

# A Readable Introduction to Real Mathematics

(VERY PRELIMINARY DRAFT)

Daniel Rosenthal, David Rosenthal and Peter  
Rosenthal

Copyright ©2012 by Daniel Rosenthal, David Rosenthal and Peter  
Rosenthal

All rights reserved.

# Contents

<b>Preface</b>	<b>vi</b>
<b>1 Introduction to the Natural Numbers</b>	<b>1</b>
1.1 Prime numbers . . . . .	2
1.2 Unanswered Questions . . . . .	6
1.3 Problems . . . . .	7
<b>2 Mathematical Induction</b>	<b>8</b>
2.1 The Principle of Mathematical Induction . . . . .	8
2.2 The Principle of Complete Mathematical Induction . . . . .	17
2.3 Problems . . . . .	22
<b>3 Modular Arithmetic</b>	<b>25</b>
3.1 The Basics . . . . .	25
3.2 Some Applications . . . . .	28
3.3 Problems . . . . .	30
<b>4 The Fundamental Theorem of Arithmetic</b>	<b>33</b>
4.1 Proof of the Fundamental Theorem of Arithmetic . . . . .	33
4.2 Problems . . . . .	36
<b>5 Fermat's Theorem and Wilson's Theorem</b>	<b>38</b>
5.1 Fermat's Theorem . . . . .	38
5.2 Wilson's Theorem . . . . .	41

---

5.3	Problems . . . . .	42
<b>6</b>	<b>Sending and Receiving Coded Messages</b>	<b>44</b>
6.1	The RSA Method . . . . .	45
6.2	Problems . . . . .	49
<b>7</b>	<b>The Euclidean Algorithm and Applications</b>	<b>50</b>
7.1	The Euclidean Algorithm . . . . .	50
7.2	Applications . . . . .	52
7.3	Problems . . . . .	61
<b>8</b>	<b>Rational Numbers and Irrational Numbers</b>	<b>64</b>
8.1	Rational Numbers . . . . .	64
8.2	Irrational Numbers . . . . .	67
8.3	Problems . . . . .	71
<b>9</b>	<b>The Complex Numbers</b>	<b>74</b>
9.1	What is a Complex Number? . . . . .	74
9.2	The Complex Plane . . . . .	77
9.3	The Fundamental Theorem of Algebra . . . . .	83
9.4	Problems . . . . .	85
<b>10</b>	<b>Sizes of Infinite Sets</b>	<b>87</b>
10.1	Cardinality . . . . .	87
10.2	Countable Sets and Uncountable Sets . . . . .	92
10.3	Comparing Cardinalities . . . . .	97
10.4	Problems . . . . .	112
<b>11</b>	<b>Fundamentals of Plane Geometry</b>	<b>115</b>
11.1	Triangles . . . . .	115
11.2	Euclidean Geometry . . . . .	120
<b>12</b>	<b>Constructability</b>	<b>126</b>
12.1	Constructions With Straightedge and Compass . . . . .	127

---

12.2 Constructible numbers . . . . .	130
12.3 Surds . . . . .	136
12.4 Constructions of Geometric Figures . . . . .	144
12.5 Problems . . . . .	151

# Preface

The fundamental purpose of this book is to teach you to understand mathematical thinking. We have tried to do that in a way that is clear, engaging and emphasizes the beauty of mathematics. You may be reading this book on your own or as a text for a course you are enrolled in. Regardless of the reason you are reading this book, we hope that you will find it understandable and interesting.

Mathematics is a huge body of knowledge, way too vast for any individual to have anything like a complete grasp of. But the essence of mathematics is thinking mathematically. It is our view that mathematical thinking can be learned by almost anyone who is willing to make a serious attempt. We invite you to make such an attempt by reading this book.

One way in which mathematics gets very complex is by building on itself; some mathematical concepts are built on a foundation of many other concepts and thus require a great deal of background to understand. That is not the case for the topics discussed in this book. For you to read this book does not require any background other than basic high school algebra and, for a small portion of it, some high school trigonometry.

A few questions, among the many, that you will be able to easily answer after reading this book are the following. Is  $13^{217} \cdot 37^{92} \cdot 41^{15} = 19^{111} \cdot 29^{145} \cdot 43^{12} \cdot 47^5$  (see Chapter 4)? Is there a largest prime number (i.e., a largest whole number whose only factors are it and itself) (Theorem 1.1.2)? If a store sells one kind of product for 9 dollars each and another kind for 16 dollars each and receives 143 dollars for the total sale of both, how many products did the store sell at each price (Example 7.2.4)? How do computers send secret messages to each other (Chapter 6)? Are there more fractions than there are whole numbers? Are there more real numbers than there are fractions? Is there a smallest infinity? Is there a largest infinity (Chapter 10)? What are complex numbers and what are they good for (Chapter 9)?

The hardest theorem we will prove concerns construction of angles using a

compass and a straightedge. (A straightedge is a ruler-like device but without measurements marked on it.) If you are given any angle, it is easy to bisect it (i.e., divide it into two equal subangles) by using a compass and a straightedge (we will show you how to do that). This and many similar results were discovered by the Ancient Greeks. The Ancient Greeks wondered whether angles could be “trisected” in the sense of being divided into three equal subangles using only a straightedge and compass. A great deal of mathematics beyond that conceived of by the Ancient Greeks was required to solve this problem; it was not solved until the 19th century. It can be proven that many angles, including angles 60 degrees, cannot be so trisected. We present a complete proof of this as an illustration of complicated but beautiful mathematical reasoning.

The most important question you’ll be able to answer after reading this book, although you would have difficulty formulating the answer in words, is: what is mathematical thinking really like? If you read and understand most of this book and do a fair number of the problems that are provided, you will certainly have a real feeling for mathematical thinking.

We hope that you read this book carefully. Reading mathematics is not like reading a novel, a newspaper or anything else. As you go along, you have to really reflect on the mathematical reasoning that we are presenting. After reading a description of an idea, think about it. When reading mathematics you should always have a pencil and paper at hand and re-work what you read.

Mathematics consists of theorems, which are statements proven to be true. We will prove a number of theorems. When you begin reading about a theorem, think about why it may be true before you read our proof. In fact, at some points you may be able to prove the theorem we state without looking at our proof at all. In any event, you should make at least a small attempt before reading the proof in the book. It is often useful to continue such attempts while in the middle of reading the proof that we present; once we have gotten you a certain way towards the result, see if you can continue on your own.

If you adopt such an approach and are patient, we are convinced that you will learn to think mathematically. We are also convinced that you will feel that much of the mathematics that you learn is beautiful, in the sense that you will find that the logical argument that establishes the theorem is what mathematicians call “elegant”.

We chose the material for this book based on the following criteria: it is beautiful mathematics that is useful in many mathematical contexts and is accessible without a great deal of mathematical background. The theorems that we prove in this book have applications to mathematics and to problems in other

subjects. Some of these applications will be presented in what follows.

Each chapter ends with a section entitled “Problems”. The problems sections are divided into three subsections. The first, “Basic Exercises”, consists of problems whose solution you should do to assure yourself that you have an understanding of the fundamentals of the material. The subsections entitled “Interesting Problems” contain problems whose solutions depend upon the material of the chapter and seem to have mathematical or other interest. The subsections labelled “Challenging Problems” contain problems that we expect you will, indeed, find to be quite challenging. You should not be discouraged if you cannot solve some of the problems. However, if you do solve problems that you find difficult at first, especially those that we have labelled “challenging”, we hope and expect that you will feel some of the pleasure and satisfaction that mathematicians feel upon discovering new mathematics.

This book was developed from lecture notes for a course that was given at the University of Toronto over a period of fifteen years. It has been greatly improved by students’ suggestions. Nonetheless, we are sure that further improvements could be made. We would appreciate your sending any comments, corrections or suggestions to any of the authors at their email addresses given below.

Daniel Rosenthal:

David Rosenthal: [rosenthd@stjohns.edu](mailto:rosenthd@stjohns.edu)

Peter Rosenthal: [rosent@math.toronto.edu](mailto:rosent@math.toronto.edu)



## Chapter 1

# Introduction to the Natural Numbers

We assume basic knowledge about the numbers that we count with; that is, the numbers 1, 2, 3, 4, 5, 6 and so on. These are called the *natural numbers*, and they do seem to be very natural, in the sense that they arose very early on in virtually all societies. There are many other names for these numbers, such as the *positive integers* and the *positive whole numbers*. Although the natural numbers are very familiar, we'll see that they have many interesting properties beyond the obvious ones. Moreover, there are many questions about the natural numbers to which nobody knows the answer. Some of these questions can be stated very simply, as we shall see, although their solution has eluded the thousands of mathematicians who have attempted to solve them.

We assume familiarity with the two basic operations on the natural numbers, addition and multiplication. The sum of two numbers will be indicated using the plus sign “+”. Multiplication will be indicated by putting a dot in the middle of the line between the numbers, or by simply writing the symbols for the numbers next to each other, or sometimes by enclosing them in parentheses. For example, the product of 3 and 2 could be denoted  $3 \cdot 2$  or  $(3)(2)$ . The product of the natural numbers represented by the symbols  $m$  and  $n$  could be denoted  $mn$ , or  $m \cdot n$ , or  $(m)(n)$ .

We also, of course, need the number 0. Moreover, we require the negative integers as well. For each natural number  $n$  there is a corresponding negative integer,  $-n$  such that  $n + (-n) = 0$ . We assume that you know how to add two negative integers and also how to add a negative integer to a positive integer. Multiplication appears to be a bit more mysterious. Most people feel comfortable

with the fact that, for  $m$  and  $n$  natural numbers, the product of  $m$  and  $(-n)$  is  $-mn$ . What some people find more mysterious is the fact that  $(-m)(-n) = mn$  for natural numbers  $m$  and  $n$ ; that is, the product of two negative integers is a positive integer. There are various possible explanations that can be provided for this, one of which is the following. Using the usual rules of arithmetic,

$$(-m)(-n) + (-m)(n) = (-m)(-n + n) = (-m)0 = 0$$

Adding  $mn$  to both sides of this equation gives

$$(-m)(-n) + (-m)(n) + mn = 0 + mn$$

or

$$(-m)(-n) + ((-m) + m)n = mn.$$

Thus

$$(-m)(-n) + 0 \cdot n = mn,$$

so

$$(-m)(-n) = mn.$$

Therefore the fact that  $(-m)(-n) = mn$  is implied by the other standard rules of arithmetic.

## 1.1 Prime numbers

One of the important concepts we will study is *divisibility*. For example, 12 is divisible by 3, which means that there is a natural number (in this case, 4) such that the product of 3 and that natural number is 12. That is,  $12 = 3 \cdot 4$ . In general, we say that *the integer  $m$  is divisible by the integer  $n$*  if there is an integer  $q$  such that  $m = nq$ . There are many other terms that are used to describe such a relationship. For example, if  $m = nq$ , we may say that  $n$  and  $q$  are *divisors* of  $m$ , and that each of  $n$  and  $q$  *divides*  $m$ . The terminology “ $q$  is the quotient when  $m$  is divided by  $n$ ” is also used. In this situation,  $n$  and  $q$  are also sometimes called *factors* of  $m$ ; the process of writing an integer as a product of two or more integers is called *factoring* the integers.

The number 1 is a divisor of every natural number since, for each natural number  $m$ ,  $m = 1 \cdot m$ . Also, every natural number  $m$  is a divisor of itself, since  $m = m \cdot 1$ .

The number 1 is the only natural number that has only one natural number divisor, namely itself. All the other natural numbers have at least two divisors,

themselves and 1. The natural numbers that have exactly two natural number divisors are called *prime numbers*. That is, a *prime number* is a natural number greater than 1 whose only natural number divisors are 1 and the number itself. We do not consider the number 1 to be a prime; the first prime number is 2. The primes continue: 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, and so on.

And so on? Is there a largest prime? Or does the sequence of primes continue without end? There is, of course, no largest natural number. For if  $n$  is any natural number, then  $n + 1$  is a natural number and  $n + 1$  is bigger than  $n$ . It is not so easy to determine if there is a largest prime number or not. If  $p$  is a prime, then  $p + 1$  is almost never a prime. Of course, if  $p = 2$ , then  $p + 1 = 3$  and  $p$  and  $p + 1$  are both primes. However, 2 is the only prime number  $p$  for which  $p + 1$  is prime. This can be proven as follows. First note that, since every even number is divisible by 2, 2 itself is the only even prime number. Therefore, if  $p$  is a prime other than 2, then  $p$  is odd and  $p + 1$  is an even number larger than 2 and is thus not prime.

Is it nonetheless true that, given any prime number  $p$ , there is a prime number larger than  $p$ ? Although we cannot get a larger prime by simply adding 1 to a given prime, there may be some other way of producing a larger prime than any given one. We will answer this question after learning a little more about primes.

A natural number, other than 1, that is not prime is said to be *composite*. (The number 1 is special, and is neither prime nor composite.) For example, 4, 68, 129 and 2010 are composites. Thus, a composite number is a natural number other than 1 that has a divisor in addition to itself and 1.

To determine if a number is prime, what potential factors must be checked to eliminate the possibility that there are factors other than the number and 1? If  $m = n \cdot q$ , it cannot be the case that both  $n$  and  $q$  are larger than the square root of  $m$ , for if two natural numbers were both larger than the square root of  $m$  then their product would be larger than  $m$ . It follows that a natural number (other than 1) that is not prime has at least 1 divisor that is larger than 1 and is no larger than the square root of that natural number. Thus, to check whether or not a natural number  $m$  is prime, you need not check whether all of the natural numbers less than  $m$  divides  $m$ . It suffices to check if  $m$  has a divisor that is larger than 1 and no larger than the square root of  $m$ . If it has such a divisor it is composite; if it has no such divisor, it is prime.

For example, we can conclude that 101 is prime since none of the numbers 2, 3, 4, 5, 6, 7, 8, 9, 10 are divisors of 101.

Using refinements of this idea and powerful computers, many very large

numbers have been shown to be prime. For example, 100,000,559 is prime, as is 22,801,763,489.

The fact that very large natural numbers have been shown to be prime does not answer the question of whether there is a largest prime. To show that there is always a prime larger than  $p$  for *every* prime number  $p$  cannot be established by computing any number of specific primes, no matter how large.

Over the centuries, mathematicians have discovered many proofs that there is no largest prime. We shall present one of the simplest and most beautiful proofs, discovered by the Ancient Greeks.

We begin by establishing a preliminary fact that is required for the proof. A statement that is proven for the purpose of being used to prove something else is called a “lemma”. We need a lemma. The lemma that we require states that every composite number has a divisor that is a prime number. (The proof that we present of the lemma is quite convincing, but we shall subsequently present a more precise proof.)

**Lemma 1.1.1.** *If  $m$  is a composite natural number, then there is a prime number that is a divisor of  $m$ .*

*Proof.* Since  $m$  is composite,  $m$  has at least one factorization  $m = n \cdot q$ , where neither  $n$  nor  $q$  is  $m$  or 1. If either of  $n$  or  $q$  is a prime number, then the lemma is established for  $m$ . If  $n$  is not prime, then it has a factorization  $n = s \cdot t$ , where  $s$  and  $t$  are natural numbers other than 1 and  $n$ . It is clear that  $s$  and  $t$  are also divisors of  $m$ . Thus, if either of  $s$  and  $t$  is a prime number, the lemma is established. If  $s$  is not prime, then it can be factored into a product where neither factor is  $s$  or 1. And so on. Continued factoring must get down to a factor that cannot itself be factored; i.e., to a prime. That prime number is a divisor of  $m$ , so the lemma is established.  $\square$

The following is the ingenious proof of the infinitude of the primes discovered by the Ancient Greeks.

**Theorem 1.1.2.** *There is no largest prime number.*

*Proof.* We must show that, given any prime number, there is a prime number larger than the given one. Let  $p$  be any prime number. We must prove that there is some prime larger than  $p$ . To do this, we will construct a larger number that we will show is either a prime larger than  $p$  or has a prime divisor larger than  $p$ . In both cases we will conclude that there is a prime number larger than  $p$ .

Here is how we construct the large number. Let  $M$  be the number obtained by taking the product of all the prime numbers up to and including the given prime  $p$  and then adding 1 to that product. That is,

$$M = (2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdot 13 \cdot 17 \cdot 19 \cdots p) + 1$$

It is possible that  $M$  is a prime number. If that is the case, then there is a prime number larger than  $p$ , since  $M$  is obviously larger than  $p$ . If  $M$  is not prime, then it is composite. We must show that there is a prime larger than  $p$  in this case as well.

Suppose, then, that  $M$  is composite. By Lemma 1.1.1, it follows that  $M$  has a prime divisor. Let  $q$  be any prime divisor of  $M$ . We will show that  $q$  is larger than  $p$ , and thus that there is a prime larger than  $p$  in this case as well.

Consider possible values of  $q$ , the prime divisor of  $M$ . Surely  $q$  is not 2, for

$$2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdot 13 \cdot 17 \cdot 19 \cdots p$$

is an even number, and thus adding 1 to that number to get  $M$  produces an odd number. That is,  $M$  is odd and is therefore not divisible by 2. Since  $q$  does divide  $M$ ,  $q$  cannot be equal to 2.

Similar reasoning shows that  $q$  cannot be 3. For

$$2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdot 13 \cdot 17 \cdot 19 \cdots p$$

is a multiple of 3, so the number obtained by adding 1, namely  $M$ , leaves a remainder of 1 when it is divided by 3. That is, 3 is not a divisor of  $M$ . Since  $q$  is a divisor of  $M$ ,  $q$  is not 3.

Exactly the same proof shows that  $q$  is not 5, since 5 is a divisor of

$$2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdot 13 \cdot 17 \cdot 19 \cdots p$$

and thus cannot be a divisor of  $M$ . In fact, the same proof establishes that  $q$  cannot be any of the factors  $2, 3, 5, \dots, p$  of the product

$$2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdot 13 \cdot 17 \cdot 19 \cdots p.$$

Since every prime number up to and including  $p$  is a factor of that product,  $q$  cannot be any of those prime numbers. Therefore  $q$  is a prime number that is not any of the prime numbers up to and including  $p$ . It follows that  $q$  is a prime number larger than  $p$ , and we have proven that there is a prime number larger than  $p$  in the case where  $M$  is composite. Therefore in both cases, the case where  $M$  is prime and the case where  $M$  is composite, we have shown that there is a prime number larger than  $p$ . This proves the theorem.  $\square$

Every mathematician would agree that the above proof is “elegant”. If you do not find this proof interesting, then you will not enjoy this book. If you do find it interesting, then you are likely to appreciate many of the other ideas that we will discuss (and much mathematics that we do not cover as well).

## 1.2 Unanswered Questions

There are some questions concerning prime numbers that no one has been able to answer. One famous question concerns what are called *twin primes*. Since 2 is the only even prime number, the only consecutive integers that are prime are 2 and 3. There are, however, many pairs of primes that are two apart, such as  $\{3, 5\}$ ,  $\{29, 31\}$ ,  $\{101, 103\}$ ,  $\{1931, 1933\}$  and  $\{104471, 104473\}$ . Such pairs are called *twin primes*. One question that remains unanswered, in spite of the efforts of thousands of mathematicians over hundreds of years, is the question of whether there is a largest pair of twin primes. Some very large pairs are known (e.g.,  $\{1000000007, 1000000009\}$  and many pairs that are even much bigger), but no one knows if there is a largest such.

Another very famous unsolved problem is whether or not the *Goldbach conjecture* is true. Several hundred years ago, Goldbach conjectured (that is, said that he thought that it was probably true) that every even natural number larger than 2 is the sum of two prime numbers. (For example,  $6 = 3 + 3$ ,  $20 = 7 + 13$ ,  $22,901,764,048 = 22,801,763,489 + 100,000,559$ .) Goldbach’s conjecture is known to be true for many very large even natural numbers, but no one has been able to prove it in general (or to show that there is an even number that cannot be written as the sum of two primes).

If you are able to solve the twin primes problem or determine the truth or falsity of Goldbach’s conjecture, you will become immediately famous throughout the world and your name will remain famous as long as civilization endures. On the other hand, it will almost undoubtedly prove to be extremely difficult to answer either of those questions. On the other “other hand”, there is a very slight possibility that one or both of those questions has a fairly simple answer that has been overlooked by the many great and not-so-great mathematicians who have thought about them. In spite of the small possibility of success you might find it interesting to think about these problems.

## 1.3 Problems

### Basic Exercises

1. Show that 68, 129, and 2010 are composite numbers.
2. Which of the following are prime numbers: 79, 153, 537, 1486?

### Interesting Problems

3. Prove that for every natural number  $n > 1$  there is a prime number between  $n$  and  $n!$ . (Recall that  $n!$  is defined to be  $n(n-1)(n-2)\cdots 2\cdot 1$ .) [Hint: There is a prime number that divides  $n! - 1$ .] Note that this gives an alternate proof that there are infinitely many prime numbers.

### Challenging Problems

4. Find a prime number  $p$  such that the number  $(2 \cdot 3 \cdot 5 \cdot 7 \cdots p) + 1$  is not prime.
5. Suppose that  $p$ ,  $p + 2$ , and  $p + 4$  are prime numbers. Prove that  $p = 3$ . [Hint: Why can't  $p$  be 5 or 7?]
6. Prove that, for every natural number  $n$ , there are  $n$  consecutive composite numbers. [Hint:  $(n + 1)! + 2$  is a composite number.]

## Chapter 2

# Mathematical Induction

There is a method for proving certain theorems that is called *mathematical induction*. We will give a number of examples of proofs that use this method. The basis for the principle of mathematical induction, however, is a statement about sets of natural numbers. Recall that the set  $\mathbb{N}$  of all natural numbers is the set  $\{1, 2, 3, \dots\}$ . The basis for the principle of mathematical induction is an alternate description of that set.

### 2.1 The Principle of Mathematical Induction

Suppose  $S$  is a set of natural numbers that has the following two properties:

- A. The number 1 is in  $S$ .
- B. Whenever a natural number is in  $S$ , the next natural number is also in  $S$ .

The second property can be stated a little more formally: If  $k$  is a natural number and  $k$  is in  $S$ , then  $k + 1$  is in  $S$ .

What can we say about a set  $S$  that has those two properties? Since 1 is in  $S$  (by property A), it follows from property B that 2 is in  $S$ . Since 2 is in  $S$ , it follows from property B that 3 is in  $S$ . Since 3 is in  $S$ , 4 is in  $S$ . Then 5 is in  $S$ , 6 is in  $S$ , 7 is in  $S$ , and so on. It seems clear that  $S$  must contain every natural number. That is, the only set of natural numbers with the above two properties is the set of all natural numbers. We state this formally:

**Definition 2.1.1** (*The Principle of Mathematical Induction*). If  $S$  is any set of natural numbers with the properties that:



A. 1 is in  $S$ , and

B.  $k + 1$  is in  $S$  whenever  $k$  is any number in  $S$ ,

then  $S$  is the set of all natural numbers.

We gave an indication above of why the principle of mathematical induction is true. A more formal proof can be based on the following different, more obvious, fact.

**Proposition 2.1.2** (The well-ordering of the natural numbers). *Every set of natural numbers that contains at least one element has a smallest element in it.*

We can establish the Principle of Mathematical Induction from the well-ordering of the natural numbers as follows. Suppose that the well-ordering of the natural numbers holds for all sets of natural numbers. Let  $S$  be any set of natural numbers that has properties A and B of the Principle of Mathematical Induction. To prove the Principle of Mathematical Induction, we must prove that  $S$  is equal to the set of all natural numbers. We will do this by showing that it is impossible that there is any natural number that is not in  $S$ . To see this, suppose that  $S$  does not contain all natural numbers. Then let  $T$  denote the set of all natural numbers that are not in  $S$ . Assuming that  $S$  is not the set of all natural numbers is equivalent to assuming that  $T$  has at least one element. If this were the case, then well-ordering would imply that  $T$  has a smallest element. We will show that this is impossible.

Suppose that  $t$  was the smallest element of  $T$ . Since 1 is in  $S$ , 1 is not in  $T$ . Therefore,  $t$  is larger than 1, so  $t - 1$  is also a natural number. Since  $t - 1$  is less than the smallest number  $t$  in  $T$ ,  $t - 1$  cannot be in  $T$ . Since  $T$  contains all the natural numbers that are not in  $S$ , it follows that  $t - 1$  is in  $S$ . This, however, leads to the following contradiction. Since  $S$  has property B,  $(t - 1) + 1$  must also be in  $S$ . But this is  $t$ , which is in  $T$  and therefore not in  $S$ . This shows that the assumption that there is a smallest element of  $T$  is not consistent with the properties of  $S$ . Thus, there is no smallest element of  $T$  and, by well-ordering, there is therefore no element in  $T$ . This proves that  $S$  is the set of all natural numbers.  $\square$

The way mathematical induction is usually explained can be illustrated by considering the following example. Suppose that we wish to prove, for every natural number  $n$ , the validity of the following formula for the sum of the first  $n$  natural numbers:

$$1 + 2 + 3 + \cdots + (n - 1) + n = \frac{n(n + 1)}{2}$$

One way to prove that this formula holds for every  $n$  is the following. First, the formula does hold for  $n = 1$ , for in this case the left-hand side is just 1 and the right-hand side is  $1(1 + 1)/2$ , which is equal to 1. To prove that the formula holds for all  $n$ , we will establish the fact that whenever the formula holds for any given natural number, the formula will also hold for the next natural number. That is, we will prove that the formula holds for  $n = k + 1$  whenever it holds for  $n = k$ . (This passage from  $k$  to  $k + 1$  is often called “the inductive step”.) If we prove this fact, then, since we know that the formula does hold for  $n = 1$ , it would follow from this fact that it holds for the next natural number, 2. Then, since it holds for  $n = 2$ , it holds for the natural number that follows 2, which is 3. Since it holds for 3, it holds for 4; and then for 5, and 6 and so on. Thus, we will conclude that the formula holds for every natural number. (This is really just the principle of mathematical induction. Let  $S$  be the set of all  $n$  for which the formula for the sum is true. Showing that  $S$  has properties A and B leads to the conclusion that  $S$  is the set of all natural numbers.)

To prove the formula in general, then, we must show that the formula holds for  $n = k + 1$  whenever it holds for  $n = k$ . Assume that the formula does hold for any particular  $n = k$ , where  $k$  is any fixed natural number. That is, we assume the formula

$$1 + 2 + 3 + \cdots + (k - 1) + k = \frac{k(k + 1)}{2}.$$

We want to derive the formula for  $n = k + 1$  from the above equation. That is easy to do, as follows. Assuming the above formula, add  $k + 1$  to both sides, getting

$$1 + 2 + 3 + \cdots + (k - 1) + k + (k + 1) = \frac{k(k + 1)}{2} + (k + 1)$$

We shall see that a little algebraic manipulation of the right-hand side of the above will produce the formula for  $n = k + 1$ . To see this, simply note that

$$\begin{aligned} \frac{k(k + 1)}{2} + (k + 1) &= \frac{k(k + 1)}{2} + \frac{2(k + 1)}{2} \\ &= \frac{k(k + 1) + 2(k + 1)}{2} \\ &= \frac{(k + 2)(k + 1)}{2} \\ &= \frac{(k + 1)(k + 2)}{2} \\ &= \frac{(k + 1)((k + 1) + 1)}{2} \end{aligned}$$

Thus,

$$1 + 2 + 3 + \cdots + (k-1) + k + (k+1) = \frac{(k+1)((k+1)+1)}{2}$$

This equation is the same as that obtained from the formula by substituting  $k+1$  for  $n$ . Therefore we have established the inductive step, so we conclude that the formula does hold for all  $n$ .

There are many very similar proofs of similar formulas.

**Theorem 2.1.3.**  $1^2 + 2^2 + 3^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6}$

*Proof.* Let  $S$  be the set of all natural numbers for which the theorem is true. We want to show that  $S$  contains all of the natural numbers. We do this by showing that  $S$  has properties A and B.

For property A, we need to check that  $1^2 = \frac{1(1+1)(2 \cdot 1 + 1)}{6}$ . This is true, so  $S$  satisfies property A. To verify property B, let  $k$  be in  $S$ . We must show that  $k+1$  is in  $S$ . Since  $k$  is in  $S$ , the theorem holds for  $k$ ; i.e.,

$$1^2 + 2^2 + 3^2 + \cdots + k^2 = \frac{k(k+1)(2k+1)}{6}$$

Using this formula, we can prove the corresponding formula for  $k+1$  as follows. Adding  $(k+1)^2$  to both sides of the above equation, we get:

$$1^2 + 2^2 + 3^2 + \cdots + k^2 + (k+1)^2 = \frac{k(k+1)(2k+1)}{6} + (k+1)^2$$

Now we do some algebraic manipulations to the right hand side to see that it is what we want.

$$\begin{aligned} \frac{k(k+1)(2k+1)}{6} + (k+1)^2 &= \frac{k(k+1)(2k+1) + 6(k+1)^2}{6} \\ &= \frac{(k+1)(k(2k+1) + 6(k+1))}{6} \\ &= \frac{(k+1)((2k^2 + k) + (6k + 6))}{6} \\ &= \frac{(k+1)(2k^2 + 7k + 6)}{6} \\ &= \frac{(k+1)(k+2)(2k+3)}{6} \end{aligned}$$

The last equation is the formula in the case when  $n = k + 1$ , so  $k + 1$  is in  $S$ . Therefore,  $S$  is the set of natural numbers by the Principle of Mathematical Induction.  $\square$

Sometimes one wants to prove something by induction that is not true for all natural numbers, but only for those bigger than a given natural number. A slightly more general principle that we can use in such situations is the following.

**Definition 2.1.4** (Generalized Principle of Mathematical Induction). Let  $m$  be a natural number. If  $S$  is a set of natural numbers with the properties that:

- A.  $m$  is in  $S$ , and
- B.  $k + 1$  is in  $S$ , whenever  $k$  is in  $S$  and is greater than or equal to  $m$ ,

then  $S$  contains every natural number greater than or equal to  $m$ .

The Principle of Mathematical Induction is the special case of this principle when  $m = 1$ . We can use induction starting at any number, not just at 1.

For example, consider the question: which is larger,  $n!$  or  $2^n$ ? (Recall that  $n! = n \cdot (n - 1) \cdot (n - 2) \cdots 3 \cdot 2 \cdot 1$ .) For  $n = 1, 2$ , and  $3$ , we see that

$$1! = 1 < 2^1 = 2$$

$$2! = 2 \cdot 1 = 2 < 2^2 = 2 \cdot 2 = 4$$

$$3! = 3 \cdot 2 \cdot 1 = 6 < 2^3 = 2 \cdot 2 \cdot 2 = 8$$

But when  $n = 4$ , the inequality is reversed, since

$$4! = 4 \cdot 3 \cdot 2 \cdot 1 = 24 > 2^4 = 2 \cdot 2 \cdot 2 \cdot 2 = 16$$

When  $n = 5$ ,

$$5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120 > 2^5 = 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 = 32$$

If you think about it a bit it is clear why eventually  $n!$  is much bigger than  $2^n$ . In both expressions we are multiplying  $n$  numbers together, but for  $2^n$  we are always multiplying by 2, whereas the numbers we multiply to build  $n!$  get larger and larger. While it is not true that  $n! > 2^n$  for every natural number (since it is not true for  $n = 1, 2$ , and  $3$ ), we can, as we now show, use the more general form of mathematical induction to prove that it is true for all natural numbers greater than or equal to 4.

**Theorem 2.1.5.**  $n! > 2^n$  for  $n \geq 4$ .

*Proof.* We use the Generalized Principle of Mathematical Induction with  $m = 4$ . Let  $S$  be the set of natural numbers for which the theorem is true. As we saw above,  $4! > 2^4$ . Therefore, 4 is in  $S$ . Thus, property A is satisfied. For property B, assume that  $k \geq 4$  and that  $k$  is in  $S$ ; i.e.,  $k! > 2^k$ . We must show that  $(k+1)! > 2^{k+1}$ . Multiplying both sides of the inequality for  $k$  (which we have assumed to be true) by  $k+1$ , gives

$$(k+1)k! > (k+1)2^k$$

The left-hand side is just  $(k+1)!$ ; therefore we have the inequality

$$(k+1)! > (k+1)2^k$$

Since  $k \geq 4$ ,  $k+1 > 2$ . Therefore, the right-hand side of the inequality,  $(k+1)2^k$ , is greater than  $2 \cdot 2^k = 2^{k+1}$ . Combining this with the above inequality, we get

$$(k+1)! > (k+1)2^k > 2^{k+1}$$

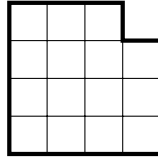
Thus,  $k+1$  is in  $S$ , which verifies property B. By the Generalized Principle of Mathematical induction,  $S$  contains all natural numbers greater than or equal to 4.  $\square$

The following is an example where mathematical induction is useful in establishing a geometric result. We will use the word “tromino” to denote an L-shaped object consisting of three squares of the same size. That is, a tromino looks like this:

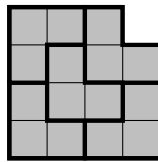


Another way of thinking of a tromino is that it is the geometric figure obtained by taking a square that is composed of four smaller squares and removing one of the smaller squares.

We are going to consider what geometric regions can be covered by trominos all of which have the same size and that do not overlap each other. As a first example, start with a square made up of 16 smaller squares (that is, a square that is “4 by 4”) and remove one small square from a corner of the square:



Can the region that is left be covered by trominos (each made up of three small squares of the same size as the small squares in the region) that do not overlap each other? It can:

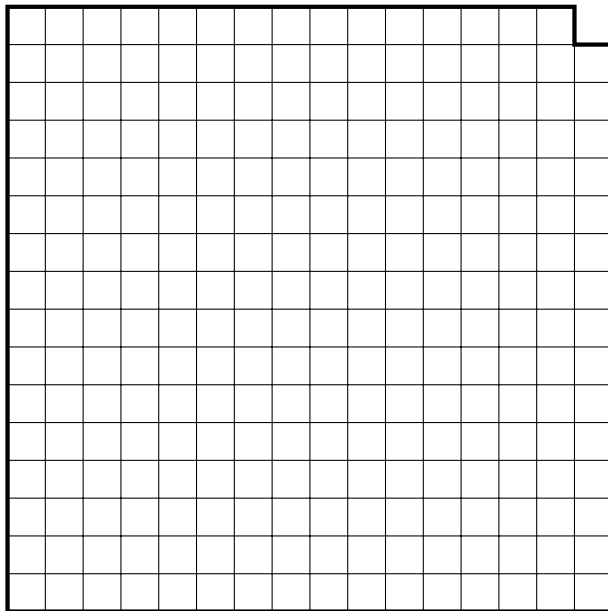


We can use mathematical induction to prove the following.

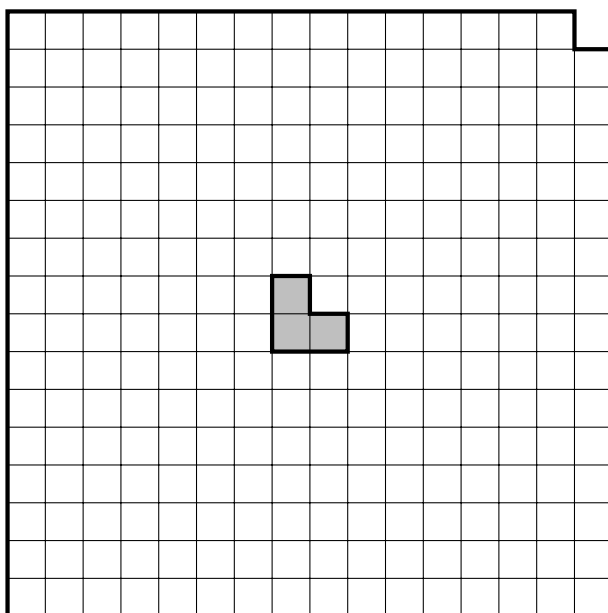
**Theorem 2.1.6.** *For each natural number  $n$ , consider a square consisting of  $2^{2n}$  smaller squares. (That is, a  $2^n \times 2^n$  square.) If one of the smaller squares is removed from a corner of the large square, then the resulting region can be completely covered by trominos (each made up of three small squares of the same size as the small squares in the region) in such a way that the trominos do not overlap.*

*Proof.* To begin a proof by mathematical induction, first note that the theorem is certainly true for  $n = 1$ ; the region obtained after removing a small corner square is a tromino, so it can be covered by one tromino.

