perfect square n^2 .

The first few sums of odd numbers can be seen in the following diagrams. Think of one square as being formed by adding dots to the edge and then count the odd number of dots that are added to the recursive square.



Draw the next square in the pattern to see that $4^2 + 9 = (1 + 3 + 5 + 7) + 9 = 5^2$.

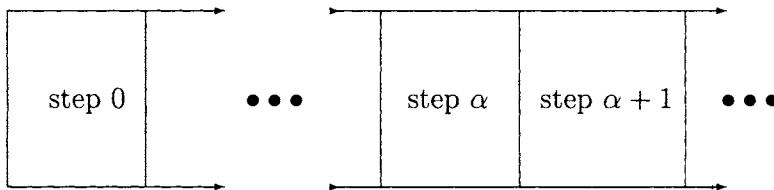
Here is an example of recursion in an infinite setting.

Example 6.3.1 Hilbert has expanded the Infinite Hotel to accommodate any and all guests. To do this he adds on *one room for each ordinal* to his Infinite Hotel using recursion.

The Intial Step is to construct room 1.

The Induction Hypothesis is to assume that there is an ordinal α for which we have constructed rooms β for each ordinal $\beta < \alpha$. Let H_α be this construction. We construct room α onto H_α thus arriving at a hotel with rooms β for ordinals $\beta < \alpha + 1$. We have constructed $H_{\alpha+1}$.

Therefore, by Transfinite Induction, we have constructed the hotel H with rooms β for each ordinal β . The result is something more. The Infinite Hotel is now a *Transfinite Hotel*. For instance, there is a room ω_0 , a room $\omega_0^{\omega_0}$, and so on.



The Transfinite Hotel is not a set, though. It is bigger than any set can be. We will consider the matter more closely later.

Mary cleans each room α and then cleans room $\alpha + 1$. Time is no longer an issue since Hilbert and the Hotel Staff signed a collective bargaining agreement that stops time. Thus, no matter how many rooms are cleaned, no time has passed. It also means that when she goes home there is no end to her leisure time since that time does not pass. In this way Mary can start with room 1, then clean all of the rooms in the Transfinite Hotel within her work day, and still have time for the grandchildren Paul and John. We will show that she does indeed clean all of the rooms.

The initial room, room 1, is cleaned shortly after she arrives at work.

Inductively, assume that there is some ordinal α such that Mary has cleaned room β for each ordinal $\beta < \alpha$.

Having cleaned rooms β for $\beta < \alpha$ Mary finds herself in front of room α . Since she is industrious, she enters the room and cleans it.

Thus room α is cleaned so that room β is cleaned for each ordinal $\beta < \alpha + 1$.

We conclude by Transfinite Induction that Mary has cleaned room γ for each ordinal γ . At this time she joins a colleague and sits and talks in the Hotel Staff Lounge.

6.4 Number Theory

The numbers considered in this section are whole numbers and *non-negative* numbers. We will not be considering negative fractions, although we will need to consider $0, 1, 2, \dots$. In this section we will use mathematical induction and recursion to give mathematical certainty to some facts we have all heard about integers.

The number $p > 1$ is said to be *prime* exactly when

$$p = ab \text{ implies that } a = p \text{ or } a = 1.$$

Another way of saying this is that p is not 1 and p is divisible only by itself and 1. A number a is a *factor* of n if there is a number b such that $ab = n$. The number a is a *prime factor* of n is a is a prime number and a is a factor of n . For example, 3 is a prime factor of 6 and 15 while 7 is a prime factor of 49. Prime factors, it seems, are everywhere.

Lemma 6.4.1 *Every number $n > 1$ has a prime factor.*

Proof: Let us assume to the contrary that there is a positive number n that has no prime factors. We take the smallest such number n . Then n is not prime (since otherwise it would be a prime factor of itself) so we can write

$$n = ab$$

for some numbers a, b not in $\{n, 1\}$. Since $a \neq n$ and $b \neq 1$ divide n , a is smaller than n .

$$a = \frac{n}{b} < n.$$

Since n is the smallest number without a prime factor, a has a prime factor p , say,

$$a = pc.$$

But then p is also a prime factor of n since

$$n = ab = p(cb).$$

This contradiction to our assumption shows us that every number n possesses a prime factor.

The next result is one that goes back to the Greek geometer Euclid (circa 300 BC). He proves that *there are infinitely many prime numbers*. That may sound obvious at first. You have known it since birth or since early in your education. In fact, you were given a known mathematical fact and you were asked to memorize it. You accepted it as a truth. But mathematics can give certainty to the arithmetic facts you learned as a child. That is what we will do here. We will give mathematical certainty to some arithmetic facts that we learned at a young and impressionable age.

Theorem 6.4.2 [Euclid] *There are infinitely many prime numbers.*

Proof: We offer a proof by contradiction. Suppose to the contrary that there is only a finite list of prime numbers. Write down that finite list.

$$p_1, p_2, \dots, p_t.$$

Consider the number

$$N = p_1 p_2 \cdots p_t + 1.$$

Since every number has a prime factor, N has a prime factor p . However, $p \neq p_1$ since division of N by p_1 leaves a remainder of 1, $p \neq p_2$ since division of N by p_2 leaves a remainder of 1, and so on. $p \neq p_t$ since division of N by p_t leaves a remainder of 1. Since p_1, p_2, \dots, p_t is supposed to be a list of all of the primes and since p is a prime not on the list, we have found our contradiction. Thus the list of primes is infinite.

Let us try to show what this result has done. Given any finite set of primes it shows us *how to locate a new prime*. For instance, the number

$$N = 2 \cdot 3 \cdot 5 \cdot 7 + 1 = 215,442$$

is divisible by a prime number p such that $p \neq 2, 3, 5, 7$. We can verify this by writing down the number N in a more familiar way.

$$N = 2 \cdot 3 \cdot 5 \cdot 7 + 1 = 121 = 11^2.$$

Thus the new prime is 11. Eleven was not on our finite list of primes, but it is surely not the last prime that exists.

In the same way the number

$$N = 17 \cdot 19 \cdot 23 \cdot 29 + 1$$

has a prime divisor p . We do not know what p is but we can say with certainty that p is not on the finite list of primes

$$17, 19, 23, 29.$$

This is what the proof of Theorem 6.4.2 has done.

Just how many times is a number n divisible by a prime? Infinitely often? Finitely often? The next lemma resolves this question.

Lemma 6.4.3 *Let n be a number with prime divisor p . There is an exponent t and a number b such that $n = p^t b$ and b is not divisible by p . That is, p divides n exactly t times and no more.*

Proof: We proceed by contradiction. Suppose there is a number n and a prime p such that p divides n infinitely often. That is, we can write

$$n = pb_1 = p^2 b_2 = \dots$$

for some integers b_1, b_2, \dots . Then

$$b_1 = \frac{n}{p} > b_2 = \frac{n}{p^2} > b_3 = \frac{n}{p^3} > \dots$$

is an infinite list of smaller and smaller positive integers. This is impossible so our supposition was incorrect. Each prime divisor of n divides n at most finitely often and no more. This completes the proof.

Theorem 6.4.4 Let $n \geq 2$ be a positive integer. There are primes p_1, \dots, p_t such that

$$n = p_1 \dots p_t. \quad (6.5)$$

Let us agree to call (6.5) a *prime factorization of n* .

Proof: We apply induction. The result is true for $n = 2$ since in this case $n = p_1 = 2$.

Assume that we have arrived at an integer $k \geq 2$ for which each number $m \leq k$ has a prime factorization like (6.5). This is our Induction Hypothesis.

We must find a prime factorization for $k + 1$. By Lemma 6.4.1, $k + 1$ has a prime factor p . Then

$$\frac{k+1}{p} < k+1$$

has a prime factorization by the Induction Hypothesis. Hence we can write

$$\frac{k+1}{p} = p_1 \dots p_s$$

for some primes p_1, \dots, p_s . It follows that

$$k+1 = p(p_1 \dots p_s),$$

which is a prime factorization of $k + 1$.

Therefore, by Mathematical Induction, each integer $n \geq 2$ has a prime factorization.

6.5 The Fundamental Theorem of Arithmetic

The above theorem shows us that each integer $n > 1$ has a factorization into primes. But we know more than that, don't we, reader? We know that there is one and only one way to factor an integer

$n > 1$. This uniqueness is where we are headed to now. Such a result is important enough that mathematicians have given it a name.

The Fundamental Theorem of Arithmetic: Let $n \geq 2$ be an integer.

1. There is a finite list of primes p_1, \dots, p_t such that $n = p_1 \dots p_t$.
2. If $n = q_1 \dots q_s$ for some other list of primes q_1, \dots, q_s , then $s = t$ and after a permutation of the list entries, $p_i = q_i$ for each $i = 1, \dots, t$. We say that n has unique prime factorization.

Proof: 1. This is Theorem 6.4.4.

2. We apply proof by contradiction. Suppose there is a number $n \geq 2$ that possesses two different prime factorizations, say

$$n = p_1 \cdots p_t = q_1 \cdots q_s$$

for some finite lists p_1, \dots, p_t and q_1, \dots, q_s of primes. Choose the smallest integer n with this property. Then the lists p_1, \dots, p_t and q_1, \dots, q_s are different. After permuting the subscripts we can assume that p_1 is not on the finite list q_1, \dots, q_s and

$$2 \leq p_1 < q_1. \tag{6.6}$$

Then $q_1 - p_1 < q_1$ so that

$$(q_1 - p_1)(q_2 \cdots q_s) < q_1(q_2 \cdots q_s) = n.$$

Since n is the smallest integer with nonunique prime factorization,

$$m = (q_1 - p_1)(q_2 \cdots q_s) \tag{6.7}$$

has a unique prime factorization.

