Language and Proofs in Algebra: An Introduction
Version 1.1

April 8, 2020

Contents

Preface

These course notes are meant to accompany the lectures of MATH 135 at the University of Waterloo. A number of faculty members in Mathematics have contributed to the writing and preparation over a number of years.

This version is a revision for first use in Fall 2018.

Chapter 1

Introduction to the Language of Mathematics

1.1 The Language

> *Mathematics* is the language of mathematicians, and a *proof* is a method of communicating a mathematical truth to another person who speaks the "language". (Solow, *How to Read and Do Proofs*)

Mathematics is an extraordinarily precise language. When we state a mathematical result, our aim is to leave no doubt about the meaning of that result, nor about what we could do in order to prove that it is true. Then, if we are able to prove that the result is true, our aim in writing a proof is to do so with no ambiguity, so there is once again no doubt about the correctness of our proof. The level of detail that is used depends on the audience, which we will assume consists of average students in the course.

In this chapter we will concentrate on the *language* that is used in stating mathematical results, and on what would be required in order to determine that these results are true.

To illustrate, we consider the following mathematical results.

Example 1 For all integers $n \geq 5$,
$$2^n > n^2.$$

Example 2 There exists at least one integer a such that
$$a^2 + 29a + 209 \leq 0.$$

The language used to state the two examples above is typical of the results that we will be considering in this course. What does this language mean?

In Example 1, we are saying that the inequality $2^n > n^2$ holds for each choice of the integer n whenever $n \geq 5$. For example, when $n = 5$, we have $2^n = 2^5 = 32$ and $n^2 = 5^2 = 25$, and $32 > 25$, so the inequality holds in this case. But in order to prove that the first result

is true, we would need to prove that the inequality holds for every possible choice of n. Verifying the inequality for values of n one at a time is not an effective way to proceed. The methods of proof that we will learn in this course will allow us to learn how to handle situations of this type, and avoid simply verifying that some fact is true for every element of a given set.

In Example 2, we are saying that the inequality $a^2 + 29a + 209 \leq 0$ holds for at least one choice of the integer a. For example, when $a = 0$, we have $a^2 + 29a + 209 = 209$, and clearly $209 > 0$, so the inequality does not hold in this case. Trying again with $a = 5$, we have $a^2 + 29a + 209 = 379$, and clearly $379 > 0$, so the inequality does not hold in this case either. Trying once more with $a = -5$, we have $a^2 + 29a + 209 = 89$, and $89 > 0$, so again the inequality does not hold in this case. How long should we try values of a at random? In order to prove that this result is true, we would only need to find one value of a for which the inequality holds. However, as for Example 1 above, checking the inequality for values of a one at a time is not in general an effective way to proceed. The methods of proof that we will learn in this course will also allow us to precisely handle situations of this slightly different type.

1.2   Sets

The mathematical results in Examples 1 and 2 involve the integers, which collectively form a *set*. Sets are fundamental in mathematics, and the way in which we refer to them forms an important part of the language of mathematics.

Definition 1.2.1
set, element

A **set** is a well-defined, unordered collection of distinct objects. Each object that appears in this collection is called an **element** (or **member**) of the set.

Sets can contain any type of object. Since this is a mathematics course, we frequently use sets of numbers. But sets could contain letters, the letters of the alphabet for example, or books, such as those in a library collection. The simplest way to describe a set is to explicitly list all of its elements between a pair of brace brackets, { and }, separating the elements in the list by commas.

Example 3

The following are examples of sets:

1. $\{2, 4, 6, 8\}$ contains all the positive even numbers less that 10.

2. $\{1, 2, \{1, 2, 3\}\}$ is a set that contains three elements: 1, 2 and the set $\{1, 2, 3\}$. Note that the set $\{1, 2, 3\}$ is considered a single element of $\{1, 2, \{1, 2, 3\}\}$, even though $\{1, 2, 3\}$ contains three elements itself.

3. $\{\clubsuit, \diamondsuit, \heartsuit, \spadesuit\}$ contains the symbols of the four suits in a deck of playing cards.

4. The set of **natural numbers**, denoted by $\mathbb{N}$, lists all the positive integers. That is,

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

Note that this notation is not used consistently - in some parts of Mathematics and Computer Science the symbol $\mathbb{N}$ means the set of non-negative integers $\{0, 1, 2, 3, \dots\}$.

5. The set of **integers**, denoted by $\mathbb{Z}$, lists all integers, whether they are negative, zero, or positive. That is,
$$\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}.$$

6. The set of **rational numbers**, denoted by $\mathbb{Q}$, contains all numbers of the form $\frac{a}{b}$, where a is an integer and b is a non-zero integer.

7. The set of **real numbers**, denoted by $\mathbb{R}$, contains all numbers in decimal form.

Many of the mathematical results that we will encounter in this course concern the specially designated sets $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{R}$ described above. Except for these specially designated sets, usually we will use uppercase letters (S, T, U, etc.) to represent sets and lowercase letters (x, y, z, etc.) to represent elements of those sets. If x is an **element of** the set S, we write $x \in S$. If x is not an element of the set S, we write $x \notin S$.

Example 4   The following examples show how this notation is used:

1. Suppose $S = \{2, 4, 6, 8\}$. Then $6 \in S$, but $7 \notin S$.

2. Let $T = \{1, 2, \{1, 2, 3\}\}$. In this case, $1 \in T$, $2 \in T$ and $\{1, 2, 3\} \in T$, but $3 \notin T$.

3. We have $3 \in \mathbb{N}$ and $8 \in \mathbb{N}$, but $-3 \notin \mathbb{N}$ and $\frac{8}{3} \notin \mathbb{N}$.

1.3   Mathematical Statements and Negation

The mathematical results in Examples 1 and 2 are written as sentences asserting that some mathematical facts hold. We say that these results are mathematical *statements*.

Definition 1.3.1   A **statement** is a sentence that has a definite state of being either true or false.

statement

Example 5   Here are some examples of sentences that are statements of a particularly simple type.

1. $2 + 3 = 6$. (A false statement.)

2. $\pi + 2 \geq 5$. (A true statement.)

3. $\sqrt{2}$ is a rational number. (A false statement.)

The mathematical results that we will encounter in this course are assertions that some statement is true. When reading such a statement, we should realize that what is being said has to be either true or false (but cannot be both), even if we do not know the truth value of that statement.

Example 6   On the other hand, the following are examples of sentences that are *not* mathematical statements.

1. Does $7 = 5$?

2. Find the smallest positive integer.

3. Let s be the length of the side of a square.

4. This statement is false.

Why are the above sentences not mathematical statements? Questions, like "Does $7 = 5$?", are never statements. We simply cannot assign a "true" or a "false" state to questions. Similarly, instructions, like "Find the smallest positive integer." or "Let s be the length of the side of a square." cannot be given truth values, and are therefore not statements. The last example is a bit of a puzzle. Can you see why it cannot be true but also cannot be false?

We now introduce some important terminology and notation for statements.

Definition 1.3.2
negation

Suppose that A is a statement. Then the **negation** of A, denoted by $\neg A$, is the statement asserting the opposite truth value to A. That is, $\neg A$ is false when A is true, and $\neg A$ is true when A is false.

Since the truth value of $\neg A$ is the opposite of the truth value of A, then the effect of taking **double negation**, that is $\neg(\neg A)$, is exactly what we expect: the truth value of $\neg(\neg A)$ is the same as that of A. In this situation, where A and $\neg(\neg A)$ are not identical as statements, but they do have the same truth values for all possible choices of statements A, we say that A and $\neg(\neg A)$ are **logically equivalent**, or that A is logically equivalent to $\neg(\neg A)$, and use the notation

$$\neg(\neg A) \equiv A. \tag{1.1}$$

(The term *logically* equivalent is used since statements are part of the study of *mathematical logic*.) We finish this section with some examples of negations of statements.

Example 7   The following statements are the negations of the statements in Example 5 above.

1. $2 + 3 \neq 6$. (A true statement.)

2. $\pi + 2 < 5$. (A false statement.)

3. $\sqrt{2}$ is not a rational number. (A true statement.)

(Note how none of the statements in this example use the negation symbol $\neg$ explicitly.)

1.4   Quantifiers and Quantified Statements

1.4.1   Universal and Existential Quantifiers

Having seen some of the simplest types of statements in the previous section, we now turn to the more complicated types of statements that we will typically encounter in this course. Recall the mathematical results given in Examples 1 and 2.

1. For all integers $n \geq 5$,
$$2^n > n^2.$$

2. There exists at least one integer a such that
$$a^2 + 29a + 209 \leq 0.$$

The above sentences are examples of *quantified statements*. Both statements involve a *variable*, "n" in the first example, and "a" in the second example. Both sentences involve a set, called the *domain* of the quantified statement, which is the set $\{5, 6, 7, \ldots\}$ for the first sentence, and the set of all integers $\mathbb{Z}$ for the second. The domain specifies the possible choices for the variable in each case.

Both sentences contain an inequality involving the variable. These inequalities are

$$2^n > n^2 \qquad\qquad \text{and} \qquad\qquad a^2 + 29a + 209 \leq 0,$$

respectively. Neither of these inequalities is a statement by itself; whether $2^n > n^2$ is true or false cannot possibly be determined, since n is a *variable*. However, once we specify a value for the variable n, then the inequality is definitely true or false. For example, when $n = 6$ we have $2^n = 2^6 = 64$, and $n^2 = 6^2 = 36$, so the inequality is true for $n = 6$ since $64 > 36$. Similarly, the inequality $a^2 + 29a + 209 \leq 0$ is not a statement by itself because whether it is true or false cannot be determined, since a is a variable. But when $a = -1$ we have $a^2 + 29a + 209 = 181$, so the inequality is false for $a = -1$ since $181 > 0$.