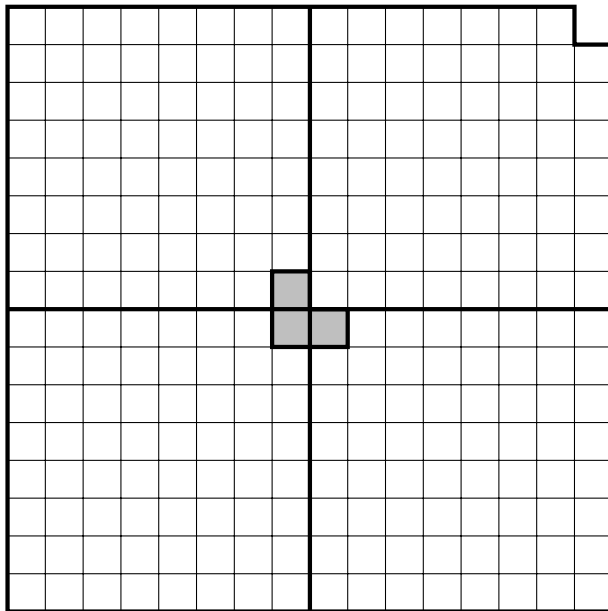
Suppose that the theorem is true for  $n = k$ . That is, we are supposing that if a small corner square is removed from any  $2^k \times 2^k$  square consisting of  $2^{2k}$  smaller squares, then the resulting region can be covered by trominos. The proof will be established by the principle of mathematical induction if we can show that the same result holds for  $n = k + 1$ . Consider, then, any  $2^{k+1} \times 2^{k+1}$  square consisting of smaller squares. Remove a corner square, getting a region that looks like this:



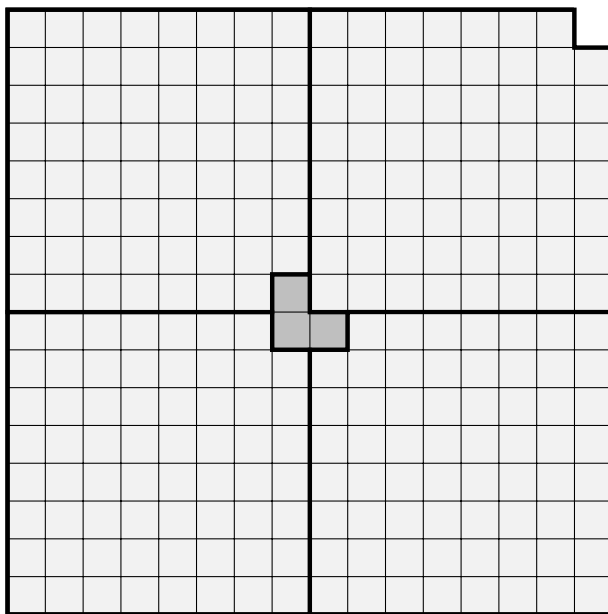
Place a tromino in the middle of this square, as illustrated below.



The region can be divided into four “medium-sized” squares that are each  $2^k \times 2^k$ , like this:



The four “medium-sized” squares of the region are each  $2^k \times 2^k$  and, because of the tromino in the middle, the “medium-sized” squares remaining to be covered each have one corner covered or missing.





By the inductive hypothesis, trominos can be used to cover the rest of each of the four “medium-sized” squares. This leads to a covering of the entire  $2^{k+1} \times 2^{k+1}$  square, thus finishing the proof by mathematical induction.  $\square$

## 2.2 The Principle of Complete Mathematical Induction

There is a variant of the Principle of Mathematical Induction that is sometimes very useful. The basis for this variant is a slightly different characterization of the set of all natural numbers.

**Definition 2.2.1** (Principle of Complete Mathematical Induction). If  $S$  is any set of natural numbers with the properties that:

- A. 1 is in  $S$ , and
- B.  $k+1$  is in  $S$  whenever  $k$  is a natural number and all of the natural numbers from 1 through  $k$  are in  $S$ ,

then  $S$  is the set of all natural numbers.

The informal and formal proofs of the Principle of Complete Mathematical Induction are virtually the same as the proofs of the Principle of (ordinary) Mathematical Induction. First consider the informal proof. If  $S$  is any set of natural numbers with properties A and B of the Principle of Complete Mathematical Induction, then, in particular, 1 is in  $S$ . Since 1 is in  $S$ , it follows from property B that 2 is in  $S$ . Since 1 and 2 are in  $S$ , it follows from property B that 3 is in  $S$ . Since 1, 2, and 3 are in  $S$ , 4 is in  $S$ . And so on. It is suggested that you write out the details of the formal proof of the Principle of Complete Mathematical Induction as a consequence of the well-ordering of the set of natural numbers. Just as for ordinary induction, the Principle of Complete Mathematical Induction can be generalized to begin at any natural number, not just 1.

**Definition 2.2.2** (Generalized Principle of Complete Mathematical Induction). If  $S$  is any set of natural numbers with the properties that:

- A.  $m$  is in  $S$ , and
- B.  $k+1$  is in  $S$  whenever  $k$  is a natural number greater than or equal to  $m$  and all of the natural numbers from  $m$  through  $k$  are in  $S$ ,

then  $S$  contains all natural numbers greater than or equal to  $m$ .

There are many situations to which it is difficult to directly apply the Principle of Mathematical Induction but easy to apply the Principle of Complete Mathematical Induction. One example of such a situation is a very precise proof of the lemma that was required to prove that there is no largest prime number.

**Lemma 2.2.3.** *If  $m$  is a composite natural number, then there is a prime number that is a divisor of  $m$ .*

The following is a statement that clearly implies the above lemma. Note that we employ the convention that a single prime number is a “product of primes” where the product has only one factor.

**Theorem 2.2.4.** *Every natural number other than 1 is a product of prime numbers.*

*Proof.* We prove this theorem using the Generalized Principle of Complete Mathematical Induction starting at 2. Let  $S$  be the set of all  $n$  that are products of primes. It is clear that 2 is in  $S$ , since 2 is a prime. Suppose that every natural number from 2 up to  $k$  is in  $S$ . We must show, in order to apply the Generalized Principle of Complete Mathematical Induction, that  $k + 1$  is in  $S$ .

The number  $k + 1$  cannot be 1. We must therefore show that either it is prime or is a product of primes. If  $k + 1$  is prime, we are done. If  $k + 1$  is not prime, then  $k + 1 = xy$  where each of  $x$  and  $y$  is a natural number strictly between 1 and  $k + 1$ . Thus  $x$  and  $y$  are each at most  $k$ , so, by the inductive hypothesis,  $x$  and  $y$  are both in  $S$ . That is,  $x$  and  $y$  are each either primes or the product of primes. Therefore  $xy$  can be written as a product of primes by writing the product of the primes comprising  $x$  (or  $x$  itself if  $x$  is prime) times the product of the primes comprising  $y$  (or  $y$  itself if  $y$  is prime). Thus by the Generalized Principle of Complete Mathematical Induction starting at 2,  $S$  contains all natural numbers greater than or equal to 2.  $\square$

We now describe an interesting theorem whose statement is a little more difficult to understand. (If you find this theorem too difficult you need not consider it; it won't be used in anything that follows. You might wish to return to it at some later time.)

We begin by describing the case where  $n = 5$ . Suppose there is a pile of 5 stones. We are going to consider the sum of certain sequences of numbers

obtained as follows. Begin one such sequence by dividing the pile into two smaller piles, a pile of 3 stones and a pile of 2 stones. Let the first term in the sum be  $3 \cdot 2 = 6$ . Repeat this process with the pile of 3 stones: divide it into a pile of 2 stones and a pile consisting of 1 stone. Add  $2 \cdot 1 = 2$  to the sum. The pile with 2 stones can be divided into 2 piles of 1 stone each. Add  $1 \cdot 1 = 1$  to the sum. Now go back to the pile of 2 stones created by the first division. That pile can be divided into 2 piles of 1 stone each. Add  $1 \cdot 1 = 1$  to the sum. The total sum that we have is 10.

Let's create another sum in a similar manner but starting a different way. Divide the original pile of 5 stones into a pile of 4 stones and a pile of 1 stone. Begin this sum with  $4 \cdot 1 = 4$ . Divide the pile of 4 stones into two piles of 2 stones each and add  $2 \cdot 2 = 4$  to the sum. The first pile of 2 stones can be divided into two piles of 1 stone each, so add  $1 \cdot 1 = 1$  to the sum. Similarly, divide the second pile of 2 into two piles of 1 each and add  $1 \cdot 1 = 1$  to the sum. The sum we get proceeding in this way is also 10.

Is it a coincidence that we got the same result, 10, for the sums we obtained in quite different ways?

**Theorem 2.2.5.** *For any natural number  $n$  greater than 1, consider a pile of  $n$  stones. Create a sum as follows: divide the given pile of stones into two smaller piles. Let the product of the number of elements in one pile and the number of elements in the other pile be the first term in the sum. Then consider the first of the smaller piles and (unless the smaller pile has only one stone in it) divide that pile into two smaller piles and let the product of the number of stones in those piles be the second term in the sum. Continue dividing, multiplying, and adding terms to the sum in all possible ways. No matter what sequence of divisions into subpiles is used, the total sum is  $n(n-1)/2$ .*

*Proof.* We prove this theorem using Generalized Complete Mathematical Induction beginning with  $n = 2$ . Given any pile of 2 stones, there is only one way to divide it, into two piles of 1 each. Since  $1 \cdot 1 = 1$ , the sum is 1 in this case. Note that  $1 = 2(2-1)/2$ , so the formula holds for the case  $n = 2$ .

Suppose now that the formula holds for all of  $n = 2, 3, 4, \dots, k$ . Consider any pile of  $k+1$  stones. We must show that, for any sequence of divisions, the resulting sum is  $(k+1)(k+1-1)/2$ .

Begin with any division of the pile into two subpiles. Call the number of stones in the subpiles  $x$  and  $y$  respectively. Consider first the situation where  $x = 1$ . Then the first term in the sum is  $1 \cdot y = y$ . The process is continued by dividing the pile of  $y$  stones. By the inductive hypothesis, the sum obtained by

completing the process on a pile of  $y$  stones, is  $y(y-1)/2$ . Thus the total sum for the original pile of  $k+1$  stones in this case is  $y + y(y-1)/2$ . This is equal to

$$\frac{2y + (y^2 - y)}{2} = \frac{y^2 + y}{2} = \frac{y(y+1)}{2}$$

Since  $x = 1$  and  $x + y = k + 1$ ,  $y = k$ . Thus, the formula we obtain for the sum in this case is  $k(k+1)/2$ , which is the correct formula in the case where  $n = k + 1$ .

If  $y = 1$ , the same proof can be given, by simply interchanging the roles of  $x$  and  $y$  in the previous paragraph.

The last, and more interesting, case is when neither  $x$  nor  $y$  is 1. In this case we have the pile of  $k+1$  stones divided into two piles, one consisting of  $x$  stones and the other consisting of  $y$  stones. The first term in the sum is then  $xy$ . Continuing the process will give a sum for the pile of  $k+1$  stones that is equal to  $xy$  plus the sum for the pile of  $x$  stones added to the sum for the pile of  $y$  stones. Thus, using the inductive hypothesis gives the sum for the pile of  $k+1$  stones as  $xy + x(x-1)/2 + y(y-1)/2$ ; call this sum  $s$ . We must show that  $s = (k+1)k/2$ , which is the correct formula for  $n = k + 1$ .

Recall that  $k+1 = x + y$ , or  $x = k+1 - y$ . Using this, we see that

$$\begin{aligned} s &= xy + \frac{x(x-1)}{2} + \frac{y(y-1)}{2} \\ &= \frac{2(k+1-y)y}{2} + \frac{(k+1-y)(k-y)}{2} + \frac{y(y-1)}{2} \\ &= \frac{2ky + 2y - 2y^2}{2} + \frac{k^2 + k - ky - ky - y + y^2}{2} + \frac{y^2 - y}{2} \\ &= \frac{k^2 + k}{2} \\ &= \frac{(k+1)k}{2} \end{aligned}$$

This completes the proof. □

Mathematics is the most precise of subjects. However, human beings are not always so precise; they must be careful not to make mistakes. See if you can figure out what is wrong with the proof of the following obviously false statement.

**False Statement.** *All human beings are the same age.*

*“Proof”*. We will present what, at first glance at least, appears to be a proof of the above statement. We begin by reformulating it as follows: For every natural number  $n$ , every set of  $n$  people consists of people the same age. The assertion that “all human beings are the same age” would clearly follow as the case where  $n$  is the present population of the earth. We proceed by mathematical induction. The case  $n = 1$  is certainly true; a set containing 1 person consists of people the same age. For the inductive step, suppose that every set of  $k$  people consisted of people the same age. Let  $S$  be any set containing  $k + 1$  people. We must show that all the people in  $S$  are the same age as each other.

List the people in  $S$ , as follows:

$$S = \{P_1, P_2, \dots, P_k, P_{k+1}\}.$$

Consider the subset  $L$  of  $S$  consisting of the first  $k$  people in  $S$ ; that is,

$$L = \{P_1, P_2, \dots, P_k\}.$$

Similarly, let  $R$  denote the subset consisting of the last  $k$  elements of  $S$ ; that is,

$$R = \{P_2, \dots, P_k, P_{k+1}\}.$$

The sets  $L$  and  $R$  each contain  $k$  people, and so by the inductive hypothesis, each consists of people who are the same age as each other. In particular, all the people in  $L$  are the same age as  $P_2$ . Also, all the people in  $R$  are the same age as  $P_2$ . But every person in the original set  $S$  is in either  $L$  or  $R$ , so all the people in  $S$  are the same age as  $P_2$ . Therefore,  $S$  consists of people the same age, and the assertion follows by the principle of mathematical induction.  $\square$

What is going on? Is it really true that all people are the same age? Not likely. Is the principle of mathematical induction flawed? Or is there something wrong with the above “proof”?

Clearly there must be something wrong with the “proof”. Please do not read further for at least a few minutes while you try to find the mistake.

Wait a minute, before you read further, please try for a little bit longer to see if you can find the mistake.

If you haven’t been able to find the error yourself, perhaps a hint will help. The proof of the case  $n = 1$  is surely valid; a set with one person in it contains a person with whatever age that person is. What about the inductive step, the going from  $k$  to  $k + 1$ ? For it to be valid, it must apply for every natural number  $k$ . To conclude that an assertion holds for all natural numbers given that it holds for  $n = 1$  requires that its truth for  $n = k + 1$  is implied by its truth for

$n = k$ , for every natural number  $k$ . In fact, there is a  $k$  for which the above derivation of the case  $n = k + 1$  from the case  $n = k$  is not valid. Can you figure out the value of that  $k$ ?

Okay, here's the mistake. Consider the inductive step when  $k = 1$ ; that is, going from 1 to 2. In this case, the set  $S$  would have the form

$$S = \{P_1, P_2\}.$$

Then,  $L = \{P_1\}$  and  $R = \{P_2\}$ .

The set  $L$  does consist of people the same age as each other, as does the set  $R$ . But there is no person who is in both sets. Thus, we cannot conclude that everyone in  $S$  is the same age. This shows that the above "proof" of the inductive step does not hold when  $k = 1$ . In fact, the following is true.

**True Statement.** If every pair of people in a given set of people consists of people the same age, then all the people in the set are the same age.

*Proof.* Let  $S$  be the given set of people; suppose  $S = \{P_1, P_2, \dots, P_n\}$ . For each  $i$  from 2 to  $n$ , the pair

$$\{P_1, P_i\}$$

consists of people the same age, by hypothesis. Thus,  $P_i$  and  $P_1$  are the same age for every  $i$ , so every person in  $S$  is the same age as  $P_1$ . Hence, everyone in  $S$  is the same age.  $\square$

## 2.3 Problems

### Basic Exercises

1. Prove by induction that, for every natural number  $n$ ,

$$1 \cdot 2 + 2 \cdot 3 + 3 \cdot 4 + \cdots + n \cdot (n + 1) = \frac{n(n + 1)(n + 2)}{3}.$$

2. Prove by induction that, for every natural number  $n$ ,

$$\frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} + \cdots + \frac{1}{n \cdot (n + 1)} = \frac{n}{n + 1}.$$

3. Prove by induction that, for every natural number  $n$ ,

$$2 + 2^2 + 2^3 + \cdots + 2^n = 2^{n+1} - 2.$$

4. Prove by induction that, for every natural number  $n$ ,

$$\frac{1}{2} + \frac{2}{2^2} + \frac{3}{2^3} + \cdots + \frac{n}{2^n} = 2 - \frac{n+2}{2^n}.$$

### Interesting Problems

5. Prove the following statement by induction: For every natural number  $n$ , every set with  $n$  elements has  $2^n$  subsets.
6. Prove by induction that, for every natural number  $n$ ,

$$1 + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}} + \cdots + \frac{1}{\sqrt{n}} < 2\sqrt{n}.$$

7. Prove the following generalization of Theorem 2.1.6:

**Theorem.** *For each natural number  $n$ , consider a square consisting of  $2^{2n}$  smaller squares. (That is, a  $2^n \times 2^n$  square.) If any one of the smaller squares is removed from the large square (not necessarily from the corner), then the resulting region can be completely covered by trominos (each made up of three small squares of the same size as the small squares in the region) in such a way that the trominos do not overlap.*

8. Prove by mathematical induction that 3 divides  $n^3 + 2n$  for every  $n$ .
9. Use mathematical induction to prove that  $3^n > n^2$  for every  $n$  larger than 2.
10. Show that for every natural number  $n$  larger than 1 and every real number  $r$  different from 1,  $1 + r + r^2 + \cdots + r^{n-1} = \frac{r^n - 1}{r - 1}$ .

### Challenging Problems

11. Prove the Principle of Complete Mathematical Induction using the well-ordering of the natural numbers.
12. Prove the well-ordering of the natural numbers using the Principle of Complete Mathematical Induction.
13. Define the  $n^{\text{th}}$  Fermat number,  $F_n = 2^{2^n} + 1$  for  $n \in \mathbb{N}$ . The first few Fermat numbers are  $F_0 = 3$ ,  $F_1 = 5$ ,  $F_2 = 17$ ,  $F_3 = 257$ . Prove by induction that  $F_0 \cdot F_1 \cdots F_{n-1} + 2 = F_n$  for  $n \geq 1$ .

14. The sequence of Fibonacci numbers is defined as follows:  $x_1 = 1$ ,  $x_2 = 1$ , and, for natural numbers  $n > 2$ ,  $x_n = x_{n-1} + x_{n-2}$ . Prove that

$$x_n = \frac{1}{\sqrt{5}} \left[ \left( \frac{1 + \sqrt{5}}{2} \right)^n - \left( \frac{1 - \sqrt{5}}{2} \right)^n \right]$$



## Chapter 3

# Modular Arithmetic

Consider the number obtained by adding 3 to the number consisting of 2 to the power 3,000,005; that is, consider the number  $3 + 2^{3,000,005}$ . This is a very big number. No computer that presently exists, or is even conceivable, would have sufficient capacity to display all the digits in that number.

When that huge number is divided by 7, what remainder is left? You can't use your calculator, or any computer, because they can't count that high. However, this and similar questions will be easily answered using a kind of "calculus" of divisibility and remainders that is called "modular arithmetic". Another application of this study will be to prove that a natural number is divisible by 9 if and only if the sum of its digits is divisible by 9. The mathematics that we develop in this chapter has numerous other consequences, including, for example, providing the basis for an extremely powerful method for sending coded messages (see Chapter 6).

### 3.1 The Basics

Recall that we say that the integer  $n$  is *divisible* by the integer  $m$  if there exists an integer  $q$  such that  $n = qm$ . In this situation, we also say that  $m$  is a *divisor* of  $n$ , or  $m$  is a *factor* of  $n$ .

The fundamental definition for modular arithmetic is the following.

**Definition 3.1.1.** For any fixed natural number  $m$  greater than 1, we say that the integer  $a$  is *congruent to the integer  $b$  modulo  $m$*  if  $a - b$  is divisible by  $m$ . We use the notation  $a \equiv b \pmod{m}$  to denote this relationship. The number  $m$  in this notation is called the *modulus*.

Here are a few examples:

$$14 \equiv 8 \pmod{3}$$

$$252 \equiv 127 \pmod{5}$$

$$3 \equiv -11 \pmod{7}$$

Congruence shares an important property with equality.

**Theorem 3.1.2.** *If  $a \equiv b \pmod{m}$  and  $b \equiv c \pmod{m}$ , then  $a \equiv c \pmod{m}$ .*

*Proof.* The hypothesis states that there is an integer  $t$  such that  $a - b = tm$  and an integer  $s$  such that  $b - c = sm$ . Thus  $a - c = a - b + b - c = tm + sm = (t + s)m$ . By definition, then,  $a \equiv c \pmod{m}$ .  $\square$

The theorem just proven shows that we can replace numbers in a congruence modulo  $m$  by any numbers congruent to them modulo  $m$ .

Although the modulus  $m$  must be bigger than 1, there is no such restriction on the integers  $a$  and  $b$ ; they could even be negative. In the case where  $a$  and  $b$  are positive integers, the relationship  $a \equiv b \pmod{m}$  can be expressed in more familiar terms.

**Theorem 3.1.3.** *In the case where  $a$  and  $b$  are non-negative integers, the relationship  $a \equiv b \pmod{m}$  is equivalent to  $a$  and  $b$  leaving equal remainders upon division by  $m$ .*

*Proof.* Consider dividing  $m$  into  $a$ ; if it “goes in evenly”, then  $m$  is a divisor of  $a$  and the remainder  $r$  is 0. In any case, there are non-negative integers  $q$  and  $r$  such that  $a = qm + r$ ;  $q$  is the quotient and  $r$  is the remainder. The non-negative number  $r$  is less than  $m$ , since it is the remainder. Similarly, divide  $b$  by  $m$ , getting  $b = q_0m + r_0$ . This yields

$$a - b = (qm + r) - (q_0m + r_0) = m(q - q_0) + (r - r_0).$$

If  $r = r_0$ , then  $a - b$  is obviously divisible by  $m$ , so  $a \equiv b \pmod{m}$ . Conversely, if  $r$  is not equal to  $r_0$ , note that  $r - r_0$  cannot be a multiple of  $m$ . (This follows from the fact that  $r$  and  $r_0$  are both non-negative numbers which are strictly less than  $m$ .) Thus  $a - b$  is a multiple of  $m$  plus a number that is not a multiple of  $m$ , and therefore is not a multiple of  $m$ . That is, it is not the case that  $a \equiv b \pmod{m}$ .  $\square$

A special case of the above theorem is that a positive number is congruent modulo  $m$  to the remainder it leaves upon division by  $m$ . The possible remainders upon division by a given natural number  $m$  are  $0, 1, 2, \dots, m-1$ .

**Theorem 3.1.4.** *For a given modulus  $m$ , each integer is congruent to exactly one of the numbers in the set  $\{0, 1, 2, \dots, m-1\}$ .*

*Proof.* Let  $a$  be an integer. If  $a$  is positive, the result follows from the fact, discussed above, that  $a$  is congruent to the remainder it leaves upon division by  $m$ . If  $a$  is not positive, choosing  $t$  big enough would make  $tm + a$  positive. For such a  $t$ ,  $tm + a$  is congruent to the remainder it leaves upon division by  $m$ . But also  $tm + a \equiv a \pmod{m}$ . It follows from Theorem 3.1.2 that  $a$  is congruent to the remainder that  $tm + a$  leaves upon division by  $m$ . An integer cannot be congruent to two different numbers in the given set  $\{0, 1, 2, \dots, m-1\}$ , since no two numbers in the set are congruent to each other (the absolute value of the difference of two numbers in the set is less than  $m$  and hence could be divisible by  $m$  only if the absolute value is 0, in which case the two numbers are the same).  $\square$

For a fixed modulus, congruences have some properties that are similar to those for equations.

**Theorem 3.1.5.** *If  $a \equiv b \pmod{m}$  and  $c \equiv d \pmod{m}$ , then*

$$(i) \quad (a + c) \equiv (b + d) \pmod{m}, \text{ and}$$

$$(ii) \quad ac \equiv bd \pmod{m}.$$

*Proof.* To prove (i), note that  $a \equiv b \pmod{m}$  means that  $a - b = sm$  for some integer  $s$ . Similarly,  $c - d = tm$  for some integer  $t$ . The conclusion we are trying to establish is equivalent to the assertion that  $(a + c) - (b + d)$  is a multiple of  $m$ . But  $(a + c) - (b + d) = (a - b) + (c - d)$ , which is equal to  $sm + tm = (s + t)m$ , so the result follows.

To prove (ii), note that, from  $a - b = sm$  and  $c - d = tm$ , we get  $a = b + sm$  and  $c = d + tm$ , so

$$ac = (b + sm)(d + tm) = bd + btm + smd + stm^2.$$

It follows that  $ac - bd = m(bt + sd + stm)$ , so  $ac - bd$  is a multiple of  $m$  and the result is established.  $\square$

Thus congruences are similar to equations in that you can add congruent numbers to both sides of a congruence or multiply both sides of a congruence by congruent numbers and preserve the congruence, as long as all the congruences are with respect to the same fixed modulus.

For example, since  $3 \equiv 28 \pmod{5}$  and  $17 \equiv 2 \pmod{5}$ , it follows that  $20 \equiv 30 \pmod{5}$  and  $51 \equiv 56 \pmod{5}$ .

Here is another example:  $8 \equiv 1 \pmod{7}$ , so  $8^2 \equiv 1^2 \pmod{7}$ , or  $8^2 \equiv 1 \pmod{7}$ . It follows that  $8^2 \cdot 8 \equiv 1 \cdot 1 \pmod{7}$ , or  $8^3 \equiv 1 \pmod{7}$ . In fact, all positive integer powers of 8 are congruent to 1 modulo 7. This is a special case of the next result.

**Theorem 3.1.6.** *If  $a \equiv b \pmod{m}$  then, for every natural number  $n$ ,  $a^n \equiv b^n \pmod{m}$ .*

*Proof.* We use mathematical induction. The case  $n = 1$  is the hypothesis. Assume that the result is true for  $n = k$ ; that is,  $a^k \equiv b^k \pmod{m}$ . Since  $a \equiv b \pmod{m}$ , using part (ii) of Theorem 3.1.5 gives  $a \cdot a^k \equiv b \cdot b^k \pmod{m}$ , or  $a^{k+1} \equiv b^{k+1} \pmod{m}$ .  $\square$

## 3.2 Some Applications

We can use the above to easily solve the problem that we mentioned at the beginning of this chapter: what is the remainder left when  $3 + 2^{3,000,005}$  is divided by 7?

First note that  $2^3 = 8$  is congruent to 1 modulo 7. Therefore, by the theorem we just proved,  $(2^3)^{1,000,000}$  is congruent to  $1^{1,000,000}$ , which is 1 modulo 7. Thus  $2^{3,000,000} \equiv 1 \pmod{7}$ . Since  $2^5 \equiv 4 \pmod{7}$  and  $2^{3,000,005} = 2^{3,000,000} \cdot 2^5$ , it follows that  $2^{3,000,005} \equiv 4 \pmod{7}$ . Thus,  $3 + 2^{3,000,005} \equiv (4 + 3) \pmod{7} \equiv 0 \pmod{7}$ . Therefore, 7 is a divisor of  $3 + 2^{3,000,005}$ . In other words, the remainder that is left when  $3 + 2^{3,000,005}$  is divided by 7 is 0.

Let's look at the next question we mentioned at the beginning of this chapter, the relationship between divisibility by 9 of a number and divisibility by 9 of the sum of the digits of the number. To illustrate, we begin with a particular example. Consider the number 73,486. What that really means is

$$7 \cdot 10^4 + 3 \cdot 10^3 + 4 \cdot 10^2 + 8 \cdot 10 + 6$$

Note that 10 is congruent to 1 modulo 9, so  $10^n$  is congruent to 1 modulo 9 for every natural number  $n$ . Thus,  $a \cdot 10^n \equiv a \pmod{9}$  for every  $a$  and every  $n$ . It

follows that  $(7 \cdot 10^4 + 3 \cdot 10^3 + 4 \cdot 10^2 + 8 \cdot 10 + 6)$  is congruent to  $(7 + 3 + 4 + 8 + 6)$  modulo 9. Thus, the number 73,486 and the sum of its digits are congruent to each other modulo 9 and therefore leave the same remainders upon division by 9. The general theorem is the following.

**Theorem 3.2.1.** *Every natural number is congruent to the sum of its digits modulo 9. In particular, a natural number is divisible by 9 if and only if the sum of its digits is divisible by 9.*

*Proof.* If  $n$  is a natural number, then we can write it in terms of its digits in the form  $a_k a_{k-1} a_{k-2} \dots a_1 a_0$ , where each  $a_i$  is one of 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 (with  $a_k \neq 0$ ). That is,  $a_0$  is the digit in the “1’s place”,  $a_1$  is the digit in the “10’s place”,  $a_2$  is the digit in the “100’s place”, and so on. (In the previous example,  $n$  was the number 73,486, so in that case  $a_4 = 7$ ,  $a_3 = 3$ ,  $a_2 = 4$ ,  $a_1 = 8$ , and  $a_0 = 6$ .) This really means that

$$n = a_k \cdot 10^k + a_{k-1} \cdot 10^{k-1} + a_{k-2} \cdot 10^{k-2} + \dots + a_2 \cdot 10^2 + a_1 \cdot 10 + a_0.$$

As shown above,  $10 \equiv 1 \pmod{9}$  implies  $10^i \equiv 1 \pmod{9}$  for every positive integer  $i$ . Therefore,  $n$  is congruent to  $(a_k + a_{k-1} + a_{k-2} + \dots + a_1 + a_0)$  modulo 9. Thus,  $n$  and the sum of its digits leave the same remainders upon division by 9. In particular,  $n$  is divisible by 9 if and only if the sum of its digits is divisible by 9.  $\square$

Congruence equations with small moduli can easily be solved by just trying all possibilities.

**Example 3.2.2.** Solve the congruence  $5x \equiv 11 \pmod{19}$ .

*Proof.* If there is a solution, then there is a solution within the set  $\{0, 1, 2, \dots, 18\}$  (by 3.1.4). If  $x = 0$ , then  $5x = 0$  so 0 is not a solution. Similarly, for  $x = 1$ ,  $5x = 5$ , for  $x = 2$ ,  $5x = 10$ , for  $x = 3$ ,  $5x = 15$ , and for  $x = 4$ ,  $5x = 20$ . None of these are congruent to 11 mod 19, so we have not yet found a solution. However, when  $x = 6$ ,  $5x = 30$ , which is congruent to 11 mod 19. Thus  $x \equiv 6 \pmod{19}$  is a solution of the congruence.  $\square$

**Example 3.2.3.** Show that there is no solution to the congruence  $x^2 \equiv 3 \pmod{5}$ .

*Proof.* If  $x = 0$ , then  $x^2 = 0$ ; if  $x = 1$ , then  $x^2 = 1$ ; if  $x = 2$ , then  $x^2 = 4$ ; if  $x = 3$ , then  $x^2 = 9$  which is congruent to 4 mod 5; and if  $x = 4$ , then  $x^2 = 16$  which is congruent to 1 mod 5. By Theorem remainder-theorem, the congruence has no solution.  $\square$

### 3.3 Problems

#### Basic Exercises


1. Find a solution  $x$  to each of the following congruences. (“Solution” means “integer solution”.)
  - (a)  $2x \equiv 7 \pmod{11}$
  - (b)  $7x \equiv 4 \pmod{11}$
  - (c)  $x^5 \equiv 3 \pmod{4}$
2. For each of the following congruences, either find a solution or prove that no solution exists. (“Solution” means “integer solution”.)
  - (a)  $39x \equiv 13 \pmod{5}$
  - (b)  $95x \equiv 13 \pmod{5}$
  - (c)  $x^2 \equiv 3 \pmod{6}$
  - (d)  $5x^2 \equiv 12 \pmod{8}$
  - (e)  $4x^3 + 2x \equiv 7 \pmod{5}$

#### Interesting Problems

3. Find the remainder when:
  - (a)  $3^{2463}$  is divided by 8.
  - (b)  $2^{923}$  is divided by 15.
  - (c)  $243^{101}$  is divided by 8.
  - (d)  $5^{2001} + (27)!$  is divided by 8.
  - (e)  $(-8)^{4124} + 6^{3101} + 7^5$  is divided by 3.
  - (f)  $7^{103} + 6^{5409}$  is divided by 3.
  - (g)  $5! \cdot 181 - 866 \cdot 332$  is divided by 6.
4. Is  $2^{598} + 3$  divisible by 15?
5. Show that there is no digit  $a$  such that the number  $2794a1$  is divisible by 8.
6. Determine whether or not  $17^{2492} + 25^{376} + 5^{782}$  is divisible by 3.

7. Determine whether or not the following congruence has an integer solution:  
 $5^x + 4 \equiv 5 \pmod{100}$
8. Prove that  $n^2 - 1$  is divisible by 8 for every odd integer  $n$ .
9. Prove that a natural number is divisible by 3 if and only if the sum of its digits is divisible by 3.
10. Prove that  $x^5 \equiv x \pmod{10}$  for every integer  $x$ . (This shows that  $x^5$  and  $x$  always have the same units' digits for every integer  $x$ .)
11. Prove that 42 divides  $n^7 - n$  for every positive integer  $n$ .
12. Prove that 21 divides  $3n^7 + 7n^3 + 11n$  for every natural number  $n$ .
13. Prove that a natural number that is congruent to 2 modulo 3 has a prime factor that is congruent to 2 modulo 3.
14. Suppose a number is written in decimal notation as  $abba$ , where  $a$  and  $b$  are integers between 1 and 9. Prove that this number is divisible by 11.

### Challenging Problems

15. Suppose that  $p$  is a prime number and  $p$  does not divide  $a$ . Prove that the congruence  $ax \equiv 1 \pmod{p}$  has a solution. (This proves that  $a$  has a “multiplicative inverse” modulo  $p$ .)
16. If  $m$  is a natural number greater than 1 and is not prime, then we know that  $m = ab$  for some  $1 < a, b < m$ . Show that there is no integer  $x$  such that  $ax \equiv 1 \pmod{m}$ . (That is,  $a$  has no “multiplicative inverse” modulo  $m$ .)
17. Prove that 5 divides  $3^{2n+1} + 2^{2n+1}$  for every natural number  $n$ .
18. Prove that 7 divides  $8^{2n+1} + 6^{2n+1}$  for every natural number  $n$ .
19. Prove that 133 divides  $11^{n+1} + 12^{2n+1}$  for every natural number  $n$ . 
20. A natural number  $r$  between 0 and  $p - 1$  is called a *quadratic residue modulo  $p$*  if there is an integer  $x$  such that  $x^2 \equiv r \pmod{p}$ . Determine all the quadratic residues modulo 11.
21. Show that there do not exist natural numbers  $x$  and  $y$  such that  $x^2 + y^2 = 4003$ . [Hint: Determine which of the numbers  $(0, 1, 2, 3)$  can be congruent to  $x^2 \pmod{4}$ . Do the same for  $y^2$  and for  $x^2 + y^2$ .]