By hypothesis $p_1 \neq q_i$ for $i = 1, \dots, s$, and p_1 does not divide $q_1 - p_1$. (Otherwise p_1 is a divisor of the prime $(q_1 - p_1) + p_1 = q_1$, contrary to (6.6).) Since (6.7) is a unique prime factorization of m , p_1 does not divide m . On the other hand, p_1 divides n so p_1 does not divide $n - m$. However,

$$\begin{aligned} n - m &= q_1 q_2 \cdots q_s - (q_1 - p_1) q_2 \cdots q_s \\ &= (q_1 - (q_1 - p_1))(q_2 \cdots q_s) \\ &= (p_1)(q_2 \cdots q_s). \end{aligned}$$

This obvious contradiction shows us that our supposition was incorrect. Thus the finite lists p_1, \dots, p_t and q_1, \dots, q_s are the same.

Subsequently the number of terms s and t on these lists are the same.

$$s = t$$

This is what we had to prove to complete the proof of part 2, and thus finish the proof of the Theorem.

Let us illustrate the above theorem with a number. The number

$$n = 215,441$$

has a unique prime factorization. At present we do not know what it is but we do know that unique prime factorization exists for n . Some work with a calculator shows us that

$$n = 17 \cdot 19 \cdot 23 \cdot 29$$

is a prime factorization of n . The Fundamental Theorem of Arithmetic shows us that this is the only prime factorization of n . If we have a prime factor p of n then the Fundamental Theorem of Arithmetic states that p is on the list

$$17, 19, 23, 29.$$

No other primes are allowed to divide n . That is the power of the Fundamental Theorem of Arithmetic. Without even knowing what p is we know that it is one of the four primes on the list.

6.6 Perfect Numbers

Euclid is also known for introducing us to *perfect numbers*. The number $m > 0$ is said to be *perfect* if m is the sum of the numbers properly dividing m . Thus 6 is a perfect number since 1, 2, 3 are the proper divisors of 6 and

$$6 = 1 + 2 + 3.$$

Six is the sum of its proper divisors. The next perfect number is 28 because 1, 2, 4, 7, 14 are the proper divisors of 24 and

$$28 = 1 + 2 + 4 + 7 + 14.$$

Other perfect numbers we know are 496, 521, 607, and 1279. These perfect numbers were found by some computer use in 1952. In those days, number problems like this were used to test a computer's accuracy. This calculation continues, though, as computers are still being used to factor numbers of the form $2^n - 1$.

So far these perfect numbers have a curious property. They are all even numbers. Are all perfect numbers even? Currently, no one knows. This uncertainty enhances interest in perfect numbers. For the even perfect numbers, there is the following theorem also due to Euclid.

Theorem 6.6.1 [Euclid] *If $2^n - 1$ is a prime number then $2^{n-1}(2^n - 1)$ is a perfect number.*

Euclid's proof of this theorem was geometric in nature since Greek algebra was under developed. The proof has evolved over the years into the following algebraic gem. Late in the eighteenth century the prodigious mathematician Leonard Euler (circa 1750 AD) showed that if m is an even perfect number then there is an integer n such that $2^n - 1$ is a prime number and $m = 2^{n-1}(2^n - 1)$. Think of the gap in time here. Euclid (300 BC) introduced perfect numbers but it took until Euler (1750 AD) to show that all even perfect numbers are encompassed by Euclid's Theorem. That makes 2050 years between first sight and last sight. This must be one hard type of number to study if it takes over two millennia to make substantial progress on the classification problem.

Thus Euclid's Theorem accounts for all of the even perfect numbers. We are left to wonder about the open question.

Open Question: Are there any odd perfect numbers? That is, is there an odd number m that is the sum of its proper divisors?

For example, 5 is *not* a perfect number since 1 is the only proper divisor of 5 and 5 is not the sum of its proper divisor 1. The odd number 25 is not the sum of its proper divisors 1, 5 because

$$25 \neq 1 + 5.$$

You can use your computer to verify that there are no odd perfect numbers less than 100.

Let us prove Euclid's Theorem.

Proof: Suppose we know that $p = 2^n - 1$ is a prime for some integer $n > 1$. Then we have a unique prime factorization

$$m = 2^{n-1}p.$$

The prime divisors of m are 2 and p so any proper divisor d of m is a multiple of 2 and p ,

$$d = 2^k p^e,$$

where k can be any integer in $\{0, 1, \dots, n-1\}$ and e is either 0 or 1. That means that the proper divisors of m that are not divisible by p are

$$1, 2, 2^2, \dots, 2^{n-1}$$

and the proper divisors divisible by p are

$$p, 2p, 2^2p, \dots, 2^{n-2}p.$$

(We do not include $2^{n-1}p = m$ since it is not a *proper* divisor of m .) No other numbers are possible. Recall the geometric identity

$$1 + x + x^2 + \dots + x^{n-1} = \frac{1 - x^n}{1 - x}$$

with common ratio x . If $x = 2$ then we have

$$1 + 2 + 2^2 + \dots + 2^{n-1} = 2^n - 1.$$

Hence adding up the proper divisors of m yields geometric identities

$$\begin{aligned} & \left. \begin{aligned} 1 + 2 + 2^2 + \dots + 2^{n-1} + \\ p + 2p + 2^2p + \dots + 2^{n-2}p \end{aligned} \right\} = \\ & \left. \begin{aligned} 1 + 2 + 2^2 + \dots + 2^{n-1} + \\ (1 + 2 + 2^2 + \dots + 2^{n-2})p \end{aligned} \right\} = \\ & (2^n - 1) + (2^{n-1} - 1)p = \\ & p + (2^{n-1} - 1)p = \\ & 2^{n-1}p = m \end{aligned}$$

This shows that m is a perfect number and completes the proof of Euclid's Theorem.

Therefore there will be an infinite number of perfect numbers if there are infinitely many numbers n such that $2^n - 1$ is a prime number. At the time of this writing, it is not known if there are infinitely many primes of the form $2^n - 1$, and so it is not known if there are infinitely many even perfect numbers.

This Page Intentionally Left Blank

Chapter 7

Prime Numbers

7.1 Prime Number Generators

Theorem 6.4.2 shows that there are infinitely many primes. The next step then would be to decide where these primes are. The pursuit of prime numbers goes back to the Greeks and is currently a source of mathematical investigation and inspiration. We will discuss the two ends of this time line.

The Greeks had a clever way of finding the prime numbers less than some fixed number. They called it *The Sieve*. Basically, if you wanted to find all of the prime numbers less than a given number, let us say 60, you would first write down all of the numbers less than 60 starting with 2. (Leave 1 out because 1 is not a prime. Remember, a prime is a number $p > 1$ that is divisible only by 1 and itself. We exclude 1.)

	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50
51	52	53	54	55	56	57	58	59	60

and so on until you write down 60. All numbers divisible by 2 are excluded from our search for prime numbers so we start with 2, and move every other space, marking these numbers out as we go. We mark them out because they are divisible by 2. You stop when you

reach the end of your Sieve. Observe.

	2	3	5	7	9
11		13	15	17	19
21		23	25	27	29
31		33	35	37	39
41		43	45	47	49
51		53	55	57	59

Instead of striking the numbers out we have deleted them from The Sieve. You can just strike them out with the slash /. The numbers that remain are not divisible by 2.

Next, we find multiples of 3. Start at 3 and count three spaces in The Sieve until The Sieve is ended. It is important that you count the spaces in The Sieve as well. The deleted numbers contribute to our search for primes. Do it this way.

	2	3	5	7	
11		13		17	19
		23	25		29
31			35	37	
41		43		47	49
			55		59

The numbers left are not divisible by 2 or by 3. The next number left on The Sieve is 5. Delete all of the numbers and spaces in The Sieve that are five spaces from 5, 5 spaces from that number, and so on. Simply slash out numbers by counting every fifth number and space, as follows.

	2	3	5	7	
11		13		17	19
		23			29
31				37	
41		43		47	49
			53		59

The remaining numbers are not divisible by 2, 3, or 5. Do the same for 7 (the next number on The Sieve) and arrive at

	2	3	5	7	
11		13		17	19
		23			29
	31			37	
41		43		47	
51		53			59

Note the disappearance of the number 49 from The Sieve. You would have deleted 14, 21, 28, 35, and 42 but these numbers were already deleted from The Sieve. They are also divisible by 2, 3, and 5. None of the remaining numbers in The Sieve is divisible by 2, 3, 5, or 7. Take 11 and strike out every eleventh number as so.

	2	3	5	7	
11		13		17	19
		23			29
	31			37	
41		43		47	
51		53			59

Nothing new is deleted. That is, because the first number encountered must be of the form 11^2 . The smaller numbers are of the form $p \cdot 11$ for $p = 2, 3, 5, 7$ and these numbers have been deleted from The Sieve. The number bigger than 11 that we would first delete is 121, which is not on our small Sieve. Thus The Sieve has filtered out all of the prime numbers less than 60. Those primes are the numbers left in The Sieve. They are

$$2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 51, 53, 59.$$

The Sieve is used to find the primes less than a given number, but this approach is quite slow. Computers especially have a hard time finding primes less than n with The Sieve when n is large, say a million. It takes a more than a week of computer time to find all of the primes less than a million.

Another method for generating primes is to use polynomials. It was observed by Euler (circa 1750 AD) that the quadratic polynomial

$$p(x) = 41 + x + x^2$$

gives a list of prime values for small x . Some of those prime numbers are given.

$$\begin{aligned} f(0) &= 41, f(1) = 43, f(2) = 47, \\ &53, 61, 71, 83, 97, 113, 151, 173, 197, 223, \\ &251, 281, 313, 347, 383, 421, 461, 503, 547, 593, 641, 691, \\ &743, 797, 853, 911, 971, 1033, 1097, 1163, 1231, 1301, \\ &1373, 1447, 1523, 1601. \end{aligned}$$

Use your desktop computer to see that each of these numbers is a prime number. Of course, the number $p(41)$ will not be a prime since it is divisible by 41, and neither is $p(40)$ a prime. Try it and see. Thus some of the values of $p(x)$ are not prime numbers.

But what a mathematical coincidence. We produce many primes from such a simple polynomial of such small degree. Why is that so? Since mathematics does not believe in mathematical coincidence, the question posed was to find a polynomial of small degree that was complex enough to yield all of the prime numbers as output.