Each of the inequalities above is called an *open sentence* – a sentence that contains a variable, where the truth of the sentence is determined by the value of the variable chosen from the domain of the variable (and, technically, the question of truth has no meaning unless a value for the variable is specified).

Finally, the variables n and a above have been introduced through the phrases "for all" and "there exists", respectively. Such phrases are called *quantifiers*, and they describe "how many" elements of the domain are claimed to make the open sentence true. In this course, we will only consider these two types of quantifiers: the **universal quantifier** ("for all"), and the **existential quantifier** ("there exists").

REMARK

To summarize the above discussion, a **quantified statement** contains four parts:

- a **quantifier** (universal, or existential);

- a **variable** (any symbol representing a quantity or mathematical object);

- a **domain** (any set);

- an **open sentence** involving the variable (that is either true or false whenever a value of the variable chosen from the domain is specified).

To help us to be completely precise when dealing with quantified statements, as well as to be compact when it is helpful, we now introduce some new mathematical symbols and notation: the symbol $\forall$ is used to denote the universal quantifier, and the symbol $\exists$ is used to denote the existential quantifier. We let $P(x)$ denote an open sentence involving the variable x. The way in which these symbols are used, together with set notation, is described in the following definition.

Definition 1.4.1

quantified statement

We consider two types of quantified statements.

1. A **universally quantified statement** is of the form

$$\forall \, x \in S, \ P(x).$$

We read the above quantified statement as "For all x in S, $P(x)$ is true" or simply as "For all x in S, $P(x)$". This quantified statement is true when $P(x)$ is true for every element x in the set S, and this quantified statement is false otherwise.

In terms of the terminology above, for this quantified statement, the *quantifier* is universal, the *variable* is x, the *domain* is S, and the *open sentence* is $P(x)$.

2. An **existentially quantified statement** is of the form

$$\exists \, x \in S, \ P(x).$$

We read the above quantified statement as "There exists at least one value of x in S for which $P(x)$ is true" or simply as "There exists an x in S such that $P(x)$". This quantified statement is true when $P(x)$ is true for at least one element x in the set S, and this quantified statement is false otherwise.

In terms of the terminology above, for this quantified statement, the *quantifier* is existential, the *variable* is x, the *domain* is S, and the *open sentence* is $P(x)$.

The use of the capital letter "P" for the open sentence $P(x)$ is because we often think of the truth of $P(x)$ as representing the fact that x has "*property*" P.

Example 8

Using this notation, the mathematical results in Examples 1 and 2 discussed above can be written as the following quantified statements.

1. $\forall \, x \in \{5, 6, 7, \ldots\}, \ 2^x > x^2$,

2. $\exists \, x \in \mathbb{Z}, \ x^2 + 29x + 209 \leq 0$.

Example 9   Determine whether the following existentially quantified statement is true or false.

$$\exists\, x \in \mathbb{Z}, \;\; x^2 + 29x + 209 \leq 0$$

Solution: When $x = -15$, we have $x^2 + 29x + 209 = 225 - 435 + 209 = -1$, and clearly $-1 \leq 0$, so the inequality
$$x^2 + 29x + 209 \leq 0$$
is true for the integer $x = -15$. We conclude that this existentially quantified statement is true.

In order to establish that the existentially quantified statement in the above example is true, we only needed to find one integer value of x for which the inequality $x^2 + 29x + 209 \leq 0$ is true, and we were able to check that $x = -15$ is such a value. Finding such a value can be quite tricky, and we will talk in more detail about how one might do so in Chapter 3, when we start to study proofs of mathematical statements. For now, as a hint, we comment that expressing the quadratic as $x^2 + 29x + 209 = (x + 15)(x + 14) - 1$ is an effective way to see that $x = -15$ and $x = -14$ are two values of x such that $x^2 + 29x + 209 \leq 0$ is true. Can you see how expressing the quadratic in this form also helps to make it clear that there are no other integer values of x for which $x^2 + 29x + 209 \leq 0$?

Example 10   Determine whether the following universally quantified statement is true or false.

$$\forall\, x \in \mathbb{R}, \;\; \sin^2 x + \cos^2 x = 1.$$

Solution: As a standard fact about trigonometry, we know that $\sin^2 x + \cos^2 x = 1$ for every real number x, so we conclude that this universally quantified statement is true.

1.4.2   Quantifiers and Language

The mathematical results that are encountered in this course involving quantified statements will seldom be expressed in precisely the form "$\forall\, x \in S, \; P(x)$" or "$\exists\, x \in S, \; P(x)$". For example, we do not always use the symbol x for the variable. Indeed, the symbol that is used for the variable has no effect on the meaning of quantified statements. Thus, the universally quantified statements

$$\forall\, x \in S, \; P(x), \qquad \forall\, a \in S, \; P(a), \qquad \forall\, n \in S, \; P(n), \qquad \forall\, \alpha \in S, \; P(\alpha),$$

in which the symbols x, a, n, α are used for the variable, all have precisely the same meaning. Similarly, the existentially quantified statements

$$\exists\, x \in S, \; P(x), \qquad \exists\, b \in S, \; P(b), \qquad \exists\, m \in S, \; P(m), \qquad \exists\, s \in S, \; P(s),$$

in which the symbols x, b, m, s are used for the variable, all have precisely the same meaning.

Another way in which we vary the form of quantified statements is to express them in sentences consisting largely of English words, but with equivalent mathematical meanings as the symbols that are omitted. For example, the following statements all have the same meaning as the universally quantified statement "$\forall\, x \in \{5, 6, 7, \ldots\}, \; 2^x > x^2$" in Example 8 above.

- For all integers $x \geq 5$, $2^x > x^2$.

- For each positive integer t in $\{5, 6, 7, \ldots\}$, $2^t > t^2$.

- For any integer $n \geq 5$, $2^n > n^2$.

- Let m be a positive integer that is at least 5. Then 2^m is greater than m^2.

- $2^x > x^2$ for all integers $x \geq 5$.

Note that in the last statement above, we even changed the order of the sentence so that the open sentence $P(x)$ came first.

Similarly, the following statements all have the same meaning as the existentially quantified statement "$\exists\, x \in \mathbb{Z},\ x^2 + 29x + 209 \leq 0$" in Example 8 above.

- There exists an integer x for which $x^2 + 29x + 209 \leq 0$.

- There is an integer n such that $n^2 + 29n + 209 \leq 0$.

- For some integer y, $y^2 + 29y + 209 \leq 0$.

- For at least one $a \in \mathbb{Z}$, $a^2 + 29a + 209 \leq 0$.

- $x^2 + 29x + 209 \leq 0$ for some integer x.

In this course, the way in which quantified statements are written will vary, so students will get a lot of practice in translating the English sentences used in this text into their precise mathematical meaning.

1.4.3   Negation of Quantifiers

In the table below, we summarize what we learned about when quantified statements are true or false in Definition 1.4.1.

Quantified Statement	True	False
$\forall\, x \in S,\ P(x)$	when $P(x)$ is true for every $x \in S$	when $P(x)$ is false for at least one $x \in S$
$\exists\, x \in S,\ P(x)$	when $P(x)$ is true for at least one $x \in S$	when $P(x)$ is false for every $x \in S$

Recall that the negation of any statement is true exactly when the statement itself is false. To apply this point of view to quantified statements, look at the rightmost column in the Table above. Then we immediately obtain the following rules for *negation* of quantified statements.

- The negation of "For all $x \in S$, $P(x)$ is true" can be written equivalently as

    "There exists some $x \in S$ for which $P(x)$ is false."

Similarly to (1.1), we can write this symbolically as the logical equivalence

$$\neg \left(\forall\, x \in S,\ P(x) \right) \equiv \left(\exists\, x \in S,\ \neg\, P(x) \right), \tag{1.2}$$

noting that the statements on both sides of the "$\equiv$" sign have the same truth values for all possible choices of domains S, and open sentences $P(x)$.

- Also, the negation of "There exists some $x \in S$ such that $P(x)$ is true" can be written equivalently as

  "For all $x \in S$, $P(x)$ is false."

Again similarly to (1.1), we can write this symbolically as the logical equivalence

$$\neg \left(\exists\, x \in S,\ P(x) \right) \equiv \left(\forall\, x \in S,\ \neg\, P(x) \right), \tag{1.3}$$

once again noting that the statements on both sides of the "$\equiv$" sign have the same truth values for all possible choices of domains S, and open sentences $P(x)$.

1.5   Nested Quantifiers

1.5.1   Two Quantifiers

Most of the statements that we will see in this course are quantified statements with more than one quantifier, each quantifier associated with a variable and a domain. The quantifiers in a statement containing more than one quantifier are called **nested quantifiers**. It is crucial to note the order in which nested quantifiers appear, and to understand how different orders can radically change the mathematical meaning of these statements.