- 
22. Let  $f(x)$  be a polynomial with integer coefficients. (That is, there exists a natural number  $n$  and integers  $a_i$  such that  $f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$ .) Let  $a$ ,  $k$ , and  $m$  be integers with  $m > 1$ . Suppose that  $f(a) \equiv k \pmod{m}$ . Prove that  $f(a+m) \equiv k \pmod{m}$ .
23. Show that the polynomial  $p(x) = x^2 - x + 41$  takes prime values for  $x = (0, 1, 2, \dots, 40)$ .
24. Show that there does not exist any polynomial such that  $p(x)$  is a prime number for all natural numbers  $x$  (except for constant polynomials).
25. Discover and prove a theorem determining whether a natural number is divisible by 11, in terms of its digits.
26. Prove that there are an infinite number of primes of the form  $4k + 3$  with  $k$  a natural number. [Hint: If  $p_1, p_2, \dots, p_n$  are  $n$  such primes, show that  $4(p_1 \cdot p_2 \cdots p_n) - 1$  has at least one prime divisor of the given form.]
27. Prove that there are an infinite number of primes of the form  $6k + 5$  with  $k$  a natural number.



## Chapter 4

# The Fundamental Theorem of Arithmetic

Is  $13^{217} \cdot 37^{92} \cdot 41^{15} = 19^{111} \cdot 29^{145} \cdot 43^{12} \cdot 47^5$ ?

We have seen that every natural number greater than 1 is either a prime or a product of primes. The above equation, if it was an equation, would express a number in two different ways as a product of primes. Does the representation of a natural number as a product of primes have to be unique? The answer is obviously “no” in one sense. For example,  $6 = 3 \cdot 2 = 2 \cdot 3$ . Thus, the same number can be written in two different ways as a product of primes if we consider different orders as “different ways”. But suppose that we don’t consider the ordering; must the factorization of a natural number into a product of primes be unique except for the order? For example, could the above equation hold? In fact, every natural number other than 1 has a factorization into a product of primes and the factorization is unique except for the order. This result is so important that it is called the *Fundamental Theorem of Arithmetic*. We will give two proofs. The second proof requires a little more development and will be given later. The first proof is short but tricky.

### 4.1 Proof of the Fundamental Theorem of Arithmetic

In order to simplify the statement of the Fundamental Theorem of Arithmetic, we use the expression “a product of primes” to include the case of a single prime number (as we did in Theorem 2.2.4).

**Theorem 4.1.1** (The Fundamental Theorem of Arithmetic). *Every natural*

*number greater than 1 can be written as a product of primes, and the expression of a number as a product of primes is unique except for the order of the factors.*

*Proof.* We have already established that every natural number greater than 1 can be written as a product of primes (see Theorem 2.2.4). That was the easy part of the Fundamental Theorem of Arithmetic; the harder part is the uniqueness. The proof of uniqueness that we present below is a proof by contradiction. That is, we will assume that there is a natural number with more than one representation as a product of primes and derive a contradiction from this assumption, thereby showing that this assumption is incorrect.

Suppose, then, that there is at least one natural number with at least two different representations as a product of primes. By the well-ordering of the natural numbers (Proposition 2.1.2), there would then be a smallest natural number with that property (that is, the smallest natural number that has at least two different such representations). Let  $N$  be that smallest such number. Write out two different factorizations of  $N$ :

$$N = p_1 p_2 \cdots p_r = q_1 q_2 \cdots q_s$$

where each of the  $p_i$  and the  $q_j$  are primes (there can be repetitions of the same prime). We first claim that no  $p_i$  could be equal to any  $q_j$ . This follows from the fact that  $N$  is the smallest number with a non-unique representation, for if  $p_i = q_j$  for some  $i$  and  $j$ , that common factor could be divided from both sides of the equation, leaving a smaller number that has at least two different factorizations. Thus, no  $p_i$  is equal to any  $q_j$ .

Since  $p_1$  is different from  $q_1$ , one of  $p_1$  and  $q_1$  is less than the other; suppose that  $p_1$  is less than  $q_1$ . (If  $q_1$  is less than  $p_1$ , the same proof could be repeated by simply interchanging the  $p$ 's and  $q$ 's.) Define  $M$  by

$$M = N - (p_1 q_2 \cdots q_s)$$

Then  $M$  is a natural number that is less than  $N$ . Substituting  $(p_1 p_2 \cdots p_r)$  for  $N$  gives

$$M = (p_1 p_2 \cdots p_r) - (p_1 q_2 \cdots q_s) = p_1 [(p_2 \cdots p_r) - (q_2 \cdots q_s)],$$

from which it follows that  $p_1$  divides  $M$ . In particular,  $M$  is not 1. Since  $M$  is less than  $N$ ,  $M$  has a unique factorization into primes.

Substituting the product  $(q_1 q_2 \cdots q_s)$  for  $N$  in the definition of  $M$  gives a different expression:

$$M = (q_1 q_2 \cdots q_s) - (p_1 q_2 \cdots q_s) = (q_1 - p_1)(q_2 \cdots q_s).$$

The unique factorization of  $M$  into primes can thus be obtained by writing the unique factorization of  $(q_1 - p_1)$  followed by the product  $(q_2 \cdots q_s)$ . On the other hand, the fact that  $p_1$  is a divisor of  $M$  implies that  $p_1$  must appear in the factorization of  $M$  into primes. Since  $p_1$  is distinct from each of  $\{q_2, \dots, q_s\}$ , it follows that  $p_1$  must occur in the factorization of  $(q_1 - p_1)$  into primes. Thus,  $q_1 - p_1 = p_1 k$  for some integer  $k$ . It follows that  $q_1 = p_1 + p_1 k = p_1(1 + k)$ , which shows that  $q_1$  is divisible by  $p_1$ . Since  $p_1$  and  $q_1$  are distinct primes, this is impossible.

Hence the assumption that there is a natural number with two distinct factorizations leads to a contradiction, so factorizations into primes are unique.  $\square$

The Fundamental Theorem of Arithmetic gives a so-called “canonical form” for expressing each natural number greater than 1.

**Corollary 4.1.2.** *Every natural number  $n$  greater than 1 has a canonical factorization into primes; that is,  $n$  has a unique representation in the form  $n = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_n^{\alpha_n}$ , where each  $p_i$  is a prime,  $p_i$  is less than  $p_{i+1}$  for each  $i$ , and each  $\alpha_i$  is a natural number.*

*Proof.* To see this, simply factor the given number as a product of primes and then collect all occurrences of the smallest prime together, then all the occurrences of the next smallest prime, and so on.  $\square$

For example, the canonical form of 60,368 is  $2^4 \cdot 7^3 \cdot 11$ . The canonical form of 19 is simply 19.

As we will see, the following corollary of the Fundamental Theorem of Arithmetic is very useful. (If the corollary below is independently established, then it is easy to derive the Fundamental Theorem of Arithmetic from it. In fact, most presentations of the proof of the Fundamental Theorem of Arithmetic use this approach rather than the shorter but trickier proof that we gave above. We will present such a proof later.)

**Corollary 4.1.3.** *If  $p$  is a prime number and  $a$  and  $b$  are natural numbers such that  $p$  divides  $ab$ , then  $p$  divides at least one of  $a$  and  $b$ . (That is, if a prime divides a product then it divides at least one of the factors.)*

*Proof.* Since  $p$  divides  $ab$ , there is some natural number  $d$  such that  $ab = pd$ . The unique factorization of  $ab$  into primes therefore contains the prime  $p$  and all the primes that divide  $d$ . On the other hand,  $a$  and  $b$  each have unique

factorizations into primes. Let the canonical factorization of  $a$  be  $q_1^{\alpha_1} q_2^{\alpha_2} \cdots q_m^{\alpha_m}$  and of  $b$  be  $r_1^{\beta_1} r_2^{\beta_2} \cdots r_n^{\beta_n}$ . Then

$$ab = (q_1^{\alpha_1} q_2^{\alpha_2} \cdots q_m^{\alpha_m})(r_1^{\beta_1} r_2^{\beta_2} \cdots r_n^{\beta_n}).$$

Since the factorization of  $ab$  into primes is unique,  $p$  must occur either as one of the  $q_i$ 's, in which case  $p$  divides  $a$ , or as one of the  $r_j$ 's, in which case  $p$  divides  $b$ . Thus  $p$  divides at least one of  $a$  and  $b$ , and the corollary is established.  $\square$

It should be noted that this corollary is not generally true for divisors that are not prime. For example, 18 divides  $3 \cdot 12$ , but 18 does not divide 3 and 18 does not divide 12.

## 4.2 Problems

### Basic Exercises

- Find the canonical factorization into primes of each of the following:
  - 72
  - 47
  - 625
  - $122 \cdot 54$
  - 112
  - 224
  - $112 + 224$
- Show that if  $p$  is a prime number and  $a_1, a_2, \dots, a_n$  are natural numbers such that  $p$  divides  $a_1 \cdot a_2 \cdots a_n$ , then  $p$  divides  $a_i$  for at least one  $a_i$ .
- Show that if  $p$  is a prime number and  $a$  and  $n$  are natural numbers such that  $p$  divides  $a^n$ , then  $p$  divides  $a$ .

### Interesting Problems

- Find the smallest natural numbers  $x$  and  $y$  such that:

(a)  $7^2x = 5^3y$

(b)  $2^5x = 10^2y$

(c)  $127x = 54y$

### Challenging Problems

5. Prove that the natural number  $m$  greater than 1 is prime if it has the property that  $m$  divides at least one of  $a$  and  $b$  whenever it divides  $ab$ .
6. Prove that  $x^2 \equiv 1 \pmod{p}$  implies  $x \equiv 1 \pmod{p}$  or  $x \equiv (p-1) \pmod{p}$  for every prime  $p$ .

## Chapter 5

# Fermat's Theorem and Wilson's Theorem

We've seen that we can add or multiply “both sides” of a congruence by congruent numbers and the result will be a congruence (see Theorem 3.1.5). What about dividing both sides of a congruence by the same natural number? For the result to have a possibility of being a congruence, the divisor must divide evenly into both sides of the congruence so that the result involves only integers, not fractions (congruences are only defined for integers). On the other hand, even that condition is not sufficient to ensure that the result will be a congruence. For example,  $6 \cdot 2$  is congruent to  $6 \cdot 1$  modulo 3, but 2 is not congruent to 1 modulo 3. This is not a surprising example, since 6 is congruent to 0 modulo 3, so “dividing both sides” of the above congruence by 6 is like dividing by 0, which gives wrong results for equations as well. However, there are also examples where dividing both sides of a congruence by a number that is not congruent to 0 leads to results that are not congruent. For example,  $12 \cdot 3$  is congruent to  $24 \cdot 3$  modulo 9, but 12 is not congruent to 24 modulo 9, in spite of the fact that 3 is not congruent to 0 modulo 9.

### 5.1 Fermat's Theorem

There are important cases in which we can divide both sides of a congruence and be assured that the result is a congruence.

**Theorem 5.1.1.** *If  $p$  is a prime and  $a$  is not divisible by  $p$ , and if  $ab \equiv ac \pmod{p}$ , then  $b \equiv c \pmod{p}$ . (That is, we can divide both of a congruence*

*modulo a prime by any natural number that divides both sides of the congruence and is not divisible by the prime.)*

*Proof.* We are given that  $p$  divides  $ab - ac$ . This is the same as saying that  $p$  divides  $a(b - c)$ . Corollary 4.1.3 shows that, since  $p$  divides  $a(b - c)$ ,  $p$  must also divide either  $a$  or  $b - c$ . Since the hypothesis states that  $a$  is not divisible by  $p$ , this implies that  $b - c$  must be divisible by  $p$ . That is the same as saying  $b \equiv c \pmod{p}$ .  $\square$

Consider any given prime number  $p$ . The possible remainders when a natural number is divided by  $p$  are the numbers  $\{0, 1, \dots, p-1\}$ . No two of these numbers are congruent to each other and by Theorem 3.1.4, every natural number (in fact, every integer) is congruent modulo  $p$  to one of those numbers. An integer is divisible by  $p$  if and only if it is congruent to 0 modulo  $p$ . Thus, each integer that is not divisible by  $p$  is congruent to exactly one of the numbers in the set  $\{1, 2, \dots, p-1\}$ . This is the basis for the proof of the following beautiful and very useful theorem.

**Theorem 5.1.2** (Fermat's Theorem). *If  $p$  is a prime number and  $a$  is any natural number that is not divisible by  $p$ , then  $a^{p-1} \equiv 1 \pmod{p}$ .*

*Proof.* Let  $p$  be any prime number and let  $a$  be any natural number that is not divisible by  $p$ . Consider the set of numbers  $\{a \cdot 1, a \cdot 2, \dots, a \cdot (p-1)\}$ . First note that no two of those numbers are congruent to each other, for if  $am \equiv an \pmod{p}$ , then, by Theorem 5.1.1,  $m \equiv n \pmod{p}$ . Since no two of the numbers in the set  $\{1, 2, \dots, p-1\}$  are congruent to each other, this shows that the same is true of numbers in the set  $\{a \cdot 1, a \cdot 2, \dots, a \cdot (p-1)\}$ . Also note that each of the numbers in the set  $\{a \cdot 1, a \cdot 2, \dots, a \cdot (p-1)\}$  is congruent to one of the numbers in  $\{1, 2, \dots, p-1\}$  since no number in either set is divisible by  $p$ . Thus, the numbers in the set  $\{a \cdot 1, a \cdot 2, \dots, a \cdot (p-1)\}$  are congruent, in some order, to the numbers in the set  $\{1, 2, \dots, p-1\}$ . This implies that the product of all of the numbers in the set  $\{a \cdot 1, a \cdot 2, \dots, a \cdot (p-1)\}$  is congruent modulo  $p$  to the product of all the numbers in  $\{1, 2, \dots, p-1\}$ . Thus,  $a \cdot 1 \cdot a \cdot 2 \cdots a \cdot (p-1)$  is congruent to  $1 \cdot 2 \cdot 3 \cdots (p-1)$  modulo  $p$ . Since the number  $a$  occurs  $p-1$  times in this congruence, this yields  $a^{p-1}(1 \cdot 2 \cdot 3 \cdots (p-1)) \equiv (1 \cdot 2 \cdot 3 \cdots (p-1)) \pmod{p}$ . Clearly,  $p$  does not divide  $1 \cdot 2 \cdot 3 \cdots (p-1)$  (by repeated application of Corollary 4.1.3). Thus, we can “divide” both sides of the above congruence by  $1 \cdot 2 \cdot 3 \cdots (p-1)$ , yielding  $a^{p-1} \equiv 1 \pmod{p}$ .  $\square$

As we shall see, Fermat's Theorem has important applications, including in establishing a method for sending coded messages. It is also sometimes useful

to apply Fermat's theorem to specific cases. For example,  $88^{100} - 1$  is divisible by 101. (Don't try to verify this on your calculator!)

The following corollary of Fermat's Theorem is sometimes useful since it doesn't require that  $a$  not be divisible by  $p$ .

**Corollary 5.1.3.** *If  $p$  is a prime number and  $a$  is any natural number, then  $a^p \equiv a \pmod{p}$ .*

*Proof.* If  $p$  does not divide  $a$ , then Fermat's Theorem (fermatsthm) states that  $a^{p-1} \equiv 1 \pmod{p}$ . Multiplying both sides of this congruence by  $a$  gives the result in this case. On the other hand, if  $p$  does divide  $a$ , then  $p$  also divides  $a^p$ , so  $a^p$  and  $a$  are both congruent to 0 mod  $p$ . □

**Definition 5.1.4.** A *multiplicative inverse modulo  $p$*  for a natural number  $a$  is a natural number  $b$  such that  $ab \equiv 1 \pmod{p}$ .

Note that if two numbers have the same multiplicative inverse, they are congruent to each other. For suppose  $ab \equiv 1 \pmod{p}$  and  $cb \equiv 1 \pmod{p}$ . Then multiplying the second congruence on the right by  $a$  yields  $cba \equiv a \pmod{p}$  and, since  $ba \equiv 1 \pmod{p}$ , this gives  $c \equiv a \pmod{p}$ .

Fermat's Theorem is one way of showing that all natural numbers that are not multiples of a given prime  $p$  have multiplicative inverses modulo  $p$ .

**Corollary 5.1.5.** *If  $p$  is a prime and  $a$  is a natural number that is not divisible by  $p$ , then there exists a natural number  $x$  such that  $ax \equiv 1 \pmod{p}$ .*

*Proof.* In the case where  $p$  is the prime 2, each such  $a$  must be congruent to 1 modulo 2, so we can take  $x = 1$ . If  $p$  is greater than 2, then, for each given  $a$ , let  $x = a^{p-2}$ . Then  $ax = a \cdot a^{p-2} = a^{p-1}$  and, by Fermat's Theorem,  $a^{p-1} \equiv 1 \pmod{p}$ . □

It turns out to be interesting and useful to know which natural numbers are congruent to their own inverses modulo  $p$ . If  $x$  is such a number, then  $x \cdot x \equiv 1 \pmod{p}$ . In other words, such an  $x$  is a solution to the congruence  $x^2 \equiv 1 \pmod{p}$ , or  $x^2 - 1 \equiv 0 \pmod{p}$ . The solutions of the equations  $x^2 - 1 = 0$  are  $x = 1$  and  $x = -1$ . The solutions of the congruence are similar.

**Theorem 5.1.6.** *If  $p$  is a prime number and  $x$  is an integer satisfying  $x^2 \equiv 1 \pmod{p}$ , then either  $x \equiv 1 \pmod{p}$  or  $x \equiv p-1 \pmod{p}$ . (Note that  $p-1 \equiv -1 \pmod{p}$ ).*



*Proof.* If  $x^2 \equiv 1 \pmod{p}$  then, by definition,  $p$  divides  $x^2 - 1$ . But  $x^2 - 1 = (x - 1)(x + 1)$ . Since  $p$  divides  $x^2 - 1$ , Corollary 4.1.3 implies that  $p$  divides at least one of  $x - 1$  and  $x + 1$ . If  $p$  divides  $x - 1$ , then  $x \equiv 1 \pmod{p}$ . If  $p$  divides  $x + 1$ , then  $x \equiv -1 \pmod{p}$ , or  $x \equiv p - 1 \pmod{p}$ .  $\square$

## 5.2 Wilson's Theorem

As we now show, these considerations lead to a proof of Wilson's Theorem, a theorem that is very beautiful although it is considerably less famous and much less useful than Fermat's Theorem.

**Theorem 5.2.1** (Wilson's Theorem). *If  $p$  is a prime number, then  $(p-1)! + 1 \equiv 0 \pmod{p}$ . (In other words, if  $p$  is prime, then  $p$  divides  $(p-1)! + 1$ .)*

*Proof.* First note that the theorem is obviously true when  $p = 2$ ; in this case, it states  $(1 + 1) \equiv 0 \pmod{2}$ . As we indicated above, a multiplicative inverse of an integer  $x$  modulo  $p$  is an integer  $y$  such that  $xy \equiv 1 \pmod{p}$ . As we have seen, every number in the set  $\{1, 2, \dots, p-1\}$  has a multiplicative inverse modulo  $p$ . Since no multiplicative inverse can be divisible by  $p$ , the multiplicative inverse of each number in  $\{1, 2, \dots, p-1\}$  is congruent to one of the numbers in  $\{1, 2, \dots, p-1\}$ . By the previous theorem, the only numbers in the set  $\{1, 2, \dots, p-1\}$  that are congruent to their own multiplicative inverses are the numbers 1 and  $p-1$ . Leave those two numbers aside for the moment. Note that if  $y$  is a multiplicative inverse of  $x$ , then  $x$  is a multiplicative inverse of  $y$ . Thus the numbers in the set  $\{2, 3, \dots, p-2\}$  each have multiplicative inverses in that same set, and each number in that set differs from its multiplicative inverse. Therefore we can arrange the numbers in the set  $\{2, 3, \dots, p-2\}$  in pairs consisting of a number and its multiplicative inverse. Since the product of a number and its multiplicative inverse is congruent to 1 modulo  $p$ , the product of all the numbers in the set  $\{2, 3, \dots, p-2\}$  is congruent to 1 modulo  $p$ . Thus  $2 \cdot 3 \cdots (p-2) \equiv 1 \pmod{p}$ . Multiplying both sides by 1 gives  $1 \cdot 2 \cdot 3 \cdots (p-2) \equiv 1 \pmod{p}$ . Now  $p-1 \equiv -1 \pmod{p}$ , so  $1 \cdot 2 \cdots (p-2) \cdot (p-1) \equiv 1 \cdot (-1) \pmod{p}$ , or  $(p-1)! \equiv -1 \pmod{p}$ , so  $(p-1)! + 1 \equiv 0 \pmod{p}$ .  $\square$

**Theorem 5.2.2.** *If  $m$  is a composite number larger than 4, then  $(m-1)! \equiv 0 \pmod{m}$  (so that  $(m-1)! + 1 \equiv 1 \pmod{m}$ .)*

*Proof.* Let  $m$  be any composite number larger than 4. We must show that  $(m-1)!$  is divisible by  $m$ . If  $m = ab$ , with  $a$  different from  $b$  and both less than  $m$ , then  $a$  and  $b$  each occur as distinct factors in  $(m-1)!$ . Thus  $m = ab$  is a

factor of  $(m-1)!$ , so  $(m-1)!$  is congruent to 0 modulo  $m$ . The only composite numbers less than  $m$  that cannot be written as a product of two distinct natural numbers less than  $m$  are those numbers that are squares of primes. (To see this, use the fact that every composite can be written as a product of primes.) Thus, the only remaining case to prove is when  $m = p^2$  for some prime  $p$ . In this case, if  $m$  is larger than 4, then  $p$  is a prime bigger than 2. In that case,  $p^2$  is greater than  $2p$ . Thus  $p^2 - 1$  is greater than or equal to  $2p$ , so  $(p^2 - 1)!$  contains the factor  $2p$  as well as the factor  $p$ . Thus  $(p^2 - 1)!$  contains the product  $2p^2$ . In particular  $m = p^2$  divides  $(m-1)!$ .  $\square$

The following is a converse of Wilson's Theorem.

**Theorem 5.2.3.** *If  $m$  is a natural number other than 1, then  $(m-1)! + 1 \equiv 0 \pmod{m}$  if and only if  $m$  is a prime number.*

*Proof.* This follows immediately from Wilson's Theorem when  $m$  is prime and from the previous theorem for composite  $m$  in all cases except for  $m = 4$ . If  $m = 4$ , then  $(m-1)! + 1 = 7$ , which is not congruent to 0 modulo 4, so the theorem holds for all  $m$ .  $\square$

It might be thought that Wilson's Theorem would provide a good way to check whether or not a given number  $m$  is prime: simply see whether  $m$  divides  $(m-1)! + 1$ . However, the fact that  $(m-1)!$  is so much larger than  $m$  makes this a very impractical way of testing primality for large values of  $m$ .

## 5.3 Problems

### Basic Exercises

- Find the remainder when  $24^{103}$  is divided by 103.
- Find a solution  $x$  to each of the following congruences.
  - $2^x \equiv 1 \pmod{103}$
  - $16!x \equiv 5 \pmod{17}$
- Find the remainder when  $99^{100} - 1$  is divided by 101.

### Interesting Problems

4. Suppose that  $p$  is a prime greater than 2 and  $a \equiv b^2 \pmod{p}$  for some natural number  $b$  that is not divisible by  $p$ . Prove that  $a^{\frac{p-1}{2}} \equiv 1 \pmod{p}$ .
5. Find three different prime factors of  $10^{12} - 1$ .
6. Let  $p$  be a prime number. Prove that  $1^2 \cdot 2^2 \cdot 3^2 \cdots (p-1)^2 - 1$  is divisible by  $p$ .
7. For each of the following congruences either find a solution or prove that no solution exists
  - (a)  $(102!)x + x \equiv 4 \pmod{103}$
  - (b)  $x^{16} - 2 \equiv 0 \pmod{17}$
8. Find the remainder when:
  - (a)  $(9! \cdot 16 + 4311)^{8603}$  is divided by 11.
  - (b)  $(42)! + 7^{28} + 66$  is divided by 29.
9. If  $a$  is a natural number and  $p$  is a prime number, show that  $p$  divides  $a^p + (p-1)!a$ .

### Challenging Problems

10. Show that the natural number  $n > 1$  is prime if and only if  $n$  divides  $[(n-2)! - 1]$ .
11. Show that if  $p$  is a prime number and  $a$  and  $b$  are natural numbers, then
$$(a+b)^p \equiv a^p + b^p \pmod{p}.$$
12. For which prime numbers  $p$  is  $(p-2)! \equiv 1 \pmod{p}$ ?
13. Prove that for all primes  $p > 3$ ,  $2(p-3)! \equiv -1 \pmod{p}$ .
14. Is there a prime number  $p$  such that  $(p-1)! + 6$  is divisible by  $p$ ?
15. Find all prime numbers  $p$  such that  $p$  divides  $(p-2)! + 6$ .
16. Suppose  $2^k + 1$  is a prime number. Prove that  $k$  has no prime divisors other than 2. (Hint: if  $k = ab$  with  $b$  odd, consider  $2^k + 1$  modulo  $2^a + 1$ .)
17. Show that  $m$  is prime if there is an integer  $a$  such that  $a^{m-1} \equiv 1 \pmod{m}$  and  $a^k \not\equiv 1 \pmod{m}$  for every natural number  $k < m-1$ .

## Chapter 6

# Sending and Receiving Coded Messages

As early as ancient times people have devised ways of sending secret messages to each other. Much of the original interest was for military purposes: commanders of one section of an army wanted to send messages to commanders of other sections of their army in such a way that the message could not be understood by enemy soldiers who might intercept it.

Much of the current interest in coded messages is still for military and other horrible purposes. However, there are also many other kinds of situations in which it is important to be able to send secret messages. For example, a huge amount of information is communicated via the internet. It is important that some of that information remain private, known only to the sender and recipient. One common situation is making withdrawals from bank accounts over the internet. If someone else was able to intercept the information being sent, that interceptor could transfer funds from the sender's bank account to the interceptor's bank account. There are many other commercial and personal communications that are sent electronically that people wish to keep secret from those who may try to intercept them.

The general problem of sending coded messages is the following. A sender encodes a message in some way so that the encoded version cannot be understood by anyone who does not know how to decode it. Of course, the intended recipient of the message does have to be able to decode it. How can the recipient get the knowledge of how to decode? If the sender has to send the decoding information to the recipient, unintended interceptors (e.g., someone who wants to transfer your money to his or her bank account) might get access to the method of

decoding as that method was being transmitted to the intended recipient.

In most procedures for sending coded messages, anyone who understands the procedure for encoding messages would also understand how to decode them. When such a method is employed, it is crucial that the intended recipient of the message receives the knowledge of how to decode the message in a way that ensures that unintended recipients are unable to receive that knowledge. This can be very hard to do.

The subject of encoding and decoding messages is called “cryptography” and the techniques of encoding and decoding messages for a given procedure are called the “keys” for that procedure.

Beginning during and after the Second World War, many people wondered whether there could be “public key cryptography.” That means, a coding procedure that has the property that everyone in the world (the “public”) is told how to send the recipient a secret message. On the other hand, the recipient must be the only one who can decode messages sent using that procedure. That is, “public key cryptography” refers to methods of sending messages that allow the person who wishes to receive messages to publicly announce the way messages should be encoded, but have the property that only the person making the announcement can decode a message. This seems to be impossible. If unintended recipients know how to encode messages, won’t they necessarily also be able to figure out how to decode messages?

## 6.1 The RSA Method

It was only in the 1970’s that a method of public key cryptography was discovered. To actually use this method requires employing very large numbers. Thus this method would not be feasible without computers. On the other hand, the only mathematics that is required to establish that the method works is Fermat’s Theorem (Theorem 5.1.2). This technique of coding is called “RSA” after three of the computer scientists who discovered it, Ron Rivest, Adi Shamir and Leonard Adleman.

Here is an outline of the method. The recipient announces to the entire world the following way to send messages. If you want to send a message, the first thing that you must do is to convert the message into a natural number. There are many ways of doing that; here is a rough description of one possibility. Write your message out as sentences in, say, the English language. Then convert the sentences into a natural number as follows. Let  $a=11$ ,  $b=12$ ,  $c=13$ , ...,  $z=36$ . Let 37 represent a space. Let 38 represent a period, 39 a comma, 40 a semicolon,

41 a full colon, 42 an exclamation point and 43 an apostrophe. If desired, other symbols could be represented by other two-digit natural numbers. Convert your English language message into a number by replacing each of the elements of your sentences by their corresponding numbers in the order that they appear. For any substantial message, this will result in a large natural number. Everyone would be able to reconstruct the English language message from that number if this aspect of the coding procedure was announced publicly. For example, the sentence “Public key cryptography is neat.” would be represented by the number

2631122219133721153537132835263025172811261835371929372415113038

Furthermore, if you read the rest of this chapter,

35253143222237212425333733183537193037332528212938

The RSA technique is a method of encoding numbers. Both the recipient and those who send messages must use computers to do the computations that are required; the numbers involved in any application of the technique that could realistically protect messages are much too large for the computations to be done by hand.

RSA encoding proceeds as follows. The person who wishes to receive messages, the recipient, chooses two very large prime numbers  $p$  and  $q$  that are different from each other, and then defines  $N$  to be  $pq$ . The recipient publicly announces the number  $N$ . However, the recipient keeps  $p$  and  $q$  secret. If  $p$  and  $q$  are large enough, there is no way that anyone other than the recipient could find  $p$  or  $q$  simply from knowing  $N$ ; factoring very large numbers is beyond the capacity of even the most powerful computers. There are some very large known prime numbers; such can easily be chosen so that the resulting  $N = pq$  is impossible to factor in any reasonable amount of time, even using the most powerful computers. The recipient announces another natural number,  $E$ , in addition to  $N$ . Below we will explain ways of choosing suitable  $E$ 's.

The recipient then instructs all those who wish to send messages to do the following. Write your message as a natural number as described above. Let's say that  $M$  is the number representing your message. For this method to work,  $M$  must be less than  $N$ . If  $M$  is greater than or equal to  $N$ , you could divide your message into several smaller messages, each of which correspond to natural numbers less than  $N$ . The method we shall describe only works when  $M$  is less than  $N$ .

“To send me messages,” the recipient announces to the world, “take your message  $M$  and compute the remainder that  $M^E$  (i.e.,  $M$  raised to the power  $E$ ) leaves upon division by  $N$ , and send me that remainder.”

In other words, to send a message  $M$ , the sender computes the  $R$  between 0 and  $N$  such that  $M^E \equiv R \pmod{N}$ . The sender then sends  $R$  to the recipient.

How can the message be decoded? That is, how can the recipient recover the original message  $M$  from  $R$ ? This will require finding a “decoder,” which will be possible for anyone who knows the factorization of  $N$  as the product  $pq$ , but virtually impossible for anyone else. We shall see that, if  $E$  is chosen properly, there is a decoder  $D$  such that, for every  $L$  between 0 and  $N$ ,  $L^{ED} \equiv L \pmod{N}$ . For such a  $D$ , since  $R \equiv M^E \pmod{N}$ , it follows that  $R^D \equiv M^{ED} \pmod{N}$ , and therefore since  $M^{ED} \equiv M \pmod{N}$ ,  $R^D \equiv M \pmod{N}$ . Thus the recipient decodes the message by finding the remainder that  $R^D$  leaves upon division by  $N$ .

Before explaining further how to find encoders  $E$  and decoders  $D$  and why this method works, let’s look at a simple example. In this example the numbers are so small that anyone could figure out what  $p$  and  $q$  are, so this example could not be used to realistically encode messages. However, it illustrates the method.

Let  $p = 7$  and  $q = 11$  be the primes; then  $N = pq = 77$ . Suppose that  $E = 13$ ; as we shall see, there are always many possible values for  $E$ . Below we will discuss the properties that  $E$  must have. There is a technique for finding  $D$ , based on knowing  $p$  and  $q$ , that we shall describe later; that technique will produce  $D = 37$  in this particular example.

In this example, the recipient announces  $N = 77$  and  $E = 13$  to the general public; the recipient keeps the values of  $p$ ,  $q$ , and  $D$  secret.

The recipient invites the world to send messages. Suppose you want to send the message  $M = 71$ . Following the encryption rule, you must compute the remainder that  $M^E = 71^{13}$  leaves upon division by 77. Let’s compute that as follows. First,  $71 \equiv -6 \pmod{77}$ , so  $M^E \equiv (-6)^{13} \pmod{77}$ . Now  $6^3 = 216$  and  $216 \equiv -15 \pmod{77}$ , so  $6^6 = (6^3)^2 \equiv (-15)^2 \equiv 225 \pmod{77}$ , which is congruent to  $-6 \pmod{77}$ . Therefore,  $6^{12} \equiv (-6)^2 \equiv 36 \pmod{77}$ , so  $6^{13} \equiv 6 \cdot 6^{12} \equiv 6 \cdot 36 \equiv 216 \equiv -15 \pmod{77}$ . Therefore  $(-6)^{13} \equiv -(6)^{13} \equiv 15 \pmod{77}$ . Thus,  $M^E = 71^{13} \equiv 15 \pmod{77}$ . It follows that the remainder upon dividing  $71^{13}$  by 77 is 15.

Thus the encoded version of your message is 15. Anyone who sees that the encoded version is 15 would be able to discover your original message if they knew the decoder. But the recipient is the only one who does know the decoder.

In this special, easy, example, the recipient receives 15 and proceeds to decode it, using the decoder 37, as follows. Your original message will be the remainder that  $15^{37}$  leaves upon division by 77. Compute:  $15^2 \equiv -6 \pmod{77}$ . Therefore,  $15^{26} \equiv (-6)^{13} \pmod{77}$ , which (as we saw above) is congruent to 15  $\pmod{77}$ . Also, from  $15^2 \equiv -6 \pmod{77}$ , we obtain  $15^4 \equiv 36 \pmod{77}$ . Thus,  $15^8 \equiv 36 \cdot 6 \cdot 6 \equiv 216 \cdot 6 \equiv (-15) \cdot 6 \equiv -90 \equiv 64 \pmod{77}$ . Now  $15^{37} \equiv 15^{26} \cdot 15^8 \cdot 15^3 \pmod{77}$  which is congruent to  $15 \cdot 64 \cdot 15^3 \pmod{77}$ , which is congruent to  $64 \cdot 15^4 \pmod{77}$ . Since  $15^2 \equiv -6 \pmod{77}$ , this is congruent to  $64 \cdot 36 \pmod{77}$ , which is congruent to  $(-13) \cdot 36$ , which equals  $-468$ . Of course,  $-468$  is congruent to  $-468$  plus any multiple of 77. Now  $7 \cdot 77 = 539$ . Hence  $15^{37} \equiv -468 \equiv -468 + 539 \equiv 71 \pmod{77}$ . Therefore we have decoded the received message, 15, and obtained the original message, 71. (The number 71 must be the original message since it is the only natural number less than  $N$  that is congruent to 71.)

The above looks somewhat complicated. We now proceed to explain and analyze the method in more detail.

For  $p$  and  $q$  distinct primes and  $N = pq$ , we use the notation  $\phi(N)$  to denote  $(p-1)(q-1)$ . (This is a particular case of a more general concept that we will introduce in the next chapter.) The theorem that underlies the RSA technique is an easy consequence of Fermat's Theorem (Theorem 5.1.2).