In 1977 a team of mathematicians found a polynomial that generated all of the primes when natural numbers were input. The polynomial does not have one variable, it has 26 variables, and its degree is not 2. It has degree 25. One has to wonder how James Jones, Daihachiro Sato, Hideo Wada, and Douglas Wiens came upon this algebraic dragon. For those wondering, that polynomial P is

$$\begin{aligned} P = & (k+2)\{1 - [wz + h + j - q]^2 \\ & - [(gk + 2g + k + 1)(h + j) + h - z]^2 \\ & - [2n + p + q + z - e]^2 \\ & - [16(k+1)^3(k+2)(n+1)^2 + 1 - f^2]^2 \\ & - [e^3(e+2)(a+1)^2 + 1 - o^2]^2 \\ & - [(a^2 - 1)y^2 + 1 - x^2]^2 - [16r^2y^4(a^2 - 1) + 1 - u^2]^2 \\ & - [((a + u^2(u^2 - a))^2 - 1)(n + 4dy)^2 + 1 - (x + cu)^2]^2 \\ & - [n + l + v - y]^2 - [(a^2 - 1)l^2 + 1 - m^2]^2 \\ & - [ai + k + 1 - l - i]^2 \\ & - [p + l(a - n - 1) + b(2an + 2a - n^2 - 2n - 2) - m]^2 \\ & - [q + l(a - n - 1) + s(2ap + 2a - p^2 - 2p - 2) - x]^2 \\ & - [z + pl(a - p) + t(2ap - p^2 - 1) - pm]^2\}. \end{aligned}$$

Try programming your desktop computer with this polynomial and integer input values to see that this polynomial puts out negative numbers, (to be ignored), and prime numbers only. You can also see that this polynomial has output 2 quite a lot, and the primes are not put out in their numerical order. But still, here is a finite formula that puts out the infinite set of prime numbers. That is the accomplishment.

7.2 The Prime Number Theorem

Now let's consider just counting the primes. For a given integer $n > 0$ let

$$\pi(n) = \text{the number of primes } p < n.$$

Here is $\pi(n)$ for some small n . Because

$$2, 3$$

are the primes less than 5,

$$\pi(5) = 2;$$

because

$$2, 3, 5, 7, 11$$

are the primes less than 12,

$$\pi(12) = 5;$$

and because

$$2, 3, 5, 7, 11, 13, 17, 19, 23, 29$$

are the primes less than 30,

$$\pi(30) = 10.$$

Using The Sieve we would conclude that

$$\pi(60) = 18.$$

There doesn't seem to be much of a pattern here, does there? And yet mathematicians have looked for and found an important pattern. It is nothing less than a closed formula for finding out how many primes are less than a given number. To talk about this milestone in mathematics we need to know a little about logarithms.

One of the fundamental constants of mathematics is the number e (in honor of Leonard Euler who first investigated this number). The decimal value of the number e is *approximately*

$$e \approx 2.718281828459.$$

This number does not repeat itself and there is no end to the seeming random pattern to the decimal values of e . One does not study the physical sciences easily without the number e . This number, it seems, is how nature put the universe together. For those familiar with chemistry, the decay of radioactive elements is based on e through the equation

$$A(t) = A_0 e^{-\lambda t}$$

where A_0 is the initial amount of material and λ is a constant called the *decay constant*. For those who have studied calculus, this function $A(t)$ comes from the differential equation

$$A'(t) = A(t),$$

which does not seem to mention e at all.

Now in possession of this mathematical constant we define the *natural logarithm* as follows. Let a and b be nonnegative real numbers. Then the *natural logarithm* of a is denoted by

$$\ln(a).$$

By definition, this logarithm is the log in the base e . That is,

$$\ln(a) = b \text{ exactly when } a = e^b.$$

Thus

$$\begin{aligned}\ln(1) &= 0 \text{ because } 1 = e^0, \\ \ln(e) &= 1 \text{ because } e = e^1, \\ \ln\left(\frac{1}{e}\right) &= -1 \text{ because } \frac{1}{e} = e^{-1} \text{ and }, \\ \ln(\sqrt{e}) &= \frac{1}{2} \text{ because } \sqrt{e} = e^{1/2}.\end{aligned}$$

In general, we have

$$\boxed{\ln(e^x) = x \text{ for any real number } x.}$$

Thus we have a rudimentary knowledge of $\ln(x)$. With this knowledge we can state the Prime Number Theorem.

The Prime Number Theorem: Given a large natural number n ,

$$\boxed{\pi(n) \approx \frac{n}{\ln(n)}}.$$

Indeed, the approximation gets better as n increases without bound. By this we mean that a comparison between $\pi(n)$ and $\frac{n}{\ln(n)}$ will show that these numbers are almost the same as n gets large. That is, the limit

$$\lim_{n \rightarrow \infty} \frac{\pi(n)}{n/\ln(n)} = 1.$$

Consider this for a moment. A fraction $\frac{A}{B} = 1$ precisely when $A = B \neq 0$. This limit implies that $\frac{\pi(n)}{n/\ln(n)}$ is very close to 1 for large n . Thus $\frac{n}{\ln(n)}$ is a very good approximation for $\pi(n)$ for large natural numbers n . Being close to $\pi(n)$ is the best we can ask for at this time.

For example, with a little calculator use we can be sure that

$$\pi(1024)$$

is close to

$$\frac{1024}{\ln(1024)} = \frac{1024}{\ln(2^{10})} = \frac{1024}{10\ln(2)} = 160.$$

Larger numbers are sure to lead to better approximations. Thus

$$\pi(e^{20})$$

is close to

$$\frac{e^{20}}{\ln(e^{20})} = \frac{e^{20}}{20} \leq \frac{3^{20}}{20} = 1.5 \times 3^{18}$$

since $e < 3$. We can get another crude approximation of the number of primes less than e^{20} by noting that $e^2 \approx 8$. That is,

$$\begin{aligned} \frac{e^{20}}{\ln(e^{20})} &= \frac{(e^2)^{10}}{20} \\ &\approx \frac{(8)^{10}}{20} \\ &= 4 \times 8^8 \\ &= 2^{26} \\ &= 64 \times (2^{10})^2 \\ &= 64 \times 1024^2 \\ &\approx 64 \times 10^6. \end{aligned}$$

There are approximately 64 million primes less than e^{20} . To see how big e^{20} is, note that $2.7 < e$ so that

$$e^{20} > (2.7)^{20} = ((2.7)^2)^{10} = (7.29)^{10} > (50)^5 = 312,500,000$$

or over 312 million.

7.3 Products of Geometric Series

We are going to discuss the most famous of mathematical problems called the *Riemann Hypothesis*. This problem is intimately connected with the *Prime Number Theorem* and involves the roots of

an infinite sum called the *Riemann Zeta Function*. What we do is probably the way that Leonard Euler went about the problem.

The *sigma notation* is shorthand for writing down sums. It is especially useful when writing down infinite sums. It takes the guesswork out of seeing which pattern is meant by Σ means “add these terms up.” For example,

$$\sum_{i=0}^2 x^i = 1 + x + x^2$$

and, more importantly, the geometric series $1 + x + x^2 + \dots$ with infinitely many terms can be written

$$\sum_{k=0}^{\infty} x^k = 1 + x + x^2 + \dots$$

We can even sum over two or more variables if we are careful. Thus, assuming that $i, j \geq 0$, we have

$$\begin{aligned} \sum_{i+j=3} x^i y^j &= x^0 y^3 + x^1 y^2 + x^2 y^1 + x^3 y^0 \\ &= y^3 + xy^2 + x^2y + x^3 \\ &= x^3 + x^2y + xy^2 + y^3. \end{aligned}$$

Another example is

$$\begin{aligned} \sum_{i+j+k=3} x^i y^j z^k &= x^3 + y^3 + z^3 \\ &\quad + x^2y + x^2z + y^2x + y^2z + z^2x + z^2y \\ &\quad + xyz. \end{aligned}$$

One more detail before we begin our multiplication. The *degree* of a term is the sum of the exponents (implied or stated) in that term. Thus

$$x^2 y^3 z^4$$

has degree

$$2 + 3 + 4 = 9,$$

while degree 3 terms look like this.

$$z^3, x^2y, xyz.$$

We discuss the product of polynomials by considering the terms of different degree in the polynomial. There is a pattern here that will help us to multiply geometric series together.

Consider the identity

$$(1 + x)(1 + y) = 1 + (x + y) + xy.$$

The term $x + y$ comes from the fact that we have multiplied all of the degree 1 terms by degree 0 terms (=constant terms).

$$1 \cdot x + y \cdot 1 = x + y.$$

The term xy comes from the fact that we have multiplied together all those terms whose product has degree 2. There is only one such product in this example.

$$x \cdot y = xy.$$

Let us do the same thing for the identity

$$\begin{aligned} (1 + x)(1 + y)(1 + z) &= 1 \\ &+ (x + y + z) \\ &+ (xy + yz + xz) \\ &+ (xyz). \end{aligned}$$

The term $x + y + z$ comes from multiplying together those terms whose degree is 1:

$$1 \cdot x + 1 \cdot y + 1 \cdot z = x + y + z.$$

The degree 2 term is found by multiplying together all those terms in this example whose product has degree 2.

$$xy + yz + xz$$

is the result. The degree 3 term is from the only product that yields a degree 3 term:

$$xyz.$$

Let us try this analysis one more time by considering the identity

$$\begin{aligned}
 (1 + x + x^2)(1 + y)(1 + z) &= 1 \\
 &+ (x + y + z) \\
 &+ (xy + yz + xz + x^2) \\
 &+ (x^2y + x^2z + xyz) \\
 &+ (x^2yz).
 \end{aligned}$$

The identity was found by using foil over and over again. Here is another way to find this identity.