For example, consider the statement with two nested quantifiers

$$\forall\, s \in \mathbb{R},\ \exists\, t \in \mathbb{R},\ t > s. \tag{1.4}$$

A statement like this is read from left to right: "For all real numbers s, there exists a real number t such that $t > s$." To see how the order matters, we can think about this as playing a game with two players, one for each variable, say player "S" for variable s, and player "T" for variable t. Player S goes first, and announces their value for the real number s. Then player T goes second, and responds by choosing a value for the real number t. Hence player T knows player S's value for s before they make their choice, and can take advantage of that knowledge when choosing a value for t. For the statement (1.4) to be true, player T must be able to make a choice of t such that $t > s$ is true, for every value $s \in \mathbb{R}$ that could be chosen by player S. Of course player T can always choose a real number t such that $t > s$ is true, since they know the value of s before they make their choice (for example, player T could choose $t = s + 1$), no matter what that real number s is. This means that statement (1.4) is true.

Now consider the related statement with two nested quantifiers

$$\exists\, t \in \mathbb{R},\ \forall\, s \in \mathbb{R},\ t > s. \tag{1.5}$$

This is almost identical to statement (1.4), except that the order of the two quantifiers has been switched. Statement (1.5) says: "There exists a real number t, such that for all real numbers s, we have $t > s$." In terms of a game with two players, here player T goes first,

and must choose a value for the real number t before player S makes a choice for s. For the statement (1.5) to be true, player T must be able to make a choice of t such that $t > s$ is true, for every value $s \in \mathbb{R}$ that could be chosen by player S. But since player T goes first, player S can always choose a real number s for which $t > s$ is *false*, since they know the value of t before they make their choice (for example, if player S chooses $s = t$). This means that statement (1.5) is false.

In summary, we have shown that statement (1.4) is true, and that statement (1.5) is false. These examples illustrate that when dealing with two nested quantifiers when one of the quantifiers is universal and the other is existential, we must be very careful about the left to right order, since changing the order of the quantifiers can change the truth value of the statement.

In fact, suppose that X and Y are any sets, and $Q(x, y)$ is any open sentence whose truth values can be determined for x chosen from the domain X and y chosen from the domain Y. Then what we have demonstrated with the different truth values of statements (1.4) and (1.5) is that the quantified statements

$$\forall \ x \in X, \ \exists \ y \in Y, \ Q(x, y) \qquad \text{and} \qquad \exists \ y \in Y, \ \forall \ x \in X, \ Q(x, y)$$

are *not* logically equivalent. The situation is simpler when there are two nested quantifiers of the same type.

Two universal quantifiers: When both quantifiers are universal, the quantified statements

$$\forall \ x \in X, \ \forall \ y \in Y, \ Q(x, y) \qquad \text{and} \qquad \forall \ y \in Y, \ \forall \ x \in X, \ Q(x, y)$$

do have the same truth values for all choices of X, Y and $Q(x, y)$, and hence are logically equivalent. That is, in the nesting order, from left to right, it makes no difference whether we select $x \in X$ before $y \in Y$, or the other way around. Indeed we might use a phrase like "for all" only once when we express such a quantified statement in English, e.g., "For all x in X and y in Y, $Q(x, y)$ is true." or, equivalently, "For all y in Y and x in X, $Q(x, y)$ is true."

Two existential quantifiers: Similarly, when both quantifiers are existential, the quantified statements

$$\exists \ x \in X, \ \exists \ y \in Y, \ Q(x, y) \qquad \text{and} \qquad \exists \ y \in Y, \ \exists \ x \in X, \ Q(x, y)$$

also have the same truth values for all choices of X, Y and $Q(x, y)$, and hence are logically equivalent. That is, in the nesting order, from left to right, it makes no difference whether we select $x \in X$ before $y \in Y$, or the other way around. Here again we might use a phrase like "for some" only once when we express such a quantified statement in English, e.g., "For some x in X and y in Y, $Q(x, y)$ is true." or, equivalently, "For some y in Y and x in X, $Q(x, y)$ is true."

1.5.2   An Arbitrary Number of Quantifiers

So far, we have only considered the case of two nested quantifiers, but of course quantified statements can involve any number of nested quantifiers. We illustrate how to deal with many nested quantifiers in general by considering the following statement involving three quantifiers, and hence three variables x, y and z and three domains X, Y and Z, in which $R(x, y, z)$ is some open sentence whose truth depends on the choices of x, y and z:

$$\exists \ x \in X, \ \forall \ y \in Y, \ \exists \ z \in Z, \ R(x, y, z). \tag{1.6}$$

Again, we read statement (1.6) from left to right, beginning with the leftmost quantified variable x. We then regard the rest of the statement as an open sentence depending on the choice of x, and proceed through the rest of the statement in a similar manner, parsing it in layers, like an onion. The following parenthesized version of (1.6) might be helpful in identifying the "layers":

$$\exists \, x \in X, \; \left(\forall \, y \in Y, \; \Big(\exists \, z \in Z, \; R(x,y,z) \Big) \right). \tag{1.7}$$

For instance, the quantified statement in (1.6) above can be written as follows.

$$\left. \begin{array}{rrl} & \exists \, x \in X, & P(x), \\ \text{where} \quad P(x) & \text{is} \quad \forall \, y \in Y, & Q(x,y), \\ \text{where} \quad Q(x,y) & \text{is} \quad \exists \, z \in Z, & R(x,y,z). \end{array} \right\} \tag{1.8}$$

Here, we think of the quantifier $\forall \, y \in Y$ as being "nested" within the open sentence $P(x)$, and the quantifier $\exists \, z \in Z$ as being "nested" within the open sentence $Q(x,y)$. This is why we refer to the quantifiers in a statement with more than one quantifier as being *nested* (like the parentheses are nested in (1.7)). A useful way to think about nested quantifiers is in terms of the nested loops that are used in computer programming.

> ## REMARK
>
> One very important setting in which nested quantifiers appear in mathematics is the study of limits in calculus. For example, consider the following definition of the limit:
>
> *Let f be a function and let $a \in \mathbb{R}$. We say that f has* **limit** *L as x approaches a, or that L is the limit of f at $x = a$, if for any positive tolerance $\epsilon > 0$, we can find a cutoff distance $\delta > 0$ such that if the distance from x to a is less than δ, and if $x \neq a$, then $f(x)$ approximates L with an error less than ϵ. That is, if $0 < |x - a| < \delta$, then $|f(x) - L| < \epsilon$.*
>
> We won't say anything further about limits themselves here, but will note how to write the above definition in the notation of this section. First, let $\mathbb{R}^+$ denote the set of all positive real numbers.
>
> *The definition above says that the* **limit** *of f approaching a is equal to L exactly when the following quantified statement is true:*
>
> $$\forall \, \epsilon \in \mathbb{R}^+, \; \exists \, \delta \in \mathbb{R}^+, \; \forall \, x \in \mathbb{R}, \; Q(\epsilon, \delta, x), \tag{1.9}$$
>
> *where $Q(\epsilon, \delta, x)$ is the open sentence:*   *If $0 < |x - a| < \delta$, then $|f(x) - L| < \epsilon$.*
>
> The open sentence Q is a special type of sentence called an *implication*, that we will study in Chapter 2. Here we'll consider only the left to right order of the nested quantifiers in (1.9), particularly for variables ϵ and δ. In terms of the two player game we have used to discuss nested quantifiers, the definition says that the arbitrary value for the universally quantified variable ϵ is known *before* a value for the existentially quantified variable δ is chosen. Hence the value of δ can depend on the value of ϵ. Indeed, finding an appropriate choice of δ in terms of ϵ is a key part in typical proofs of limits using the $\epsilon - \delta$ definition in calculus.

1.5.3   Negation of Nested Quantifiers

So far, we have learned how to negate quantified statements with a single quantifier (and single variable). We have also learned how to interpret quantified statements with nested

quantifiers - more than one quantifier (and more than one variable) in a given left to right order. Negating a statement with nested quantifiers needs to be done with care - the order of the quantifiers is very important. However, it is straightforward if we simply negate each quantifier layer-by-layer starting from the left.

For example, consider the statement in (1.6), which was written in layers in (1.8). Using the rules (1.2) and (1.3) for negating quantified statements, logically equivalent expressions for the negations of these three layers are easily obtained, given below.

$$\left. \begin{array}{llll} & \neg \left(\exists\, x \in X,\; P(x)\right) & \equiv \left(\forall\, x \in X,\; \neg\, P(x)\right), \\ \text{where} & \neg\, P(x) & \equiv\; \neg \left(\forall\, y \in Y,\; Q(x,y)\right) & \equiv \left(\exists\, y \in Y,\; \neg\, Q(x,y)\right), \\ \text{where} & \neg\, Q(x,y) & \equiv\; \neg \left(\exists\, z \in Z,\; R(x,y,z)\right) & \equiv \left(\forall\, z \in Z,\; \neg\, R(x,y,z)\right). \end{array} \right\}$$

Now, putting the negations of the layers back together, we obtain

$$\neg \left(\exists\, x \in X,\; \forall\, y \in Y,\; \exists\, z \in Z,\; R(x,y,z)\right) \equiv \left(\forall\, x \in X,\; \exists\, y \in Y,\; \forall\, z \in Z,\; \neg\, R(x,y,z)\right),$$

as a logically equivalent expression for the negation of the triply nested statement in (1.6). Note that in the above logical equivalence, the variables (and their domains) appear in exactly the same left to right order on both sides; however, the existential quantifiers for x and z in (1.6) have been *switched* to universal quantifiers, and the universal quantifier for y has been *switched* to an existential quantifier; also the open sentence $R(x,y,z)$ has been *switched* to its negation $\neg\, R(x,y,z)$.