**Theorem 6.1.1.** *Let  $N = pq$ , where  $p$  and  $q$  are distinct prime numbers, and let  $\phi(N) = (p-1)(q-1)$ . If  $k$  and  $a$  are any natural numbers, then  $a \cdot a^{k\phi(N)} \equiv a \pmod{N}$ .*

*Proof.* The conclusion of the theorem is equivalent to the assertion that  $N$  divides the product of  $a$  and  $a^{k(p-1)(q-1)} - 1$ . Since  $N$  is the product of the distinct primes  $p$  and  $q$ , this is equivalent to the assertion that  $p$  divides  $a \cdot (a^{k(p-1)(q-1)} - 1)$  and  $q$  divides  $a \cdot (a^{k(p-1)(q-1)} - 1)$ .

Consider  $p$  (obviously the same proof works for  $q$ ). There are two cases. First, if  $p$  divides  $a$ , then  $p$  certainly divides  $a \cdot (a^{k(p-1)(q-1)} - 1)$ . If  $p$  does not divide  $a$  then, by Fermat's Theorem (Theorem 5.1.2),  $a^{p-1} \equiv 1 \pmod{p}$ . Raising both sides of this congruence to the power  $k(q-1)$  shows that  $p$  divides  $a^{k(p-1)(q-1)} - 1$ , so it also divides  $a \cdot (a^{k(p-1)(q-1)} - 1)$ . This establishes the result in the case that  $p$  does not divide  $a$ . Thus, in both cases,  $p$  divides  $(a \cdot a^{k(p-1)(q-1)} - a)$ . Therefore  $a \cdot a^{k\phi(N)} \equiv a \pmod{N}$ .  $\square$

How does this theorem apply to RSA coding? We pick as an encoder  $E$  any natural number that does not have any factor in common with  $\phi(N)$ . As we shall see in the next chapter, this implies that there is a natural number  $D$  such



that  $ED$  is equal to the sum of 1 and a multiple of  $\phi(N)$ ; that is, there is a  $D$  such that  $ED = 1 + k\phi(N)$  for some natural number  $k$ . The theorem we have just proven shows that  $D$  is a decoder, as follows. Suppose that  $M$  is the original message, so that  $R \equiv M^E \pmod{N}$  is its encoding. Since  $R$  is congruent to  $M^E \pmod{N}$ ,  $R^D$  is congruent to  $M^{ED} \pmod{N}$ . But  $ED = 1 + k\phi(N)$ , so  $R^D$  is congruent to  $M^{ED}$ , which is congruent to  $M^{1+k\phi(N)}$ . This is congruent to the product of  $M$  and  $M^{k(p-1)(q-1)}$ , which is congruent to  $M$  by the above theorem. (Of course,  $M$  is a natural number less than  $N$ , which uniquely determines it.)

To explain how to find decoders requires some additional mathematical tools that we develop in the next chapter. If  $n$  is very small, decoders can be found simply by trial and error.

## 6.2 Problems

### Basic Exercises

1. You are to receive a message using the RSA system. You choose  $p = 5$ ,  $q = 7$  and  $E = 5$ . Verify that  $D = 5$  is a decoder. The encoded message you receive is 17. What is the actual (decoded) message?
2. Use the RSA system with  $N = 21$  and the public key  $E = 5$ .
  - (a) Encrypt the message  $M = 7$ .
  - (b) Verify that  $D = 5$  is a decoder (the multiplicative inverse of  $E$  modulo  $\phi(N)$ ).
  - (c) Decrypt the encrypted form of the message.
3. A person tries to receive messages without you being able to decode them. The person announces  $N = 15$  and  $E = 7$  to the world; the person uses such low numbers assuming that you don't understand RSA. A coded message  $R = 8$  is sent. By trial and error, find the decoder,  $D$ , and use it to find the original message.

## Chapter 7

# The Euclidean Algorithm and Applications

### 7.1 The Euclidean Algorithm

Each pair of natural numbers has a greatest common divisor; i.e., a largest natural number that is a factor of both of the numbers in the pair. For example, the greatest common divisor of 27 and 15 is 3, the greatest common divisor of 36 and 48 is 12, the greatest common divisor of 257 and 101 is 1, the greatest common divisor of 4 and 20 is 4, the greatest common divisor of 7 and 7 is 7, and so on.

**Definition 7.1.1.** The *greatest common divisor* of the natural numbers  $m$  and  $n$  is denoted  $\gcd(m, n)$ .

Thus  $\gcd(27, 15) = 3$ ,  $\gcd(36, 48) = 12$ ,  $\gcd(7, 21) = 7$ , and so on. One way to find the greatest common divisor of a pair of natural numbers is by factoring the numbers into primes. Then the greatest common divisor of the two numbers is obtained in the following way: for each prime that occurs as a factor of both numbers, find the highest power of that prime that is a common factor of both numbers, and then multiply all those primes to all those powers together to get the greatest common divisor. For example, since  $48 = 2^4 \cdot 3$  and  $56 = 2^3 \cdot 7$ ,  $\gcd(24, 56) = 2^3 = 8$ . As another example, note that  $\gcd(1292, 14440) = 76$ , since  $1292 = 2^2 \cdot 17 \cdot 19$  and  $14440 = 2^3 \cdot 5 \cdot 19^2$  and  $2^2 \cdot 19 = 76$ .

Another way of finding the greatest common divisor of two natural numbers is by using what is called the “Euclidean Algorithm”. One advantage of this method is that it provides a way of expressing the greatest common divisor

as a combination of the two original numbers in a way that can be extremely useful. In particular, this technique will allow us to compute a decoder for each encoder chosen for RSA coding. As we shall see, other applications of the Euclidean Algorithm include a method for finding integer solutions of linear equations in two variables (“Diophantine equations”) and a different proof of the Fundamental Theorem of Arithmetic.

The Euclidean Algorithm is based on the ordinary operation of division of natural numbers, allowing for a remainder. We can express that concept of division as follows (the term “non-negative integer” refers to either a natural number or 0): if  $a$  and  $b$  are any natural numbers, then there exist non-negative integers  $q$  and  $r$  such that  $a = bq + r$  and  $0 \leq r < b$ . (The number  $q$  is called the “quotient” and the number  $r$  is called the “remainder” in this equation.) If  $b$  divides  $a$ , then, of course,  $r = 0$ .

Let  $a$  and  $b$  be natural numbers whose greatest common divisor is  $d$ . The Euclidean Algorithm for finding  $d$  is the following technique. If  $b = a$ , then clearly the greatest common divisor is  $a$ . Suppose that  $b$  is less than  $a$ . (If  $b$  is greater than  $a$ , interchange the roles of  $a$  and  $b$  below.) Divide  $a$  by  $b$  as described above to get  $q$  and  $r$  satisfying  $a = bq + r$  with  $0 \leq r < b$ . If  $r = 0$ , then clearly the greatest common divisor of  $a$  and  $b$  is  $b$  itself. If  $r$  is not 0, divide  $b$  into  $r$ , to get  $b = r_1q_1 + r_2$ , where  $0 \leq r_2 < r_1$ . If  $r_2 = 0$ , stop here. If  $r_2$  is different from 0, divide  $r_1$  into  $r_2$ , to get  $r_2 = r_1q_2 + r_3$  where  $0 \leq r_3 < r_2$ . Continue this process until there is the remainder 0. (That will have to occur eventually since the remainders are all non-negative integers and each one is less than the preceding one.) Thus there is a sequence of equations as follows:

$$\begin{aligned} a &= bq + r \\ b &= r_1q_1 + r_2 \\ r &= r_1q_2 + r_3 \\ r_1 &= r_2q_3 + r_4 \\ &\vdots \\ r_{k-3} &= r_{k-2}q_{k-1} + r_{k-1} \\ r_{k-2} &= r_{k-1}q_k + r_k \\ r_{k-1} &= r_kq_{k+1} \end{aligned}$$

It follows that  $r_k$  is the greatest common divisor of the original  $a$  and  $b$ . To see this, note first that  $r_k$  is a common divisor of  $a$  and  $b$ . This can be seen by

“working your way up” the equations. Replacing  $r_{k-1}$  by  $r_k q_{k+1}$  in the next to last equation gives  $r_{k-2} = r_k q_{k+1} q_k + r_k = r_k (q_{k+1} q_k + 1)$ . Thus  $r_k$  divides  $r_{k-2}$ . The equation for  $r_{k-3}$  can then be rewritten

$$r_{k-3} = [r_k (q_{k+1} q_k + 1)] q_{k-1} + r_k q_{k+1} = r_k [(q_{k+1} q_k + 1) q_{k-1} + q_{k+1}]$$

Thus  $r_{k-3}$  is also divisible by  $r_k$ . Continuing to work upwards eventually shows that  $r_k$  divides  $r$ , then  $b$ , and then  $a$ . Thus  $r_k$  is a common divisor of  $a$  and  $b$ .

To show that  $r_k$  is the greatest common divisor of  $a$  and  $b$ , we show that every other common divisor of  $a$  and  $b$  divides  $r_k$ . Suppose, then, that  $d$  is a natural number that divides both  $a$  and  $b$ . The equation  $a = bq + r$  shows that  $d$  also divides  $r$ . Since  $d$  divides both  $b$  and  $r$ , it divides  $r_1$ ; since it divides  $r$  and  $r_1$ , it divides  $r_2$ ; and so on. Eventually, we see that  $d$  also divides  $r_k$ . Hence every common divisor of  $a$  and  $b$  divides  $r_k$ , so  $r_k$  is the greatest common divisor of  $a$  and  $b$ .

Let’s look at an example. Suppose we want to use the Euclidean Algorithm to find the greatest common divisor of 33 and 24. We begin with  $33 = 24 \cdot 1 + 9$ . Then,  $24 = 9 \cdot 2 + 6$ . Then,  $9 = 6 \cdot 1 + 3$ . Then,  $6 = 3 \cdot 2$ . Thus the greatest common divisor of 33 and 24 is 3.

**Definition 7.1.2.** We say that the integer  $d$  is a *linear combination of the integers  $a$  and  $b$*  if there exist integers  $x$  and  $y$  such that  $ax + by = d$ .

Obtaining the greatest common divisor by the Euclidean Algorithm allows us to express the greatest common divisor as a linear combination of the original numbers, as follows. First consider the above example. From the next to last equation, we get  $3 = 9 - 6 \cdot 1$ . Substituting the expression for 6 obtained from the previous equation into this one gives

$$3 = 9 - (24 - 9 \cdot 2) \cdot 1 = 9 \cdot 3 - 24.$$

Then solve for 9 in the first equation,  $9 = 33 - 24 \cdot 1$  and substitute this into the above equation to get  $3 = (33 - 24 \cdot 1) \cdot 3 - 24 = 33 \cdot 3 - 24 \cdot 4$ . Therefore  $3 = 33 \cdot 3 + 24(-4)$ . The greatest common divisor of the numbers 33 and 24, 3, is expressed in the last equation as a linear combination of 33 and 24.

## 7.2 Applications

In general, a linear combination of the integers  $a$  and  $b$  is an expression of the form  $ax + by$  where  $x$  and  $y$  are integers. The Euclidean Algorithm can always

be used as in the above example to write the greatest common divisor of two natural numbers as a linear combination of those numbers. That is, given natural numbers  $a$  and  $b$  with greatest common divisor  $d$ , there exists integers  $x$  and  $y$  such that  $d = ax + by$ . This can be seen by working upwards in the sequence of equations that constitute the Euclidean Algorithm. The next to last equation can be used to write the greatest common divisor,  $r_k$ , as a linear combination of  $r_{k-1}$  and  $r_{k-2}$ ; simply solve the next to last equation for  $r_k$ . Solving for  $r_{k-1}$  in the previous equation and substituting represents  $r_k$  as a linear combination of  $r_{k-2}$  and  $r_{k-3}$ . By continuing to work our way up the ladder of equations in the Euclidean Algorithm, we eventually obtain  $r_k$  as a linear combination of the given numbers  $a$  and  $b$ .

**Definition 7.2.1.** The integers  $m$  and  $n$  are said to be *relatively prime* if their only common divisor is 1; that is, if  $\gcd(m, n) = 1$ .

By the above-described consequence of the Euclidean Algorithm,  $\gcd(m, n) = 1$  implies that there exist integers  $s$  and  $t$  such that  $sm + tn = 1$ , or  $sm = 1 - tn$ . If  $m$ ,  $n$  and  $s$  are all positive, this equation clearly implies that  $t$  is negative. These facts are exactly what we need in order to find decoders in the RSA system.

Before establishing this in general, let's illustrate it in the case of one of the particular examples given in the previous section. In that example, we started with  $p = 7$  and  $q = 11$ , so that  $N = 77$  and  $\phi(N) = 6 \cdot 10 = 60$ . We took the encoder  $E = 13$ . The crucial property of the encoder is that it is relatively prime to  $\phi(N)$ . That is true in this case; clearly the only common factor of 13 and 60 is 1. Since  $\gcd(13, 60) = 1$ , the consequence of the Euclidean Algorithm discussed above implies that there exist integers  $s$  and  $t$  such that  $1 = 13s + 60t$ , or  $13s = 1 - 60t$ . Note that if  $s$  and  $t$  satisfy this equation then, for every  $m$ ,  $13(s + 60m) = 1 - 60(t - 13m)$ , since this latter equation is obtained from the previous one by simply adding  $13 \cdot 60m$  to both sides of the equation. Thus, if the original  $s$  was negative we could choose a positive  $m$  large enough so that  $s + 60m$  is positive. Therefore without loss of generality we can assume that  $s$  is positive, which forces  $t$  to be negative, in the equation  $13s = 1 - 60t$ . Replace  $-t$  by  $u$ ; then  $13s = 1 + 60u$ , with  $s$  and  $u$  both positive integers. We will find such  $s$  and  $u$  using the Euclidean Algorithm. First, however, note that any such  $s$  is a decoder. To see this, recall that  $M^{13}$  is congruent to the encoded version of the message  $M$ . Thus  $(M^{13})^s = M^{13s} = M^{1+60u} = M \cdot M^{60u}$  which is congruent modulo 77 to  $M$  (by Theorem 6.1.1).

Let's find a decoder for this example. We begin by using the Euclidean

Algorithm to find  $\gcd(13, 60)$ :

$$60 = 4 \cdot 13 + 8$$

$$13 = 1 \cdot 8 + 5$$

$$8 = 1 \cdot 5 + 3$$

$$5 = 1 \cdot 3 + 2$$

$$3 = 1 \cdot 2 + 1$$

$$2 = 2 \cdot 1$$

Thus the greatest common divisor of 13 and 60 is 1. Of course, we knew that already; we chose 13 to be relatively prime to 60. The point of using the Euclidean Algorithm is that it allows us to express 1 as a linear combination of 13 and 60, as follows. From the above equation  $3 = 1 \cdot 2 + 1$  we get  $1 = 3 - 2$ , so substituting by using the above equation  $5 = 1 \cdot 3 + 2$  yields

$$1 = 3 - (5 - 3).$$

Continuing by working our way up and collecting coefficients gives the following:

$$\begin{aligned} 1 &= 2 \cdot 3 - 5 \\ &= 2(8 - 5) - 5 \\ &= 2 \cdot 8 - 3 \cdot 5 \\ &= 2 \cdot 8 - 3(13 - 8) \\ &= 5 \cdot 8 - 3 \cdot 13 \\ &= 5(60 - 4 \cdot 13) - 3 \cdot 13 \\ &= 5 \cdot 60 - 23 \cdot 13 \end{aligned}$$

Equivalently,  $1 - 5 \cdot 60 = -(23 \cdot 13)$ . We are not done. We must find positive integers  $k$  and  $D$  such that  $1 + k \cdot 60 = 13 \cdot D$ . For any integer  $m$ , adding  $-13 \cdot 60m$  to both sides of the above equation gives  $1 - (5 + 13m)60 = (-23 - 60m)13$ . Taking  $m = -1$  in this equation gives  $1 + 8 \cdot 60 = 37 \cdot 13$ . Thus 37 is a decoder.

We have illustrated and proven the RSA technique. The following is a statement of what we have established.

**Theorem 7.2.2** (RSA Technique for Encoding Messages). *The recipient chooses (very large) distinct prime numbers  $p$  and  $q$  and lets  $N = p \cdot q$  and  $\phi(N) = (p - 1)(q - 1)$ . The recipient then chooses an  $E$ , any natural number greater than 1 that is relatively prime to  $\phi(N)$ . The recipient announces the numbers  $N$  and  $E$  and states that any message  $M$  consisting of a natural number less than  $N$  can be sent as follows: Compute the natural number  $R$  less than  $N$  such that  $M^E \equiv R \pmod{N}$ . The encoded message that is sent is the natural number  $R$ . The recipient decodes the message by using the Euclidean Algorithm to find natural numbers  $m$  and  $D$  such that  $1 + m\phi(N) = ED$ . The recipient then recovers the original message  $M$  as the natural number less than  $N$  that is congruent to  $M^{ED} \pmod{N}$ .*

The technique that we used to find decoders can be used to solve many other practical problems.

**Definition 7.2.3.** A *Linear Diophantine equation* is an equation of the form  $ax + by = c$  for which we seek solutions  $(x, y)$  where  $x$  and  $y$  are integers.

Note that finding decoders involves solving certain Diophantine equations. There are other practical applications of solutions of Diophantine equations.

**Example 7.2.4.** A store sells two different kinds of boxes of candies. One kind sells for 9 dollars a box and the other kind for 16 dollars a box. At the end of the day, the store has received 143 dollars from the sale of boxes of candy. How many boxes did the store sell at each price?

How can we approach this problem? If  $x$  is the number of the less expensive boxes sold and  $y$  is the number of the more expensive boxes sold then the information we are given is

$$9x + 16y = 143$$

There are obviously an infinite number of pairs  $(x, y)$  of real numbers that satisfy this equation; the graph in the plane of the set of solutions is a straight line. However, we know more about  $x$  and  $y$  than simply that they satisfy that equation. We also know that they must both be non-negative integers. Are there non-negative integral solutions? Are there any integral solutions at all? Since 9 and 16 are relatively prime, the Euclidean Algorithm tells us that there exist integers  $s$  and  $t$  satisfying  $9s + 16t = 1$ . Multiplying through by 143 gives  $9(143s) + 16(143t) = 143$ . Therefore, there are integral solutions. However, it is not immediately clear whether there are non-negative integral solutions, which the actual problem requires. Let's investigate.

We'll use the Euclidean Algorithm to find integral solutions to the equation  $9s + 16t = 1$ . This proceeds as follows. We first use the Euclidean Algorithm to find the greatest common divisor (although we know it):

$$16 = 9 \cdot 1 + 7$$

$$9 = 7 \cdot 1 + 2$$

$$7 = 2 \cdot 3 + 1$$

$$2 = 2 \cdot 1$$

Working our way back upward to express 1 as a linear combination of 9 and 16 gives

$$\begin{aligned} 1 &= 7 - 3 \cdot 2 \\ &= 7 - 3(9 - 7) \\ &= 4 \cdot 7 - 3 \cdot 9 \\ &= 4(16 - 9) - 3 \cdot 9 \\ &= 16 \cdot 4 - 9(7) \end{aligned}$$

Therefore  $9(-7) + 16 \cdot 4 = 1$ . Multiplying by 143 yields  $9(-7 \cdot 143) + 16(4 \cdot 143) = 143$ . Note that  $7 \cdot 143 = 1001$  and  $4 \cdot 143 = 572$ . For any integer  $m$ , we can add and subtract  $16 \cdot 9m$ ; thus, for every integer  $m$ ,

$$9(-1001 - 16m) + 16(572 + 9m) = 143$$

This gives infinitely many integer solutions; what about non-negative solutions?

We require that  $-1001 - 16m$  be at least 0. That is equivalent to  $16m \leq -1001$ , or  $m \leq \frac{-1001}{16}$ . Thus  $m \leq -62.5625$ . The largest  $m$  satisfying this inequality is  $m = -63$ . When  $m = -63$ ,  $-1001 - 16m = 7$  and  $572 + 9m = 5$ . Thus one pair of non-negative solutions to the original equation is  $x = 7$  and  $y = 5$ . Are there other non-negative solutions? We will show that all the solutions of this equation are of the form  $x = -1001 - 16m$  and  $y = 572 + 9m$  for some integer  $m$  (see Example 7.2.8 below). To show that the only non-negative solution is (7,5) we reason as follows. If we take the next largest  $m$ ,  $m = -64$ , then the  $y$  we get is  $572 - 9 \cdot 64 = -4$ . Obviously, if  $m$  is even smaller,  $572 + 9m$  will be even more negative. Therefore the only pair of non-negative solutions to the original equation is (7,5). Thus the store sold 7 of the cheaper boxes and 5 of the more expensive boxes of candy.



The basic theorem about solutions of linear Diophantine equations is the following.

**Theorem 7.2.5.** *The Diophantine equation  $ax + by = c$ , with  $a$ ,  $b$ , and  $c$  integers, has integral solutions if and only if  $\gcd(a, b)$  divides  $c$ .*

*Proof.* Let  $d = \gcd(a, b)$ . If there is a pair of integers  $(x, y)$  satisfying the equation, then  $ax + by = c$  and, since  $d$  divides both of  $a$  and  $b$ , it follows that  $d$  divides  $c$ . This proves the easy part of the theorem.

The converse is also easy, but only because of what we learned about the Euclidean Algorithm. In fact, we used the Euclidean Algorithm to prove that there exists a pair  $(s, t)$  satisfying  $as + bt = d$ . If  $d$  divides  $c$ , then there is a  $k$  satisfying  $c = dk$ . Let  $x = sk$  and  $y = tk$ . Then clearly  $ax + by = c$   $\square$

As we've seen in the example where we determined the number of boxes of each kind of candy sold, it is sometimes important to be able to determine all the solutions of a Diophantine equation. In both the decoder and the candy examples above, we use the very easy fact that  $(x + bm, y - am)$  is a solution of  $ax + by = c$  whenever  $x, y$  is a solution. (This follows since  $a(x + bm) + b(y - am) = ax + abm + by - abm = ax + by$ .) This shows that a Diophantine equation has an infinite number of solutions if it has any solution at all. In finding decoders, we don't care if the decoder that we find is only one of a number of possible decoders. However in other situations, such as the problem about determining the number of different kinds of boxes of candy that were sold, it is important to have a unique solution that satisfies some other condition of the problem (such as that both of  $x$  and  $y$  be non-negative). The following theorem precisely describes all the solutions of given linear Diophantine equation.

We require a lemma that generalizes the fact that if a prime divides a product then it divides at least one of the factors (Corollary 4.1.3).

**Lemma 7.2.6.** *If  $s$  divides  $tu$  and  $s$  is relatively prime to  $u$ , then  $s$  divides  $t$*

*Proof.* The hypothesis implies that there exists an  $r$  such that  $tu = rs$ . Write the canonical factorization of  $u$  into primes:  $u = p_1^{\alpha_1} \cdot p_2^{\alpha_2} \cdots p_k^{\alpha_k}$ ; then  $tp_1^{\alpha_1} \cdot p_2^{\alpha_2} \cdots p_k^{\alpha_k} = rs$ . Imagine factoring both sides of this equation into products of primes. By the Fundamental Theorem of Arithmetic (Theorem 4.1.1), the factorization of the left hand side into primes has to be the same as the factorization of the right hand side. Since  $s$  is relatively prime to  $u$ , none of the primes comprising  $s$  are among the  $p_i$ . Thus all the primes in  $s$  occur to at least same power in the factorization of  $t$ , and thus  $s$  divides  $t$ . This proves the lemma.  $\square$

**Theorem 7.2.7.** *Let  $\gcd(a, b) = d$ . The Diophantine equation  $ax + by = c$  has a solution if and only if  $d$  divides  $c$ . If  $d$  does divide  $c$  and  $(x_0, y_0)$  is a solution, then the integral solutions of the equation consist of all the pairs  $(x_0 + m\frac{b}{d}, y_0 - m\frac{a}{d})$  where  $m$  assumes all integral values.*

*Proof.* We already established the first assertion, the criterion for the existence of a solution (7.2.5). It is easy to see that each of the other pairs is a solution, for

$$a(x_0 + m\frac{b}{d}) + b(y_0 - m\frac{a}{d}) = ax_0 + m\frac{ab}{d} + by_0 - m\frac{ab}{d} = c + 0 = c$$

All that remains to be proven is that there are no solutions other than those described in the theorem. To see this, suppose that  $(x_0, y_0)$  is a solution and that  $(x, y)$  is any other solution of  $ax + by = c$ . Since  $ax_0 + by_0 = c$  we can subtract the first equation from the second to conclude that

$$a(x - x_0) + b(y - y_0) = 0$$

Bring one of the terms to the other side and divide both sides of this equation by  $d$  to get

$$\frac{a}{d}(x - x_0) = -\frac{b}{d}(y - y_0)$$

Note that  $\frac{a}{d}$  and  $\frac{b}{d}$  are relatively prime. (For if  $c$  was a common factor greater than 1, then  $dc$  would be a greater common divisor of  $a$  and  $b$  than  $d$ .) Hence, by Lemma 7.2.6,  $\frac{a}{d}$  divides  $(y - y_0)$  and  $\frac{b}{d}$  divides  $(x - x_0)$ . That is, there are integers  $k$  and  $l$  such that  $y - y_0 = k\frac{a}{d}$  and  $x - x_0 = l\frac{b}{d}$ . Equivalently  $y = y_0 + k\frac{a}{d}$  and  $x = x_0 + l\frac{b}{d}$ . For  $(x, y)$  to be a solution, we must have

$$a(x_0 + l\frac{b}{d}) + b(y_0 + k\frac{a}{d}) = c$$

Thus

$$ax_0 + al\frac{b}{d} + by_0 + bk\frac{a}{d} = c$$

Since  $ax_0 + by_0 = c$ , we get  $al\frac{b}{d} + bk\frac{a}{d} = 0$ . Thus  $l = -k$ . Call this common value  $m$ ; then

$$x = x_0 + m\frac{b}{d}$$

$$y = y_0 - m\frac{a}{d}$$

This proves the theorem. □

**Example 7.2.8.** The uniqueness of the solution to the “candy boxes problem” (Example 7.2.4) follows from this theorem. In that example,  $\gcd(9, 16) = 1$ , so all the solutions are indeed of the form  $(-1001 - 16m, 572 + 9m)$ .

There are many other interesting applications of the theorem concerning solutions of linear Diophantine equations - see, for example, the problems at the end of this chapter.

Recall that we used the notation  $\phi(N)$  to denote  $(p-1)(q-1)$  when we were describing the RSA technique with  $N = pq$  and  $p$  and  $q$  distinct prime numbers. This is a special case of a general notation for a useful general concept.

**Definition 7.2.9.** For any natural number  $m$ , the *Euler  $\phi$  function*  $\phi(m)$  is defined to be the number of numbers in  $\{1, 2, \dots, m-1\}$  that are relatively prime to  $m$ .

**Example 7.2.10.** To compute  $\phi(8)$ , we consider the set  $\{1, 2, 3, 4, 5, 6, 7\}$ . Then  $\phi(8) = 4$  since 1, 3, 5, 7 are the numbers in the set that are relatively prime to 8. Similarly  $\phi(7) = 6$ , and  $\phi(12) = 4$ .

**Theorem 7.2.11.** *If  $p$  is prime, then  $\phi(p) = p - 1$ .*

*Proof.* Since  $p$  is prime, every number in  $\{1, 2, \dots, p-1\}$  is relatively prime to  $p$ , so  $\phi(p) = p - 1$ .  $\square$

In discussing the RSA technique, we used the notation  $\phi(pq) = (p-1)(q-1)$  when  $p$  and  $q$  are distinct primes. This is consistent with the definition of  $\phi$  we are now using.

**Theorem 7.2.12.** *If  $p$  and  $q$  are distinct primes, then  $\phi(pq) = (p-1)(q-1)$ .*

*Proof.* Suppose that  $p$  and  $q$  are primes with  $p$  less than  $q$  (since they are different, one of them is less than the other), and  $N = pq$ . Consider the set  $S = \{1, 2, 3, \dots, p, \dots, q, \dots, pq-1\}$ . To find  $\phi(N)$ , we must determine how many numbers in this set are relatively prime to  $N$ . If a number is not relatively prime to  $N$ , then it must be divisible by either  $p$  or  $q$  or both. There are a total of  $pq - 1$  numbers in  $S$ ; how many multiples of  $p$  are there in  $S$ ? There is  $p, 2p, 3p$  and so on, up to  $(q-1)p$ , since  $qp$  is not in  $S$ . Thus there are  $q-1$  multiples of  $p$  in  $S$ . Similarly there are  $p-1$  multiples of  $q$  in  $S$ . Thus there is a total of  $(q-1) + (p-1) = p+q-2$  numbers in  $S$  that are not relatively prime to  $N$ . Since there are  $pq-1$  numbers in  $S$ , the number of numbers in  $S$  that are relatively prime to  $N$  is

$$\phi(N) = pq - 1 - (p + q - 2) = pq - p - q + 1$$

But  $pq - p - q + 1 = (p - 1)(q - 1)$ . Therefore  $\phi(N) = (p - 1)(q - 1)$ .  $\square$

There is a formula for  $\phi(m)$  for any natural number  $m$  greater than 1, in terms of the canonical factorization of  $m$  into a product of primes - see problem 23 at the end of this chapter.

Fermat's beautiful theorem that  $a^{p-1} \equiv 1 \pmod{p}$ , for primes  $p$  and numbers  $a$  that are not divisible by  $p$ , can be generalized to composite moduli. We require the following lemma.

**Lemma 7.2.13.** *If  $a$  is relatively prime to  $m$  and  $ax \equiv ay \pmod{m}$ , then  $x \equiv y \pmod{m}$ .*

*Proof.* We are given that  $m$  divides  $(ax - ay)$ . That is,  $m$  divides  $a(x - y)$ . By Lemma 7.2.6,  $m$  divides  $(x - y)$ . Thus  $x \equiv y \pmod{m}$ .  $\square$

**Theorem 7.2.14** (Euler's Theorem). *If  $m$  is a natural number greater than 1, and  $a$  is a natural number that is relatively prime to  $m$ , then  $a^{\phi(m)} \equiv 1 \pmod{m}$ .*

*Proof.* The proof is very similar to the proof of Fermat's Theorem (5.1.2). Let  $S = \{r_1, r_2, \dots, r_{\phi(m)}\}$  be the set of numbers in  $\{1, 2, \dots, m - 1\}$  that are relatively prime to  $m$ . Then let  $T = \{ar_1, ar_2, \dots, ar_{\phi(m)}\}$ . Clearly, no two of the numbers in  $S$  are congruent to each other, since they are distinct numbers all of which are less than  $m$ . Note also that no two of the numbers in  $T$  are congruent to each other, since  $ar_i \equiv ar_j \pmod{m}$  would imply, by Lemma 7.2.13, that  $r_i \equiv r_j \pmod{m}$ . Thus the numbers in  $\{ar_1, ar_2, \dots, ar_{\phi(m)}\}$  are congruent, in some order, to the numbers in  $\{r_1, r_2, \dots, r_{\phi(m)}\}$ . It follows, as in the proof of Fermat's Theorem, that the product of all the numbers in  $T$  is equal to the product of all the numbers in  $S$ . That is,

$$a \cdot r_1 \cdot a \cdot r_2 \cdots a \cdot r_{\phi(m)} \equiv r_1 \cdot r_2 \cdots r_{\phi(m)} \pmod{m}$$

Since  $r_1 \cdot r_2 \cdots r_{\phi(m)}$  is relatively prime to  $m$ , we can divide both sides of this congruence by that product (see 7.2.13), getting

$$a^{\phi(m)} \equiv 1 \pmod{m}$$

$\square$

Fermat's Theorem is a special case of Euler's.

**Corollary 7.2.15** (Fermat's Theorem). *If  $p$  is a prime and  $p$  does not divide  $a$ , then  $a^{p-1} \equiv 1 \pmod{p}$ .*

*Proof.* Since  $p$  is prime, the fact that  $p$  does not divide  $a$  means that  $a$  and  $p$  are relatively prime. Also,  $\phi(p) = p - 1$ . Thus Fermat's Theorem follows from Euler's.  $\square$

## 7.3 Problems

### Basic Exercises

1. Find the greatest common divisor of each of the following pairs of integers in two different ways, by using the Euclidean Algorithm and by factoring both numbers into primes.
  - (a) 252 and 198
  - (b) 291 and 573
  - (c) 1800 and 240
  - (d) 52 and 135
2. For each of the pairs in problem 1 above, write the greatest common divisor as a linear combination of the given numbers.
3. Find integers  $x$  and  $y$  such that  $3x - 98y = 12$ .
4.
  - (a) Find a formula for all integer solutions of the Diophantine equation  $3x + 4y = 14$
  - (b) Find all pairs of natural numbers that solve the above equation
5. Let  $\phi$  be Euler's  $\phi$ -function. Find:
  - (a)  $\phi(12)$
  - (b)  $\phi(26)$
  - (c)  $\phi(21)$
  - (d)  $\phi(36)$
  - (e)  $\phi(97)$
  - (f)  $\phi(73)$
  - (g)  $\phi(101 \cdot 37)$
  - (h)  $\phi(3^{100})$
6. Use the Euclidean Algorithm to find the decoders in problems 1, 2 and 3 in the previous chapter.

**Interesting Problems**

7. Use the Euclidean Algorithm (and a calculator) to find the greatest common divisor of each of the following pairs of natural numbers
  - (a) 47,295 and 297
  - (b) 77,777 and 2,891
8. Tickets for reserved seats for a concert are \$50 each and general admission tickets to the concert are \$24 each. The total revenue from sales of tickets to the concert is \$6,620. How many of each kind of ticket were sold?
9. Find the smallest natural number  $x$  such that  $24x$  leaves a remainder of 2 upon division by 59.
10. Suppose that  $7^{22}$  is written out in the ordinary way. What is its last digit?
11. A small theater has a student rate of \$3 per ticket and a regular rate of \$10 per ticket. Last night there was \$243 collected from the sale of tickets. There were more than 50 but less than 60 tickets sold. How many student tickets were sold?
12. Let  $a$  and  $b$  and  $n$  be natural numbers. Prove that if  $a^n$  and  $b^n$  are relatively prime, then  $a$  and  $b$  are relatively prime.
13. Let  $a$ ,  $b$ ,  $m$  and  $n$  be natural numbers with  $m$  and  $n$  greater than 1. Assume that  $m$  and  $n$  are relatively prime. Prove that if  $a \equiv b \pmod{m}$  and  $a \equiv b \pmod{n}$ , then  $a \equiv b \pmod{mn}$ .
14. Let  $a$  and  $b$  be natural numbers.
  - (a) Suppose there exist integers  $m$  and  $n$  such that  $am + bn = 1$ . Prove that  $a$  and  $b$  are relatively prime.
  - (b) Prove that  $5a + 2$  and  $7a + 3$  are relatively prime.
15. Let  $p$  be a prime number. Prove that  $\phi(p^2) = p^2 - p$ .

**Challenging Problems**

16. Suppose that  $a$  and  $b$  are relatively prime natural numbers such that  $ab$  is a perfect square. Show that  $a$  and  $b$  are each perfect squares.