The degree 0 term is found by multiplying together all the terms that add up to a degree 0 term. There is just one:

$$1 \cdot 1 \cdot 1 = 1.$$

The degree 1 term is found by multiplying all of the terms together that result in a degree 1 term:

$$1 \cdot x + 1 \cdot y + 1 \cdot z = x + y + z.$$

The next would be the degree 2 term. Multiply all of the terms together that result in degree 2. They are

$$xy + yz + xz + x^2.$$

The degree 3 term is next. The products resulting in degree 3 are degree 1 times degree 2 or three degree 1 terms together. They are

$$x^2y, x^2z, xyz$$

so the degree 3 term is their sum.

$$x^2y + x^2z + xyz.$$

The degree 4 term is the only product that results in degree 4, namely, degree 2, degree 1, and degree 1:

$$x^2yz.$$

This completes the product.

So when multiplying we ask which products result in a degree 0 term, then a degree 1 term, then a degree 2 term, then a degree 3 term, and so on until the terms in the product are all used.

We call an infinite sum

$$1 + x + x^2 + x^3 + \dots$$

a *geometric series*. This geometric series is a number when $-1 < x < 1$ and only when x is in this domain, so it is important that we do not allow x to stray from this *interval of convergence*. We will multiply two geometric series together using the above pattern.

$$G = (1 + x + x^2 + \dots)(1 + y + y^2 + \dots).$$

Operating as we did above, the degree 0 term is

$$1 \cdot 1 = 1$$

and the degree 1 term is formed by multiplying degree 0 and degree 1 terms together. The result is

$$1 \cdot y + x \cdot 1 = x + y.$$

The degree 2 term is next. Combine all terms together whose degree is 2:

$$1 \cdot y^2 + x \cdot y + x^2 \cdot 1 = x^2 + xy + y^2.$$

The degree 3 term is found by counting degrees.

$$1 \cdot y^3 + x \cdot y^2 + x^2 \cdot y + x^3 \cdot 1 = x^3 + x^2y + xy^2 + y^3.$$

One more, the degree 4 term.

$$\begin{aligned} 1 \cdot y^4 + x \cdot y^3 + x^2 \cdot y^2 + x^3 \cdot y + x^4 \cdot 1 \\ = x^4 + x^3y + x^2y^2 + xy^3 + y^4. \end{aligned}$$

We have thus found the first 4 terms of the product G .

$$\begin{aligned} G &= 1 \\ &+ (x + y) \\ &+ (x^2 + xy + y^2) \\ &+ (x^3 + x^2y + xy^2 + y^3) \\ &+ (x^4 + x^3y + x^2y^2 + xy^3 + y^4) \\ &\vdots \end{aligned}$$

In general, we can say that the n th term in G is found by adding together all of the products whose sum is n . In terms of the sigma notation we have the following.

$$\text{The } n\text{th term of } G = \sum_{i+j=n} x^i y^j.$$

It is worth noting that if $(1 + x + x^2 + \dots)$ and $(1 + y + y^2 + \dots)$ converge (i.e., if $-1 < x, y < 1$) then the product G also converges. However, as we will show presently, in case there are infinitely many geometric series multiplied together, then the product of them need not converge to a number.

This is how we extend our products from two to three or more geometric series. Use the sigma notation to write down a geometric series. The rest of the notation is due to Isaac Newton. Let

$$\begin{aligned} A &= \sum_{i=0}^{\infty} a^i = 1 + a + a^2 + \dots, \\ B &= \sum_{j=0}^{\infty} b^j = 1 + b + b^2 + \dots, \\ C &= \sum_{k=0}^{\infty} c^k = 1 + c + c^2 + \dots, \\ &\vdots \end{aligned}$$

be an infinite implied list of geometric sums. The product

$$\Gamma = A \cdot B \cdot C \cdot \dots$$

is the infinite series whose n th term is

$$\text{nth term of } \Gamma = \sum_{i+j+k+\dots=n} (a^i b^j c^k \dots).$$

For example, the degree 0 term is from sums $i + j + k + \dots = 0$, which happens only when $i = j = k = \dots = 0$. Thus the degree 0 term is

The degree 1 term of Γ is the sum of all terms whose exponents add to 1. That is, the degree 1 term is the infinite sum

$$a^1 + b^1 + c^1 + \dots = a + b + c + \dots$$

The degree 2 term, which is all we are going to do, is formed by taking products of degree 1 terms like ab or bc , and products of square terms with the only degree 0 term that exists here, 1. The degree 2 term is the infinite sum

$$\sum_{r+s=2} x^r y^s.$$

In this case $x, y \in \{a, b, c, \dots\}$. Some of the terms in the degree 2 term of the product Γ are

$$\begin{aligned} \sum_{r+s=2} x^r y^s &= a^2 + b^2 + c^2 + \dots \\ &\quad + ab + bc + ac + \dots \end{aligned}$$

Somewhere in the product Γ is a factor

$$D = \sum_{\ell=0}^{\infty} d^\ell = 1 + d + d^2 + \dots$$

It also contributes to the infinite sum that is Γ . Terms like d^5 , abd^2 and $a^2c^3d^7$ occur when we write down Γ in all of its detail. The sigma notation helps to simplify our representation of Γ .

$$\boxed{\Gamma(a, b, c, \dots) = \sum_{n=0}^{\infty} \sum_{i+j+k+\dots=n} (a^i b^j c^k \dots)}$$

We now have enough algebra to understand a little about the *Riemann Zeta Function*.

7.4 The Riemann Zeta Function

In this section, we investigate a special value of Γ called the *Riemann Zeta Function*. We will replace a, b, c, \dots with specific values or numbers, and the resulting product will be untangled. Our stopping place will be what is considered to be the most important mathematics problem of the last 150 years. It is called the *Riemann Hypothesis*.

Let $s > 1$ be a real number. As in the previous section let us substitute $\frac{1}{2^s}$ for a in A , and call the result

$$A\left(\frac{1}{2^s}\right) = 1 + \frac{1}{2^s} + \frac{1}{2^{2s}} + \frac{1}{2^{3s}} + \dots$$

Let $b = \frac{1}{3^s}$ in B and write

$$B\left(\frac{1}{3^s}\right) = 1 + \frac{1}{3^s} + \frac{1}{3^{2s}} + \frac{1}{3^{3s}} + \dots$$

Let $c = \frac{1}{5^s}$ in C and write

$$C\left(\frac{1}{5^s}\right) = 1 + \frac{1}{5^s} + \frac{1}{5^{2s}} + \frac{1}{5^{3s}} + \dots$$

Continue inductively, forming reciprocals of s -powers of prime numbers

$$\frac{1}{p^s}$$

and substituting them into the associated geometric series:

$$D\left(\frac{1}{p^s}\right) = 1 + \frac{1}{p^s} + \frac{1}{p^{2s}} + \frac{1}{p^{3s}} + \dots$$

Each geometric series $1 + x + x^2 + \dots$ can be written in closed form as

$$1 + x + x^2 + \dots = \frac{1}{1 - x}$$

when $-1 < x < 1$. For primes p , $-1 < \frac{1}{p^s} < 1$, so

$$D\left(\frac{1}{p^s}\right) = 1 + \frac{1}{p^s} + \frac{1}{p^{2s}} + \dots = \frac{1}{1 - \frac{1}{p^s}}.$$

Making this substitution for A, B, C, \dots yields the equation

$$\begin{aligned}\Gamma\left(\frac{1}{2^s}, \frac{1}{3^s}, \frac{1}{5^s}, \dots\right) &= A\left(\frac{1}{2^s}\right)B\left(\frac{1}{3^s}\right)C\left(\frac{1}{5^s}\right)\dots \\ &= \prod_{\text{all primes } p} \frac{1}{1 - \frac{1}{p^s}}\end{aligned}$$

where the symbol

$$\prod_{\text{all primes } p}$$

means to take the product over all primes p . The first little bit of this formula is

$$\frac{1}{1 - \frac{1}{2^s}} \cdot \frac{1}{1 - \frac{1}{3^s}} \cdot \frac{1}{1 - \frac{1}{5^s}}.$$

and continues on indefinitely, one factor for each prime.

This infinite product is called the *Riemann Zeta Function*:

$$\zeta(s) = \prod_{\text{all primes } p} \frac{1}{1 - \frac{1}{p^s}}.$$

This function $\zeta(s)$ is defined at all complex numbers except $s = 1$. There is a deep connection between the roots or zeros of $\zeta(s)$ and the number of primes $\pi(n)$ less than n . The details of that connection are beyond the scope of this book, so we will just write down certain facts about $\zeta(s)$.

We can just as easily use complex numbers as real numbers in $\zeta(s)$ so we do just that. Let

$$s = u + iv,$$

where u and v are real numbers, and $i = \sqrt{-1}$, the imaginary unit. The next question that we come upon is simple enough. You have often been asked to find the zeros of a polynomial function, and that is what we ask here. *What are the zeros of the Riemann Zeta Function $\zeta(s)$?*

The Riemann Hypothesis: The zeros of the Riemann Zeta Function have real part $\frac{1}{2}$. That is, if $\zeta(s) = 0$ and if $s = u + iv$ then

$$u = \frac{1}{2}.$$

As of March, 2006 an answer to this question has escaped the most careful of analysis. All of the zeros that have been found satisfy $u = \frac{1}{2}$ but that does not preclude the possibility that there might be a very large number $s = u + iv$ out there somewhere that is a zero of $\zeta(s)$ but that has $u \neq \frac{1}{2}$. The best computer models show that if $\zeta(s) = 0$ and if $u < 10^{14}$ then $u = \frac{1}{2}$. But this does not prove that all zeros of $\zeta(s)$ have real part $\frac{1}{2}$.

There is a story that circulates among the graduate students during the coffee hour at mathematics departments in universities all over the world. It seems that the Devil came to Riemann and said that he would give him his heart's desire for his soul. You know the contract. It has become immortal since we read *The Devil and Daniel Webster*. Riemann, being mathematically astute, said to the Devil that he would give up his soul provided that the Devil could give Riemann a correct proof of the Riemann Hypothesis in one year. The Devil agreed, and vanished in a puff of smoke. Riemann went about his work. One year later to the day, the Devil returned to Riemann's office. He sadly announced that despite using all of the resources in his realm he was unable to solve the problem Riemann had given him. "But," says he, "if I could just prove this one lemma." Moral of the story is that there are no world class mathematicians in hell.