This "switching" pattern for the quantifiers and for the open sentence always occurs in the negation of a quantified statement with nested quantifiers. The proof of this fact can be obtained by writing the quantified statement in layers, as we have done in the example above. We won't give the proof here, nor will we give an exact formula for an arbitrary number of variables. Instead, to help make clear what happens in general, we will state the rules for negating statements with two nested quantifiers. Suppose that X and Y are any sets, and $Q(x,y)$ is any open sentence depending on $x \in X$ and $y \in Y$. Here are the general rules for negating two nested quantifiers:

- $\neg\,(\forall\, x \in X,\; \forall\, y \in Y,\; Q(x,y)) \equiv (\exists\, x \in X,\; \exists\, y \in Y,\; \neg\, Q(x,y))$

- $\neg\,(\forall\, x \in X,\; \exists\, y \in Y,\; Q(x,y)) \equiv (\exists\, x \in X,\; \forall\, y \in Y,\; \neg\, Q(x,y))$

- $\neg\,(\exists\, x \in X,\; \forall\, y \in Y,\; Q(x,y)) \equiv (\forall\, x \in X,\; \exists\, y \in Y,\; \neg\, Q(x,y))$

- $\neg\,(\exists\, x \in X,\; \exists\, y \in Y,\; Q(x,y)) \equiv (\forall\, x \in X,\; \forall\, y \in Y,\; \neg\, Q(x,y))$

We finish this chapter with an example of negating two nested quantifiers.

Example 11   Consider the quantified statement

$$\forall\, s \in \mathbb{R},\; \exists\, t \in \mathbb{R},\; t > s, \tag{1.10}$$

that previously appeared as statement (1.4) on page 14 in Section 1.5.1. Applying the above rules for negating two nested quantifiers, the negation of this quantified statement is

$$\exists\, s \in \mathbb{R},\; \forall\, t \in \mathbb{R},\; t \leq s, \tag{1.11}$$

noting that the negation of "$t > s$" is given by "$t \leq s$". Of course, this means that the quantified statements (1.10) and (1.11) have opposite truth values. Now, on page 14, we showed that statement (1.10) is true, so statement (1.11) must be false.

To provide an independent check of this, we'll now demonstrate directly that statement (1.11) is false, using the setting of two players that appeared in the discussion on page 14.

Statement (1.11) says: "There exists a real number s, such that for all real numbers t, we have $t \leq s$." In terms of a game with two players, here player S goes first, and must choose a value for the real number s before player T makes a choice for t. For the statement (1.11) to be true, player S must be able to make a choice of s such that $t \leq s$ is true, for every value $t \in \mathbb{R}$ that could be chosen by player T. But since player S goes first, player T can always choose a real number t for which $t \leq s$ is *false*, since they know the value of s before they make their choice (for example, if player T chooses $t = s + 1$). This demonstrates directly that statement (1.11) is indeed false.

Chapter 2

Logical Analysis of Mathematical Statements

In this chapter we will make a more detailed study of the mathematical statements that were introduced in Chapter 1.

2.1   Truth Tables and Negation

We begin by recalling that the *negation* of a statement A, denoted by $\neg A$, is the statement asserting the opposite truth value to A. We will also use "not A" to have the same logical meaning as $\neg A$. In this context, we regard "$\neg$", or equivalently "not", as a **logical operator** that acts on the statement A, and we regard $\neg A$ as a **logical expression**. A simple way to describe the truth value of a logical expression is a **truth table**, which lists the **truth value** for each statement involved, using "T" for "true" and "F" for "false".

Definition 2.1.1

not, negation

The truth value for "**not** A", written symbolically as $\neg A$, is defined by the truth table given below.

A	$\neg A$
T	F
F	T

We also refer to $\neg A$ as the **negation** of A.

The first column of the truth table above, headed by "A", gives the possible truth values for the statement A, and the second column, headed by "$\neg A$", gives the corresponding truth values for the statement $\neg A$. We could also add other columns to such a table; for example we have added a column headed by "$\neg(\neg A)$" below.

A	$\neg A$	$\neg(\neg A)$
T	F	T
F	T	F

Now, we have defined a statement to have a definite truth value. That is, any given statement A is either true or false, and hence can only correspond to a single row in a truth table. So, how do we interpret the multiple rows and the use of "A" in the truth tables above? We answer these questions in the following Remark.

REMARK

As in the study of mathematical logic, we consider the "A" in these truth tables to be a symbol or a variable that we will refer to as a **statement variable**. Once we have a truth table, we can determine the truth value corresponding to the logical expression for any column, where the statement variable "A" is replaced by any given statement. The row of the truth table that we use to do so depends on the truth value of the given statement.

In Chapter 1 we discussed the fact that A and $\neg(\neg A)$ have the same truth value for all choices of statement A, and hence we said that A and $\neg(\neg A)$ are **logically equivalent**, which we wrote symbolically using an "$\equiv$" sign as

$$\neg(\neg A) \equiv A. \tag{2.1}$$

For similar reasons as in a truth table, in a logical equivalence of logical expressions such as (2.1), we also regard A as a statement variable. Note that, in terms of the truth table, the logical equivalence (2.1) means precisely that the columns headed "A" and "$\neg(\neg A)$" are identical (i.e., they have T's and F's in exactly the same rows).

2.2   Conjunction and Disjunction

In this section, we introduce two logical operators, "and" and "or", which combine a pair of statements. Statements formed in this way are called *compound statements*.

Definition 2.2.1

compound, component

A **compound statement** is a statement composed of several individual statements, each of which is called a **component statement**.

Let A and B be statements. We now define the compound statement "A and B". Note that there are four rows, one row for each of the possible pairs of truth values for the statement variables A and B.

Definition 2.2.2

and, conjunction

The truth value for "A **and** B", written symbolically as $A \wedge B$, is defined by the truth table

A	B	$A \wedge B$
T	T	T
T	F	F
F	T	F
F	F	F

We also refer to $A \wedge B$ as the **conjunction** of A and B.

The truth table above tells us that the compound statement $A \wedge B$ is true when A and B are both true, and that $A \wedge B$ is false otherwise.

Example 1   The compound statement
$$(2 + 3 = 6) \wedge (\pi + 2 \geq 5)$$
is *false*, since the component statement "$2 + 3 = 6$" is false. On the other hand, the statement
$$(2 + 4 = 6) \wedge (\pi + 2 \geq 5)$$
is *true*, since both of the components "$2 + 4 = 6$" and "$\pi + 2 \geq 5$" are true.

More typically, conjunctions are embedded in a quantified statement, as in the following example.

Example 2   Determine whether the following universally quantified statement is true or false.

For all real numbers x, $x^2 \geq 0$ and $\sin^2 x + \cos^2 x = 1$.

Solution: In symbols, this statement can be written as
$$\forall \, x \in \mathbb{R}, \, (x^2 \geq 0) \wedge \left(\sin^2 x + \cos^2 x = 1 \right).$$
In Chapter 1, we learned that the universally quantified statement
$$\forall \, x \in \mathbb{R}, \, P(x)$$
is true when $P(x)$ is true for all elements x in $\mathbb{R}$. Here the open sentence $P(x)$ is the conjunction
$$(x^2 \geq 0) \wedge \left(\sin^2 x + \cos^2 x = 1 \right),$$
and we have learned above that this conjunction is true only when both of its components "$x^2 \geq 0$" and "$\sin^2 x + \cos^2 x = 1$" are true.

But, as a standard fact about the reals, we know that $x^2 \geq 0$ for all real numbers x. Also, as a standard fact about trigonometry, we know that $\sin^2 x + \cos^2 x = 1$ for all real numbers x. Hence, $P(x)$ is true for all real numbers x, and we conclude that this universally quantified statement is true.

Similarly, we now define the compound statement "A or B".

Definition 2.2.3   The truth value for "A **or** B", written symbolically as $A \vee B$, is defined by the truth table

or, disjunction

A	B	$A \vee B$
T	T	T
T	F	T
F	T	T
F	F	F

We also refer to $A \vee B$ as the **disjunction** of A and B.

The truth table above tells us that the compound statement $A \lor B$ is true when one of A or B or both are true, and that $A \lor B$ is false otherwise (when A and B are both false). Hence this logical operator "or" is *inclusive*.

Example 3   The compound statement
$$\left(2 + 3 = 6\right) \lor \left(\pi + 2 \geq 5\right)$$
is *true*, since the component statement "$\pi + 2 \geq 5$" is true. On the other hand, the statement
$$\left(2 + 3 = 6\right) \lor \left(\pi + 2 < 5\right)$$
is *false*, since both of the components "$2 + 3 = 6$" and "$\pi + 2 < 5$" are false.

Disjunctions, like conjunctions, are more typically embedded in a quantified statement, as in the following example.

Example 4   Determine whether the following universally quantified statement is true or false.

For all real numbers x, $x^3 \leq 0$ or $x \geq 0$.

Solution: In symbols, this statement can be written as
$$\forall\, x \in \mathbb{R},\; (x^3 \leq 0) \lor \left(x \geq 0\right).$$
This is a universally quantified statement of the form
$$\forall\, x \in \mathbb{R},\; P(x)$$
where the open sentence $P(x)$ is the disjunction
$$(x^3 \leq 0) \lor \left(x \geq 0\right),$$
and we have learned above that this disjunction is true precisely when either the component "$x^3 \leq 0$" is true, or the component "$x \geq 0$" is true, or both of these are true.