- 
17. Show that if  $m$  and  $n$  are relatively prime and  $a$  and  $b$  are any integers, then there is an integer  $x$  that simultaneously satisfies the two congruences  $x \equiv a \pmod{m}$  and  $x \equiv b \pmod{n}$ .
18. Generalize the previous problem, as follows (this result is called the “Chinese Remainder Theorem”): If  $\{m_1, m_2, \dots, m_k\}$  is a collection of natural numbers greater than 1, each pair of which is relatively prime, and if  $\{a_1, a_2, \dots, a_k\}$  is any collection of integers, then there is an integer  $x$  that simultaneously satisfies all of the congruences  $x \equiv a_j \pmod{m_j}$ . Moreover, if  $x_1$  and  $x_2$  are both simultaneous solutions of all of those congruences, then  $x_1 \equiv x_2 \pmod{m_1 \cdot m_2 \cdots m_k}$ .
19. Using problem 13 in Chapter 2, prove that each pair of distinct Fermat numbers is relatively prime. (Note that this gives another proof that there are infinitely many prime numbers.)
20. Let  $p$  be an odd prime and let  $m = 2p$ . Prove that  $a^{m-1} \equiv a \pmod{m}$  for all natural numbers  $a$ .
21. Let  $a$  and  $b$  be relatively prime natural numbers greater than or equal to 2. Prove that  $a^{\phi(b)} + b^{\phi(a)} \equiv 1 \pmod{ab}$ .
22. For  $p$  a prime and  $k$  a natural number, show that  $\phi(p^k) = p^k - p^{k-1}$ .
23. If the canonical factorization of the natural number  $n$  into primes is  $n = p_1^{k_1} \cdot p_2^{k_2} \cdots p_m^{k_m}$ , prove that
- $$\phi(n) = (p_1^{k_1} - p_1^{k_1-1})(p_2^{k_2} - p_2^{k_2-1}) \cdots (p_m^{k_m} - p_m^{k_m-1}).$$
24. Suppose that all of  $a, b$  and  $c$  are natural numbers. Prove that there are at most a finite number of pairs of natural numbers  $(x, y)$  that satisfy  $ax + by = c$ .

## Chapter 8

# Rational Numbers and Irrational Numbers

So far, the only numbers that we have been discussing are the “whole numbers”; that is, the integers. There are many other interesting things that can be said about the integers but, for now, we will move on to consider other numbers, the “fractions”, or “the rational numbers”, and then the “real numbers”.

### 8.1 Rational Numbers

**Definition 8.1.1.** A *rational number* is a number of the form  $\frac{m}{n}$  where  $m$  and  $n$  are integers and  $n \neq 0$ .

Some examples of rational numbers are:  $\frac{3}{4}$ ,  $\frac{-7}{23}$ ,  $\frac{12}{-36}$ ,  $\frac{1}{2}$  and  $\frac{2}{4}$ .

Wait a minute. Are  $\frac{1}{2}$  and  $\frac{2}{4}$  different rational numbers? They are not; they are two different expressions representing the same number. Similarly  $\frac{12}{48} = \frac{1}{4}$ ,  $\frac{-7}{3} = \frac{7}{-3}$ ,  $\frac{16}{2} = \frac{8}{1}$ , and so on. The condition under which two different expressions as quotients of integers represent the same rational number is the following.

**Definition 8.1.2.** We define  $\frac{m_1}{n_1}$  to be equal to  $\frac{m_2}{n_2}$  when  $m_1 n_2 = m_2 n_1$ .

Thus when we use the representation  $\frac{1}{2}$  we recognize that we are representing a number that could also be denoted  $\frac{2}{4}$ ,  $\frac{-3}{-6}$  and so on.

Why don't we allow 0 denominators in the expressions for rational numbers? If we did allow 0 denominators, the arithmetic would be very peculiar. For example  $\frac{7}{0}$  would equal  $\frac{-12}{0}$  since  $7 \cdot 0 = -12 \cdot 0$ . In fact, we would have  $\frac{a}{0} = \frac{b}{0}$



for all integers  $a$  and  $b$ . It is not at all useful to have such peculiarities as part of our arithmetic, so we do not allow 0 to be a denominator of any rational number.

**Definition 8.1.3.** The set of all rational numbers is denoted  $\mathbb{Q}$ .

The operations of multiplication and addition of rational numbers can be defined in terms of the operations on integers.

**Definition 8.1.4.** The product of the rational numbers  $\frac{m_1}{n_1}$  and  $\frac{m_2}{n_2}$ , denoted  $\frac{m_1}{n_1} \cdot \frac{m_2}{n_2}$  or simply  $\frac{m_1 m_2}{n_1 n_2}$ , is defined to be  $\frac{m_1 m_2}{n_1 n_2}$ . The sum of the rational numbers  $\frac{m_1}{n_1}$  and  $\frac{m_2}{n_2}$  is defined by

$$\frac{m_1}{n_1} + \frac{m_2}{n_2} = \frac{m_1 n_2 + m_2 n_1}{n_1 n_2}.$$

We can think of the integers as the rational numbers whose denominator is 1; we invariably write them without the denominator. For example, we write -17 for  $\frac{-17}{1}$  (and also, of course, for  $\frac{-34}{2}$  and so on). In particular, we write 0 for  $\frac{0}{1}$  and 1 for  $\frac{1}{1}$ . Note that 0 and 1 are, respectively, additive and multiplicative identities for the rational numbers, as they are for the integers. That is,  $\frac{m}{n} + 0 = \frac{m}{n}$  and  $\frac{m}{n} \cdot 1 = \frac{m}{n}$  for every rational number  $\frac{m}{n}$ . Also note that, as is the case with the set of integers, every rational number has an additive inverse:  $\frac{m}{n} + \frac{-m}{n} = \frac{0}{n} = 0$ .

**Definition 8.1.5.** A *multiplicative inverse* for the number  $x$  is a number  $y$  such that  $xy = 1$ .

Of course, if  $xy = 1$  then  $y = \frac{1}{x}$ . If  $x$  and  $y$  are both integers, then, since  $\frac{1}{x}$  is an integer,  $x$  must be 1 or -1. Hence the only integers that have multiplicative inverses within the set of integers are the numbers 1 and -1. In the set of rational numbers, the situation is very different.

**Theorem 8.1.6.** If  $\frac{m}{n}$  is a rational number other than 0,  $\frac{m}{n}$  has a multiplicative inverse.

*Proof.* If  $\frac{m}{n} \neq 0$ , then  $m \neq 0$ . Therefore  $\frac{n}{m}$  is also a rational number and  $\frac{m}{n} \cdot \frac{n}{m} = \frac{mn}{nm} = \frac{1}{1} = 1$ . Therefore  $\frac{n}{m}$  is a multiplicative inverse for  $\frac{m}{n}$ .  $\square$

**Definition 8.1.7.** A *polynomial with integer coefficients* is an expression of the form

$$a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0$$

where  $n$  is a non-negative integer and the  $a_i$  are integers with  $a_n$  different from 0. The number  $x_0$  is a *root* (or *zero*) of a polynomial if the value of the polynomial obtained by replacing  $x$  by  $x_0$  is 0.

**Example 8.1.8.** The polynomial  $x^5 + x - 1$  does not have any rational roots.

*Proof.* Suppose that  $\frac{m}{n}$  was a rational root. Without loss of generality we can assume that  $m$  and  $n$  are relatively prime (if  $m$  and  $n$  had a common factor, that common factor could be divided out from  $m$  and  $n$ , getting an equivalent fraction). Substituting  $\frac{m}{n}$  in the polynomial would yield  $(\frac{m}{n})^5 + \frac{m}{n} - 1 = 0$ . This gives  $m^5 + mn^4 = n^5$  or  $m(m^4 + n^4) = n^5$ . It follows that any prime divisor of  $m$  is a divisor of  $n^5$  and hence also of  $n$ . Since  $m$  and  $n$  are relatively prime,  $m$  has no prime divisors. Thus  $m$  is either 1 or  $-1$ . Similarly, the above equation yields  $m^5 = n(n^4 - mn^3)$  from which it follows that any prime divisor of  $m$  would divide  $n$ . Thus  $n$  does not have any prime divisors, so  $n$  is either 1 or  $-1$ . Therefore the only possible values of  $\frac{m}{n}$  are 1 or  $-1$ . That is, the only possible rational roots of the polynomial are 1 and  $-1$ . However, it is clear that neither 1 nor  $-1$  is a root. Thus the polynomial does not have any rational roots.  $\square$

There is a general theorem, whose proof is similar to the above example, that is often useful in determining whether or not polynomials have rational roots and may also be used to find such roots.

**Theorem 8.1.9.** (*The Rational Roots Theorem*) If  $\frac{p}{q}$  is a rational root of the polynomial

$$a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0$$

and  $p$  and  $q$  are relatively prime, then  $p$  divides  $a_0$  and  $q$  divides  $a_n$ .

*Proof.* Assuming that  $\frac{p}{q}$  is a root gives

$$a_n \left(\frac{p}{q}\right)^n + a_{n-1} \left(\frac{p}{q}\right)^{n-1} + \dots + a_1 \left(\frac{p}{q}\right) + a_0 = 0.$$

Multiplying both sides of this equation by  $q^n$  produces the equation

$$a_n p^n + a_{n-1} p^{n-1} q + \dots + a_1 p q^{n-1} + a_0 q^n = 0.$$

It follows that

$$p(a_n p^{n-1} + a_{n-1} p^{n-2} q + \dots + a_1 q^{n-1}) = -a_0 q^n.$$

Since  $p$  and  $q$  are relatively prime,  $p$  and  $q^n$  are also relatively prime. On the other hand,  $p$  divides  $-a_0 q^n$ . Thus by Lemma 7.2.6,  $p$  divides  $a_0$ .

Similarly,

$$a_n p^n = -(a_{n-1} p^{n-1} q + \dots + a_1 p q^{n-1} + a_0 q^n),$$

so

$$a_n p^n = -q(a_{n-1} p^{n-1} + \cdots + a_1 p q^{n-2} + a_0 q^{n-1}).$$

Since  $q$  is relatively prime to  $p$  and  $q$  divides  $a_n p^n$ , it follows (by 7.2.6) that  $q$  divides  $a_n$ . This proves the theorem.  $\square$

**Example 8.1.10.** Find all the rational roots of the polynomial  $2x^3 - x^2 + x - 6$ .

By the Rational Roots Theorem, any rational root  $\frac{p}{q}$  in lowest terms satisfies  $q$  divides 2 and  $p$  divides 6. Thus the only possible values of  $q$  are  $1, -1, 2, -2$  and the only possible values of  $p$  are  $3, -3, 2, -2, 1, -1$ . The possible values of the quotient  $\frac{p}{q}$  are therefore  $\frac{3}{2}, -\frac{3}{2}, 3, -3, 2, -2, 1$  and  $-1$ . We can determine which of these possible roots actually are roots by simply substituting them for  $x$  and seeing if the result is 0. In this example the only root is  $\frac{3}{2}$ .

## 8.2 Irrational Numbers

In a sense, all actual computations, by human or electronic computers, are done with rational numbers. However, it is important, within mathematics itself and in using mathematics to understand the world, to have other numbers as well.

**Example 8.2.1.** Suppose that you walk one mile due east and then one mile due north. How far are you from your starting point? The straight line from your starting point to your final position is the hypotenuse of a right triangle whose legs are each one mile long. The length of the hypotenuse is the distance that you are from your starting point. If  $x$  denotes that distance, then Pythagoras tells us that  $x^2 = 2$ .

It is obviously useful to have *some* number that denotes that distance. Is there a rational number  $x$  such that  $x^2 = 2$ ? This question can be rephrased: are there integers  $m$  and  $n$  with  $n \neq 0$  such that  $(\frac{m}{n})^2 = 2$ ? This, of course, is equivalent to the question of whether there are integers  $m$  and  $n$  different from 0 that satisfy the equation  $m^2 = 2n^2$ . This is a very concrete question about integers; what is the answer?

**Theorem 8.2.2.** *There do not exist integers  $m$  and  $n$  with  $n \neq 0$  such that  $(\frac{m}{n})^2 = 2$ .*

*Proof.* Suppose that there did exist such  $m$  and  $n$ . We will show that that assumption leads to a contradiction. From  $(\frac{m}{n})^2 = 2$  it would follow that  $m^2 = 2n^2$ . The equation  $m^2 = 2n^2$  implies that  $m^2$  is an even number, since it is the

product of 2 and another number. What about  $m$  itself? If  $m$  was odd then  $m - 1$  would have to be even, so  $m - 1 = 2k$  for some integer  $k$ , or  $m = 2k + 1$ . It would follow from this that  $m^2 = (2k + 1)^2 = 4k^2 + 4k + 1 = 2(2k^2 + 2k) + 1$ , which is an odd number (since it is 1 more than a multiple of 2). Thus if  $m$  was odd,  $m^2$  would have to be odd. Since  $m^2$  is even, we conclude that  $m$  is even. Therefore  $m = 2s$  for some integer  $s$ , from which it follows that  $m^2 = 4s^2$ . Substituting  $4s^2$  for  $m^2$  in the equation  $m^2 = 2n^2$ , gives  $4s^2 = 2n^2$ , or  $2s^2 = n^2$ . Thus  $n^2$  is an even number and, reasoning as we did above for  $m$ , it follows that  $n$  itself is an even number.

What have we proven so far? We have proven that  $m^2 = 2n^2$  implies that both  $m$  and  $n$  are even. But if  $\frac{m}{n}$  is any rational number with  $m$  and  $n$  both even integers, then the common factor of 2 can be “divided out” from both  $m$  and  $n$ , which gives an expression of the number with numerator and denominator each half of the corresponding part of the original representation of the number. This process of dividing out 2’s can be repeated until at least one of the numerator and denominator is odd. That is, there exists  $m_0$  and  $n_0$  different from 0 such that at least one of  $m_0$  and  $n_0$  is odd, and  $\frac{m}{n} = \frac{m_0}{n_0}$ . Then  $(\frac{m_0}{n_0})^2 = 2$ , so the above reasoning would imply that both of  $m_0$  and  $n_0$  are even, which contradicts the fact that at least one of them is odd.  $\square$

We have proven that there is no rational number that satisfies the equation  $x^2 = 2$ . Is there any number that satisfies this equation? It would obviously be useful to have such a number, for the purpose of specifying how far a person in Example 8.2.1 is from the person’s starting point and for many other purposes. Mathematicians have developed what are called “the real numbers”; the real numbers include numbers for every possible distance. The real numbers can be put into correspondence with the points on a line by labeling one point “0” and marking points to the right of 0 with the distances that they are from 0 (using any fixed units). Points on the line to the left of 0 are labeled with corresponding negative real numbers. The resulting “real number line” looks like:

The set of real numbers and the arithmetical operations on them can be precisely constructed in terms of rational numbers. In fact, there are several ways to do that. None of the ways of constructing the real numbers in terms of the rational numbers are easy; they all require substantial development. There are two main approaches, one using “Cauchy Sequences” and the other using “Dedekind Cuts”. If you are interested in learning about these constructions you can simply search the internet for “construction of the real numbers”. We will not describe any construction of the real numbers, we will simply assume that the real numbers exist and the arithmetical operations on them have the

usual properties.

**Definition 8.2.3.** The set of all real numbers is commonly denoted by  $\mathbb{R}$ .

It can be shown that there is a positive real number  $x$  such that  $x^2 = 2$ . This number is denoted  $\sqrt{2}$  or  $2^{\frac{1}{2}}$ .

**Definition 8.2.4.** A real number that is not a rational number is said to be *irrational*.

The theorem that we just proved can be rephrased as follows.

**Theorem 8.2.5.** *The number  $\sqrt{2}$  is irrational.*

*Proof.* The proof of Theorem 8.2.2 shows that  $\sqrt{2}$  is not a rational number.  $\square$

The symbol  $\sqrt{3}$  represents the positive real number satisfying  $(\sqrt{3})^2 = 3$ ; is  $\sqrt{3}$  irrational too?

We can establish a more general result.

**Theorem 8.2.6.** *If  $p$  is a prime number, then  $\sqrt{p}$  is irrational.*

*Proof.* The proof will be similar to that of the special case  $p = 2$ . Suppose that  $n \neq 0$  and  $m$  and  $n$  are integers satisfying  $(\frac{m}{n})^2 = p$ . Then  $m^2 = pn^2$ . Since  $m^2 = pn^2$ ,  $p$  divides  $m^2$ . Thus  $p$  divides the product  $m \cdot m$ , from which it follows that  $p$  divides at least one of the factors (see 4.1.3); that is,  $p$  divides  $m$ . Therefore there is an integer  $s$  such that  $m = ps$ , which gives  $(ps)^2 = pn^2$ . Dividing both sides of this equation by  $p$  gives  $ps^2 = n^2$ . Thus  $p$  divides the product  $n \cdot n$  and we conclude that  $p$  divides  $n$ . This shows that whenever  $\frac{m}{n}$  is a rational number with  $(\frac{m}{n})^2 = p$ , both  $m$  and  $n$  are divisible by  $p$ . As in the case where  $p = 2$  (see 8.2.2), the fact that common factors of numerators and denominators of fractions can be “divided out” leads to a contradiction.  $\square$

Of course, some natural numbers do have rational square roots. For example,  $\sqrt{1} = 1$ ,  $\sqrt{4} = 2$  and  $\sqrt{289} = 17$ . What about  $\sqrt{6}$ ? More generally is there a natural number  $m$  such that  $\sqrt{m}$  is rational but  $\sqrt{m}$  is not an integer? To answer this question it is useful to begin with the following.

**Lemma 8.2.7.** *A natural number other than 1 is a perfect square (i.e., is the square of a natural number) if and only if every prime number in its canonical factorization occurs to an even power.*

*Proof.* Let  $n$  be a natural number. If the canonical factorization of  $n$  (see 4.1.2) is  $n = p_1^{\alpha_1} \cdot p_2^{\alpha_2} \cdots p_k^{\alpha_k}$ , then  $n^2 = p_1^{2\alpha_1} \cdot p_2^{2\alpha_2} \cdots p_k^{2\alpha_k}$ . The uniqueness of the factorization into primes implies that this expression is the canonical factorization of  $n^2$ . All the exponents are obviously even. This proves that the square of every natural number has the property that every exponent in its canonical factorization is even. The converse is even easier. For if  $n^2 = p_1^{2\alpha_1} \cdot p_2^{2\alpha_2} \cdots p_k^{2\alpha_k}$ , then obviously  $n = (p_1^{\alpha_1} \cdot p_2^{\alpha_2} \cdots p_k^{\alpha_k})$ .  $\square$

**Theorem 8.2.8.** *If the square root of a natural number is rational, then the square root is an integer.*

*Proof.* Suppose that  $N$  is a natural number and that the square root of  $N$  is rational. The case  $N = 1$  poses no difficulties. If  $N$  is greater than 1, let its canonical factorization be  $p_1^{\alpha_1} \cdot p_2^{\alpha_2} \cdots p_t^{\alpha_t}$ . Since  $\sqrt{N}$  is rational, there exist natural numbers  $m$  and  $n$  such that  $\sqrt{N} = \frac{m}{n}$ . Let the canonical factorizations of  $m$  and  $n$  respectively be  $m = q_1^{\beta_1} \cdot q_2^{\beta_2} \cdots q_u^{\beta_u}$  and  $n = r_1^{\gamma_1} \cdot r_2^{\gamma_2} \cdots r_v^{\gamma_v}$ . Since  $N = \frac{m^2}{n^2}$ , it follows that  $n^2 N = m^2$ . In terms of the canonical factorizations of  $N, n$  and  $m$ , this yields

$$(r_1^{\gamma_1} \cdot r_2^{\gamma_2} \cdots r_v^{\gamma_v})^2 p_1^{\alpha_1} \cdot p_2^{\alpha_2} \cdots p_t^{\alpha_t} = (q_1^{\beta_1} \cdot q_2^{\beta_2} \cdots q_u^{\beta_u})^2.$$

It follows that

$$r_1^{2\gamma_1} \cdot r_2^{2\gamma_2} \cdots r_v^{2\gamma_v} \cdot p_1^{\alpha_1} \cdot p_2^{\alpha_2} \cdots p_t^{\alpha_t} = q_1^{2\beta_1} \cdot q_2^{2\beta_2} \cdots q_u^{2\beta_u}.$$

We want to prove that each of the  $\alpha_i$  is even. By the uniqueness of the factorization into primes, each  $p_i$  is one of the  $q_j$  for some  $j$ . Since  $2\beta_j$  is even,  $p_i$  occurs to an even power on both sides of the equation. Of course,  $p_i$  could be one of the  $r$ 's. If so, since the powers of all the  $r$ 's are even, the total power that  $p_i$  occurs to on the left hand side of the equation is the sum of  $\alpha_i$  and an even number. Since this sum must be the even number  $2\beta_j$ , it follows that  $\alpha_i$  is even. Thus every  $\alpha_i$  is even and the above lemma (8.2.7) implies that  $N$  is the square of an integer.  $\square$

**Example 8.2.9.** The number  $\sqrt[3]{4}$  is irrational.

*Proof.* If  $\sqrt[3]{4} = \frac{m}{n}$  with  $m$  and  $n$  integers, then  $4n^3 = m^3$ . Write this equation in terms of the canonical factorizations of  $m$  and  $n$ , getting

$$4(p_1^{\alpha_1} \cdot p_2^{\alpha_2} \cdots p_r^{\alpha_r})^3 = (q_1^{\beta_1} \cdot q_2^{\beta_2} \cdots q_s^{\beta_s})^3$$

So

$$2^2 \cdot p_1^{3\alpha_1} \cdot p_2^{3\alpha_2} \cdots p_r^{3\alpha_r} = q_1^{3\beta_1} \cdot q_2^{3\beta_2} \cdots q_s^{3\beta_s}$$

The prime 2 must occur to a power that is a multiple of 3, since every prime on the right hand side of this equation occurs to such a power. On the other hand, 2 occurs on the left hand side of the equation to a power that is two more than a multiple of 3. The uniqueness of the factorization into primes implies that no such equation is possible.  $\square$

**Example 8.2.10.** The number  $\sqrt{3} + \sqrt{5}$  is irrational.

*Proof.* Suppose that  $\sqrt{3} + \sqrt{5} = r$  with  $r$  a rational number. Then  $\sqrt{3} = r - \sqrt{5}$ . Squaring both sides of this equation gives

$$3 = (r - \sqrt{5})^2 = r^2 - 2\sqrt{5}r + 5$$

From this it would follow that  $2\sqrt{5}r = r^2 + 2$  or  $\sqrt{5} = \frac{r^2+2}{2r}$ . But  $r$  rational implies that  $\frac{r^2+2}{2r}$  is rational, which contradicts the fact that  $\sqrt{5}$  is irrational.  $\square$

The following is a question with an interesting answer: Do there exist two irrational numbers such that one of them to the power of the other is rational? That is, can  $x^y$  be rational if  $x$  and  $y$  are both irrational? A case that appears to be simple is that of  $(\sqrt{3})^{\sqrt{2}}$ ? In fact, however, it is not at all easy to determine whether or not  $(\sqrt{3})^{\sqrt{2}}$  is rational. Nonetheless, this example can still be used to prove that the general question has an affirmative answer, as follows. Either  $(\sqrt{3})^{\sqrt{2}}$  is rational or it is irrational. If it is rational, it furnishes an example showing that the answer to the question is affirmative. If  $(\sqrt{3})^{\sqrt{2}}$  is irrational let  $x = (\sqrt{3})^{\sqrt{2}}$  and  $y = \sqrt{2}$ . Then  $x^y$  is an irrational number to an irrational power. But  $x^y = ((\sqrt{3})^{\sqrt{2}})^{\sqrt{2}} = \sqrt{3}^{\sqrt{2}\sqrt{2}} = \sqrt{3}^2 = 3$ . This gives an affirmative answer in this case as well. In other words  $\sqrt{3}^{\sqrt{2}}$  answers our original question, whether or not it itself is rational or irrational. In fact,  $(\sqrt{3})^{\sqrt{2}}$  is irrational, as follows from the Gelfond-Schneider Theorem, a theorem that is very difficult to prove.

## 8.3 Problems

### Basic Exercises

1. Use the Rational Roots Theorem to find all the rational roots of each of the following polynomials (some may not have any rational roots at all).

- (a)  $x^2 + 5x + 2$
  - (b)  $2x^2 - 5x^2 + 14x - 35$
  - (c)  $x^{10} - x + 1$
2. Show that  $\sqrt[3]{5}$  is irrational.
3. Show that  $\sqrt{\frac{1}{2}}$  is irrational.
4. Must the sum of an irrational number and a rational number be irrational?
5. Must an irrational number to a rational power be irrational?
6. Must the sum of two irrational numbers be irrational?
7. If  $y$  is irrational and  $x$  is any rational number other than 0, show that  $xy$  is irrational.

**Interesting Problems**

8. Use the Rational Roots Theorem to find all the rational roots of each of the following polynomials (some may not have any rational roots at all).
- (a)  $x^9 - x^8 + x^7 - x^2 + x - 1$
  - (b)  $7x^5 - x^4 - x^3 - x^2 - 2x - 2$
  - (c)  $2x^4 + 7x^3 + 5x^2 + 7x + 3$
9. Determine whether each of the following numbers is rational or irrational and prove that your answer is correct:

(a)  $32^{\frac{2}{3}}$

(b)  $28^{\frac{2}{5}}$

(c)  $\frac{\sqrt{7}}{\sqrt{5}}$

(d)  $\frac{\sqrt{7}}{\sqrt[3]{15}}$

(e)  $\frac{\sqrt{63}}{\sqrt{28}}$

(f)  $\sqrt{\frac{3}{8}}$

(g)  $\sqrt[7]{\frac{8}{9}}$

**Challenging Problems**

10. Prove that the following numbers are irrational:

(a)  $\sqrt{5} + \sqrt{7}$



(b)  $\sqrt[3]{4} + \sqrt{10}$

(c)  $\sqrt[3]{5} + \sqrt{3}$

(d)  $\sqrt{3} + \sqrt{5} + \sqrt{7}$

11. Suppose that  $a$  and  $b$  are odd natural numbers and  $a^2 + b^2 = c^2$ . Prove that  $c$  is irrational.
12. Prove that  $\sqrt{3} - \frac{\sqrt{5}}{17}$  is irrational.
13. Prove that if  $a$  and  $b$  are natural numbers and  $k$  is a natural number such that  $k^{\frac{a}{b}}$  is rational, then  $k^{\frac{a}{b}}$  is a natural number.

## Chapter 9

# The Complex Numbers

The set of real numbers is rich enough to be useful in a wide variety of situations. In particular, it provides a number for every distance. There are, however, some situations where additional numbers are required.

### 9.1 What is a Complex Number?

Let's consider the problem of finding roots for polynomial equations. Recall that polynomials are expressions such as  $7x^2 + 5x - 3$ , and  $\sqrt{2}x^3 + \frac{5}{7}x$ , and  $x^7 - 1$ . The general definition is the following.

**Definition 9.1.1.** A *polynomial* is an expression of the form

$$a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0$$

where  $n$  is a natural number and the  $a_i$  are numbers with  $a_n \neq 0$ . The  $a_i$  are called the *coefficients* of the polynomial. The natural number  $n$ , the highest power to which  $x$  occurs in the polynomial, is called the *degree* of the polynomial. We allow the case where  $n = 0$ ; i.e., where the polynomial is simply  $a_0$ . (Such a polynomial is called a *constant polynomial*.)

Note that in the definition of polynomial we used  $x$  as the variable; this is very standard. However it is often the case that other variables are used as well. For example  $z^3 - 4z + 3$  would be a polynomial in the variable  $z$ .

A polynomial defines a function; whenever any specific number is substituted for  $x$ , the resulting expression is some number. The values of  $x$  for which the polynomial is 0 have special significance.

**Definition 9.1.2.** A *root* or *zero* of the polynomial  $a_nx^n + a_{n-1}x^{n-1} + \dots + a_1x + a_0$  is a number that when substituted for  $x$  makes

$$a_nx^n + a_{n-1}x^{n-1} + \dots + a_1x + a_0 = 0$$

For example, 2 is a root of the polynomial  $x^2 - 4$ , 3 is a root of the polynomial  $5x^2 - 2x - 39$ ,  $-\frac{7}{5}$  is a root of the polynomial  $5x + 7$ , and so on.

A very natural question is: Which polynomials have roots? All polynomials of degree 1 have roots: the polynomial  $a_1x + a_0$  has the root  $-\frac{a_0}{a_1}$ . What about polynomials of degree two? A simple example is the polynomial  $x^2 + 1$ . No real number is a root of that polynomial, since  $x^2$  is non-negative for every real number  $x$ , and therefore  $x^2 + 1$  is strictly greater than 0 for every real  $x$ . If the polynomial  $x^2 + 1$  is to have a root, it would have to be in a larger number system than that of the real numbers. Such a system was invented by mathematicians hundreds of years ago.

We use the symbol  $i$  to denote a root of the polynomial  $x^2 + 1$ . That is, we define  $i^2$  to be equal to  $-1$ . We then combine this symbol  $i$  with real numbers, using standard manipulations of algebra in the usual ways, to get the complex numbers. The definition is the following.

**Definition 9.1.3.** A *complex number* is an expression of the form  $a + bi$  where  $a$  and  $b$  are real numbers. The real number  $a$  is called the *real part* of  $a + bi$  and the real number  $b$  is called the *imaginary part* of  $a + bi$ . We sometimes use the notation  $\text{Re}(z)$  and  $\text{Im}(z)$  to denote the real and imaginary parts of the complex number  $z$  respectively. Addition of complex numbers is defined by

$$(a + bi) + (c + di) = (a + c) + (b + d)i.$$

Multiplication of complex numbers is defined by

$$(a + bi)(c + di) = ac + adi + bic + bdi^2 = ac + bdi^2 + (ad + bc)i = (ac - bd) + (ad + bc)i,$$

where we replaced  $i^2$  by  $-1$  to get the last equation.

**Example 9.1.4.**

$$(6 + 2i) + (-4 + 5i) = 2 + 7i$$

$$(-\sqrt{12} + \sqrt{6}i) + (4 + \pi i) = (-\sqrt{12} + 4) + (\sqrt{6} + \pi)i$$

$$(7 + 2i)(3 - 4i) = 21 + 6i - 28i - 8i^2 = 21 - 22i - 8(-1) = 21 + 8 - 22i = 29 - 22i$$

We use the symbol  $0$  as an abbreviation for the complex number  $0+0i$ . More generally, we use  $a$  as an abbreviation for the complex number  $a+0i$ . Similarly we use  $bi$  as an abbreviation for the complex number  $0+bi$ . When  $r(=r+0i)$  is a real number,  $r(a+bi)$  is simply  $ra+rb i$ .

Note that every complex number has an additive inverse (that is, a complex number that gives  $0$  when added to the given number). For example, the additive inverse of  $-7+\sqrt{2}i$  is  $7-\sqrt{2}i$ . In general, the additive inverse of  $(a+bi)$  is  $-a+(-b)i$ .

There is an important relationship between the complex number  $a+bi$  and the complex number  $a-bi$ .

**Definition 9.1.5.** The number  $a-bi$  is called the *complex conjugate* of the number  $a+bi$ ; the complex conjugate of a complex number is often denoted by placing a horizontal bar over the complex number:  $\overline{a+bi} = a-bi$ .

**Example 9.1.6.** The complex conjugate of  $2+3i$  is  $2-3i$ , or  $\overline{2+3i} = 2-3i$ . Similarly,  $-\sqrt{3}-5i = -\sqrt{3}+5i$ , and  $\frac{2}{9}-\sqrt{2}i = \frac{2}{9}+\sqrt{2}i$ .

The product of a complex number and its conjugate is important.

**Theorem 9.1.7.** For any complex number  $a+bi$ ,  $(a+bi)(a-bi) = a^2+b^2$ .

*Proof.* Simply multiplying gives the result. □

**Definition 9.1.8.** The *modulus* of the complex number  $a+bi$  is  $\sqrt{a^2+b^2}$ ; it is often denoted  $|a+bi|$ .

Thus  $(a+bi)(\overline{a+bi}) = |a+bi|^2$ .

Do complex numbers have multiplicative inverses? That is, given  $a+bi$ , is there a complex number  $c+di$  such that  $(a+bi)(c+di) = 1$ ? Of course, the complex number  $0$  cannot have a multiplicative inverse since its product with any complex number is  $0$ . What about other complex numbers?

Given a complex number  $a+bi$ , let's try to compute a multiplicative inverse  $c+di$  for it. Suppose that  $(a+bi)(c+di) = 1$ . Multiplying both sides of this equation by  $\overline{a+bi}$  and using the fact that  $(\overline{a+bi})(a+bi) = a^2+b^2$  yields  $(a^2+b^2)(c+di) = a-bi$ . Since  $a^2+b^2$  is a real number, this implies (unless  $a^2+b^2 = 0$ ) that  $c+di = \frac{a}{a^2+b^2} - \frac{b}{a^2+b^2}i$ . Note that if  $a^2+b^2 = 0$ , then  $a=b=0$ , so the number  $a+bi$  is  $0$ . Thus if  $a+bi$  has a multiplicative inverse, that multiplicative inverse must be  $\frac{a}{a^2+b^2} - \frac{b}{a^2+b^2}i$ . In fact, as we now show, that expression is a multiplicative inverse for  $a+bi$ .

**Theorem 9.1.9.** If  $a + bi \neq 0$ , then  $\frac{a}{a^2+b^2} - \frac{b}{a^2+b^2}i$  is a multiplicative inverse for  $a + bi$ .

*Proof.* We verify this by simply multiplying:  $(a + bi) \left( \frac{a}{a^2+b^2} - \frac{b}{a^2+b^2}i \right)$  is equal to  $\frac{a^2}{a^2+b^2} + \frac{b^2}{a^2+b^2} - \frac{ab}{a^2+b^2}i + \frac{ab}{a^2+b^2}i$ , which simplifies to  $\frac{a^2+b^2}{a^2+b^2} = 1$ .  $\square$

## 9.2 The Complex Plane

It is very useful to represent complex numbers in a coordinatized plane. We let the complex number  $a + bi$  correspond to the point  $(a, b)$  in the ordinary  $xy$  plane. Note that the modulus  $|a + bi|$  is the distance from  $(a, b)$  to the origin.

**Definition 9.2.1.** For a complex number  $a + bi$  other than 0, the angle that the line from  $(0, 0)$  to  $(a, b)$  makes with the positive  $x$  axis is called the *argument* of  $a + bi$ .

In day to day life, angles are usually measured in degrees: a right angle is  $90^\circ$ , a straight angle is  $180^\circ$ , and an angle of  $37^\circ$  is  $\frac{37}{180}$  of a straight angle. For doing mathematics, however, it is almost always more convenient to measure angles differently.

**Definition 9.2.2.** The *radian measure* of the angle  $\theta$  is the length of the arc of a circle of radius 1 that is cut off by an angle  $\theta$  at the centre of the circle. (See the diagram below.)

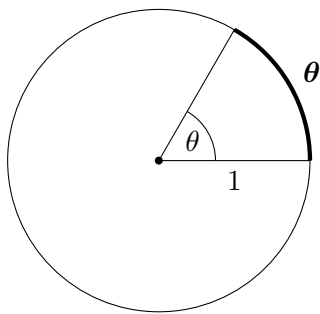


Figure 9.1

Thus, since a circle of radius 1 has circumference  $2\pi$ , the radian measure of a right angle is  $\frac{\pi}{2}$ , of a straight angle is  $\pi$ , of an angle of  $60^\circ$  is  $\frac{\pi}{3}$  and so on. Note that  $2\pi$  is a full revolution. Therefore, for any natural number  $k$ , the angle  $\theta + 2\pi k$  measured from the positive  $x$  axis ends up at the same position as  $\theta$ . We will use the radian measure of angles for the rest of this chapter.

We require the basic properties of the trigonometric functions sine, cosine, and tangent.