Let us try a more traditional form of $\zeta(s)$. We can realize $\zeta(s)$ as an infinite sum as long as we begin with a real number $s > 1$. Recall that $\zeta(s)$ is a product of geometric series

$$\zeta(s) = A\left(\frac{1}{2^s}\right) \cdot B\left(\frac{1}{3^s}\right) \cdot C\left(\frac{1}{5^s}\right) \dots$$

where there is one geometric series $D\left(\frac{1}{p^s}\right)$ for each prime number p . The terms of the series $\zeta(s)$ are 1, the sum of the reciprocal primes

$$\sum_{\text{primes } p} \frac{1}{p^s},$$

and the sum of the degree 2 terms

$$\sum_{\text{primes } p} \frac{1}{p^{2s}} + \sum_{\text{primes } p \neq q} \frac{1}{(pq)^s}.$$

The higher degree terms are just too complicated to write down here.

We approach the sum from the opposite direction.

Consider the sum

$$\sum_{n=1}^{\infty} \frac{1}{n^s}.$$

We could add up these numbers by taking $n = 1, 2, 3, \dots$ or we could add them up by taking their prime factorizations. Add up 1 and the numbers that have just one prime dividing them.

$$1 + \sum_{\text{primes } p} \frac{1}{p^s}$$

Next, take the numbers that are of the form p^2 or that have exactly two primes dividing them. Such an n would look like

$$n = pq$$

for two primes $p \neq q$:

$$1 + \sum_{\text{primes } p} \frac{1}{p^s} + \sum_{\text{primes } p} \frac{1}{p^{2s}} + \sum_{\text{primes } p \neq q} \frac{1}{(pq)^s}.$$

These sums are the initial sums of $\zeta(s)$. We have therefore argued that

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}. \quad (7.1)$$

That is, the infinite product $\zeta(s)$ is an infinite sum.

A most unusual mathematical coincidence occurs here. Remember that I have warned you about manipulating infinite sums that do not converge. Well here is a good example of that. The geometric series

$$D\left(\frac{1}{p}\right) = 1 + \frac{1}{p} + \frac{1}{p^2} + \frac{1}{p^3} + \dots$$

converges since

$$-1 < \frac{1}{p} < 1 \text{ for primes } p.$$

Does the infinite product

$$\zeta(1) = A\left(\frac{1}{2}\right) \cdot B\left(\frac{1}{3}\right) \cdot C\left(\frac{1}{5}\right) \dots$$

of convergent geometric series converge? No. You see, if we use the infinite series representation (7.1) of $\zeta(s)$, we see that

$$\zeta(1) = \sum_{n=1}^{\infty} \frac{1}{n},$$

which is the divergent *harmonic series*. That should raise a few eyebrows. Certain infinite products of numbers do not converge. In this way, infinite products of numbers are similar to infinite sums of numbers. Some converge, some diverge.

7.5 Real Numbers

The previous section shows us how mathematicians handle infinite sets of integers in a finite manner. In this section, we will see how mathematicians handle the continuum of the real numbers in a finite manner. For example, there is an infinite set called a *sequence*. An example of a sequence is

$$\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \dots$$

The sequence is infinite but it takes place in a finite interval $[0, 1]$. Moreover, there is a number 0 such that each *open neighborhood* of 0 contains infinitely many of the terms in this sequence. An *open neighborhood of 0* is an open interval $(-a, a)$ that contains 0. Other open neighborhoods of 0 are $(-1, 1)$, $(-2, 2)$, and $(\frac{-1}{2}, \frac{1}{2})$.

Another sequence is formed by allowing

$$x_n = \frac{(-1)^n}{2^n}$$

for each integer $n \geq 0$. The first few terms are

$$\frac{1}{1}, \frac{-1}{2}, \frac{1}{4}, \frac{-1}{8}, \dots$$

They oscillate between positive and negative numbers. The entire sequence is contained in the finite interval $[-1, 1]$. Moreover, there is a number 0 not in the sequence with a special property. Given any open neighborhood of 0, say, $(-a, a)$, there is a point beyond which all of the terms in the sequence are in this neighborhood. That is, there is an integer $N > 0$ such that

$$\frac{(-1)^n}{2^n} \in (-a, a) \text{ for all integers } n > N.$$

An open neighborhood of 0 is like having an infinite queue of people and saying that each of them lives on the same block of the city. A smaller open neighborhood of 0 would be one home. We would be saying that almost all of these people live in the same home. An even smaller open neighborhood would be a room in the home. All but finitely many of these people live in the same room. I won't get into the social implications that must come up when the open neighborhood of 0 is a chair in the room.

Let us be a bit more precise. A sequence in the closed interval $[-1, 1]$ is a function $f : \mathbb{N} \longrightarrow [-1, 1]$ from the natural numbers into $[-1, 1]$. The images of the function or sequence is usually denoted with a subscript as in

$$f(n) = x_n \text{ for each } n \in \mathbb{N}.$$

For instance, $f(1) = x_1$, $f(2) = x_2$, and so on. The sequence with terms $\frac{(-1)^n}{2^n}$ is given by the function $f : \mathbb{N} \longrightarrow [-1, 1]$ with rule

$$f(n) = \frac{(-1)^n}{2^n}.$$

Let us write down an arbitrary sequence in $[-1, 1]$.

$$x_0, x_1, x_2, \dots,$$

of different terms. This is an infinite subset of $[-1, 1]$ so there should be something finite that we can say about the sequence. Cut the interval in half. Is it possible that both halves contain a finite number of sequence elements in them? No. That would mean that the sequence is finite, which it is not. Thus one half contains an infinite number of sequence elements. So let I_1 be the half of the interval that contains infinitely many sequence elements. Now cut I_1 in half. As we have argued above, one half must contain infinitely many elements of the sequence since otherwise I_1 contains only finitely many sequence elements, a contradiction. Hence we let $I_2 \subset I_1$ be the half that contains infinitely many of the sequence elements. Continue in this way, constructing a chain of sets

$$I_1 \supset I_2 \supset \dots,$$

each term half as big as the previous term. The lengths of the intervals form a sequence

$$1, \frac{1}{2}, \frac{1}{2^2}, \frac{1}{2^3}, \dots$$

Observe that these lengths converge or approach 0 as n is allowed to grow.

Now, we can write

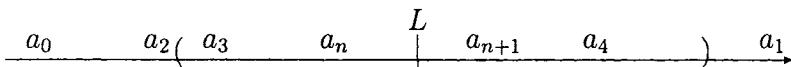
$$I_k = (a_k, b_k) \text{ for some real numbers } a_k < b_k.$$

In this case we have a sequence of real numbers

$$a_1 < a_2 < a_3 < \dots < b_3 < b_2 < b_1. \quad (7.2)$$

Then the sequence $\{a_1, a_2, a_3, \dots\}$ is a sequence bounded above by b_1 . That is, there is only one direction for the sequence, namely, upward, and there is a number larger than all of the sequence elements. We say that $\{a_n \mid n \in \mathbb{N}\}$ is a *bounded monotonic sequence*. One of the primary properties of the real line is that such a sequence $\{a_n \mid n \in \mathbb{N}\}$ has a limit. That is, there is a number L such that each open neighborhood \mathcal{O} of L contains all but finitely many of the a_n . Pictorially we have

Almost all sequence elements are
in this open neighborhood of L .



The picture shows us that almost all (i.e., all but a finite number) of the sequence elements a_n are in the open neighborhood \mathcal{O} of L . The sequence $\{b_n \mid n \in \mathbb{N}\}$ has the same property around L . Thus given an open interval \mathcal{O} about L there is a point where

$$(a_n, b_n) \subset \mathcal{O}$$

for almost all of the open intervals $(a_n, b_n) = I_n$. Since I_n is chosen so that I_n contains infinitely many elements of $\{x_n \mid n \in \mathbb{N}\}$, we see that \mathcal{O} contains infinitely many elements in $\{x_n \mid n \in \mathbb{N}\}$. We call such a point L a *cluster point* of $\{x_n \mid n \in \mathbb{N}\}$.

At this point we summarize.

Theorem 7.5.1 [Bolzano-Weierstrauss] *Let $\{x_n \mid n \in \mathbb{N}\} \subset [-1, 1]$ be a sequence. There is a cluster point L in $[-1, 1]$ for the sequence.*

What did we know about the sequence $\{x_n \mid n \in \mathbb{N}\}$? Only that it was a sequence in $[-1, 1]$. That is precious little to know about a sequence before we determine that it has a cluster point, a point that is very close to infinitely many of the elements in the sequence. Very little hypothesis gave us a powerful result. That is one of the aims or goals of mathematical research. When less gives us more we know that we are onto something special.

Let me share some stories surrounding Karl Weierstrauss (circa 1880 AD). Unlike many mathematical giants Karl did not start out as a strong mathematics student. Karl was a late bloomer. He spent his college life in pursuit of personal joy. Most of his young life was misspent fencing and quaffing beers in bars. His pursuit of fun distracted him from his studies. He was a party animal. Because of his party attitude he barely graduated college. Upon graduation he was denied a job at the local institutions and accepted a job instead teaching at an elementary school. It was here that he hit upon the sense of rigor and creativity that we remember him for today. In a change of image he set new standards for mathematical rigor that are still in effect more than 100 years later. Just as curious is the fact that in contrast to his earlier work, Karl's lectures and teaching are renowned for the care taken in their presentation and preparation. His gifts to his students are another unique quality for

Karl. He would begin a mathematical gem and then allow a student or a colleague to finish it, thus giving them the credit. This is a very generous and quite uncommon behavior. Mathematics cannot underestimate Karl's foundational contributions to teaching, analysis, and rigor in mathematics.