But, as a standard fact about the reals, we know that $x^3 \leq 0$ for all real numbers x such that $x \leq 0$. Hence, $P(x)$ is true for all real numbers x: when $x < 0$, the component "$x^3 \leq 0$" is true; when $x > 0$, the component "$x \geq 0$" is true; when $x = 0$, both of the components "$x^3 \leq 0$" and "$x \geq 0$" are true. We conclude that this universally quantified statement is true.

2.3   Logical Operators and Algebra

We have now defined three logical operators: $\neg$, $\land$, and $\lor$. We can use them in combination with a number of component statement variables to form more complicated logical expressions, and then determine truth values for these logical expressions from a truth table.

For example, the truth table below has columns headed by a number of different expressions involving two statement variables A and B. These include the negations $\neg(A \land B)$

and $\neg(A \vee B)$. Note that the brackets used in these expressions serve the same purpose as they do in normal arithmetic: they specify the *order of operations*. For example, when determining truth values for $\neg(A \wedge B)$, we first consider $A \wedge B$, and then apply the negation operator $\neg$.

A	B	$A \wedge B$	$\neg(A \wedge B)$	$A \vee B$	$\neg(A \vee B)$	$\neg A$	$\neg B$	$(\neg A) \vee (\neg B)$	$(\neg A) \wedge (\neg B)$
T	T	T	F	T	F	F	F	F	F
T	F	F	T	T	F	F	T	T	F
F	T	F	T	T	F	T	F	T	F
F	F	F	T	F	T	T	T	T	T

In this truth table, note that the column headed by $\neg(A \wedge B)$ is identical to the column headed by $(\neg A) \vee (\neg B)$. Similarly, the column headed by $\neg(A \vee B)$ is identical to the column headed by $(\neg A) \wedge (\neg B)$. The corresponding logical equivalences, known as **De Morgan's Laws**, are recorded below.

> **REMARK (De Morgan's Laws (DML))**
>
> For statement variables A and B, we have
>
> 1. $\neg(A \wedge B) \equiv (\neg A) \vee (\neg B)$
>
> 2. $\neg(A \vee B) \equiv (\neg A) \wedge (\neg B)$

Now, we have all had lots of experience in using the algebraic operations of *addition* ($+$) and *multiplication* ($\times$) on *real numbers*. These satisfy the rules of *algebra*, which are certain basic facts that we use routinely. For example, for variables x and y representing two real numbers, the *commutative* laws give

$$x + y = y + x, \qquad \text{and} \qquad x \times y = y \times x,$$

and for variables x, y and z representing three real numbers, the *associative* laws give

$$x + (y + z) = (x + y) + z, \qquad \text{and} \qquad x \times (y \times z) = (x \times y) \times z.$$

Also, to use the addition operator "$+$" in combination with the multiplication operator "$\times$", the *distributive* law gives
$$x \times (y + z) = (x \times y) + (x \times z).$$

It turns out that similar rules hold for using the logical operations of *conjunction* ($\wedge$) and *disjunction* ($\vee$) on statement variables, where the "$\equiv$" sign takes the place of the "$=$" sign. We record these rules below.

> **REMARK**
>
> For statement variables A, B and C, the following rules hold for the logical operators $\wedge$ and $\vee$:

Commutative Laws:

- $A \wedge B \equiv B \wedge A$

- $A \vee B \equiv B \vee A$

Associative Laws:

- $A \wedge (B \wedge C) \equiv (A \wedge B) \wedge C$

- $A \vee (B \vee C) \equiv (A \vee B) \vee C$

Distributive Laws:

- $A \wedge (B \vee C) \equiv (A \wedge B) \vee (A \wedge C)$

- $A \vee (B \wedge C) \equiv (A \vee B) \wedge (A \vee C)$

Each of the six rules above can be verified in a straightforward way using a truth table.

Example 5   We use the following truth table to verify the second Distributive Law. Note that there are eight rows, one row for each of the possible triples of truth values for the statement variables A, B and C.

A	B	C	$B \wedge C$	$A \vee (B \wedge C)$	$A \vee B$	$A \vee C$	$(A \vee B) \wedge (A \vee C)$
T	T	T	T	T	T	T	T
T	T	F	F	T	T	T	T
T	F	T	F	T	T	T	T
T	F	F	F	T	T	T	T
F	T	T	T	T	T	T	T
F	T	F	F	F	T	F	F
F	F	T	F	F	F	T	F
F	F	F	F	F	F	F	F

Since the columns headed by $A \vee (B \wedge C)$ and $(A \vee B) \wedge (A \vee C)$ are identical, these two expressions are logically equivalent. That is, $A \vee (B \wedge C) \equiv (A \vee B) \wedge (A \vee C)$.

Self Check 1   As in the previous example, use truth tables to verify the Commutative Laws, the Associative Laws, and the first Distributive Law.

By applying the above Laws, together with DeMorgan's Law, we can prove the logical equivalence of a large variety of logical expressions involving the operators $\neg$, $\wedge$ and $\vee$ without needing to use a truth table.

When we do this, we need to use one additional fact: If we have $S_1 \equiv S_2$ and $S_2 \equiv S_3$, then we can immediately conclude that $S_1 \equiv S_3$ (what we actually do is to write $S_1 \equiv S_2 \equiv S_3$,

and then conclude that $S_1 \equiv S_3$; hence we treat "$\equiv$" for logical expressions just like we treat the equal sign "$=$" for algebraic expressions). We refer to this as the *transitivity* of logical equivalence, and will use this fact many times in the course.

To see how this works, consider the following result.

Example 6   For statement variables A and B, prove that

$$\neg\big(A \wedge (\neg B)\big) \equiv \big(B \vee (\neg A)\big)$$

without using a truth table.

Solution: Starting with the logical expression to the left of the $\equiv$ sign, we have

$$\begin{aligned} \neg\big(A \wedge (\neg B)\big) &\equiv (\neg A) \vee \big(\neg(\neg B)\big), && \text{using De Morgan's Laws,} \\ &\equiv (\neg A) \vee B, && \text{using double negation,} \\ &\equiv B \vee (\neg A), && \text{using Commutative Laws.} \end{aligned}$$

Note that the final logical expression is $B \vee (\neg A)$, exactly what appears on the right of the $\equiv$ sign. Since we managed to arrive at $\big(B \vee (\neg A)\big)$ from $\neg\big(A \wedge (\neg B)\big)$ with the help of established logical equivalences, we conclude that $\neg\big(A \wedge (\neg B)\big) \equiv \big(B \vee (\neg A)\big)$.

Mathematicians refer to manipulations like those in the example above as *Boolean* algebra, named after George Boole, a 19th-century mathematician and logician.

2.4   Implication

We now turn to the compound statement "A implies B", which will be featured in virtually all of the mathematical results that we encounter in this course.

Definition 2.4.1

implies,
implication,
hypothesis,
conclusion

The truth value for "A **implies** B", written symbolically as $A \implies B$, is defined by the truth table

A	B	$A \implies B$
T	T	T
T	F	F
F	T	T
F	F	T

We also refer to $A \implies B$ as an **implication**; component A is referred to as the **hypothesis** for this implication, and component B is referred to as the **conclusion**.

The truth table above tells us that $A \implies B$ is false when A is true and B is false, and that $A \implies B$ is true otherwise. In other words, for an implication to be true, the truth of the hypothesis must be accompanied by the truth of the conclusion.

REMARK

We often refer to an implication $A \implies B$ via an implicit timeline: that is, we *start* with the hypothesis A, and *then proceed* towards the logical conclusion B. From this point of view, for an implication to be true, the truth of the hypothesis must lead logically to (i.e., *implies*) the truth of the conclusion. Hence, when writing the implication $A \implies B$ in English, we use conditional sentences such as

$$\text{If } A \text{ is true, then } B \text{ must be true,}$$

or, more often, simply
$$\text{If } A \text{ then } B.$$

To see how this works, we immediately turn to a pair of examples in which an implication is embedded in a universally quantified statement.

Example 7   Determine whether the following universally quantified statement is true or false.

$$\text{For all real numbers } x, \text{ if } x > 4, \text{ then } x^2 > 9.$$

Solution: In symbols, this statement can be written as

$$\forall\, x \in \mathbb{R},\ (x > 4) \implies (x^2 > 9).$$

In Chapter 1, we learned that the universally quantified statement

$$\forall\, x \in \mathbb{R},\ P(x)$$

is true when $P(x)$ is true for all elements x in $\mathbb{R}$. Here the open sentence $P(x)$ is the implication
$$(x > 4) \implies (x^2 > 9),$$

and we have learned above that this implication is false only when the hypothesis "$x > 4$" is true and the conclusion "$x^2 > 9$" is false.

Now we make a detailed study of all values of x in the domain of the universally quantified statement (which is the set $\mathbb{R}$ of real numbers). First, note that every real number x belongs to exactly one of the following four intervals labelled A, B, C, D:

- $A:$ $x < -3,$
- $B:$ $-3 \leq x \leq 3,$
- $C:$ $3 < x \leq 4,$
- $D:$ $4 < x.$

Then, using standard facts about the reals, we know the following:

- for values of x in A and C, the hypothesis "$x > 4$" is false and the conclusion "$x^2 > 9$" is true,

- for values of x in B, the hypothesis "$x > 4$" is false and the conclusion "$x^2 > 9$" is false,

- for values of x in D, the hypothesis "$x > 4$" is true and the conclusion "$x^2 > 9$" is true.

Hence, since there is no real number x for which the hypothesis is true and the conclusion is false, the implication

$$(x > 4) \implies (x^2 > 9)$$

is true for all values of $x \in \mathbb{R}$. We conclude that this universally quantified statement is true.