If the complex number  $a+bi$  has modulus  $r$  and argument  $\theta$ , then  $a = r \cos \theta$ , and  $b = r \sin \theta$ . To see this, first consider the case where both  $a$  and  $b$  are greater than or equal to 0, which is equivalent to  $\theta$  being an angle between 0 and  $\frac{\pi}{2}$ . Then the situation is as in Figure 9.2 below. The fact that the cosine of an angle of a right triangle is its adjacent side divided by its hypotenuse gives  $\cos \theta = \frac{a}{r}$ , or  $a = r \cos \theta$ . Similarly, the fact that sine of  $\theta$  is the opposite side divided by the hypotenuse gives  $\sin \theta = \frac{b}{r}$ , or  $b = r \sin \theta$ . Similar analysis yields the same equations when 1 or more of  $a$  and  $b$  is negative, and thus the conclusion holds for any  $\theta$ . Each complex number is determined by its modulus  $r$  and its argument  $\theta$ ; that is,  $a+bi = r(\cos \theta + i \sin \theta)$ . (The only complex number whose modulus is 0 is the number 0, and 0 is the only complex number whose argument is not defined.)

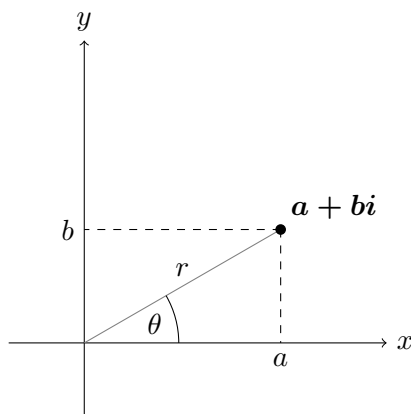


Figure 9.2

**Definition 9.2.3.** The *polar form* of the complex number with modulus  $r$  and argument  $\theta$  is  $r(\cos \theta + i \sin \theta)$ .

One reason that the polar form is important is because there is a neat description of multiplication of complex numbers in terms of their moduli and arguments.

**Theorem 9.2.4.** *The modulus of the product of two complex numbers is the product of their moduli. The argument of the product of two complex numbers is the sum of their arguments.*

*Proof.* Simply multiplying the two complex numbers  $r_1(\cos \theta_1 + i \sin \theta_1)$  and

$r_2(\cos \theta_2 + i \sin \theta_2)$  and collecting terms yields

$$r_1 r_2 ((\cos \theta_1 \cos \theta_2 - \sin \theta_1 \sin \theta_2) + i(\cos \theta_1 \sin \theta_2 + \sin \theta_1 \cos \theta_2)).$$

Recall the addition formulae for cosine and sine:

$$\cos(\theta_1 + \theta_2) = \cos \theta_1 \cos \theta_2 - \sin \theta_1 \sin \theta_2$$

and

$$\sin(\theta_1 + \theta_2) = \sin \theta_1 \cos \theta_2 + \sin \theta_2 \cos \theta_1.$$

Using these addition formulae on the right hand side of the above equation shows that the product is equal to

$$r_1 r_2 (\cos(\theta_1 + \theta_2) + i \sin(\theta_1 + \theta_2)).$$

This proves the theorem.  $\square$

Thus to multiply two complex numbers we can simply multiply their moduli and add their arguments. In particular, the case where the two complex numbers are equal shows that the square of a complex number is obtained by squaring its modulus and doubling its argument. One application of this fact is the following.

**Theorem 9.2.5.** *Every complex number has a complex square root.*

*Proof.* To show that any given complex number has a square root, write the number in polar form, say  $z = r(\cos \theta + i \sin \theta)$ . Let  $w$  equal  $\sqrt{r}(\cos \frac{\theta}{2} + i \sin \frac{\theta}{2})$ . By the previous theorem,  $w^2 = z$ .  $\square$

It is also easy to compute powers higher than the second.

**Theorem 9.2.6** (De Moivre's Theorem). *For every natural number  $n$ ,*

$$(r(\cos \theta + i \sin \theta))^n = r^n(\cos n\theta + i \sin n\theta).$$

*Proof.* This is easily established by induction on  $n$ . The case  $n = 1$  is clear. Suppose that the formula holds for  $n = k$ ; that is, suppose

$$(r(\cos \theta + i \sin \theta))^k = r^k(\cos k\theta + i \sin k\theta).$$

Multiplying both sides of this equation by  $r(\cos \theta + i \sin \theta)$  and using Theorem 9.2.4 gives

$$\begin{aligned} (r(\cos \theta + i \sin \theta))^{k+1} &= r^k(\cos k\theta + i \sin k\theta)r(\cos \theta + i \sin \theta) \\ &= r \cdot r^k(\cos(k\theta + \theta) + i \sin(k\theta + \theta)) \\ &= r^{k+1}(\cos((k+1)\theta) + i \sin((k+1)\theta)). \end{aligned}$$

This is the formula for  $n = k + 1$ , so the theorem is established by mathematical induction. □

De Moivre's Theorem leads to some very nice computations, such as the following.

**Example 9.2.7.** We can compute  $(1 + i)^8$  as follows. First,  $|1 + i| = \sqrt{2}$ . Plotting  $1 + i$  as the point (1,1) in the plane makes it apparent that the argument of  $1 + i$  is  $\frac{\pi}{4}$ . Thus by De Moivre's Theorem (9.2.6), the modulus of  $(1 + i)^8$  is  $\sqrt{2}^8 = 2^4 = 16$ . The argument of  $(1 + i)^8$  is  $8 \cdot \frac{\pi}{4} = 2\pi$ . It follows that

$$(1 + i)^8 = 16(\cos 2\pi + i \sin 2\pi) = 16.$$

Therefore  $(1 + i)^8 = 16$ .

The following is a very similar computation.

**Example 9.2.8.**

$$(1 + i)^{100} = [\sqrt{2}(\cos \frac{\pi}{4} + i \sin \frac{\pi}{4})]^{100} = 2^{50}(\cos 25\pi + i \sin 25\pi).$$

Since the angle with the positive  $x$  axis of  $25\pi$  is in the same position as the angle of  $\pi$  radians, it follows that

$$(1 + i)^{100} = 2^{50}(\cos \pi + i \sin \pi) = 2^{50}(-1 + 0) = -2^{50}.$$

It is interesting to compute the roots of the complex number 1. The number 1 is sometimes called "unity" in this context.

**Example 9.2.9** (Square Roots of Unity). Obviously,  $1^2 = 1$  and  $(-1)^2 = 1$ . Are there any other complex square roots of 1? To compute the square roots of 1 we can proceed as follows. Let  $z = r(\cos \theta + i \sin \theta)$  then  $z^2 = r^2(\cos 2\theta + i \sin 2\theta)$ . If  $z^2 = 1$  then  $r^2$  must be the modulus of 1; i.e.,  $r^2 = 1$ . Since  $r$  is non-negative,

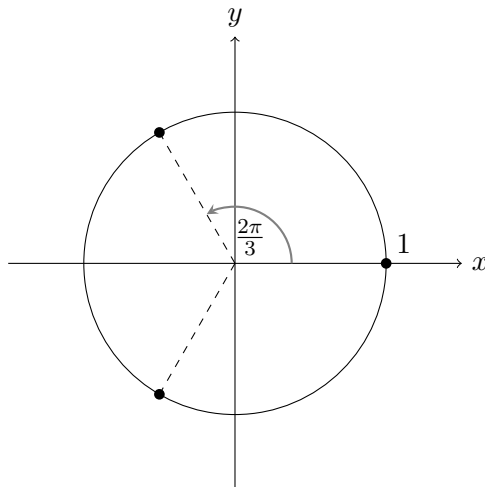


it follows that  $r = 1$ . Also,  $(\cos 2\theta + i \sin 2\theta) = 1$ . Therefore  $\cos 2\theta = 1$  and  $i \sin 2\theta = 0$ . What are the possible values of  $\theta$ ? Clearly  $\theta = 0$  is one solution, as is  $\theta = \pi$ ; the corresponding values of  $z$  are  $z = \cos 0 + i \sin 0 = 1$  and  $z = \cos \pi + i \sin \pi = -1$ . Are there any other possible values of  $\theta$ ? Of course there are:  $\theta$  could be  $2\pi$  or  $3\pi$  or  $4\pi$  or  $5\pi$ . If  $\theta$  is any multiple of  $\pi$  then  $\cos 2\theta = 1$  and  $\sin 2\theta = 0$ . However, we do not get any new values of  $z$  by using those other values of  $\theta$ . We only get  $z = 1$  or  $z = -1$  depending upon whether we have an even or an odd multiple of  $\pi$ . It is easily seen that only the multiples of  $\pi$  simultaneously satisfy the equations  $\cos 2\theta = 1$  and  $\sin 2\theta = 0$ . This follows from the fact that  $\cos \phi = 1$  only when  $\phi$  is a multiple of  $2\pi$ , so  $\cos 2\theta = 1$  only when  $\theta$  is a multiple of  $\pi$ . Thus the only complex square roots of 1 are 1 and  $-1$ .

Cube roots of unity are more interesting. The only real number  $z$  satisfying  $z^3 = 1$  is  $z = 1$ . However, there are other complex numbers satisfying this equation.

**Example 9.2.10** (Cube Roots of Unity). Suppose that  $z = r(\cos \theta + i \sin \theta)$  and  $z^3 = 1$ . Then clearly  $r = 1$ . By De Moivre's Theorem (9.2.6),  $z^3 = \cos 3\theta + i \sin 3\theta$ . From  $z^3 = 1$  we get  $\cos 3\theta = 1$  and  $i \sin 3\theta = 0$ . These equations are, of course, satisfied by  $\theta = 0$ , which gives  $z = \cos 0 + i \sin 0 = 1$ , the obvious cube root of 1. But also  $\cos 3\theta = 1$  and  $\sin 3\theta = 0$  when  $3\theta = 2\pi$ . That is, when  $\theta = \frac{2\pi}{3}$ . Thus  $z = \cos \frac{2\pi}{3} + i \sin \frac{2\pi}{3} = -\frac{1}{2} + \frac{\sqrt{3}}{2}i$  is another cube root of 1. (Once it is conjectured that  $-\frac{1}{2} + \frac{\sqrt{3}}{2}i$  is a cube root of 1, that could be verified by simply computing  $(-\frac{1}{2} + \frac{\sqrt{3}}{2}i)^3$ ). There is another cube root of 1. If  $3\theta = 4\pi$ , then  $\cos 3\theta = 1$  and  $\sin 3\theta = 0$ . Thus  $z = \cos \frac{4\pi}{3} + i \sin \frac{4\pi}{3} = -\frac{1}{2} - \frac{\sqrt{3}}{2}i$  is another cube root of 1. Therefore we have found three cube roots of 1: 1,  $-\frac{1}{2} + \frac{\sqrt{3}}{2}i$  and  $-\frac{1}{2} - \frac{\sqrt{3}}{2}i$ . Are there any other cube roots of 1? If  $3\theta = 6\pi$  then  $\cos 3\theta = 1$  and  $\sin 3\theta = 0$ . When  $3\theta = 6\pi$ ,  $\theta = 2\pi$ . Thus  $\cos \theta + i \sin \theta$  is simply 1, so we are not getting an additional cube root. More generally, for every integer  $k$ ,  $\cos 2k\pi = 1$  and  $\sin 2k\pi = 0$ . However, if  $3\theta = 2k\pi$ , then there are only the three different values given above for  $\cos \theta + i \sin \theta$ , since all the values of  $\theta$  obtained from other values of  $k$  differ from one of  $0$ ,  $\frac{4\pi}{3}$  and  $\frac{2\pi}{3}$  by a multiple of  $2\pi$ .

It is interesting to plot the three cube roots of unity in the plane.



The three cube roots of unity are obtained by starting at the point 1 on the circle of radius 1 and then moving in a counterclockwise direction  $\frac{2\pi}{3}$  to get the next cube root, and then moving an additional  $\frac{2\pi}{3}$  to get the third cube root.

For each natural number  $n$ , the complex  $n^{\text{th}}$  roots of 1 can be obtained by starting at 1 and successively moving around the unit circle in a counterclockwise direction through angles  $\frac{2\pi}{n}$ .

**Example 9.2.11** ( $n^{\text{th}}$  Roots of Unity). For each natural number  $n$ , the complex  $n^{\text{th}}$  roots of 1 are the numbers  $1, \cos \frac{2\pi}{n} + i \sin \frac{2\pi}{n}, \cos \frac{4\pi}{n} + i \sin \frac{4\pi}{n}, \cos \frac{6\pi}{n} + i \sin \frac{6\pi}{n}, \cos \frac{8\pi}{n} + i \sin \frac{8\pi}{n}, \dots, \cos \frac{2\pi(n-1)}{n} + i \sin \frac{2\pi(n-1)}{n}$ .

To see this, first note that, for any natural number  $k$ ,

$$\left( \cos \frac{2\pi k}{n} + i \sin \frac{2\pi k}{n} \right)^n = \cos 2\pi k + i \sin 2\pi k$$

by De Moivre's Theorem (9.2.6). Since  $\cos 2\pi k + i \sin 2\pi k = 1$ , this shows that each of  $\cos \frac{2\pi k}{n} + i \sin \frac{2\pi k}{n}$  is an  $n^{\text{th}}$  root of unity. To show that these are the only  $n^{\text{th}}$  roots of unity we proceed as follows. Suppose that  $z = \cos \theta + i \sin \theta$  and  $z^n = 1$ . Then  $\cos n\theta + i \sin n\theta = 1$ , so  $\cos n\theta = 1$  and  $\sin n\theta = 0$ . Thus  $n\theta = 2\pi k$  for some integer  $k$ . It follows that  $\theta = \frac{2\pi k}{n}$ . Taking  $k = 0, 1, \dots, n-1$  gives the  $n^{\text{th}}$  roots that we have listed. Taking other values of  $k$  gives different values for  $\frac{2\pi k}{n}$  but each of them differs from one of the listed values by a multiple of  $2\pi$  and therefore gives a value for  $\cos \theta + i \sin \theta$  that we already have. Thus the  $n$  roots that we listed are all of the  $n^{\text{th}}$  roots of unity.

Roots of other complex numbers can be computed.

**Example 9.2.12.** All of the solutions of the equation  $z^3 = 1 + i$  can be found as follows. First note that  $|1 + i| = \sqrt{2}$  and the argument of  $1 + i$  is  $\frac{\pi}{4}$ . That is,  $1 + i = \sqrt{2}(\cos \frac{\pi}{4} + i \sin \frac{\pi}{4})$ . Suppose that  $z = r(\cos \theta + i \sin \theta)$  and  $z^3 = 1 + i$ . Then  $z^3 = r^3(\cos 3\theta + i \sin 3\theta)$ . Therefore  $r^3 = \sqrt{2}$ , so  $r = 2^{\frac{1}{6}}$ , and  $3\theta$  is  $\frac{\pi}{4}$  or  $\frac{\pi}{4} + 2\pi$  or  $\frac{\pi}{4} + 4\pi$ . Therefore  $\theta$  itself can be  $\frac{\pi}{12}$ ,  $\frac{3\pi}{4}$ , or  $\frac{17\pi}{12}$ . This gives the three solutions of the equation  $z^3 = 1 + i$ :  $2^{\frac{1}{6}}(\cos \frac{\pi}{12} + i \sin \frac{\pi}{12})$ ,  $2^{\frac{1}{6}}(\cos \frac{3\pi}{4} + i \sin \frac{3\pi}{4})$  and  $2^{\frac{1}{6}}(\cos \frac{17\pi}{12} + i \sin \frac{17\pi}{12})$ .

### 9.3 The Fundamental Theorem of Algebra

One reason for introducing complex numbers was to provide a root for the polynomial  $x^2 + 1$ . There are many other polynomials that do not have any real roots. For example, if  $p(x)$  is any polynomial, then the polynomial obtained by writing out  $[p(x)]^2 + 1$  has no real roots, since its value is at least 1 for every value of  $x$ .

Does every such polynomial have a complex root? More generally, does every polynomial have a complex root? There is a trivial sense in which the answer to this question is “no”, since constant polynomials other than 0 clearly do not have any roots of any kind. For other polynomials, the answer is not so simple. It is a remarkable and very useful fact that every non-constant polynomial, with real coefficients, or even with complex coefficients, has a complex root.

**Theorem 9.3.1** (The Fundamental Theorem of Algebra). *Every non-constant polynomial with complex coefficients has a complex root.*

There are a number of different proofs of the Fundamental Theorem of Algebra. They all rely on mathematical concepts that we do not develop in this book. We will therefore simply discuss implications of this theorem without proving it.

How many roots does a polynomial have?

**Example 9.3.2.** The only root of the polynomial  $p(z) = z^2 - 6z + 9$  is  $z = 3$ . This follows from the fact that  $p(z) = (z - 3)(z - 3)$ . Since the product of two complex numbers is 0 only if at least one of the numbers is 0, the only solution to  $p(z) = 0$  is  $z = 3$ . In some sense, however, this polynomial has 3 as a “double root”; we’ll discuss this a little more below.

To explore the question of the number of roots that a polynomial can have, we need to use the concept of the division of one polynomial by another. This

concept of division is very similar to “long division” of one multi-digit integer into another. Actually, we only need a special case of this concept, the case where the polynomial divisor is linear. We begin with an example.

**Example 9.3.3.** To divide  $z - 3$  into  $z^4 + 5z^3 - 2z + 1$  we proceed as follows:

INSERT DIAGRAM

The only consequence of the division of one polynomial by another that we need for present purposes is the following.

**Theorem 9.3.4.** *If  $r$  is a complex number and  $p(z)$  is a non-constant polynomial with complex coefficients, then there exists a polynomial  $q(z)$  and a constant  $c$  such that*

$$p(z) = (z - r)q(z) + c$$

INSERT PROOF

**Definition 9.3.5.** The polynomial  $f(z)$  is a *factor* of the polynomial  $p(z)$  if there exists a polynomial  $q(z)$  such that  $p(z) = f(z)q(z)$

**Theorem 9.3.6** (The Factor Theorem). *The complex number  $r$  is a root of a polynomial  $p(z)$  if and only if  $(z - r)$  is a factor of  $p(z)$ .*

*Proof.* If  $(z - r)$  is a factor of  $p(z)$ , then  $p(z) = (z - r)q(z)$  implies that  $p(r) = (r - r)q(r) = 0 \cdot q(r) = 0$ . Conversely, suppose that  $r$  is a root of  $p(z)$ . By Theorem 9.3.4,  $p(z) = (z - r)q(z) + c$  for some constant  $c$ . Substituting  $r$  for  $z$  and using the fact that  $r$  is a root gives  $0 = (r - r)q(r) + c$ , so  $0 = 0 + c$ , from which it follows that  $c = 0$ . Hence  $p(z) = (z - r)q(z)$  and  $z - r$  is a factor of  $p(z)$ .

□

**Example 9.3.7.** The complex number  $2i$  is a root of the polynomial  $iz^3 + z^2 - 4$  (as can be seen by simply substituting  $2i$  for  $z$  in the expression for the polynomial and noting that the result is 0). It follows from the Factor Theorem that  $z - 2i$  is a factor of the given polynomial. Doing “long division” gives  $iz^3 + z^2 - 4 = (z - 2i)(iz^2 - z - 2i)$ .

We can use the Factor Theorem to determine the maximum number of roots that a polynomial may have.

**Theorem 9.3.8.** *A polynomial of degree  $n$  has  $n$  complex roots “counting multiplicity”.*

*Proof.* Let  $p(z)$  be a polynomial of degree  $n$ . If  $n$  is at least 1 then  $p(z)$  has a root, say  $r_1$ , by the Fundamental Theorem of Algebra (9.3.1). By the Factor Theorem (9.3.6), there exists a polynomial  $q_1(z)$  such that  $p(z) = (z - r_1)q_1(z)$ . The degree of  $q_1$  is clearly  $n - 1$ . If  $n - 1 > 0$  then  $q_1(z)$  has a root, say  $r_2$ . It follows from the Factor Theorem that there is a polynomial  $q_2(z)$  such that  $q_1(z) = (z - r_2)q_2(z)$ . The degree of  $q_2(z)$  is  $n - 2$ , and

$$p(z) = (z - r_1)(z - r_2)q_2(z)$$

This process can continue until a quotient is simply a constant, say  $k$ . Then

$$p(z) = k(z - r_1)(z - r_2) \cdots (z - r_n)$$

If the  $r_i$  are all different, the polynomial will have  $n$  roots. If some of the  $r_i$  coincide, collecting all the terms where  $r_i$  is equal to a given  $r$  produces a factor of the form  $(z - r)^m$  where  $m$  is the number of times that  $r$  occurs in the factorization. In this situation, we say that  $r$  is a root “of multiplicity  $m$ ” of the polynomial. Thus a polynomial of degree  $n$  has at most  $n$  distinct roots. If the roots are counted according to their multiplicities, then a polynomial of degree  $n$  has exactly  $n$  roots.  $\square$

## 9.4 Problems

### Basic Exercises

1. Write the following complex numbers in  $a + bi$  form.

(a)  $\left(\frac{1}{\sqrt{2}} + \frac{i}{\sqrt{2}}\right)^{10}$

(b)  $\left(\frac{1}{\sqrt{2}} + \frac{i}{\sqrt{2}}\right)^{106}$

(c)  $\left(-\frac{\sqrt{3}}{2} + \frac{i}{2}\right)^{11}$

2. Show that the real part of  $(1 + i)^{10}$  is 0.

3. Find both square roots of the following numbers.

(a)  $-i$

(b)  $-15 - 8i$

4. Find the cube roots of the following numbers.

(a)  $2$

(b)  $8\sqrt{3} + 8i$

### Interesting Problems

5. Find a polynomial  $p$  with integer coefficients such that  $p(3 + i\sqrt{7}) = 0$ .

6. Find all the complex roots of  $z^6 + z^3 + 1$ .

7. Find all the complex roots of the polynomial  $z^7 - z$ .

### Challenging Problems

8. Find a polynomial whose complex roots are  $\{2 - i, 2 + i, 7\}$ .

9. Find all the complex solutions of  $\frac{z^3 + 1}{z^3 - 1} = i$ .

10. Let  $p$  be a polynomial with real coefficients. Prove that the complex conjugate of each root of  $p$  is also a root of  $p$ .

11. Show that every non-constant polynomial with real coefficients can be factored into a product of linear and quadratic polynomials, each of which also has real coefficients.

## Chapter 10

# Sizes of Infinite Sets

How many natural numbers are there? How many even natural numbers are there? How many odd natural numbers are there? How many rational numbers are there? How many real numbers are there? How many points are there in the plane? How many sets of natural numbers are there? One answer to all the above questions would be: there are an infinite number of them. But there are more precise answers that can be given; there are, in a sense that we will explain, an infinite number of different size infinities.

### 10.1 Cardinality

**Definition 10.1.1.** By a *set* we simply mean any collection of things; the things are called *elements of the set*. (As will be discussed at the end of this chapter, a general definition of set such as we are using is problematic in certain senses.)

For example, the set of all words on this page is a set. The set containing the letters a, b, and c is a set: it could be denoted  $\mathcal{S} = \{a, b, c\}$ . The set of numbers greater than 4 could be written

$$\{x : x > 4\}.$$

The fact that something is an element of a set is often denoted with the Greek letter Epsilon. We write  $x \in \mathcal{S}$ . to represent the fact that  $x$  is an element of the set  $\mathcal{S}$ .

**Definition 10.1.2.** If  $\mathcal{S}$  is a set, a *subset* of  $\mathcal{S}$  is a set all of whose elements are elements of the set  $\mathcal{S}$ . The *empty set* is the set that has no elements at all. It is denoted  $\emptyset$ . The empty set is, by definition, a subset of every set. The *union*

of a collection of sets is the set consisting of all elements that occur in any of the given sets. The *intersection* of a collection of sets is the set containing all elements that are in every set in the given collection. If the intersection of two sets is the empty set, the sets are said to be *disjoint*.

How should we define the concept that two sets have the same number of elements? For finite sets, we count the number of elements in each set. When we count the number of elements in a set, we assign the number 1 to one of the elements of the set, then assign the number 2 to another element of the set, then 3 to another element of the set, and so on, until we have counted every element in the set. If the set has  $n$  elements, when we finish counting we will have assigned a number in the set  $\{1, 2, 3, \dots, n\}$  to each element of the set and will not have assigned two different numbers to the same element in the set. That is, counting that a set has  $n$  elements produces a pairing of the elements of the set  $\{1, 2, 3, \dots, n\}$  with the elements of the set that we are counting. A set whose elements can be paired with the elements of the set  $\{1, 2, 3, \dots, n\}$  is said to have  $n$  elements.

More generally, we can say that two sets have the same number of elements if the elements of those two sets can be paired with each other.

**Example 10.1.3.** Pairs of running shoes are manufactured in a given factory. Each day, some number of pairs is manufactured. Without knowing how many pairs were manufactured in a given day, we can still conclude that the same number of left shoes was manufactured as the number of right shoes that was manufactured, since they are manufactured in pairs. If, for example, the number of left shoes was determined to be 1,012, then it could be concluded that the number of right shoes was also 1,012. This could be established as follows: since the set  $\{1, 2, 3, \dots, 1,012\}$  can be paired with the set of left shoes, it could also be paired with the set of right shoes, simply by pairing each right shoe to the number assigned to the corresponding left shoe in the pair.

The above discussion suggests the general definition that we shall use. In the following, the phrase “have the same cardinality” is the standard mathematical terminology for what might colloquially be expressed “have the same size”.

**Definition 10.1.4.** (rough definition) The sets  $\mathcal{S}$  and  $\mathcal{T}$  are said to *have the same cardinality* if their elements can be paired with each other's.

We need to be able to precisely define what we mean by a “pairing” of the elements of two sets. This can be specified in terms of functions. A function from a set  $\mathcal{S}$  into a set  $\mathcal{T}$  is simply an assignment of an element of  $\mathcal{T}$  to each element of



$\mathcal{S}$ . For example, if  $\mathcal{S} = \{a, b, d, e\}$  and  $\mathcal{T} = \{+, \pi\}$ , then one particular function taking  $\mathcal{S}$  to  $\mathcal{T}$  is the function  $f$  defined by  $f(a) = \pi$ ,  $f(b) = \pi$ ,  $f(d) = +$ , and  $f(e) = \pi$ .

**Definition 10.1.5.** The notation  $f : \mathcal{S} \rightarrow \mathcal{T}$  is used to denote a function  $f$  taking the set  $\mathcal{S}$  into the set  $\mathcal{T}$ ; that is, a mapping of each element of  $\mathcal{S}$  to an element of  $\mathcal{T}$ . The set  $\mathcal{S}$  is called the *domain* of the function. The *range* of a function is the set of all its values; that is, the range  $f : \mathcal{S} \rightarrow \mathcal{T}$  is  $\{f(s) : s \in \mathcal{S}\}$ .

**Definition 10.1.6.** A function  $f : \mathcal{S} \rightarrow \mathcal{T}$  is said to be *one-to-one* (or *injective*) if  $f(s_1) \neq f(s_2)$  whenever  $s_1 \neq s_2$ . That is, a function is one-to-one if it does not send two different elements to the same element.

We also require another property that functions may have.

**Definition 10.1.7.** A function  $f : \mathcal{S} \rightarrow \mathcal{T}$  is *onto* (or *surjective*) if for every  $t \in \mathcal{T}$  there is an  $s \in \mathcal{S}$  such that  $f(s) = t$ ; that is, the range of  $f$  is all of  $\mathcal{T}$ .

Note that a one-to-one onto function from a set  $\mathcal{S}$  onto a set  $\mathcal{T}$  gives a pairing of the elements of  $\mathcal{S}$  with those of  $\mathcal{T}$ .

The formal definition of when sets are to be considered to have the same size can be stated as follows.

**Definition 10.1.8.** The sets  $\mathcal{S}$  and  $\mathcal{T}$  *have the same cardinality* if there is a function  $f : \mathcal{S} \rightarrow \mathcal{T}$  that is one-to-one and is onto all of  $\mathcal{T}$ .

We require the concept of the inverse of a function. If  $f$  is a one-to-one function mapping a set  $\mathcal{S}$  onto a set  $\mathcal{T}$ , then there is a function mapping  $\mathcal{T}$  onto  $\mathcal{S}$  that “sends elements back to where they came from via  $f$ ”.

**Definition 10.1.9.** If  $f$  is a one-to-one function mapping  $\mathcal{S}$  onto  $\mathcal{T}$ , then the *inverse of  $f$* , often denoted  $f^{-1}$ , is the function mapping  $\mathcal{T}$  onto  $\mathcal{S}$  defined by  $f^{-1}(t) = s$  when  $f(s) = t$ .

With respect to the definition above, note that  $f$  must be onto for  $f^{-1}$  to be defined on all of  $\mathcal{T}$ . Also,  $f$  must be one-to-one, otherwise for some  $t$  there will be more than one  $s$  for which  $f(s) = t$  and therefore  $f^{-1}(t)$  will not be determined.

Let’s consider some examples.

**Example 10.1.10.** The set of even natural numbers and the set of odd natural numbers have the same cardinality.

*Proof.* To prove this, write the set of even natural numbers as  $\mathcal{E} = \{2, 4, \dots, 2n, \dots\}$  and the set of odd natural numbers as  $\mathcal{O} = \{1, 3, \dots, 2n+1, \dots\}$ . We can define a function  $f$  taking  $\mathcal{E} \rightarrow \mathcal{O}$  by letting  $f(k) = k - 1$  for each  $k$  in  $\mathcal{E}$ . To see that this  $f$  is one-to-one, simply note that  $k_1 - 1 = k_2 - 1$  implies  $k_1 = k_2$ . Also,  $f$  is clearly onto. Thus the sets  $\mathcal{E}$  and  $\mathcal{O}$  have the same cardinality.  $\square$

It is not very surprising that the set of even natural numbers and the set of odd natural numbers have the same cardinalities. The following example is a little more unexpected.

**Example 10.1.11.** The set of even natural numbers has the same cardinality as the set of all natural numbers.

*Proof.* This seems surprising at first because it seems that the set of even numbers should have half the number of elements as the set of all natural numbers does. However, it is easy to prove that these sets,  $\mathcal{E}$  and  $\mathbb{N}$ , have the same cardinality. Simply define the function  $f : \mathbb{N} \rightarrow \mathcal{E}$  by  $f(n) = 2n$  for  $n = 1, 2, 3, \dots$ . It is easily seen that  $f$  is one-to-one: if  $f(n_1) = f(n_2)$ , then  $2n_1 = 2n_2$ , so  $n_1 = n_2$ . The function  $f$  is onto since every even number is of the form  $2k$ . Therefore  $\mathbb{N}$  and  $\mathcal{E}$  have the same cardinality.  $\square$

Thus, in the sense of the definition we are using, the subset  $\mathcal{E}$  of  $\mathbb{N}$  has the same size as the entire set  $\mathbb{N}$  has. This shows that, with respect to cardinality, it is not necessarily the case that “the whole is greater than any of its parts”.

Another example showing that “the whole” can have the same cardinality as “one of its parts” is the following.

**Example 10.1.12.** The set of natural numbers and the set of non-negative integers have the same cardinality.

*Proof.* The set of natural numbers is  $\mathbb{N} = \{1, 2, 3, \dots\}$ . Let  $\mathcal{S}$  denote the set  $\{0, 1, 2, 3, \dots\}$  of non-negative integers. We want to construct a one-to-one function  $f$  taking  $\mathcal{S}$  onto  $\mathbb{N}$ . We can simply define  $f$  by  $f(n) = n + 1$  for each  $n$  in  $\mathcal{S}$ . Clearly  $f$  maps  $\mathcal{S}$  onto  $\mathbb{N}$ . Also,  $f(n_1) = f(n_2)$  implies  $n_1 + 1 = n_2 + 1$ , which gives  $n_1 = n_2$ . That is,  $f$  does not send two different integers to the same natural number, so  $f$  is one-to-one. Therefore  $\mathbb{N}$  and  $\mathcal{S}$  have the same cardinality.  $\square$

The following notation is useful.

**Definition 10.1.13.** We use the notation  $|\mathcal{S}| = |\mathcal{T}|$  to mean that  $\mathcal{S}$  and  $\mathcal{T}$  have the same cardinality.

Thus, as shown above,  $|\mathcal{O}| = |\mathcal{E}| = |\mathbb{N}|$ .

How does the size of the set of all positive rational numbers, which we'll denote by  $\mathbb{Q}^+$ , compare to the size of the set of natural numbers? The subset of  $\mathbb{Q}^+$  consisting of those rational numbers with numerator 1 can obviously be paired with  $\mathbb{N}$ : simply pair  $\frac{1}{n}$  with  $n$  for each  $n$  in  $\mathbb{N}$ . But then there are all the rational numbers with numerator 2, and with numerator 3, and so on. It seems that there are many more positive rational numbers than there are natural numbers. However, we now prove that  $|\mathbb{N}| = |\mathbb{Q}^+|$ .

**Theorem 10.1.14.** *The set of natural numbers and the set of positive rational numbers have the same cardinality.*

*Proof.* To prove this theorem, we first describe a way of displaying all the positive rational numbers. We imagine writing all the rational numbers with numerator 1 in one line, and then, underneath that, the rational numbers with numerator 2 in a line, and under that the rational numbers with numerator 3 in a line, and so on. That is, we consider the following array:

$$\begin{array}{ccccccc}
 \frac{1}{1} & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \frac{1}{6} & \frac{1}{7} & \cdots \\
 \\ 
 \frac{2}{1} & \frac{2}{2} & \frac{2}{3} & \frac{2}{4} & \frac{2}{5} & \frac{2}{6} & \frac{2}{7} & \cdots \\
 \\ 
 \frac{3}{1} & \frac{3}{2} & \frac{3}{3} & \frac{3}{4} & \frac{3}{5} & \frac{3}{6} & \frac{3}{7} & \cdots \\
 \\ 
 \frac{4}{1} & \frac{4}{2} & \frac{4}{3} & \frac{4}{4} & \frac{4}{5} & \frac{4}{6} & \frac{4}{7} & \cdots \\
 \\ 
 \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & 
 \end{array}$$

Imagining the positive rational numbers arranged as above, we can show that the natural numbers can be paired with them. That is, we will define a one to one function  $f$  taking  $\mathbb{N}$  onto  $\mathbb{Q}^+$ . As we define the function, you should keep looking back at the array to see the pattern that we are using.

Define  $f(1) = \frac{1}{1}$  and  $f(2) = \frac{1}{2}$ . (We can't continue by  $f(3) = \frac{1}{3}$ ,  $f(4) = \frac{1}{4}$ ,  $\dots$ , for then  $f$  would only map onto those rational numbers with numerator 1.) Define  $f(3) = \frac{2}{1}$  and  $f(4) = \frac{3}{1}$ . We can't just keep going down in our array; we must include the numbers above as well. We need not include  $\frac{2}{2}$  however,

since  $\frac{2}{2} = \frac{1}{1}$  which is already paired with 1. Thus we let  $f(5) = \frac{1}{3}$ ,  $f(6) = \frac{1}{4}$ ,  $f(7) = \frac{2}{3}$ ,  $f(8) = \frac{3}{2}$ ,  $f(9) = \frac{4}{1}$ , and  $f(10) = \frac{5}{1}$ . We need not consider  $\frac{4}{2}$ , since  $\frac{4}{2} = \frac{2}{1}$  and we need not consider  $\frac{3}{3} = \frac{1}{1}$  or  $\frac{2}{4} = \frac{1}{2}$ . Thus  $f(11)$  is defined to be  $\frac{1}{5}$  and  $f(12) = \frac{1}{6}$ . It is apparent that a pairing of the natural numbers and the positive rational numbers is indicated by continuing to label rational numbers with natural numbers in this manner ("zig-zagging", you might say, through the above array). Therefore  $|\mathbb{Q}^+| = |\mathbb{N}|$ .  $\square$

## 10.2 Countable Sets and Uncountable Sets

You may be wondering whether or not every infinite set can be paired with the set of natural numbers. If the elements of a set can be paired with the natural numbers, then the elements can be listed in a sequence. For example, if we let  $s_1$  be the element of the set corresponding to the natural number 1,  $s_2$  be the element of the set corresponding to the natural number 2,  $s_3$  to 3, and so on, then the set could be displayed

$$\{s_1, s_2, s_3 \dots\}$$

Pairing the elements of a set with the set of natural numbers is, in a sense, "counting the elements of the set".