Let us use this kind of recursive argument to hunt tigers on the Arabian Desert. Your job is to find a Tiger in the otherwise barren Arabian Desert. You have all of the material you need, so you first erect an impassible fence around the Desert. Call the contained area S_1 . You divide S_1 into two equal regions using the impassible fencing. To find the Tiger, you send men out into each half with orders to engage the Tiger in a fight. The men on one side return saying they could not find the Tiger and the men on the other side do not come back. Call their side S_2 . It contains the Tiger. Divide S_2 in half using impassible fencing and send the men out again. The side S_3 is the one whose men do not return. The Tiger is there. We have started an induction, constructing regions

$$S_1 \supset S_2 \supset S_3 \supset \dots,$$

whose areas are given by a sequence

$$1 > \frac{1}{2} > \frac{1}{2^2} > \frac{1}{2^3} > \dots$$

that becomes vanishingly small. Eventually we will construct a region S_n that contains the Tiger and whose area is $\frac{1}{2^n}$. This power of $\frac{1}{2}$ will be so small that we can see the Tiger in the region S_n . At this point we can say that we have corralled the Tiger.

Some might say that we have no way of knowing that S_n is small enough to cage the Tiger. There is less than 10^{12} square miles in the Arabian Desert, so on the 41st attempt to find the Tiger we have a region of area

$$\begin{aligned} \frac{1}{2^{40}} 10^{12} &= \frac{1}{2^{40}} (1000)^4 \\ &< \frac{1}{2^{40}} (1024)^4 \end{aligned}$$

$$\begin{aligned} &= \frac{1}{2^{40}} (2^{10})^4 \\ &= 2^{-40} \cdot 2^{40} \\ &= 1 \text{ mile.} \end{aligned}$$

Another few fences and we have corralled the Tiger.

Chapter 8

Logic and Meta-Mathematics

8.1 The Collection of All Sets

Let

F

denote the *Collection of all Finite Sets*. Since F contains $\{1\}$, $\{1, 2\}$, $\{1, 2, 3\}$, ..., the collection F is infinite. Put another way:

The collection F of all finite sets is not finite.

While this is not a surprise, it prepares us for similar ideas concerning deeper implications about more abstract collections.

Example 8.1.1 The following is a classic example due to Bertrand Russell. Let

C

be the *Collection of All Sets*. We will prove the following.

Theorem 8.1.2 \mathbf{C} is not a set.

Proof: Assume for the sake of contradiction that \mathbf{C} is a set. Then \mathbf{C} has the rather unsettling property

$$\mathbf{C} \in \mathbf{C}.$$

That is, \mathbf{C} is an element of itself. This is like saying that a bag of sand is itself a grain of sand (bag and all), that this book is contained on one page of this book, or that a crowd of people is a person. And yet, although unsettling, $\mathbf{C} \in \mathbf{C}$ is a consequence of the assumption that \mathbf{C} is a set. This unsettling statement will lead us to a wonderful contradiction. Picture (6.5) will help with the argument that follows.

The Collection of All Sets \mathbf{C} (8.1)

		Empty overlap		
• \mathbf{C}			• \emptyset	
\mathcal{W}			\mathcal{W}'	
$S \in S$ for these sets S			$S \notin S$ for these sets S	

To produce a contradiction we will do something that may strike you as familiar. Define a set $\mathcal{W} = \{ \text{sets } S \mid S \in S \}$. That is,

\mathcal{W} = the set of all sets S that contain themselves as an element.

The complement of \mathcal{W}

$$\mathcal{W}' = \{\text{sets } S \mid S \notin S\}.$$

By our assumption \mathbf{C} is a set such that $\mathbf{C} \in \mathbf{C}$ so that $\mathbf{C} \in \mathcal{W}$. Thus \mathcal{W} is nonempty. The empty set \emptyset satisfies $\emptyset \notin \emptyset$ so $\emptyset \in \mathcal{W}'$. In fact, most sets S satisfy $S \notin S$ so \mathcal{W}' is a *nonempty set*.

We observe that given a set S either $S \in S$ or $S \notin S$ so either $S \in \mathcal{W}$ or $S \in \mathcal{W}'$ but not both. Thus \mathcal{W} and \mathcal{W}' are sets. Let us determine where \mathcal{W}' resides.

Assume that $\mathcal{W}' \in \mathcal{W}'$. Then, because \mathcal{W} is the set of all sets S such that $S \in S$, we must have $\mathcal{W}' \in \mathcal{W}$. This is a contradiction to the fact that *no set* can be in both \mathcal{W} and \mathcal{W}' , so our assumption is in error.

Alternatively, assume that $\mathcal{W}' \notin \mathcal{W}'$. Because \mathcal{W}' is the set of all sets S such that $S \notin S$, we must have $\mathcal{W}' \in \mathcal{W}'$. This is another mathematical impossibility.

Therefore \mathcal{W}' is not a set, which is clearly a contradiction to our choice of \mathcal{W} and \mathcal{W}' . This shows us that our assumption that \mathbf{C} is a set is in error. We conclude that

\mathbf{C} is not a set.

This is a good time to demonstrate that $\text{card}(\{\bullet\})$ is not a set since it contains a copy of \mathbf{C} : it contains $\{X\}$ for each set X .

Now that we have established that \mathbf{C} is not a set, here are several mind expanding shockers.

Theorem 8.1.3 *There are no functions $f : \mathbf{C} \rightarrow B$ for any sets B .*

Proof: Functions are defined on sets and sets only. Since \mathbf{C} is not a set we cannot define a function on it.

Theorem 8.1.4 *\mathbf{C} has no cardinality.*

Proof: Suppose to the contrary that \mathbf{C} has cardinality \aleph_0 . Then there is a set B and a bijection $f : \mathbf{C} \rightarrow B$ such that $\text{card}(B) = \aleph_0$. This is contrary to the previous theorem. This contradiction shows us that \mathbf{C} has no cardinality.

Even though \mathbf{C} has no cardinality there is a more philosophical way of seeing that \mathbf{C} is larger than any mathematical construction. Let A be a set. Form the subcollection

$$\mathcal{A} = \{\{x\} \mid x \in A\}.$$

Then \mathcal{A} is a set and

$$\text{card}(A) = \text{card}(\mathcal{A}).$$

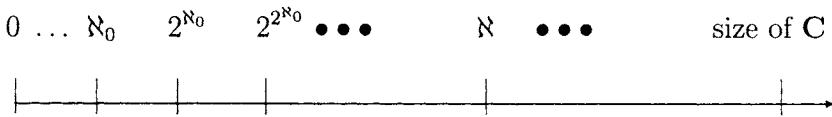
Furthermore,

\mathcal{A} is a subcollection of elements in \mathbf{C}

so that in an intuitive but compelling way

each set A is smaller than \mathbf{C} .

If we did not know about Theorem 8.1.4 we might be tempted to say that the cardinality of \mathbf{C} is larger than every cardinality. In pictures



8.2 Other Than True or False

Have you ever engaged in logical games in which you had to decide who was telling the truth given some information or statements made by the people involved. My favorite is the classic chestnut *The Hidden Tiger Puzzle*. It goes like this. You are on an island where the island ruler asks you to make a choice. Behind one door is a Golden Key that will make you the ruler of the island. Behind the other door is a yellow and black striped tiger who will eat you if the door is opened. The two doors are guarded by two men who know where the Golden Key is located. You may ask one question of the guards in order to find the Golden Key. Here is the twist. One of these guards tells the truth all of the time and the other one lies all of the time. What one question can you ask that will lead

you to the Golden Key? The answer to this question is given on page 281.

The answer to *The Hidden Tiger Puzzle* relies on the fact that most statements have exactly one of two logical states. Namely, they are true or false, but not both. No third alternative is considered. In this section we will give examples of statements that will lead us to conclude that certain statements in the English language have a third possible logical state.

Example 8.2.1 [5] We will show that there is a word W in the English language such that the sentence

“ W is a W word.”

is neither true nor false. We conclude that it must possess some alternative logical state.

Proof: Assume for the sake of contradiction that the statement “ W is a W word” is either true or false, and not both. We will say that a word W is *self-descriptive* if the statement “ W is a W word” is true. The word W is said to be *non-self-descriptive* if the statement “ W is a W word” is false. By our assumption, a word W is either a self-descriptive word or it is a non-self-descriptive word but not both.

For example, since the word *pentasyllabic* has five syllables the sentence

“Pentasyllabic is a pentasyllabic word.”

is true. Thus *pentasyllabic* is a self-descriptive word. This demonstrates that there are self-descriptive words. There are many non-self-descriptive words. For example, since *monosyllabic* means having one syllable, the sentence

“Monosyllabic is a monosyllabic word.”

is false, whence *monosyllabic* is a non-self-descriptive word.

So what kind of word do you think *non-self-descriptive* is? Make your guess and then read on.

Non-self-descriptive is the contradiction we seek. We will show that the sentence

NSD “Non-self-descriptive is a non-self-descriptive word.”

is neither true nor false, thus contradicting our assumption that the statement “*W* is a *W* word” is either true or false but not both.

Suppose that statement *NSD* is true. By the definition of *non-self-descriptive* sentence *NSD* must be false. This is not possible as our initial assumption says that this sentence can have only one logical state.

So it must be that *NSD* is false. By the definition of *non-self-descriptive* the word non-self-descriptive is a *non-self-descriptive* word. That is, sentence *NSD* is true, contrary to our initial assumption.

Since each logical state leads us to a contradiction we conclude that *the logical state of sentence NSD is something other than true or false*. We have found a simple statement that seems to suggest true and false. These statements are also quite simple, leading us to believe that this kind of counterintuitive logical behavior is actually quite common.

Consider what the above example does for the set of words in the English dictionary. Let **S** denote the collection of *self-descriptive* words and let **NS** denote the collection of *non-self-descriptive* words. According to the above example **S** and **NS** do not account for all of the words in the English language. There is a third collection:

O = words *W* for which “*W* is a *W* word” has a logical state *other than true or false*, including the nonsense state.