Example 8   Determine whether the following universally quantified statement is true or false.

For all real numbers x, if $x > 2$, then $x^2 > 9$.

Solution: In symbols, this statement can be written as

$$\forall\ x \in \mathbb{R},\ (x > 2) \implies (x^2 > 9).$$

As in Example 7 above, this is a universally quantified statement of the form

$$\forall\ x \in \mathbb{R},\ P(x)$$

where the open sentence $P(x)$ is the implication

$$(x > 2) \implies (x^2 > 9),$$

and this implication is false only when the hypothesis "$x > 2$" is true and the conclusion "$x^2 > 9$" is false.

Now we make a detailed study of all values of x in the domain $\mathbb{R}$. First, note that every real number x belongs to exactly one of the following four intervals labelled A, B, C, D:

- $A:$ $x < -3$,

- $B:$ $-3 \leq x \leq 2$,

- $C:$ $2 < x \leq 3$,

- $D:$ $3 < x$.

Then, using standard facts about the reals, we know the following:

- for values of x in A, the hypothesis "$x > 2$" is false and the conclusion "$x^2 > 9$" is true,

- for values of x in B, the hypothesis "$x > 2$" is false and the conclusion "$x^2 > 9$" is false,

- for values of x in C, the hypothesis "$x > 2$" is true and the conclusion "$x^2 > 9$" is false,

- for values of x in D, the hypothesis "$x > 2$" is true and the conclusion "$x^2 > 9$" is true.

Hence the implication

$$(x > 2) \implies (x^2 > 9)$$

is false for all values of x in C, and we conclude that this universally quantified statement is false.

REMARK

The examples above give typical situations in which implications occur. Note that in both examples, there are values of x in the domain of the universally quantified statement for which the hypothesis doesn't hold (i.e., is false). But these results state only that for values of x in the domain for which the hypothesis does hold, then the conclusion must also hold. These results say *nothing* about values of x in the domain for which the hypothesis doesn't hold; moreover, when the hypothesis doesn't hold, it is *irrelevant* to the truth of these results whether the conclusion holds or not. That is why we put a "T" in rows 3 and 4 of the truth table defining the implication $A \implies B$. This implication cannot possibly be made "false" by anything that happens when the hypothesis does not hold, and so we say the implication $A \implies B$ is "true" whenever the hypothesis A is "false".

Note that we determined the truth values of both the hypothesis and conclusion for every $x \in \mathbb{R}$ in Examples 7 and 8 above. We did this because this was our first encounter with implications, so we wanted to make clear what the possibilities were for the four rows of the truth table that defined an implication. However, as we made clear in the remark above, the third and fourth rows of the truth table for implication are irrelevant. In other words, when the hypothesis does not hold, there is no need to determine whether the conclusion holds or not.

REMARK

When determining the truth value of an implication, from now on we will only consider situations in which the hypothesis holds, and determine whether or not the conclusion holds for these. In Chapter 3, when we learn how to prove implications, this will be referred to as *assuming the hypothesis*.

Note also that, in order to demonstrate that the universally quantified statement in Example 8 is false, we did not need to identify all values in the domain for which the hypothesis "$x > 2$" is true and the conclusion "$x^2 > 9$" is false. It would have been sufficient to find a single value for which the hypothesis is true and the conclusion is false. Hence any value of x in C would work. For example, when $x = 3$, the hypothesis "$x > 2$" is true and the conclusion "$x^2 > 9$" is false, and this single value of x in the domain demonstrates that the universally quantified statement in Example 8 is false.

There are many variants of implications that arise in mathematics. Either the hypothesis or the conclusion, or both, can themselves be compound statements of various types. For example, for statements A, B, C, we may be asked to consider implications like

1. $(A \wedge B) \implies C$,

2. $(A \vee B) \implies C$,

3. $A \implies (B \wedge C)$,

4. $A \implies (B \vee C)$,

or even more complicated ones than these. When we encounter complicated implications like these, we will need to be able to analyze their truth values. One of these that will arise, and often creates difficulties, is implication 2. $(A \vee B) \implies C$ above, and we consider it in the following truth table.

A	B	C	$A \vee B$	$(A \vee B) \implies C$	$A \implies C$	$B \implies C$	$(A \implies C) \wedge (B \implies C)$
T	T	T	T	T	T	T	T
T	T	F	T	F	F	F	F
T	F	T	T	T	T	T	T
T	F	F	T	F	F	T	F
F	T	T	T	T	T	T	T
F	T	F	T	F	T	F	F
F	F	T	F	T	T	T	T
F	F	F	F	T	T	T	T

In this truth table, note that the column headed by $(A \vee B) \implies C$ is identical to the column headed by $(A \implies C) \wedge (B \implies C)$. We record the corresponding (and perhaps surprising!) logical equivalence below.

REMARK

For statement variables A, B and C, we have

$$\big((A \vee B) \implies C\big) \equiv \big((A \implies C) \wedge (B \implies C)\big). \tag{2.2}$$

This logical equivalence means that to prove that a statement of the form $(A \vee B) \implies C$ is true, we can instead prove that both of the implications $A \implies C$ and $B \implies C$ are true. We will encounter an example where this is useful in one of the proofs in Chapter 3.

2.5   Converse and Contrapositive

Another variant of an implication is to interchange the hypothesis and the conclusion.

Definition 2.5.1
converse

The implication $B \implies A$ is called the **converse** of $A \implies B$.

Consider the following truth table, which has columns headed by the implication $A \implies B$ and by its converse $B \implies A$.

A	B	$A \implies B$	$B \implies A$
T	T	T	T
T	F	F	T
F	T	T	F
F	F	T	T

Note that the columns headed by $A \implies B$ and $B \implies A$ are *not* identical. In particular, they differ both in rows 2 and 3. Hence, the converse of an implication is *not* logically equivalent to the implication itself.

REMARK

A common mistake is to think that the implication $A \implies B$ and its converse $B \implies A$ are logically equivalent. They are not!

To see why one needs to be careful when dealing with the converse of an implication, recall that in Example 7, we showed that the universally quantified statement

For all real numbers x, if $x > 4$, then $x^2 > 9$.

is *true*. In the following example, we show that the universally quantified statement obtained by replacing the implication above by its converse is actually *false*.

Example 9   Determine whether the following universally quantified statement is true or false.

For all real numbers x, if $x^2 > 9$, then $x > 4$.

Solution: In symbols, this statement can be written as

$$\forall \, x \in \mathbb{R}, \; P(x),$$

where the open sentence $P(x)$ is the implication

$$(x^2 > 9) \implies (x > 4).$$

But from the solution to Example 7, we know that for all real x such that A: $x < -3$, and for all real x such that C: $3 < x \le 4$, the hypothesis "$x^2 > 9$" is true, and the conclusion "$x > 4$" is false. Hence the implication

$$(x^2 > 9) \implies (x > 4)$$

is false for all values of x in A and C, and we conclude that this universally quantified statement is false.

Yet another variant of an implication is obtained by using negation.

Definition 2.5.2

contrapositive

The implication $(\neg B) \implies (\neg A)$ is called the **contrapositive** of $A \implies B$.

Consider the following truth table, which has columns headed by the implication $A \implies B$ and by its contrapositive $(\neg B) \implies (\neg A)$.

A	B	$A \implies B$	$\neg B$	$\neg A$	$(\neg B) \implies (\neg A)$
T	T	T	F	F	T
T	F	F	T	F	F
F	T	T	F	T	T
F	F	T	T	T	T

Note that the columns headed by $A \implies B$ and $(\neg B) \implies (\neg A)$ are identical. This logical equivalence is recorded below.

REMARK

For statement variables A and B, we have

$$(A \implies B) \equiv ((\neg B) \implies (\neg A)). \tag{2.3}$$

In other words, an implication is logically equivalent to its contrapositive.

To see how we can use this logical equivalence, recall again that in Example 7, we showed that the following universally quantified statement involving an implication is true.

> For all real numbers x, if $x > 4$, then $x^2 > 9$.

Then (noting that the negation of "$x > 4$" is "$x \leq 4$", and the negation of "$x^2 > 9$" is "$x^2 \leq 9$"), the logical equivalence of an implication with its contrapositive implies immediately that the following related universally quantified statement is also true.

> For all real numbers x, if $x^2 \leq 9$, then $x \leq 4$.

Finally, an implication is also logically equivalent to another type of compound statement, as shown in the following truth table.

A	B	$A \implies B$	$\neg A$	$(\neg A) \vee B$
T	T	T	F	T
T	F	F	F	F
F	T	T	T	T
F	F	T	T	T

Note that the column headed by $A \implies B$ is identical to the column headed by $(\neg A) \vee B$. The corresponding logical equivalence is recorded below, together with its negation, which follows from DeMorgan's Laws. When we are dealing with the negation of an implication, this equivalence can be particularly useful.

REMARK

For statement variables A and B, we have

$$(A \implies B) \equiv ((\neg A) \vee B), \qquad \neg(A \implies B) \equiv (A \wedge (\neg B)). \qquad (2.4)$$

2.6   If and Only If

We end Chapter 2 with our final compound statement.

Definition 2.6.1

if and only if, iff

The truth value for "A **if and only if** B", written symbolically as $A \iff B$, is defined by the truth table

A	B	$A \iff B$
T	T	T
T	F	F
F	T	F
F	F	T

Sometimes we concisely write A if and only if B as "A **iff** B".