**Definition 10.2.1.** A set is *countable* (sometimes called *denumerable*, or *enumerable*) if it is either finite or has the same cardinality as the set of natural numbers. A set is said to be *uncountable* if it is not countable.

One example of an uncountable set is the following.

**Theorem 10.2.2.** *The set of all real numbers between 0 and 1 is uncountable.*

*Proof.* We must show that there is no way of pairing the set of natural numbers with the set of real numbers between 0 and 1. Let  $\mathcal{S}$  denote the set of real numbers between 0 and 1:  $\mathcal{S} = \{x : 0 \leq x \leq 1\}$ . We will show that every pairing of natural numbers with elements of  $\mathcal{S}$  fails to include some members of  $\mathcal{S}$ . In other words, we will show that there does not exist any function that maps  $\mathbb{N}$  onto  $\mathcal{S}$ .

Note that the elements of  $\mathcal{S}$  can be written as infinite decimals; that is, in the form  $.c_1c_2c_3\dots$  where each  $c_i$  is a digit between 0 and 9. Some numbers have two different such representations. For example,  $.1999\dots$  is the same number

as .20000... For the rest of this proof, let us agree that we choose the representation involving an infinite string of 9's rather than the representation involving an infinite string of 0's for all numbers that have two different representations.

Suppose, then, that  $f$  is any function taking  $\mathbb{N}$  to  $\mathcal{S}$ . To show that  $f$  cannot be onto, we imagine writing out all the values of  $f$  in a list, as follows:

$$\begin{aligned} f(1) &= .a_{11}a_{12}a_{13}a_{14}a_{15}\dots \\ f(2) &= .a_{21}a_{22}a_{23}a_{24}a_{25}\dots \\ f(3) &= .a_{31}a_{32}a_{33}a_{34}a_{35}\dots \\ f(4) &= .a_{41}a_{42}a_{43}a_{44}a_{45}\dots \\ f(5) &= .a_{51}a_{52}a_{53}a_{54}a_{55}\dots \\ &\vdots \qquad \qquad \vdots \end{aligned}$$

We now construct a number in  $\mathcal{S}$  that is not in the range of the function  $f$ . We do that by showing how to choose digits  $b_j$  so that the number  $x = .b_1b_2b_3b_4\dots$  is not in the range of  $f$ . Begin by choosing  $b_1 = 3$  if  $a_{11} \neq 3$  and  $b_1 = 4$  if  $a_{11} = 3$ . No matter what digits we choose for the  $b_j$  for  $j \geq 2$ , the number  $x$  will be different from  $f(1)$  since its first digit is different from the first digit of  $f(1)$ . Then choose  $b_2 = 3$  if  $a_{22} \neq 3$  and  $b_2 = 4$  if  $a_{22} = 3$ . This insures that  $x \neq f(2)$ . We continue in this manner, choosing  $b_j = 3$  if  $a_{jj} \neq 3$  and  $b_j = 4$  if  $a_{jj} = 3$ . The number  $x$  that is so constructed differs from  $f(j)$  in its  $j$ th digit. Therefore  $f(j) \neq x$  for all  $j$ , so  $x$  is not in the range of  $f$ . Thus we have proven that there is no function (one-to-one or otherwise) taking  $\mathbb{N}$  onto  $\mathcal{S}$ , so we conclude that  $\mathcal{S}$  has cardinality different from that of  $\mathbb{N}$ .  $\square$

Of course, any given function  $f$  in the above proof could be modified so as to produce a function whose range does include the specific number  $x$  that we constructed in the course of the proof. For example, given any  $f$ , define the function  $g : \mathbb{N} \rightarrow \mathcal{S}$  by defining  $g(1) = x$  and  $g(n) = f(n-1)$  for  $n = 2, 3, 4, \dots$ . The range of  $g$  includes  $x$  and also includes the range of  $f$ . However,  $g$  does not map  $\mathbb{N}$  onto  $\mathcal{S}$ , for the above proof could be used to produce a different  $x$  that is not in the range of  $g$ .

**Definition 10.2.3.** The *closed interval from  $a$  to  $b$* , with  $a$  and  $b$  real numbers and  $a \leq b$ , is the set of all real numbers between  $a$  and  $b$ , including  $a$  and  $b$ . It is denoted  $[a, b]$ . That is,  $[a, b] = \{x : a \leq x \leq b\}$ .

Thus the theorem we have just proven asserts that the closed unit interval,  $[0, 1]$ , is uncountable. How does the cardinality of other closed intervals compare to that of  $[0, 1]$ ?

**Theorem 10.2.4.** *If  $a$  and  $b$  are real numbers and  $a < b$ , then  $[a, b]$  and  $[0, 1]$  have the same cardinality.*

*Proof.* The theorem will be established if we can construct a function  $f : [0, 1] \rightarrow [a, b]$  that is one-to-one and onto. That is easy to do. Simply define  $f$  by  $f(x) = a + (b - a)x$ . Then  $f(0) = a$  and  $f(1) = b$ . Moreover, the function  $f$  increases from  $a$  to  $b$  as  $x$  increases from 0 to 1. If  $y \in [a, b]$ , let  $x = \frac{y-a}{b-a}$ . Then  $x \in [0, 1]$  and  $f(x) = y$ . This shows that  $f$  is onto. To show that  $f$  is one-to-one, assume that  $a + (b - a)x_1 = a + (b - a)x_2$ . Subtracting  $a$  from both sides of this equation and then dividing both sides by  $(b - a)$  yields  $x_1 = x_2$ . This shows that  $f$  is one-to-one. Thus  $f$  is a pairing of the elements of  $[0, 1]$  with the elements of  $[a, b]$ , so  $|[0, 1]| = |[a, b]|$   $\square$

There are other intervals that frequently arise in mathematics.

**Definition 10.2.5.** If  $a$  and  $b$  are real numbers and  $a < b$ , then the *open interval between  $a$  and  $b$* , denoted  $(a, b)$ , is defined by

$$(a, b) = \{x : a < x < b\}$$

The *half open intervals* are defined by

$$(a, b] = \{x : a < x \leq b\} \text{ and } [a, b) = \{x : a \leq x < b\}$$

How does the size of a half open interval compare to the size of the corresponding closed interval?

**Theorem 10.2.6.** *The intervals  $[0, 1]$  and  $(0, 1]$  have the same cardinality.*

*Proof.* We want to construct a one-to-one function  $f$  taking  $[0, 1]$  onto  $(0, 1]$ . We will define  $f(x) = x$  for most  $x$  in  $[0, 1]$  but we need to make a place for 0 to go to in the half open interval. For each natural number  $n$ , the rational number  $\frac{1}{n}$  is in both intervals. Define  $f$  on those numbers by  $f(\frac{1}{n}) = \frac{1}{n+1}$  for  $n = 1, 2, 3, \dots$ . In particular,  $f(1) = \frac{1}{2}$ . Note that the number 1, which is in  $(0, 1]$ , is not in the range of  $f$  as defined so far. We define  $f(0)$  to be 1. We define  $f$  on the rest of  $[0, 1]$  by  $f(x) = x$ . That is,  $f(x) = x$  for those  $x$  other than 0 that are not of the form  $\frac{1}{n}$  with  $n$  a natural number. It is easy to see that we have constructed a one-to-one function mapping  $[0, 1]$  onto  $(0, 1]$ .  $\square$

Suppose that  $|\mathcal{S}| = |\mathcal{T}|$  and  $|\mathcal{T}| = |\mathcal{U}|$ ; must  $|\mathcal{S}| = |\mathcal{U}|$ ? If this was not the case, we would be using the “equals” sign in a very peculiar way.

**Theorem 10.2.7.** *If  $|\mathcal{S}| = |\mathcal{T}|$  and  $|\mathcal{T}| = |\mathcal{U}|$ , then  $|\mathcal{S}| = |\mathcal{U}|$ .*

*Proof.* By hypothesis, there exists one-to-one functions  $f$  and  $g$  mapping  $\mathcal{S}$  onto  $\mathcal{T}$  and  $\mathcal{T}$  onto  $\mathcal{U}$  respectively. That is,  $f : \mathcal{S} \rightarrow \mathcal{T}$  and  $g : \mathcal{T} \rightarrow \mathcal{U}$ . Let  $h = g \circ f$  be the composition of  $g$  and  $f$ . In other words,  $h$  is the function defined on  $\mathcal{S}$  by  $h(s) = g(f(s))$ . We must show that  $h$  is a one-to-one function taking  $\mathcal{S}$  onto  $\mathcal{U}$ . Let  $u$  be any element of  $\mathcal{U}$ . Since  $g$  is onto, there exists a  $t$  in  $\mathcal{T}$  such that  $g(t) = u$ . Since  $f$  is onto, there is a  $s$  in  $\mathcal{S}$  such that  $f(s) = t$ . Then  $h(s) = g(f(s)) = g(t) = u$ . Thus  $h$  is onto.

To see that  $h$  is one-to-one, suppose that  $h(s_1) = h(s_2)$ ; we must show that  $s_1 = s_2$ . Now  $g(f(s_1)) = g(f(s_2))$ , so  $f(s_1) = f(s_2)$  since  $g$  is one-to-one. But  $f$  is also one-to-one, and therefore  $s_1 = s_2$ . We have shown that  $h$  is one-to-one and onto, from which it follows that  $|\mathcal{S}| = |\mathcal{U}|$ .  $\square$

**Theorem 10.2.8.** *If  $a, b, c$  and  $d$  are real numbers with  $a$  less than  $b$  and  $c$  less than  $d$ , then the half open intervals  $(a, b]$  and  $(c, d]$  have the same cardinality.*

*Proof.* The function  $f$  defined by  $f(x) = a + (b - a)x$  is a one-to-one function mapping  $(0, 1]$  onto  $(a, b]$ , as can be seen by a proof almost exactly the same as that in 10.2.4. Hence  $|(0, 1]| = |(a, b]|$ . Similarly the function  $g$  defined by  $g(x) = c + (d - c)x$  is a one-to-one function mapping  $(0, 1]$  onto  $(c, d]$ , so  $|(0, 1]| = |(c, d]|$ . It follows from 10.2.7 that  $|(a, b]| = |(c, d]|$ .  $\square$

Are there more positive real numbers than there are numbers in  $[0, 1]$ ? The, perhaps surprising, answer is “no”.

**Theorem 10.2.9.** *The cardinality of the set of non-negative real numbers is the same as the cardinality of the unit interval  $[0, 1]$ .*

*Proof.* We begin by showing that the set  $\mathcal{S} = \{x : x \geq 1\}$  has the same cardinality as  $(0, 1]$ . Note that the function  $f$  defined by  $f(x) = \frac{1}{x}$  maps  $\mathcal{S}$  into  $(0, 1]$ . For if  $x \geq 1$ ,  $\frac{1}{x} \leq 1$ . Also,  $f$  maps  $\mathcal{S}$  onto  $(0, 1]$ . For if  $y \in (0, 1]$ , then  $\frac{1}{y} \geq 1$ , and  $f(\frac{1}{y}) = y$ .

To see that  $f$  is one-to-one, suppose that  $f(x_1) = f(x_2)$ . Then  $\frac{1}{x_1} = \frac{1}{x_2}$ , so  $x_1 = x_2$ . Hence  $f$  is one-to-one and onto, and it follows that  $|\mathcal{S}| = |(0, 1]|$ .

Now let  $\mathcal{T} = \{x \in \mathbb{R} : x \geq 0\}$ . Define the function  $g$  by  $g(x) = x - 1$ . Then  $g$  is obviously a one-to-one function mapping  $\mathcal{S}$  onto  $\mathcal{T}$ . Hence  $|\mathcal{T}| = |\mathcal{S}|$ .

Therefore, by 10.2.7,  $|\mathcal{T}| = |(0, 1]|$ . But, by Theorem 10.2.6,  $|[0, 1]| = |(0, 1]|$ . It follows that  $|\mathcal{T}| = |[0, 1]|$ .  $\square$

Must the union of two countable sets be countable? A much stronger result is true.

**Theorem 10.2.10.** *The union of a countable number of countable sets is countable.*

*Proof.* This can be proven using ideas similar to those used in the proof of the fact that the set of positive rational numbers is countable (see 10.1.14). Recall that “countable” means either finite or having the same cardinality as  $\mathbb{N}$ . We will prove this theorem for the cases where all the sets are infinite; you should be able to see how to modify the proof if some or all of the cardinalities are finite. Suppose then that we have a countable collection  $\{\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3, \dots\}$  of sets, each of which is itself countably infinite. We can label the elements of  $\mathcal{S}_i$  so that  $\mathcal{S}_i = \{a_{i1}, a_{i2}, a_{i3}, \dots\}$ . We can imagine displaying the sets in the following array:

$$\begin{array}{l} \mathcal{S}_1 = \{a_{11}, a_{12}, a_{13}, a_{14}, a_{15}, a_{16}, a_{17} \dots\} \\ \mathcal{S}_2 = \{a_{21}, a_{22}, a_{23}, a_{24}, a_{25}, a_{26}, a_{27} \dots\} \\ \mathcal{S}_3 = \{a_{31}, a_{32}, a_{33}, a_{34}, a_{35}, a_{36}, a_{37} \dots\} \\ \mathcal{S}_4 = \{a_{41}, a_{42}, a_{43}, a_{44}, a_{45}, a_{46}, a_{47} \dots\} \\ \vdots \\ \vdots \\ \vdots \end{array}$$

Let  $\mathcal{S}$  denote the union of the  $\mathcal{S}_i$ 's. To show that  $\mathcal{S}$  is countable, we show that we can list all of its elements. Proceed as follows. First, list  $a_{11}$  then  $a_{12}$ . Then look at  $a_{21}$ . It is possible that  $a_{21}$  is one of  $a_{11}$  or  $a_{12}$ , in which case we do not, of course, list it again. If, however,  $a_{21}$  is neither  $a_{11}$  nor  $a_{12}$ , we list it next. Then look at  $a_{31}$ ; if it is not yet listed, list it next. Then go back up to  $a_{22}$ , then  $a_{13}$ , and so on. In this way, we sweep through the entire array and list all the elements of  $\mathcal{S}$ . It follows that  $\mathcal{S}$  is countable.  $\square$



## 10.3 Comparing Cardinalities

When two sets have different cardinalities, the question arises of whether we can say that one set has cardinality that is less than the cardinality of the other set. What should we mean by saying that the cardinality of one set is less than that of another set? It is easiest to begin with a definition of “less than or equal to” for cardinalities.

**Definition 10.3.1.** If  $\mathcal{S}$  and  $\mathcal{T}$  are sets, we say that  $\mathcal{S}$  *has cardinality less than or equal to the cardinality of*  $\mathcal{T}$ , and write  $|\mathcal{S}| \leq |\mathcal{T}|$ , if there is a subset  $\mathcal{T}_0$  of  $\mathcal{T}$  such that  $|\mathcal{S}| = |\mathcal{T}_0|$ . This is equivalent to saying that there is a one-to-one function mapping  $\mathcal{S}$  into (not necessarily onto)  $\mathcal{T}$ . For if  $f$  is a one-to-one function mapping  $\mathcal{S}$  onto  $\mathcal{T}_0$ , we can regard  $f$  as a function taking  $\mathcal{S}$  into  $\mathcal{T}$ . Conversely, if  $f$  is a one-to-one function mapping  $\mathcal{S}$  into  $\mathcal{T}$ , and if  $\mathcal{T}_0$  is the range of  $f$ , then  $f$  gives a pairing of  $\mathcal{S}$  and  $\mathcal{T}_0$ .

**Example 10.3.2.** The function  $f : \mathbb{N} \rightarrow [0, 1]$  defined by  $f(n) = \frac{1}{n}$  establishes that  $|\mathbb{N}| \leq |[0, 1]|$ , since  $f$  is one-to-one.

Note that  $|\mathcal{S}_0| \leq |\mathcal{S}|$  whenever  $\mathcal{S}_0$  is a subset of  $\mathcal{S}$ , since the function  $f : \mathcal{S}_0 \rightarrow \mathcal{S}$  defined by  $f(s) = s$  for all  $s$  in  $\mathcal{S}_0$  is clearly one-to-one.

We have defined “ $\leq$ ” for cardinalities; how should we define “ $<$ ”? The following definition is very natural.

**Definition 10.3.3.** We say that  $\mathcal{S}$  *has cardinality less than that of*  $\mathcal{T}$ , and write  $|\mathcal{S}| < |\mathcal{T}|$ , if  $|\mathcal{S}| \leq |\mathcal{T}|$  and  $|\mathcal{S}| \neq |\mathcal{T}|$ .

**Example 10.3.4.** If  $\mathbb{N}$  is the set of natural numbers and  $[0, 1]$  is the unit interval, then  $|\mathbb{N}| < |[0, 1]|$ .

*Proof.* By example 10.3.2,  $|\mathbb{N}| \leq |[0, 1]|$ , and, by 10.2.2,  $|\mathbb{N}| \neq |[0, 1]|$ , so the result follows.  $\square$

Thus, in the sense of the definitions we are using, there are more real numbers in the interval  $[0, 1]$  than there are natural numbers.

There is a question that immediately arises from the definition of “less than or equal to” for cardinalities: If  $\mathcal{S}$  and  $\mathcal{T}$  are sets such that  $|\mathcal{S}| \leq |\mathcal{T}|$  and  $|\mathcal{T}| \leq |\mathcal{S}|$ , must  $|\mathcal{S}| = |\mathcal{T}|$ ? The language suggests that this question should have an affirmative answer but that language doesn’t prove anything. What does this question come down to? We are given the fact that  $|\mathcal{S}| \leq |\mathcal{T}|$ . That is equivalent to the existence of a one-to-one function  $f : \mathcal{S} \rightarrow \mathcal{T}$ . Similarly,

$|\mathcal{T}| \leq |\mathcal{S}|$  gives a one-to-one function  $g : \mathcal{T} \rightarrow \mathcal{S}$ . To say that  $|\mathcal{S}| = |\mathcal{T}|$  is equivalent to saying that there exists a function  $h : \mathcal{S} \rightarrow \mathcal{T}$  that is both one-to-one and onto. The question, therefore, is whether we can show the existence of such a function  $h$  from the existence of the functions  $f$  and  $g$ .

In addition to being important in justifying the above terminology, the following theorem is often very useful in proving that given sets have the same cardinalities.

**Theorem 10.3.5.** (*The Cantor-Bernstein Theorem*) *If  $\mathcal{S}$  and  $\mathcal{T}$  are sets such that  $|\mathcal{S}| \leq |\mathcal{T}|$  and  $|\mathcal{T}| \leq |\mathcal{S}|$ , then  $|\mathcal{S}| = |\mathcal{T}|$ .*

*Proof.* The hypotheses imply that there exist one-to-one functions  $f : \mathcal{S} \rightarrow \mathcal{T}$  and  $g : \mathcal{T} \rightarrow \mathcal{S}$ . We must construct a one-to-one function  $h$  that takes  $\mathcal{S}$  onto  $\mathcal{T}$ . To do this we will break  $\mathcal{S}$  up into three subsets and then define  $h$  to be the function  $f$  on two of those subsets and the function  $g^{-1}$  on the third subset.

Consider any element  $s$  of  $\mathcal{S}$ . Such an  $s$  may or may not be in the range of  $g$ . If it is in the range of  $g$ , then there is exactly one element  $t_0$  in  $\mathcal{T}$  such that  $g(t_0) = s$ , since  $g$  is one-to-one. Call such an element  $t_0$  the “immediate ancestor” of  $s$ . Similarly, if  $t$  is in  $\mathcal{T}$  and  $f(s_0) = t$  for some  $s_0$  in  $\mathcal{S}$ , we say that  $s_0$  is the “immediate ancestor” of  $t$ . Thus elements of  $\mathcal{S}$  have immediate ancestors in  $\mathcal{T}$  if they are in the range of  $g$ , and elements of  $\mathcal{T}$  have immediate ancestors in  $\mathcal{S}$  if they are in the range of  $f$ . It is possible that some elements do not have any immediate ancestors.

We will say that an immediate ancestor of an immediate ancestor of an element  $s$  in  $\mathcal{S}$  is an “ancestor” of the element  $s$ . That is, if  $s$  in  $\mathcal{S}$  has an immediate ancestor  $t_0$  in  $\mathcal{T}$  and  $t_0$  has an immediate ancestor  $s_0$  in  $\mathcal{S}$ , then  $s_0$  is an ancestor of  $s$ . Similarly, if  $t_1$  in  $\mathcal{T}$  has an immediate ancestor  $s_1$  in  $\mathcal{S}$  and  $s_1$  has an immediate ancestor  $t_2$  in  $\mathcal{T}$ , we say that  $t_2$  is an ancestor of  $t_1$ . We continue backwards whenever possible. That is, we keep on finding immediate ancestors unless and until we reach an element that does not have an immediate ancestor. All the ancestors in such a chain of ancestors are called ancestors of the original element of  $\mathcal{S}$  or  $\mathcal{T}$ .

For each given element that we start with, there are three possibilities. One possibility is that there is no element in the chain of ancestors which itself does not have any ancestor. That is, it could be that we can keep on going back and back and back indefinitely in the ancestry of a given element. Let  $\mathcal{S}_\infty$  denote the set of all those elements  $s$  in  $\mathcal{S}$  for which we can keep on finding ancestors without stopping. Similarly, let  $\mathcal{T}_\infty$  denote the set of all  $t$  in  $\mathcal{T}$  for which we can keep on finding ancestors without stopping. (It might be noted that it is possible

that we can keep on finding ancestors indefinitely but nonetheless there are only a finite number of distinct ancestors. For example, it would be possible that, for some  $s$  in  $\mathcal{S}$  and  $t$  in  $\mathcal{T}$ ,  $f(s) = t$  and  $g(t) = s$ . Then the immediate ancestor of  $s$  would be  $t$ , the immediate ancestor of  $t$  would be  $s$ , the immediate ancestor of  $s$  would be  $t$ , and so on. Thus there would be no stopping the process of finding ancestors, in spite of the fact that each of  $s$  and  $t$  has only two distinct ancestors,  $s$  and  $t$ . In this situation,  $s \in \mathcal{S}_\infty$  and  $t \in \mathcal{T}_\infty$ .)

Those elements of  $\mathcal{S}$  and  $\mathcal{T}$  that are not in either of  $\mathcal{S}_\infty$  or  $\mathcal{T}_\infty$  have what might be called “ultimate ancestors”. That is, since the chain of ancestors comes to a stop, there is a most distant ancestor. Of course, one possibility is that the element has no ancestors at all, in which case we say that it itself is its ultimate ancestor. The ultimate ancestor of any given element is either in  $\mathcal{S}$  or in  $\mathcal{T}$ . Let  $\mathcal{S}_\mathcal{S}$  denote the set of all elements of  $\mathcal{S}$  whose ultimate ancestor is in  $\mathcal{S}$  and let  $\mathcal{S}_\mathcal{T}$  denote the set of all elements of  $\mathcal{S}$  whose ultimate ancestor is in  $\mathcal{T}$ . Similarly, let  $\mathcal{T}_\mathcal{S}$  and  $\mathcal{T}_\mathcal{T}$  denote the sets of elements of  $\mathcal{T}$  whose ultimate ancestors are in  $\mathcal{S}$  and  $\mathcal{T}$  respectively.

Thus we have divided  $\mathcal{S}$  into three subsets:  $\mathcal{S}_\infty$ ,  $\mathcal{S}_\mathcal{S}$  and  $\mathcal{S}_\mathcal{T}$ . Every element of  $\mathcal{S}$  is in exactly one of those subsets. Similarly, every element of  $\mathcal{T}$  is in exactly one of the subsets  $\mathcal{T}_\infty$ ,  $\mathcal{T}_\mathcal{S}$  or  $\mathcal{T}_\mathcal{T}$ . (Of course, any of the subsets may be empty.)

We can now define the function  $h$ . For  $s$  in  $\mathcal{S}$ , we define  $h(s)$  to be  $f(s)$  if  $s$  is in either  $\mathcal{S}_\infty$  or  $\mathcal{S}_\mathcal{S}$ , and we define  $h(s)$  to be  $g^{-1}(s)$  if  $s$  is in  $\mathcal{S}_\mathcal{T}$ . Note that  $g^{-1}(s)$  is defined for all  $s \in \mathcal{S}_\mathcal{T}$  since all the elements of  $\mathcal{S}_\mathcal{T}$  have immediate ancestors in  $\mathcal{T}$ . We will show that  $h$  is a one-to-one function taking  $\mathcal{S}$  onto  $\mathcal{T}$ .

Let’s first show that  $h$  is one-to-one. Suppose that  $h(s_1) = h(s_2)$  for  $s_1$  and  $s_2$  in  $\mathcal{S}$ . We must show that  $s_1 = s_2$ . If both of  $s_1$  and  $s_2$  are in the union of  $\mathcal{S}_\infty$  and  $\mathcal{S}_\mathcal{S}$ , then  $h(s_1) = f(s_1)$  and  $h(s_2) = f(s_2)$ . Therefore  $f(s_1) = f(s_2)$ . Since  $f$  is one-to-one, it follows that  $s_1 = s_2$  in this case. Similarly, if both of  $s_1$  and  $s_2$  are in  $\mathcal{S}_\mathcal{T}$ , then  $h(s_1) = g^{-1}(s_1)$  and  $h(s_2) = g^{-1}(s_2)$ . Therefore  $g^{-1}(s_1) = g^{-1}(s_2)$ . Applying  $g$  to both sides of this equation gives  $s_1 = s_2$  in this case.

One case remains; the case where one of  $s_1$  and  $s_2$  is in the union of  $\mathcal{S}_\mathcal{S}$  and  $\mathcal{S}_\infty$  and the other is in  $\mathcal{S}_\mathcal{T}$ . Suppose that  $s_1 \in \mathcal{S}_\infty \cup \mathcal{S}_\mathcal{S}$  and  $s_2 \in \mathcal{S}_\mathcal{T}$ . Then  $h(s_1) = f(s_1)$  and  $h(s_2) = g^{-1}(s_2)$ . Therefore  $f(s_1) = g^{-1}(s_2)$ . We show that this case cannot arise. For  $f(s_1) = g^{-1}(s_2)$  implies that  $s_1$  is an immediate ancestor of  $g^{-1}(s_2)$ . Thus  $s_1$  is an ancestor of  $s_2$ . But  $s_2$  is in  $\mathcal{S}_\mathcal{T}$  so has an ultimate ancestor in  $\mathcal{T}$ . Since  $s_1$  is an ancestor of  $s_2$ , the ultimate ancestor of  $s_2$  is the ultimate ancestor of  $s_1$ . But  $s_1$  being in  $\mathcal{S}_\infty \cup \mathcal{S}_\mathcal{S}$  implies that  $s_1$  either has no ultimate ancestor or has an ultimate ancestor in  $\mathcal{S}$ . This is inconsistent

with having an ultimate ancestor in  $\mathcal{T}$ , so this case does not arise.

We have proven that the function  $h$  that we constructed is one-to-one. We must show that  $h$  maps  $\mathcal{S}$  onto  $\mathcal{T}$ .

Each  $t$  in  $\mathcal{T}$  is in one of  $\mathcal{T}_{\mathcal{S}}$ ,  $\mathcal{T}_{\infty}$  or  $\mathcal{T}_{\mathcal{T}}$ . We must show that, wherever  $t$  lies, there is an  $s$  in  $\mathcal{S}$  such that  $h(s) = t$ . Suppose first that  $t \in \mathcal{T}_{\mathcal{S}}$ . Since  $t$  has an ultimate ancestor in  $\mathcal{S}$ , in particular we know that  $t$  is in the range of  $f$ , so we can consider  $f^{-1}(t)$ . The ancestors of  $f^{-1}(t)$  are also ancestors of  $t$ , from which it follows that the ultimate ancestor of  $f^{-1}(t)$  is in  $\mathcal{S}$ . That is,  $f^{-1}(t)$  is in  $\mathcal{S}_{\mathcal{S}}$ . The function  $h$  is defined to be  $f$  on  $\mathcal{S}_{\mathcal{S}}$ , so  $h(f^{-1}(t)) = f(f^{-1}(t)) = t$ . This shows that the range of  $h$  contains every element of  $\mathcal{T}_{\mathcal{S}}$ .

Now consider any  $t$  in  $\mathcal{T}_{\infty}$ . Such a  $t$  has an immediate ancestor in  $\mathcal{S}$ ,  $f^{-1}(t)$ . Since the ancestors of  $f^{-1}(t)$  are also ancestors of  $t$ ,  $f^{-1}(t)$  has no ultimate ancestor. That is,  $f^{-1}(t)$  is in  $\mathcal{S}_{\infty}$ . The function  $h$  was defined to be the function  $f$  on  $\mathcal{S}_{\infty}$ , so  $h(f^{-1}(t)) = f(f^{-1}(t)) = t$ . This proves that the range of  $h$  contains  $\mathcal{T}_{\infty}$ .

All that remains to be shown is that the range of  $h$  includes  $\mathcal{T}_{\mathcal{T}}$ . Suppose, then, that  $t$  is in  $\mathcal{T}_{\mathcal{T}}$ . Let  $s = g(t)$ . Then  $t$  is the immediate ancestor of  $s$ . Thus the ultimate ancestor of  $t$  is the ultimate ancestor of  $s$ . Since the ultimate ancestor of  $t$  is in  $\mathcal{T}$ , the ultimate ancestor of  $s$  is in  $\mathcal{T}$ . In other words,  $s$  is in  $\mathcal{S}_{\mathcal{T}}$ . On elements of  $\mathcal{S}_{\mathcal{T}}$ ,  $h$  is defined to be  $g^{-1}$ . Thus  $h(s) = g^{-1}(s)$  and, since  $s = g(t)$ ,  $h(s) = g^{-1}(g(t)) = t$ . This establishes that the range of  $h$  includes  $\mathcal{T}_{\mathcal{T}}$ .

We have therefore shown that, for every  $t$  in  $\mathcal{T}$ , whatever subset of  $\mathcal{T}$   $t$  lies in, there is an  $s$  in  $\mathcal{S}$  such that  $h(s) = t$ . This proves that  $h$  is onto.

Therefore  $h$  is a one-to-one function mapping  $\mathcal{S}$  onto  $\mathcal{T}$ , and we conclude that  $|\mathcal{S}| = |\mathcal{T}|$ .  $\square$

**Corollary 10.3.6.** *If  $\mathcal{S}$  is a subset of  $\mathcal{T}$  and there exists a function  $f : \mathcal{T} \rightarrow \mathcal{S}$  that is one-to-one, then  $\mathcal{S}$  and  $\mathcal{T}$  have the same cardinality.*

*Proof.* Since  $\mathcal{S}$  is a subset of  $\mathcal{T}$ ,  $|\mathcal{S}| \leq |\mathcal{T}|$ . Since there is a one-to-one function mapping  $\mathcal{T}$  into  $\mathcal{S}$ , it follows that  $|\mathcal{T}| \leq |\mathcal{S}|$ . By the Cantor-Bernstein Theorem (10.3.5),  $|\mathcal{S}| = |\mathcal{T}|$ .  $\square$

The Cantor-Bernstein Theorem can often be used to simplify proofs that given sets have the same cardinalities.

**Theorem 10.3.7.** *If  $a < b$ , then  $|[a, b]| = |(a, b)| = |(a, b]| = |[a, b)|$ .*

*Proof.* Clearly,  $|(a, b)| \leq |[a, b]|$ . Note that  $[a + \frac{b-a}{3}, b - \frac{b-a}{3}]$  is contained in  $(a, b)$ , so  $|[a + \frac{b-a}{3}, b - \frac{b-a}{3}]| \leq |(a, b)|$ . But, by Theorem 10.2.8,  $|[a + \frac{b-a}{3}, b - \frac{b-a}{3}]| = |[a, b]|$ . Therefore  $|[a, b]| \leq |(a, b)|$ . So, by the Cantor-Bernstein Theorem (10.3.5),  $|[a, b]| = |(a, b)|$ .

The proofs for the semi-open intervals are almost exactly the same as the above proof for the open interval.  $\square$

What is the cardinality of the set of all real numbers?

**Theorem 10.3.8.** *The cardinality of the set of all real numbers is the same as the cardinality of the unit interval  $[0, 1]$ .*

*Proof.* Let  $\mathbb{R}$  denote the set of all real numbers. We will “patch together” some of the results that we have already proven to show that  $|\mathbb{R}| \leq |[0, 1]|$ .

As we have seen, the set, call it  $\mathcal{S}$ , of non-negative real numbers has the same cardinality as  $[0, 1]$  (see 10.2.9). Thus there exists a one-to-one function  $f$  mapping  $\mathcal{S}$  onto  $[0, 1]$ . The set of negative real numbers obviously has the same cardinality as the set of positive real numbers, as can be seen by using the mapping that takes  $x$  to  $-x$ . The positive real numbers can be mapped in a one-to-one way into  $[0, 1]$ . Since  $|[0, 1]| = |[3, 4]|$ , it follows that the positive numbers can be mapped in a one-to-one way into  $[3, 4]$ . Then, using the equivalence of the positive and negative real numbers, we conclude that there is a function  $g$  mapping the negative real numbers into  $[3, 4]$ . We now define a function  $h$  mapping  $\mathbb{R}$  into  $[0, 1] \cup [3, 4]$  by letting  $h$  be  $f$  on the non-negative numbers and  $g$  on the negative numbers. Then  $h$  is one-to-one function mapping  $\mathbb{R}$  into a subset of  $[0, 1] \cup [3, 4]$  which is a subset of  $[0, 4]$ . It follows that  $|\mathbb{R}| \leq |[0, 4]|$ . On the other hand,  $[0, 4]$  is a subset of  $\mathbb{R}$ , so  $|[0, 4]| \leq |\mathbb{R}|$ , and, by the Cantor-Bernstein Theorem (10.3.5),  $|\mathbb{R}| = |[0, 4]|$ . Since  $|[0, 4]| = |[0, 1]|$  (the function  $k(t) = \frac{t}{4}$  is a one-to-one mapping of  $[0, 4]$  onto  $[0, 1]$ ), the theorem follows.  $\square$

There is a theorem that can often be used to give very easy proofs that sets are countable. The next several results form the basis for that theorem.

**Theorem 10.3.9.** *A subset of a countable set is countable.*

*Proof.* Let  $\mathcal{S}$  be a countable set. If  $\mathcal{S}$  is finite, then the result is clear. If  $\mathcal{S}$  is infinite, then there exists a one-to-one function  $f$  mapping the set of natural numbers onto  $\mathcal{S}$ . Thus the elements of  $\mathcal{S}$  can be listed in a sequence:  $(s_1, s_2, s_3, s_4, \dots)$  where  $s_j = f(j)$  for all  $j$ . If  $\mathcal{S}_0$  is a subset of  $\mathcal{S}$ , then the elements of  $\mathcal{S}_0$  correspond to some of the elements in the sequence. Thus the

elements of  $\mathcal{S}_0$  can be listed as well, and hence  $\mathcal{S}_0$  is either finite or has the same cardinality as  $\mathbb{N}$ .  $\square$

**Corollary 10.3.10.** *If  $\mathcal{S}$  is any set and there exists a one-to-one function  $f$  mapping  $\mathcal{S}$  into the set of natural numbers, then  $\mathcal{S}$  is countable.*

*Proof.* Let  $f$  be a one-to-one function taking  $\mathcal{S}$  into  $\mathbb{N}$ . The range of  $f$  is some subset  $\mathbb{N}_0$  of  $\mathbb{N}$ . Since  $f$  is a one-to-one function taking  $\mathcal{S}$  onto  $\mathbb{N}_0$ , it follows that  $|\mathcal{S}| = |\mathbb{N}_0|$ . By the previous theorem,  $\mathbb{N}_0$  is countable, and therefore so is  $\mathcal{S}$ .  $\square$

**Definition 10.3.11.** A *finite sequence* of elements of a set  $\mathcal{S}$  is an ordered collection of elements of  $\mathcal{S}$  of the form  $(s_1, s_2, s_3, \dots, s_k)$ .