If we wish to say that a word *W* is not in **S** we will say that

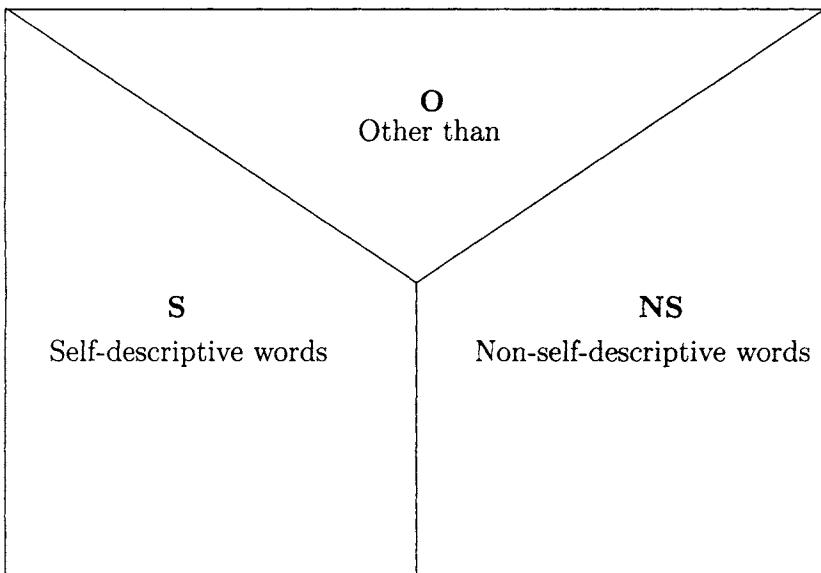
We cannot say that *W* is a self-descriptive word.

This kind of wording allows for the fact that W might be in **NS** (non-self-descriptive words) or it might be in **O** (other than self-descriptive or non-self-descriptive words).

What went wrong here? Why should the simple statement NSD give us so much trouble regarding its logical state? The number of words for which statement NSD is true is not large. It is something you might program your desktop computer to find if it could determine when a word was self-descriptive or non-self-descriptive. Just download the CD that contains the entire Oxford English Dictionary and perform a search. Since this CD is finite the number of words for which statement NSD is true is finite. However, there is a word N such that the statement “ N is an N word” is neither true nor false. Its logical state is some unknown third alternative state. It is possible that the missing third state is that the statement NSD has *no logical state whatever*. We will not pursue the matter here.

Picture (8.2) of the English dictionary might help envision what we are talking about.

Dictionary of the English Language (8.2)



The following classic story is another example of a statement whose logical state is neither true nor false. We will update the story.

The Epimenides Paradox: A philosophy professor named Epi walks into a room of college professors and announces that

"I cannot tell the truth. I am lying."

Let us try to determine the logical state of Epi's simple sentence *I am lying*.

Assume that *I am lying* is a true statement. This is contrary to the fact that *Epi* cannot tell the truth. This is a contradiction in logical states, so that *I am lying* is **not** a true statement.

Alternatively, we assume that *I am lying* is a false statement. Then Epi is telling the truth, contrary to the fact that she cannot tell the truth. This contradiction shows us that *I am lying* is **not** a false statement.

Since it is neither true nor false we conclude that the logical state of *I am lying* is some kind of alternative third logical state.

Try Epi's sentence on your friends. Establish yourself as someone who cannot tell the truth and then announce *I am lying*.

A diagram of English might help you understand how we have partitioned the statements in the language. See the next page.

The point in the previous two examples is that there are sentences in the English language whose logical state cannot be decided. We suggest that their truth state is some logical state that is *other than* true or false.

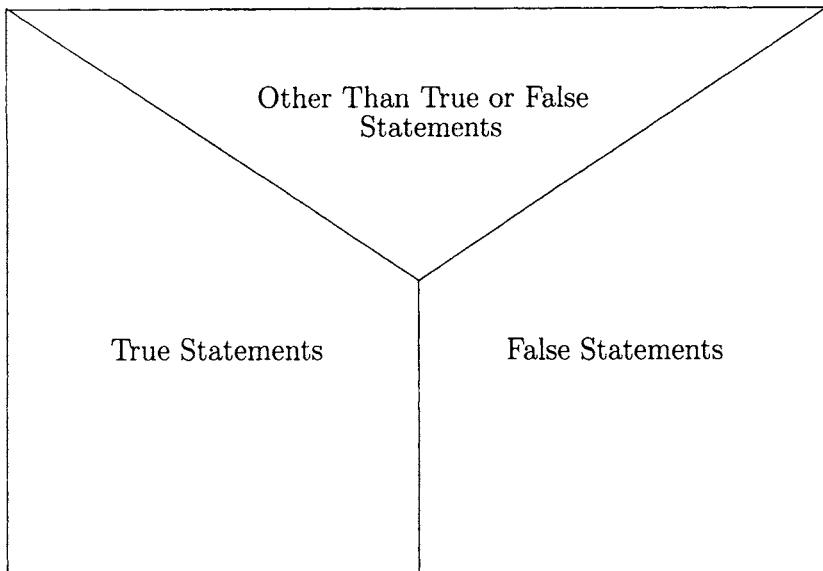
Here is another example of alternative logical states that will really mess with your mind.

Example 8.2.2 Consider the following two sentences.

- #1. Sentence 2 is false.
#2. Sentence 1 is true.

Let us investigate the logical state of Sentence #1.

Statements of the English Language (8.3)



Begin by assuming that Sentence #1 is true. That is, *Sentence 2 is false* is a true sentence. Then sentence #2 is false, or in other words *Sentence 1 is true* is false. It must be then that sentence #1 is false. This is a contradiction in logical states for sentence #1 so that our assumption is incorrect.

Alternatively we assume that sentence #1 is false. That is, *Sentence 2 is false* is a false sentence. I leave it to the reader to produce the contradiction in logical states for sentence #1.

We conclude that the logical state of sentence #1 is some alternative third logical state.

Statements #1 and #2 are not terribly complicated and yet they quickly lead us to a difficult logical situation. We conclude that not all of our language can be analyzed with traditional binary logic. That is, we cannot investigate the language without recognizing the existence of simple statements whose logical states are something other than true or false.

Investigate the logical state of statement #1 in the following exercises.

(1)

- #1. Sentence 2 is false.
 - #2. Sentence 1 is false.

(2)

- #1. Sentence 2 is true.
 - #2. Sentence 3 is true.
 - #3. Sentence 1 is true.

(3)

- #1. Sentence 2 is false.
 - #2. Sentence 3 is false.
 - #3. Sentence 1 is false.

(4)

- #1. Sentence 2 is false.
 - #2. Sentence 3 is false.
 - #3. Sentence 1 is true.

(5)

- #1. Sentence 2 is false.
 - #2. Sentence 3 is true.
 - #3. Sentence 1 is true.

(6)

- #1. Sentence 4 is false.
- #2. Sentence 3 is true.
- #3. Sentence 2 is false.
- #4. Sentence 1 is true.

(7)

- #1. Sentence 4 is false.
- #2. Sentence 3 is true.
- #3. Sentence 2 is true.
- #4. Sentence 1 is true.

(8)

- #1. Sentence 4 is false.
- #2. Sentence 3 is false.
- #3. Sentence 2 is true.
- #4. Sentence 1 is true.

The solution to the Hidden Tiger Puzzle: To find the door that contains the Golden Key choose one of the guards. It doesn't matter which one. Ask him, *Which door would the other guard say that the Golden Key is behind?*. That is one question. Then open the other door. That is where you will find the Golden Key.

Now here is how it works. Label the guards 1 and 2. Suppose you ask the one question of guard 1. If he lies all of the time then guard 2 tells the truth all of the time. Guard 1 will change guard 2's answer from the truth to a falsehood. Guard 1 will answer with the wrong door, the door hiding the Tiger. Open the other door.

On the other hand, suppose guard 1 tells the truth all the time. Then guard 2 lies all the time. Guard 2's answer to the question

would be the door that does not have the Golden Key behind it. Guard 1, being the one who tells the truth, will answer with guard 2's answer, which is the door that hides the Tiger. Open the other door.

Bibliography

- [1] Georg Cantor, *Contributions to the Founding of the Theory of Transfinite Numbers*, Dover Press, New York, 1955.
- [2] K. Devlin, *Mathematics: The New Golden Age*, Pelican Books, New York, 1988.
- [3] H. Eves, *An Introduction to the History of Mathematics*, Saunders College Publishing, New York, 1990.
- [4] C. Faith, *Algebra I: Algebra and Categories*, Springer-Verlag, New York, 1974.
- [5] Douglas R. Hofstader, *Godel-Escher-Bach: An Eternal Golden Braid*, Vintage Press, New York, 1980.
- [6] Nathan Jacobson, *Basic Algebra 1*, W.H. Freeman and Company, New York, 1985.
- [7] Edward Kasner and James Newman, *Mathematics and the Imagination*, Simon-Schuster, New York, 1940.