The above truth table tells us that $A \iff B$ is true when A and B have the same truth values, and is false when they have opposite truth values.

Now, for another way of looking at if and only if statements, consider the following truth table.

A	B	$A \iff B$	$A \implies B$	$B \implies A$	$(A \implies B) \wedge (B \implies A)$
T	T	T	T	T	T
T	F	F	F	T	F
F	T	F	T	F	F
F	F	T	T	T	T

Note that the columns headed by $A \iff B$ and $(A \implies B) \wedge (B \implies A)$ are identical, so we conclude that $A \iff B$ is logically equivalent to the conjunction of the implication $A \implies B$ and its converse $B \implies A$. We record this logical equivalence below, together with the way in which it is often used when embedded in a universally quantified statement.

REMARK

1. For statement variables A and B, we have

$$(A \iff B) \equiv ((A \implies B) \wedge (B \implies A)). \qquad (2.5)$$

2. We also have the logical equivalence for universally quantified statements

$$\left(\forall\, x \in X,\ P(x) \iff Q(x)\right) \qquad (2.6)$$
$$\equiv \left((\forall\, x \in X,\ P(x) \implies Q(x)) \wedge (\forall\, x \in X,\ Q(x) \implies P(x))\right).$$

To see how if and only if statements work we conclude with a pair of examples, both embedded in universally quantified statements.

Example 10   Determine whether the following universally quantified statement is true or false.

For all real numbers x, $x > 4$ if and only if $x^2 > 9$.

Solution: In symbols, this statement can be written as

$$\forall\, x \in \mathbb{R},\ (x > 4) \iff (x^2 > 9),$$

and from the logical equivalence (2.6), this statement has the same truth value as

$$\Big(\forall\, x \in \mathbb{R},\ (x > 4) \implies (x^2 > 9)\Big) \wedge \Big(\forall\, x \in \mathbb{R},\ (x^2 > 9) \implies (x > 4)\Big).$$

Now in Example 7, we showed that

$$\forall\, x \in \mathbb{R},\ (x > 4) \implies (x^2 > 9)$$

is true. But in Example 9 we showed that the converse

$$\forall\, x \in \mathbb{R},\ (x^2 > 9) \implies (x > 4)$$

is false. Hence we conclude that the universally quantified statement is false.

Example 11   Determine whether the following universally quantified statement is true or false.

For all real numbers x, $\big((x < -3)$ or $(x > 3)\big)$ if and only if $x^2 > 9$.

Solution: In symbols, this statement can be written as

$$\forall\, x \in \mathbb{R},\ \big((x < -3) \vee (x > 3)\big) \iff (x^2 > 9).$$

To make a detailed study of all values of x in the domain $\mathbb{R}$, note that every real number x belongs to exactly one of the following three intervals labelled A, B, C:

- $A:$ $\quad x < -3$,
- $B:$ $\quad -3 \le x \le 3$,
- $C:$ $\quad 3 < x$.

Then, using standard facts about the reals, we know the following:

- for values of x in A and C, both "$\big((x < -3) \vee (x > 3)\big)$" and "$x^2 > 9$" are true,
- for values of x in B, both "$\big((x < -3) \vee (x > 3)\big)$" and "$x^2 > 9$" are false.

In other words, "$\big((x < -3) \vee (x > 3)\big)$" and "$x^2 > 9$" have the same truth values for all x in A, B and C, and hence for all $x \in \mathbb{R}$. Thus the if and only if statement

$$\big((x < -3) \vee (x > 3)\big) \iff (x^2 > 9)$$

is true for all $x \in \mathbb{R}$, and we conclude that this universally quantified statement is true.

Note that in Example 11, we could also have shown that the statement is true by considering the pair of universally quantified implications

- $\forall\, x \in \mathbb{R}, \;\; \big((x < -3) \vee (x > 3)\big) \implies (x^2 > 9)$,

- $\forall\, x \in \mathbb{R}, \;\; (x^2 > 9) \implies \big((x < -3) \vee (x > 3)\big)$,

and deducing that these are both true from the truth values of "$\big((x < -3) \vee (x > 3)\big)$" and "$x^2 > 9$" for x in the intervals A, B and C given in the solution above.

Chapter 3

Proving Mathematical Statements

What do mathematicians prove? Mathematicians prove statements, which we have learned a great deal about in Chapters 1 and 2.

A mathematical statement has a definite truth value. Given a statement, however, it is not always obvious whether the statement is true or false. We **prove** a statement when we demonstrate that it is true, and the argument that we create to do so is called a **proof**. We **disprove** the statement when we demonstrate that it is false (and we could call the argument to do so a *disproof*, though this term is used less frequently). Throughout this course we will encounter statements like this. Such statements are known as *propositions*.

> **REMARK**
>
> A **proposition** is a mathematical claim posed in the form of a statement that either needs to be proven true or demonstrated false by a valid argument. You will encounter several variations on the word proposition. A **theorem** is a particularly significant proposition. A **lemma** is a subsidiary proposition, or more informally, a "helper" proposition, that is used in the proof of a theorem. A **corollary** is a proposition that follows almost immediately from a theorem.

A *proof* is simply a series of convincing arguments that leaves no doubt that a given proposition is true. Proofs work by connecting our assumed knowledge from known facts, definitions and previously proven statements, in a mathematically accurate way to deduce a result that establishes the proposed truth. In this chapter, we will consider how to use the language and logical analysis developed in the previous chapter to create such proofs.

A major aim of this course is for students to learn how to create their own proofs, and there will be lots of opportunities to practice this on assignments, tests and exams. Sometimes it will be helpful to distinguish between discovering an argument, and writing it down clearly and effectively. First, you need to convince yourself of the mathematical and logical validity of each step. Then you need to organize and phrase the argument so that it convinces someone else that a given statement is true. Assume other students in the course are reading the proofs that you create. Facts they all should know can be used without proof. This audience should be able to follow your arguments without having to question any facts that you use.

Mathematicians also need to be skilled in the critical analysis of proofs. As you encounter proofs in these notes and in the course, read them with a critical eye to ensure you believe they are correct.

3.1   Proving Universally Quantified Statements

Typically, the mathematical statements that we consider involve quantifiers, so we begin by considering how to prove the universally quantified statement "$\forall\, x \in S,\; P(x)$". To do so, as we learned in Chapter 1, we need to justify that each element x of the set S satisfies the property $P(x)$. One way to do so would be to go through each and every element of S, and check that the open sentence P is always true.

However, unless the domain S has only a few elements, this is not an effective way to proceed. Mathematicians have developed various ways to effectively prove such statements, and we will consider a number of them in this section. We begin with the method of *direct proof*, in which we proceed as follows.

Proof Method

To prove the universally quantified statement "$\forall\, x \in S,\; P(x)$":

Choose a representative mathematical object $x \in S$. This cannot be a specific object. It has to be a placeholder, that is, a variable, so that our argument would work for any specific member of the domain S.

Then, show that the open sentence P must be true for our representative x, using known facts about the elements of S.

How do we choose a representative object x from S? We simply declare "Let x be an arbitrary element of S", or state "Let $x \in S$". Next, we start using the symbol x as if it has all the characteristics of a typical member of S. The philosophy here is that we could replace x by any particular element from S, and all our steps would be correct. Then, when we symbolically show that $P(x)$ must be satisfied, we are guaranteed that the open sentence is true for all elements of S.

We refer to this as a *direct* proof because we begin by choosing a representative element x of the set S, and then, using only properties that we know to be true for any element of S, we prove that the statement $P(x)$ is true. We always end with the truth of $P(x)$, we never begin with, or *assume*, the truth of $P(x)$. We call this a direct proof because we are directly proving what we are asked to prove - that, starting with any element of S, we can demonstrate that $P(x)$ is always true.

To see how this works, consider the following proposition.

Proposition 1

For every real number x, $x^2 + 1 \geq 2x$.

Let us read our first proof.

Proof of Proposition 1: Let x be an arbitrary real number. Therefore, $x - 1$ must also be a real number, and hence
$$(x - 1)^2 \geq 0.$$

Expanding the terms on the left side, we get $x^2 - 2x + 1 \geq 0$. Adding $2x$ to both sides yields $x^2 + 1 \geq 2x$. $\square$

Note that in the given proof, we did not specify any particular number, rather we worked with the algebraic symbol x. The significance of using x is that it establishes the rule in general, for an arbitrary real number x. The arguments that are used in the proof are just applications of known facts about real numbers and inequalities. Since we followed a valid line of reasoning, the last expression must hold true for the x we started with, thus proving the statement.

Note also that the proof did *not* consist entirely of mathematical symbols and inequalities. Instead, it was written in sentences consisting largely of English words, together with some mathematical symbols. The sentences provided an explanation of how we were proceeding, and the order of the different steps. In fact, this was also true of the way in which the proposition itself was stated. For example, we didn't use the universal quantifier symbol "$\forall$" in the statement of the proposition. Instead of writing "$\forall\ x \in \mathbb{R}$", we used some of the language introduced in Chapter 1, and wrote "For every real number x".

We continue with another example of a direct proof, this one involving the set of natural numbers $\mathbb{N} = \{1, 2, 3, 4, \ldots\}$.

Example 1   Prove the following statement.

For all natural numbers n, $3n + 5 \geq 1$.

Solution: Let n be an arbitrary natural number. Since the smallest natural number is 1, we must have
$$n \geq 1.$$
Multiplying this inequality by 3 gives
$$3n \geq 3,$$
and adding 5 on both sides of this inequality then gives
$$3n + 5 \geq 8.$$
Finally, we know that $8 \geq 1$, so we have
$$3n + 5 \geq 8 \geq 1,$$
and we conclude that $3n + 5 \geq 1$.