For example, one finite sequence of rational numbers is  $(-\frac{1}{2}, -7, \frac{22}{7}, 0)$ .

**Theorem 10.3.12.** *The set of all finite sequences of natural numbers is countable.*

*Proof.* Let  $\mathbb{N}$  denote the set of natural numbers and let  $\mathcal{S}$  denote the set of all finite sequences of natural numbers. By the above corollary (10.3.10), it suffices to show that there is a one-to-one function  $g$  mapping  $\mathcal{S}$  into  $\mathbb{N}$ . Here is a description of one such function. Define  $g$  by defining the value of  $g$  at each given finite sequence of natural numbers to be the number whose digits are 1's and 0's, determined as follows (where the 0's are used for the commas): begin with the number of 1's equal to the first number in the given finite sequence, follow that by a 0, then follow that by the number of 1's equal to the second number in the sequence, then another 0, then the number of 1's corresponding to the third number in the sequence, then a 0, and so on, ending with the number of 1's corresponding to the last number in the sequence. For example,

$$g((2, 3, 7)) = 11011101111111$$

and

$$g((5, 1)) = 1111101$$

The function  $g$  is one-to-one since you can recover the unique sequence corresponding to any number in the range of  $g$  by using the definition of  $g$ . For example, the number 11110101111101111111 corresponds to the sequence  $(4, 1, 6, 8)$ . Since  $g$  is one-to-one and maps  $\mathcal{S}$  into  $\mathbb{N}$ ,  $\mathcal{S}$  is countable.  $\square$

**Corollary 10.3.13.** *If  $\mathcal{L}$  is any countable set, then the set of all finite sequences of elements of  $\mathcal{L}$  is countable.*

*Proof.* This follows easily from the above theorem. By hypothesis, there exists a one-to-one function  $f$  mapping  $\mathcal{L}$  into  $\mathbb{N}$ . Then a one-to-one function  $F$  mapping sequences of elements of  $\mathcal{L}$  into sequences of elements of  $\mathbb{N}$  can be obtained by defining

$$F(a_1, a_2, a_3, \dots, a_k) = (f(a_1), f(a_2), f(a_3), \dots, f(a_k))$$

Thus the previous theorem implies the corollary.  $\square$

The following definition will be useful.

**Definition 10.3.14.** Let  $\mathcal{S}$  and  $\mathcal{T}$  be any sets. We will say that  $\mathcal{T}$  *can be labeled* by the set  $\mathcal{S}$  if there is a way of assigning a finite sequence of elements of  $\mathcal{S}$  to each element of  $\mathcal{T}$  so that each finite sequence corresponds to at most one element of  $\mathcal{T}$ .

**Example 10.3.15.** The set  $\mathbb{Q}$  of rational numbers can be labeled by the set

$$\mathcal{L} = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, +, -, /\}$$

To label a given rational number, simply write it in the usual way using symbols from the set  $\mathcal{L}$ .

The following theorem is useful in many situations.

**Theorem 10.3.16** (The Enumeration Principle). *Every set that can be labeled by a countable set is countable.*

*Proof.* Let  $\mathcal{S}$  be a set that is labeled by the countable set  $\mathcal{L}$ . The fact that no two elements of  $\mathcal{S}$  have the same label implies that there is a one-to-one function  $f$  mapping  $\mathcal{S}$  into the set of labels. Thus there is a one-to-one function mapping  $\mathcal{S}$  into the set of finite sequences of elements of  $\mathcal{L}$ , which is a countable set by the above corollary (10.3.13). It follows from corollary 10.3.10 that  $\mathcal{S}$  is countable.  $\square$

Any set that can be proven to be countable by using the enumeration principle could, of course, also be proven to be countable without using this principle. However, the Enumeration Principle often leads to very simple proofs.

**Theorem 10.3.17.** *The set of all rational numbers is countable.*

*Proof.* As indicated above (example 10.3.15), the set of rational numbers can be labeled by the set

$$\mathcal{L} = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, +, -, /\}$$

The Enumeration Principle gives the result.  $\square$

**Corollary 10.3.18.** *The set of integers is countable.*

*Proof.* A subset of a countable set is countable (Theorem 10.3.9), so this follows from the previous theorem (10.3.17).  $\square$

You may have heard the assertion that  $\pi$  is a “transcendental” number; what does that mean?

**Definition 10.3.19.** The real number  $x_0$  is said to be *algebraic* if it is the root of a polynomial with integer coefficients. The real number  $x_0$  is said to be *transcendental* if there is no polynomial with integer coefficients that has  $x_0$  as a root.

For example, the number  $-\frac{3}{4}$  is algebraic, since it is a root of the polynomial  $4x + 3$ . More generally, every rational number  $\frac{m}{n}$  is algebraic since it is a root of the polynomial  $mx - n$ . There are also many irrational numbers that are algebraic, such as  $\sqrt{2}$  which is a root of the polynomial  $x^2 - 2$ , and  $(\frac{3}{4})^{\frac{1}{5}}$  which is a root of the polynomial  $4x^5 - 3$ .

It is not so easy to prove the existence of transcendental numbers. It is true that  $\pi$  is transcendental, but it is very difficult to prove that fact. It is somewhat easier, but still quite difficult, to prove that  $e$ , the base for natural logarithms, is transcendental. It is a very surprising and beautiful fact that it is much easier to prove that most real numbers are transcendental than it is to prove that any specific real number is transcendental. This is a corollary of the following.

**Theorem 10.3.20.** *The set of algebraic numbers is countable.*

*Proof.* We show that the set of algebraic numbers can be labeled by the set of integers and a comma; the Enumeration Principle then gives the result. We can label an algebraic number as follows. Specify the degree,  $n$ , of the polynomial of least degree with integer coefficients which has the number as a root. Then put a comma. Among the polynomials of degree  $n$  that have the number as a root, choose the one that has the smallest natural number coefficient of the  $x^n$ . Then specify the coefficients of that polynomial, in the order corresponding to descending powers of  $x$ , each separated by commas. Then put the integer 1, or 2, or 3 and so on to indicate that the algebraic number is the smallest, next to smallest, third from smallest, and so on, of the roots of that polynomial. In this manner, we label every algebraic number by a finite sequence of integers. Since the set of integers is countable (10.3.18), it follows from the Enumeration Principle that the set of algebraic numbers is countable.  $\square$



The above establishes the existence of transcendental numbers.

**Corollary 10.3.21.** *There exist transcendental numbers.*

*Proof.* Since the set of algebraic numbers is countable (10.3.20), and the set of all real numbers is uncountable (10.3.8), there are some real numbers that are not algebraic and thus are transcendental.  $\square$

The cardinality of a finite set consisting of  $n$  elements is said to be  $n$ . We now introduce some standard notation for the sizes of some infinite sets.

**Definition 10.3.22.** We say that the set  $\mathcal{S}$  has *cardinality*  $\aleph_0$  (which we read “aleph nought”) if the cardinality of  $\mathcal{S}$  is the same as that of the natural numbers. In this case we write  $|\mathcal{S}| = \aleph_0$ .

For example,  $|\mathbb{Q}| = \aleph_0$ .

There is also a standard notation for the cardinality of the set of real numbers.

**Definition 10.3.23.** We say that the set  $\mathcal{S}$  has *cardinality*  $c$  if the cardinality of  $\mathcal{S}$  is the same as the cardinality of the set of real numbers;  $c$  is sometimes said to be “the cardinality of the continuum”.

For example,  $|[3, 9]| = c$ .

Note that  $\aleph_0 < c$ , in the sense that every set with cardinality  $\aleph_0$  has cardinality less than every set with cardinality  $c$ .

It is important to note that  $\aleph_0$  is the smallest infinite cardinal number, in the following sense.

**Theorem 10.3.24.** *If  $\mathcal{S}$  is an infinite set, then  $\aleph_0 \leq |\mathcal{S}|$ .*

*Proof.* To establish this, we must show that  $\mathcal{S}$  has a subset  $\mathcal{S}_0$  of cardinality  $\aleph_0$ . We proceed as follows. Since  $\mathcal{S}$  is infinite, it surely contains some element, say  $s_1$ . Similarly,  $\mathcal{S} \setminus \{s_1\}$  (that is, the set obtained from  $\mathcal{S}$  by removing  $s_1$ ) contains some element, say  $s_2$ . Similarly,  $\mathcal{S} \setminus \{s_1, s_2\}$  contains some element  $s_3$ . Proceeding in this manner creates an infinite sequence  $\{s_1, s_2, s_3, \dots\}$  of elements of  $\mathcal{S}$ . Let  $\mathcal{S}_0 = \{s_1, s_2, s_3, \dots\}$ . Then clearly  $|\mathcal{S}_0| = |\mathbb{N}| = \aleph_0$ . Since  $\mathcal{S}_0$  is a subset of  $\mathcal{S}$ , it follows that  $\aleph_0 \leq |\mathcal{S}|$ .  $\square$

Thus  $\aleph_0$  is the smallest infinite cardinal number. Is there a largest cardinal number?

**Definition 10.3.25.** If  $\mathcal{S}$  is any set, then the set of all subsets of  $\mathcal{S}$  is called the *power set of  $\mathcal{S}$*  and is denoted  $\mathcal{P}(\mathcal{S})$ .

The terminology “power set of  $\mathcal{S}$ ” comes from the following theorem.

**Theorem 10.3.26.** *If  $\mathcal{S}$  is a finite set with  $n$  elements, then the cardinality of  $\mathcal{P}(\mathcal{S})$  is  $2^n$ .*

*Proof.* First note that this is true for  $n = 0$ . For the only set with 0 elements is the empty set,  $\emptyset$ . The set  $\emptyset$  has one subset, namely itself. Since  $2^0 = 1$ , the theorem holds for  $n = 0$ .

We proceed by mathematical induction. Suppose that every set with  $k$  elements has  $2^k$  subsets and let  $\mathcal{S}$  be a set with  $k + 1$  elements. Suppose that  $s_0$  is any element of  $\mathcal{S}$  and let  $\mathcal{S}_0$  be the subset  $\mathcal{S} \setminus s_0$  of  $\mathcal{S}$  obtained by removing  $s_0$ . Then  $\mathcal{S}_0$  has  $k$  elements and, by the inductive hypothesis,  $|\mathcal{P}(\mathcal{S}_0)| = 2^k$ . Suppose that  $\mathcal{T}$  is any subset of  $\mathcal{S}_0$ . Then  $\mathcal{T}$  is also a subset of  $\mathcal{S}$ . The set  $\mathcal{T} \cup \{s_0\}$  is a different subset of  $\mathcal{S}$ . Thus for each subset  $\mathcal{T}$  of  $\mathcal{S}_0$ , there are two subsets of  $\mathcal{S}$ ,  $\mathcal{T}$  and  $\mathcal{T} \cup s_0$ . It follows that there are twice as many subsets of  $\mathcal{S}$  as there are of  $\mathcal{S}_0$ . That is,

$$|\mathcal{P}(\mathcal{S})| = 2 \cdot |\mathcal{P}(\mathcal{S}_0)| = 2 \cdot 2^k = 2^{k+1}$$

The theorem follows by mathematical induction.  $\square$

What is the relationship between  $|\mathcal{S}|$  and  $|\mathcal{P}(\mathcal{S})|$  when  $\mathcal{S}$  is an infinite set?

**Theorem 10.3.27.** *For every set  $\mathcal{S}$ ,  $|\mathcal{S}| < |\mathcal{P}(\mathcal{S})|$ .*

*Proof.* It is easy to see that  $|\mathcal{S}| \leq |\mathcal{P}(\mathcal{S})|$ . For among the subsets of  $\mathcal{S}$  are the “singleton sets”; i.e., sets of the form  $\{s\}$  for  $s \in \mathcal{S}$ . The collection  $\mathcal{P}_0$  of all singleton subsets of  $\mathcal{S}$  is a subset of  $\mathcal{P}(\mathcal{S})$ . A one-to-one function  $f$  mapping  $\mathcal{S}$  into  $\mathcal{P}(\mathcal{S})$  can be defined by  $f(s) = \{s\}$  for all  $s$  in  $\mathcal{S}$ . Thus  $|\mathcal{S}| = |\mathcal{P}_0|$ , so  $|\mathcal{S}| \leq |\mathcal{P}(\mathcal{S})|$ .

To show that  $|\mathcal{S}| < |\mathcal{P}(\mathcal{S})|$ , we must show that there is no one-to-one function  $f$  taking  $\mathcal{S}$  onto  $\mathcal{P}(\mathcal{S})$ . The proof will use a “diagonal argument” similar to the proof that we gave that  $[0, 1]$  is uncountable (10.2.2).

Suppose, then, that  $f$  is any function taking  $\mathcal{S}$  into  $\mathcal{P}(\mathcal{S})$ . We will show that  $f$  cannot be onto; that is, that there is an element of  $\mathcal{P}(\mathcal{S})$  (i.e., a subset of  $\mathcal{S}$ ) that is not in the range of  $f$ .

For each  $s \in \mathcal{S}$ ,  $f(s)$  is a subset of  $\mathcal{S}$ . Define the subset  $\mathcal{S}_0$  of  $\mathcal{S}$  by

$$\mathcal{S}_0 = \{s \in \mathcal{S} : s \notin f(s)\}$$

That is, the subset  $\mathcal{S}_0$  of  $\mathcal{S}$  is defined to consist of all of those elements  $s$  of  $\mathcal{S}$  that are not in the subset of  $\mathcal{S}$  that  $f$  assigns to  $s$ .

The set  $\mathcal{S}_0$  is an element of  $\mathcal{P}(\mathcal{S})$ . We will show that it is not in the range of  $f$ . To prove this by contradiction, suppose that there was some  $s_0 \in \mathcal{S}$  such that  $f(s_0) = \mathcal{S}_0$ . We show that this is impossible by asking the question: is  $s_0$  in  $\mathcal{S}_0$ ? We will see that this question does not have an answer. For suppose that  $s_0 \notin \mathcal{S}_0$ . The definition of  $\mathcal{S}_0$  is that it contains those elements of  $\mathcal{S}$  that are not in the subsets they are sent to by  $f$ . Thus if  $s_0$  is not in  $f(s_0)$ ,  $s_0$  is in  $\mathcal{S}_0$ . In other words,  $s_0 \notin \mathcal{S}_0$  implies  $s_0 \in \mathcal{S}_0$ , which is a contradiction.

In the other direction, if  $s_0$  is in  $\mathcal{S}_0$ , then the definition of  $\mathcal{S}_0$  implies that  $s_0$  is not in  $f(s_0)$ . Since  $f(s_0) = \mathcal{S}_0$ ,  $s_0 \notin \mathcal{S}_0$ . Thus  $s_0 \in \mathcal{S}_0$  implies  $s_0 \notin \mathcal{S}_0$ , which is also a contradiction. But if there was an  $s_0$  satisfying  $f(s_0) = \mathcal{S}_0$ , then  $s_0$  would either be in  $\mathcal{S}_0$  or not be in  $\mathcal{S}_0$ . Therefore there is no  $s_0$  satisfying  $f(s_0) = \mathcal{S}_0$ , and the theorem is proven.  $\square$

One of the consequences of the theorem we have just established is that there is no largest cardinal number. For if  $\mathcal{S}$  is any set, there is a set whose cardinality is bigger than that of  $\mathcal{S}$ , namely  $\mathcal{P}(\mathcal{S})$ .

In particular, for  $\mathbb{R}$  the set of real numbers, the cardinality of  $\mathcal{P}(\mathbb{R})$ , the set of all sets of real numbers, is greater than  $c$ . Because of the analogy to the case of finite sets, it is standard to write  $|\mathcal{P}(\mathbb{R})| = 2^c$ .

Similarly,  $2^{\aleph_0}$  denotes the cardinality of  $\mathcal{P}(\mathbb{N})$ . Of course,  $\aleph_0 < 2^{\aleph_0}$ . Also, as we have seen,  $\aleph_0 < c$ . What is the relationship between  $2^{\aleph_0}$  and  $c$ ?

**Theorem 10.3.28.** *The cardinality of the set of all sets of natural numbers is the same as the cardinality of the set of real numbers. That is,  $|\mathcal{P}(\mathbb{N})| = c$ , or  $2^{\aleph_0} = c$ .*

*Proof.* Since  $|[0, 1]| = |\mathbb{R}|$  (10.3.8), it suffices to prove that  $|[0, 1]| = |\mathcal{P}(\mathbb{N})|$ . It will be convenient to introduce another set. Let  $\mathcal{S}$  denote the set of all infinite sequences of 0's and 1's (typical elements of  $\mathcal{S}$  are  $\{1, 0, 1, 0, 1, 0, \dots\}$ ,  $\{1, 1, 0, 1, 1, 1, \dots\}$  and so on). We begin by showing that  $|\mathcal{S}| = |\mathcal{P}(\mathbb{N})|$ . For this, we define a function  $f$  taking  $\mathcal{S}$  into  $\mathcal{P}(\mathbb{N})$  by

$$f(\{a_1, a_2, a_3, \dots\}) = \{i : a_i = 1\}$$

That is,  $f$  takes a sequence of 0's and 1's to the set of those natural numbers consisting of the places where the sequence has 1's. It is clear that  $f$  is one-to-one, for two different sequences would have at least one place where one has a 0 and the other has a 1, and the number corresponding to that place would be in

the subset corresponding to the second sequence but not the first. The function  $f$  is also onto, for if  $\mathcal{T}$  is any subset of  $\mathbb{N}$ , define a sequence  $\{a_i\}$  by letting  $a_i = 1$  if  $i$  is in  $\mathcal{T}$  and  $a_i = 0$  if  $i$  is not in  $\mathcal{T}$ . Then  $f(\{a_i\}) = \mathcal{T}$ . Thus  $|\mathcal{S}| = |\mathcal{P}(\mathbb{N})|$  and the theorem will be proven if we establish that  $|\mathcal{S}| = |[0, 1]|$ .

We define the function  $g$  mapping  $\mathcal{S}$  into  $[0, 1]$  by

$$g(\{a_1, a_2, a_3, \dots\}) = .a_1a_2a_3\dots$$

Since the  $a_i$ 's are 0's, and 1's, the range of  $g$  is contained in  $[0, 1]$ . Since two different sequences of 0's and 1's are sent by  $g$  to two different numbers,  $g$  is one-to-one. Thus  $|\mathcal{S}| \leq |[0, 1]|$ .

For the reverse inequality, we must produce a one-to-one function  $h$  that takes  $[0, 1]$  into  $\mathcal{S}$ . To do that, we represent the elements of  $[0, 1]$  as "binary decimals". That is, every element of  $[0, 1]$  can be written as an infinite sum

$$\frac{a_1}{2} + \frac{a_2}{2^2} + \frac{a_3}{2^3} + \frac{a_4}{2^4} + \dots$$

where each  $a_i$  is 0 or 1. (Of course, as with ordinary decimal representation, some elements of  $[0, 1]$  have more than one such representation. For example,  $\frac{1}{2} + \frac{0}{2^2} + \frac{0}{2^3} + \frac{0}{2^4} \dots$  represents the same number as  $\frac{0}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \frac{1}{2^4} \dots$ . In such ambiguous cases, choose either representation.)

After representing the elements of  $[0, 1]$  as binary decimals as above, define the mapping  $h$  taking  $[0, 1]$  into  $\mathcal{S}$  by

$$h\left(\frac{a_1}{2} + \frac{a_2}{2^2} + \frac{a_3}{2^3} + \frac{a_4}{2^4} + \dots\right) = \{a_1, a_2, a_3, a_4, \dots\}$$

The function  $h$  is a one-to-one mapping of  $[0, 1]$  into  $\mathcal{S}$ , so  $|[0, 1]| \leq |\mathcal{S}|$ . By the Cantor-Bernstein Theorem (10.3.5),  $|[0, 1]| = |\mathcal{S}|$ , proving the theorem.  $\square$

**Definition 10.3.29.** The *unit square in the plane* is the subset of the plane consisting of all points whose  $x$  and  $y$  coordinates are both between 0 and 1. That is, the unit square is the set  $\mathcal{S}$  defined by

$$\mathcal{S} = \{(x, y) : 0 \leq x \leq 1, 0 \leq y \leq 1\}$$

**Theorem 10.3.30.** *The cardinality of the unit square in the plane is  $c$ .*

*Proof.* Let  $\mathcal{S}$  denote the unit square. It is clear that  $|\mathcal{S}| \geq c$  since  $\mathcal{S}$  contains the subset

$$\mathcal{S}_0 = \{(x, 0) : 0 \leq x \leq 1\}$$

and there is an obvious pairing of  $\mathcal{S}_0$  with  $[0, 1]$ .

To establish the reverse inequality, we will construct a one-to-one function  $f$  mapping  $\mathcal{S}$  into  $[0, 1]$ . We represent the coordinates of points in the unit square as infinite decimals. In ambiguous cases (i.e., where a representation of a number could end in either a string of 0's or a string of 9's), we choose the representation ending in a string of 9's. We then define the function  $f$  by

$$f((.a_1a_2a_3\dots, .b_1b_2b_3\dots)) = .a_1b_1a_2b_2a_3b_3\dots$$

We claim that  $f$  is one-to-one. This follows since  $f((x, y)) = .c_1c_2c_3\dots$  implies that  $x = .c_1c_3c_5\dots$  and  $y = .c_2c_4c_6\dots$ . Thus  $|\mathcal{S}| \leq |[0, 1]|$ , so the Cantor-Bernstein Theorem gives  $|\mathcal{S}| = |[0, 1]|$ .  $\square$

It can be interesting to determine the cardinality of various sets of functions. We present one example now; many other examples are given in the problems. The following definition will be useful.

**Definition 10.3.31.** If  $\mathcal{S}$  is a set and  $\mathcal{S}_0$  is a subset of  $\mathcal{S}$ , then the *characteristic function* of  $\mathcal{S}_0$  as a subset of  $\mathcal{S}$  is the function  $f$  with domain  $\mathcal{S}$  defined by  $f(s) = 1$  if  $s \in \mathcal{S}_0$  and  $f(s) = 0$  if  $s \notin \mathcal{S}_0$ .

The following is a very easy but very useful fact.

**Theorem 10.3.32.** For any set  $\mathcal{S}$ , the set of all characteristic functions with domain  $\mathcal{S}$  has the same cardinality as  $\mathcal{P}(\mathcal{S})$ .

*Proof.* As indicated in the definition above of characteristic function, each subset does have a characteristic function. On the other hand, if two characteristic functions are equal as functions, they must be characteristic functions of the same subset (the subset consisting of all elements of the set on which the functions have value 1). Thus the correspondence between the set of subsets of  $\mathcal{S}$  and characteristic functions with domain  $\mathcal{S}$  is one to one and onto.  $\square$

**Theorem 10.3.33.** The cardinality of the set of all functions mapping  $[0, 1]$  into  $[0, 1]$  is  $2^c$ .

*Proof.* Among the functions are those that take on values contained in  $\{0, 1\}$ ; i.e., the characteristic functions with domain  $[0, 1]$ . By the previous theorem (10.3.32), this set of characteristic functions has cardinality  $2^c$ . Thus the set of all functions mapping  $[0, 1]$  into  $[0, 1]$  has cardinality at least  $2^c$ .

To prove the reverse inclusion, recall that every function is determined by its graph. The graph of a function  $f$  from  $[0, 1]$  to  $[0, 1]$  is  $\{(x, f(x)) : x \in [0, 1]\}$ ,

which is a subset of the unit square. Thus the set of functions we are considering corresponds to a collection of some of the subsets of the unit square, and hence has cardinality at most equal to that of the set of all subsets of the unit square. We have seen (10.3.30) that the cardinality of the unit square is  $c$ . It follows that the cardinality of the set of *all* subsets of the unit square is  $2^c$ . Therefore the cardinality of the set of graphs of functions is at most  $2^c$ . By the Cantor-Bernstein Theorem (10.3.5), the cardinality of the set of functions is  $2^c$ .  $\square$

There are some serious deficiencies in the general approach to set theory that we have been describing. The following illustrates some of the problems.

**Example 10.3.34.** (Cantor's Paradox) Let  $\mathcal{S}$  denote the set of all sets. Then every subset of  $\mathcal{S}$  is an element of  $\mathcal{S}$ , since each subset is a set. That is,  $\mathcal{P}(\mathcal{S})$  is a subset of  $\mathcal{S}$ . Hence  $|\mathcal{P}(\mathcal{S})| \leq |\mathcal{S}|$ . On the other hand, (10.3.27) implies that  $|\mathcal{S}| < |\mathcal{P}(\mathcal{S})|$ . The Cantor-Bernstein Theorem (10.3.5) proves that this is a contradiction.

What does this contradiction mean? If there is a contradiction, then something is false; but what? The only assumption that we have made is that there *is* a set consisting of the set of all sets. This contradiction shows that there cannot be such a set. To avoid Cantor's Paradox, the definition of set has to be more restrictive.

There is another paradox similar to Cantor's.

**Example 10.3.35.** (Russell's Paradox) Define a set to be *ordinary* if it is not an element of itself. (That is,  $\mathcal{S} \notin \mathcal{S}$ .) All of the sets that we have discussed so far except for the set of all sets are ordinary sets. Each set is, of course, a *subset* of itself, but that is very different from being a member of itself. (For example, the set of natural numbers is not a natural number.)

Let  $\mathcal{T}$  denote the set of all ordinary sets. We now ask the question: is  $\mathcal{T}$  an ordinary set? If  $\mathcal{T}$  was an ordinary set then, since  $\mathcal{T}$  is the set of all ordinary sets,  $\mathcal{T} \in \mathcal{T}$ . But then  $\mathcal{T}$  would not be an ordinary set, since it is an element of itself. On the other hand, if  $\mathcal{T}$  is not an ordinary set, then  $\mathcal{T} \in \mathcal{T}$ . But every element of  $\mathcal{T}$  is an ordinary set, so it would follow that  $\mathcal{T}$  is ordinary. That is, if  $\mathcal{T}$  is ordinary, it is not ordinary; if  $\mathcal{T}$  is not ordinary, it is ordinary. There cannot be such a set.

Note that the Cantor and Russell paradoxes are related in the following sense: the set of all sets, if it existed, would be a set that is not an ordinary set.

When mathematicians first observed the Cantor and Russell paradoxes, over a hundred years ago, they were very concerned. Why aren't "the set of all sets"

and “the set of all ordinary sets” themselves sets? What other “sets” are not really sets?

Mathematicians have developed several different “axiomatic set theories” in which the concept of “set” is restricted so that the Cantor and Russell paradoxes do not arise. In these set theories, there are no sets that are elements of themselves. The most popular of the axiomatic set theories is called “Zermelo-Fraenkel Set Theory”. The development of axiomatic set theories is fairly complicated and we will not discuss it here. However, the theorems that we presented in this chapter are also theorems in “Zermelo-Fraenkel Set Theory” although the formal proofs are slightly different.

The following is a very natural question: is there any set  $\mathcal{S}$  whose cardinality is greater than  $\aleph_0$  and less than  $c$ ? If there is such a set, it could be mapped into  $\mathbb{R}$ . That is, if there is any such set, there is a subset of  $\mathbb{R}$  with that property. The question can therefore be reformulated: if  $\mathcal{S}$  is an uncountable subset of  $\mathbb{R}$ , must the cardinality of  $\mathcal{S}$  be  $c$ ? This appears to be a very concrete question. It can be made even more concrete, as follows: if  $\mathcal{S}$  is a subset of  $\mathbb{R}$  and there is no one-to-one function taking  $\mathcal{S}$  into  $\mathbb{N}$ , must there exist a one-to-one function taking  $\mathcal{S}$  onto  $\mathbb{R}$ ?

**Definition 10.3.36.** The *Continuum Hypothesis* is the assertion that there is no set with cardinality strictly between  $\aleph_0$  and  $c$ .

It is very surprising that it is not known whether the Continuum Hypothesis is true or false. It is even more surprising that it has been proven that the Continuum Hypothesis is an undecidable proposition, in the following sense: it has been established that the Continuum Hypothesis can neither be proven nor disproven within standard set theories such as “Zermelo-Fraenkel Set Theory”. Mathematicians disagree about the full implications of this. It is our view that it is possible that someone will prove the Continuum Hypothesis in a way that would convince all mathematicians, in spite of its being undecidable within Zermelo-Fraenkel Set Theory. That is, someone might begin a proof as follows: “Let  $\mathcal{S}$  be an uncountable subset of  $\mathbb{R}$ . We construct a one-to-one function  $f$  mapping  $\mathcal{S}$  onto  $\mathbb{R}$  by first ...”. Any such proof would have to use something that was not part of Zermelo-Fraenkel Set Theory, since it has been proven that the Continuum Hypothesis cannot be decided within Zermelo-Fraenkel Set Theory. On the other hand, it is our opinion that it is possible that a proof could be found that would be based on properties of the set of real numbers that virtually every mathematician would agree are true, in spite of the fact that at least one of them would not be part of Zermelo-Fraenkel Set Theory. Some mathematicians believe that Zermelo-Fraenkel Set Theory captures all the reasonable

properties of the real numbers and thus conclude that no such proof is possible. We invite you to try to prove that those mathematicians are wrong by proving the Continuum Hypothesis.

## 10.4 Problems

### Basic Exercises

1. What is the cardinality of  $\mathbb{R}^2$  (the plane)?
2. What is the cardinality of the set of all complex numbers?
3. What is the cardinality of the unit cube (i.e.  $\{(x, y, z) : x \in [0, 1], y \in [0, 1], z \in [0, 1]\}$ )?
4. What is the cardinality of  $\mathbb{R}^3$  ( $\{(x, y, z) : x \in \mathbb{R}, y \in \mathbb{R}, z \in \mathbb{R}\}$ )?
5. Prove that the set of all finite subsets of  $\mathbb{Q}$  is countable.
6. What is the cardinality of the set of all functions from  $\mathbb{N}$  to  $\{1, 2\}$ ?
7. Let  $S$  and  $T$  be finite sets and let  $C = \{f : S \rightarrow T\}$  be the set of all functions from  $S$  to  $T$ . Show that if  $|T| > 1$ , then  $|C| \geq 2^{|S|}$ .

### Interesting Problems

8. What is the cardinality of the set of all functions from  $\{1, 2\}$  to  $\mathbb{N}$ ?
9. Find the cardinality of the set of all points in  $\mathbb{R}^3$  all of whose coordinates are rational.
10. Let  $\mathcal{S}$  be the set of all functions mapping the set  $\{\sqrt{2}, \sqrt{3}, \sqrt{5}, \sqrt{7}\}$  into  $\mathbb{Q}$ . What is the cardinality of  $\mathcal{S}$ ?
11. Find the cardinality of the set  $\{(x, y) \mid x \in \mathbb{R}, y \in \mathbb{Q}\}$ .
12. What is the cardinality of the set of all numbers in the interval  $[0, 1]$  which have decimal expansions with a finite number of non-zero digits?
13. What is the cardinality of the set of all numbers in the interval  $[0, 1]$  which have decimal expansions that end with an infinite sequence of 7's?



14. Let  $t$  be a transcendental number. Prove that  $t^4 + 7t + 2$  is also transcendental.
15. Suppose that the sets  $\mathcal{S}$ ,  $\mathcal{T}$  and  $\mathcal{U}$  satisfy  $\mathcal{S} \subset \mathcal{T} \subset \mathcal{U}$ , and that  $|\mathcal{S}| = |\mathcal{U}|$ . Show that  $\mathcal{T}$  has the same cardinality as  $\mathcal{S}$ .
16. Suppose that  $\mathcal{T}$  is an infinite set and  $\mathcal{S}$  is a countable set. Show that  $\mathcal{S} \cup \mathcal{T}$  has the same cardinality as  $\mathcal{T}$ .
17. Let  $A$  and  $B$  be countable sets. Prove that the cardinality of the Cartesian product of  $A$  and  $B$ ,  $A \times B = \{(a, b) \mid a \in A, b \in B\}$ , is countable.
18. Show that the set of all polynomials with rational coefficients is countable.
19. Let  $a$ ,  $b$  and  $c$  be distinct real numbers. Find the cardinality of the set of all functions mapping  $\{a, b, c\}$  into the set of real numbers.

### Challenging Problems

20. (a) Prove directly that the cardinality of the closed interval  $[0, 1]$  is equal to the cardinality of the open interval  $(0, 1)$  by constructing a function  $f : [0, 1] \rightarrow (0, 1)$  that is one-to-one and onto.  
(b) More generally, show that if  $S$  is an infinite set and  $\{a, b\} \subset S$ , then  $|S| = |S \setminus \{a, b\}|$ . (Hint: use the fact that  $S$  has a countably infinite subset containing  $a$  and  $b$ .)
21. Call a complex number *complex-algebraic* if it is a root of a polynomial with integer coefficients. Prove that the set of all complex-algebraic numbers is countable.
22. Assume that  $|A_1| = |B_1|$  and  $|A_2| = |B_2|$ . Prove:
  - (a)  $|A_1 \times A_2| = |B_1 \times B_2|$ .
  - (b) If  $A_1$  is disjoint from  $A_2$  and  $B_1$  is disjoint from  $B_2$ , then  $|A_1 \cup A_2| = |B_1 \cup B_2|$ .
23. Suppose that  $\mathcal{S}$  and  $\mathcal{T}$  each have cardinality  $c$ . Show that  $\mathcal{S} \cup \mathcal{T}$  also has cardinality  $c$ .
24. What is the cardinality of the set of all finite subsets of  $\mathbb{R}$ ?
25. Find the cardinality of the set of all lines in the plane.

- 
26. Show that the set of all functions mapping  $\mathbb{R} \times \mathbb{R}$  into  $\mathbb{Q}$  has cardinality  $2^c$ .
27. Let  $\mathcal{S}$  be the set of all real numbers that have a decimal representation using only the digits 2 and 6. Show that the cardinality of  $\mathcal{S}$  is  $c$ .
28. Let  $\mathcal{S}$  denote the collection of all circles in the plane. Is the cardinality of  $\mathcal{S}$  equal to  $c$  or  $2^c$ ?
29. Prove that if  $\mathcal{S}$  is uncountable and  $\mathcal{T}$  is a countable subset of  $\mathcal{S}$ , then the cardinality of  $\mathcal{S} \setminus \mathcal{T}$  is the same as the cardinality of  $\mathcal{S}$ .
30. Let  $\mathbb{Q}(\sqrt{2})$  be the set of real numbers of the form  $a + b\sqrt{2}$ , where  $a$  and  $b$  are rational numbers. Find the cardinality of  $\mathbb{Q}(\sqrt{2})$ .
31. Let  $S$  be the set of real numbers  $t$  such that  $\cos(t)$  is algebraic. Prove that  $S$  is countably infinite.
32. Let  $p(\mathbb{R})$  be the set of polynomials with real coefficients. That is,  $p(\mathbb{R})$  is the set of expressions functions of the form  $f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$ , where  $n$  is a non-negative integer (which depends on the function) and  $a_0, a_1, \dots, a_n$  are real numbers. Find the cardinality of  $p(\mathbb{R})$ .
33. Prove that the union of  $c$  sets which each have cardinality  $c$  has cardinality  $c$ .
34. Prove that the set of all sequences of real numbers has cardinality  $c$ .