Index

$(0, 1)$ 46
 $\{0, 1\}^A$ 142

\cap 7

\emptyset 10

\leq for cardinals 117

| 5

A

\aleph (aleph) 144
 \aleph_0 (aleph naught) 102, 145
Abel, Niels 218
absolute zero 3
abstraction 62
Aladdin 227
Aristotelian Logic 186
Ars Magna 217
at least 29, 33
axiomatic method 186

B

β^α 144, 145, 160
 B^A 159
bijection 65
binary decimal 132
binary expansion 45
binary sequences 137
binary series 45
bounded monotonic sequence 263

C

C 266
 $\text{card}(A)$ 23
Cardano 217
cardinality 22, 50
Cartesian product 17
CH 189
cluster point 264
codomain 39, 48
Cohen, Paul J. 186 collection of all finite sets 266
collection of all sets 266
commutative property 147
composite numbers 7
composition 48
continuum 148
Continuum Hypothesis 189
countable 104
countable families 27
countable union 109

D

decimal number 137
degree 251
dense subset 105
discrete subset 105
disjoint 146
domain 39, 48

E**E** 5

- elements 3
- empty set 3
- Epimenides Paradox 276
- equal cardinals 98
- equal sets 11
- equivalent sets 98
- Euclid 186
- Evariste Galois 218

F

- factor 235
- factorial 229
- families 27
- Field's Medal 186
- finite cardinals 139
- for all 29, 33, 59
- for some 29, 33, 34
- function 38

G

- Cantor, Georg 105, 189
- Gauss, Carl F. 92
- GCH** 194
- Generalized Continuum Hypothesis 193
- geometric series 253
- googol 81
- googolplex 81

H

- harmonic series 261
- Hidden Tiger Puzzle 272, 279

Hilbert, David 188

Hilbert's Infinite Hotel 85

Hofstader, Douglas R. 227

I

- image 38, 53
- implied list 4
- incomparable 170
- Induction Hypothesis 205
- Induction Step 205
- infinite family 28
- infinite set 84
- Initial Step 205
- integers 4
- intersection 7
- interval of convergence 253
- inverse functions 68
- invertible 68
- irrational 129
- is an element of 3
- is contained in 8
- is not an element of 3, 14

L

- light year 76
- limit 199
- line segments 118
- linear order 184

M

- magnitude 77
- Mathematical Induction 203
- Minimum Property 168
- mole of matter 80
- multiples of cardinals 154

multiplication of cardinals 155

N

\mathbb{N} 5, 75

NSD 270

natural logarithm 248

natural number 4, 75

nonnegative real numbers 15

non-self-descriptive 270

notCH 189

O

ω_0 (omega naught) 172

one-to-one 62

onto 62

open neighborhood 261

P

$\pi(n)$ 247

\mathbb{P} 5

$\mathcal{P}(A)$ 21, 132

pairs of natural numbers 19

perfect numbers 240

positive irrational numbers 15

positive rational numbers 6, 104

postage rate function 41

power set 21, 132

predecessor element 176

predicate 5

preimage 53

prime number function 247

Prime Number Theorem 249

prime numbers 32, 235

proof by contradiction 16

Pythagoreans 130

Q

\mathbb{Q} 5, 112

\mathbb{Q}^+ 5

quadratic formula 217

R

\mathbb{R} 5

radius of convergence 202

rational numbers 5

real numbers 5

recursion 229

Riemann Hypothesis 256

Riemann Zeta Function 250, 256, 258

Russell, Bertrand 140

S

scientific notation 78

self-descriptive 270

sequence 76, 261

set 2, 3

Set Theory 2

sieve 245

sigma notation 250

Student Induction 228

subset 8

successor element 163

such that 5

T

$\tan(\theta)$ 40, 55, 101, 121

tangent function 40, 55

Tartaglia 217

there exists 59
Transfinite Induction 220
triangle number 230
Trichotomy Property 117, 166, 168,
184

U

\cup 7
 \mathcal{U} 14
uncountable 117, 126
Unexpected Termination, The 218
union 7
unique prime factorization 238
unit cube 123
unit interval 46
unit square 123
universal set 14

W

Well Ordered Property, The 163
well ordered set 221
whole numbers 4

X

\times 19, 114, 115, 116, 155

Z

\mathbb{Z} 4
ZFC Axioms 186

PURE AND APPLIED MATHEMATICS

A Wiley-Interscience Series of Texts, Monographs, and Tracts

Consulting Editor: David A. Cox

Founded by RICHARD COURANT

Editors Emeriti: MYRON B. ALLEN III, DAVID A. COX, PETER HILTON, HARRY HOCHSTADT, PETER LAX, JOHN TOLAND

ADÁMEK, HERRLICH, and STRECKER—Abstract and Concrete Categories

ADAMOWICZ and ZBIERSKI—Logic of Mathematics

AINSWORTH and ODEN—A Posteriori Error Estimation in Finite Element Analysis

AKIVIS and GOLDBERG—Conformal Differential Geometry and Its Generalizations

ALLEN and ISAACSON—Numerical Analysis for Applied Science

*ARTIN—Geometric Algebra

AUBIN—Applied Functional Analysis, Second Edition

AZIZOV and IOKHVIDOV—Linear Operators in Spaces with an Indefinite Metric

BERG—The Fourier-Analytic Proof of Quadratic Reciprocity

BERMAN, NEUMANN, and STERN—Nonnegative Matrices in Dynamic Systems

BERKOVITZ—Convexity and Optimization in \mathbb{R}^n

BOYARINTSEV—Methods of Solving Singular Systems of Ordinary Differential Equations

BURK—Lebesgue Measure and Integration: An Introduction

*CARTER—Finite Groups of Lie Type

CASTILLO, COBO, JUBETE, and PRUNEDA—Orthogonal Sets and Polar Methods in Linear Algebra: Applications to Matrix Calculations, Systems of Equations, Inequalities, and Linear Programming

CASTILLO, CONEJO, PEDREGAL, GARCÍA, and ALGUACIL—Building and Solving Mathematical Programming Models in Engineering and Science

CHATELIN—Eigenvalues of Matrices

CLARK—Mathematical Bioeconomics: The Optimal Management of Renewable Resources, Second Edition

COX—Galois Theory

†COX—Primes of the Form $x^2 + ny^2$: Fermat, Class Field Theory, and Complex Multiplication

*CURTIS and REINER—Representation Theory of Finite Groups and Associative Algebras

*CURTIS and REINER—Methods of Representation Theory: With Applications to Finite Groups and Orders, Volume I

CURTIS and REINER—Methods of Representation Theory: With Applications to Finite Groups and Orders, Volume II

DINCULEANU—Vector Integration and Stochastic Integration in Banach Spaces

*DUNFORD and SCHWARTZ—Linear Operators

Part 1—General Theory

Part 2—Spectral Theory, Self Adjoint Operators in Hilbert Space

Part 3—Spectral Operators

FARINA and RINALDI—Positive Linear Systems: Theory and Applications

FATICONI—The Mathematics of Infinity: A Guide to Great Ideas

FOLLAND—Real Analysis: Modern Techniques and Their Applications

*Now available in a lower priced paperback edition in the Wiley Classics Library.

†Now available in paperback.

- FRÖLICHER and KRIEGL—Linear Spaces and Differentiation Theory
 GARDINER—Teichmüller Theory and Quadratic Differentials
 GILBERT and NICHOLSON—Modern Algebra with Applications, Second Edition
 *GRIFFITHS and HARRIS—Principles of Algebraic Geometry
 GRILLET—Algebra
 GROVE—Groups and Characters
 GUSTAFSSON, KREISS and OLIGER—Time Dependent Problems and Difference Methods
 HANNA and ROWLAND—Fourier Series, Transforms, and Boundary Value Problems,
 Second Edition
 *HENRICI—Applied and Computational Complex Analysis
 Volume 1, Power Series—Integration—Conformal Mapping—Location
 of Zeros
 Volume 2, Special Functions—Integral Transforms—Asymptotics—
 Continued Fractions
 Volume 3, Discrete Fourier Analysis, Cauchy Integrals, Construction
 of Conformal Maps, Univalent Functions
 *HILTON and WU—A Course in Modern Algebra
 *HOCHSTADT—Integral Equations
 JOST—Two-Dimensional Geometric Variational Procedures
 KHAMSI and KIRK—An Introduction to Metric Spaces and Fixed Point Theory
 *KOBAYASHI and NOMIZU—Foundations of Differential Geometry, Volume I
 *KOBAYASHI and NOMIZU—Foundations of Differential Geometry, Volume II
 KOSHY—Fibonacci and Lucas Numbers with Applications
 LAX—Functional Analysis
 LAX—Linear Algebra
 LOGAN—An Introduction to Nonlinear Partial Differential Equations
 MARKLEY—Principles of Differential Equations
 MORRISON—Functional Analysis: An Introduction to Banach Space Theory
 NAYFEH—Perturbation Methods
 NAYFEH and MOOK—Nonlinear Oscillations
 PANDEY—The Hilbert Transform of Schwartz Distributions and Applications
 PETKOV—Geometry of Reflecting Rays and Inverse Spectral Problems
 *PRENTER—Splines and Variational Methods
 RAO—Measure Theory and Integration
 RASSIAS and SIMSA—Finite Sums Decompositions in Mathematical Analysis
 RENELET—Elliptic Systems and Quasiconformal Mappings
 RIVLIN—Chebyshev Polynomials: From Approximation Theory to Algebra and Number
 Theory, Second Edition
 ROCKAFELLAR—Network Flows and Monotropic Optimization
 ROITMAN—Introduction to Modern Set Theory
 ROSSI—Theorems, Corollaries, Lemmas, and Methods of Proof
 *RUDIN—Fourier Analysis on Groups
 SENDOV—The Averaged Moduli of Smoothness: Applications in Numerical Methods
 and Approximations
 SENDOV and POPOV—The Averaged Moduli of Smoothness
 SEWELL—The Numerical Solution of Ordinary and Partial Differential Equations,
 Second Edition
 SEWELL—Computational Methods of Linear Algebra, Second Edition
 *SIEGEL—Topics in Complex Function Theory
 Volume 1—Elliptic Functions and Uniformization Theory
 Volume 2—Automorphic Functions and Abelian Integrals
 Volume 3—Abelian Functions and Modular Functions of Several Variables

*Now available in a lower priced paperback edition in the Wiley Classics Library.

†Now available in paperback.

- SMITH and ROMANOWSKA—Post-Modern Algebra
ŠOLÍN—Partial Differential Equations and the Finite Element Method
STADE—Fourier Analysis
STAKGOLD—Green's Functions and Boundary Value Problems, Second Edition
STAHL—Introduction to Topology and Geometry
STANOVYEVITCH—Introduction to Numerical Ordinary and Partial Differential
Equations Using MATLAB®
*STOKER—Differential Geometry
*STOKER—Nonlinear Vibrations in Mechanical and Electrical Systems
*STOKER—Water Waves: The Mathematical Theory with Applications
WATKINS—Fundamentals of Matrix Computations, Second Edition
WESSELING—An Introduction to Multigrid Methods
†WHITHAM—Linear and Nonlinear Waves
†ZAUDERER—Partial Differential Equations of Applied Mathematics, Second Edition

*Now available in a lower priced paperback edition in the Wiley Classics Library.

†Now available in paperback.