Now we consider a direct proof involving the set of integers $\mathbb{Z} = \{\ldots, -1, 0, 1, \ldots\}$. The proof uses a *case analysis*, in which we prove the universally quantified statement for different parts of the domain separately.

Example 2   Prove the following statement.

For every integer k, $k(k - 1) \geq 0$.

Solution: Every integer k is either less than or equal to -1, equal to 0 or 1, or greater than or equal to 2. We now prove the result separately for these three cases.

- **Case 1:** For $k \leq -1$, we have $k - 1 \leq -2$, and multiplying these inequalities (and reversing the inequality for the negative signs), we obtain $k(k - 1) \geq 2$. Hence we have $k(k - 1) \geq 0$ in this case.

- **Case 2:** For $k = 0$ or 1, we have $k = 0$ or $k - 1 = 0$, which gives $k(k - 1) = 0$. Hence we have $k(k - 1) \geq 0$ in this case.

- **Case 3:** For $k \geq 2$, we have $k - 1 \geq 1$, and multiplying these inequalities, we obtain $k(k - 1) \geq 2$. Hence we have $k(k - 1) \geq 0$ in this case.

Since every integer k is accounted for by one of these three cases, we have thus proved this universally quantified statement.

The case analysis in the proof of Example 2 allowed us to use different lines of reasoning in the various cases to prove the given result. This is typical of situations where a case analysis might be considered.

REMARK

In trying to prove a universally quantified statement, suppose that you are unable to find a single line of reasoning that works for all elements in the domain. Then you might consider a **case analysis**, in which use different lines of reasoning are used for different parts of the domain. Be certain that your cases cover all elements in the domain.

Our next example of a direct proof involves trigonometry. For the proof, it will be helpful to recall the following *double angle formulas* for sine and cosine.

$$\sin 2x = 2 \sin x \cos x \tag{3.1}$$
$$\cos 2x = \cos^2 x - \sin^2 x \tag{3.2}$$

Example 3   Prove the following statement.

For all real numbers x in the open interval $\left(-\frac{\pi}{4}, \frac{\pi}{4}\right)$,

$$\tan 2x = \frac{2 \tan x}{1 - \tan^2 x}.$$

Solution: Let x be an arbitrary real number such that $-\frac{\pi}{4} < x < \frac{\pi}{4}$. Then we have

$$
\begin{aligned}
\tan 2x &= \frac{\sin 2x}{\cos 2x} \\
&= \frac{2 \sin x \cos x}{\cos^2(x) - \sin^2 x}, && \text{from (3.1) and (3.2),} \\
&= \frac{2 \frac{\sin x}{\cos x}}{1 - \frac{\sin^2 x}{\cos^2 x}}, && \text{dividing top and bottom by } \cos^2 x, \\
&= \frac{2 \tan x}{1 - \tan^2 x}.
\end{aligned}
$$

Note that since $-\frac{\pi}{4} < x < \frac{\pi}{4}$, then we have $\cos 2x \neq 0$, and $\cos x \neq 0$, so there has been no division by 0 at any stage of the above proof.

The type of result given in Example 3 is often referred to as a *trigonometric identity*. Note that in our proof, we started with $\tan 2x$, the expression on the left hand side of the identity. Then we used known facts from trigonometry to give a sequence of equalities which ended with the expression on the right hand side of the identity.

We did *not* start by assuming the truth of the identity. However, in some high schools students do learn to prove identities by starting with the truth of the identity, and then to do the same thing to both sides until the identity $0 = 0$ is obtained. This method often causes students to give incorrect proofs, and is not used in this course.

REMARK

When proving an identity, do *not* start by assuming the truth of that identity.

So far, we have been asked to prove that a mathematical statement involving a universal quantifier is true. But of course not all such statements are true. Often we are asked to consider a mathematical statement and to determine whether it is true or false. If it is true, we are asked to prove it, and if it is false, we are asked to disprove it.

In the case of a universally quantified statement, in equation (1.2) on page 14 of Chapter 1 we have the logical equivalence

$$\neg\big(\forall\; x \in S,\; P(x)\big) \equiv \big(\exists\; x \in S,\; \neg P(x)\big).$$

Hence to disprove "$\forall\; x \in S,\; P(x)$" (i.e., to demonstrate that this universally quantified statement is false), it is equivalent to prove its negation "$\exists\; x \in S,\; \neg P(x)$" (i.e., to demonstrate that this existentially quantified statement is true). Hence we need to demonstrate that there exists an $x \in S$ for which "$\neg P(x)$" is true (i.e., $P(x)$ is false). Typically we prove the existence of such an x by identifying it, which we call "finding" it. This general procedure is stated as the proof method below.

Proof Method

To disprove the universally quantified statement "$\forall\; x \in S,\; P(x)$":

Find an element $x \in S$ for which the open sentence $P(x)$ is false. This process is called finding a **counter-example**.

To see how this works, consider the following example.

Example 4

Prove or disprove the following statement.

For every integer a, $3a + 5 \geq 1$.

Solution: This statement is false, and we disprove it by finding a counter-example. Suppose we let $a = -4$. Then we have

$$3a + 5 = -12 + 5 = -7 < 1,$$

so there exists an integer a for which the inequality $3a + 5 \geq 1$ is false. This means that the inequality is not true for all integers a, and we have disproved this statement (proved that it is false) by finding the counter-example $a = -4$.

Note that to disprove a universally quantified statement, we only need to find a single counter-example, but in general there may be many counter-examples since our usage of "there exists" has the same English meaning as "there is at least one". In fact, there is an infinite number of counter-examples of the statement above, since the inequality $3a + 5 \geq 1$ is false for every integer $a \leq -2$.

In Example 4 above, it was easy to find a counter-example, but in general, it might be trickier. We'll discuss a general strategy for proofs with the existential quantifier in the next section.

3.2   Proving Existentially Quantified Statements

In this section we consider how to prove the existentially quantified statement "$\exists\, x \in S,\ P(x)$". To do so, as we learned in Chapter 1, we need to justify that there is at least one element x in S for which $P(x)$ is true. Typically, we *find* such an element x (just like *finding* a counter-example in Section 3.1), described below.

Proof Method

To prove the existentially quantified statement "$\exists\, x \in S,\ P(x)$":

Provide an explicit value of x from the domain S, and show that $P(x)$ is true for this value of x. In other words, find an element of S that satisfies property P.

Of course, we could try to find such a value of $x \in S$ by **trial-and-error**, substituting various elements x from the domain S in the open sentence P, until we find an x for which $P(x)$ is true. This approach works well when the domain is small, and it is easy to verify whether the property is true or false for each element of the domain. However, usually our domains are large, typically infinite, and we proceed more methodically, as illustrated below for the following proposition.

Proposition 2

For some real number x, $x^2 + 3 = 4x$.

Proof of Proposition 2: To prove this result, we wish to find a real number x for which $x^2 + 3 = 4x$ is true. That is, such a real number x is a solution to the equation

$$x^2 + 3 = 4x. \tag{3.3}$$

Subtracting $4x$ from both sides, the equation becomes

$$x^2 - 4x + 3 = 0.$$

Now we factor the left hand side, to obtain the equation

$$(x - 3)(x - 1) = 0. \tag{3.4}$$

The only real numbers satisfying equation (3.4) are $x = 1$ and $x = 3$.

Now we check to make sure that these are not extraneous solutions to equation (3.3). Substituting $x = 3$ into the expression on the left hand side of (3.3), we obtain $3^2 + 3 = 12$, and substituting $x = 3$ into the expression on the right hand side of (3.3), we obtain $4 \cdot 3 = 12$, which confirms that $x = 3$ is indeed a solution to (3.3). This proves the result. (There is now no need to further consider $x = 1$ once we have found that $x = 3$ is a solution.) $\qquad\square$

Note that in the above proof of Proposition 2, we did not randomly try values of x in the domain $\mathbb{R}$ and check if the statement $P(x)$ is true. Instead, we *narrowed our search* by only looking for values of x for which $P(x)$ is true. In Proposition 2 above, the statement that $P(x)$ is true is the equation (3.3) involving x, and hence we narrowed our search by looking only for *solutions* to that equation. Of course, to solve the equation, we used only known facts from our mathematical background, and followed a valid line of reasoning. This involved transforming equation (3.3) to equation (3.4), and then solving (3.4).

In general, when we transform one equation to another, we may create additional solutions to the original equation (these are usually called **extraneous solutions**). One way to deal with this is the method we have used in the above proof. That is, confirm that the solutions to the transformed equation are not extraneous solutions to the original equation by substituting them into the original equation to check that they satisfy it.

Another way to deal with the possibility of extraneous solutions when transforming equations is to check that the transformations at every stage are *invertible* – a transformation to an equation is invertible when the solutions to the original equation are exactly the same as the solutions to the transformed equation. Equivalently, a transformation is invertible when for all values x in the domain, x is a solution to the original equation if and only if x is a solution to the transformed equation.

REMARK

It is important to understand that narrowing the search for *one or more* values of $x \in S$ for which $P(x)$ is true, is *not* the same as assuming the truth of $P(x)$ for *all* values of $x \in S$. (We already know that the latter is not allowed when trying to prove the universally quantified statement "$\forall\ x \in S,\ P(x)$".)

It is also important to understand that to prove Proposition 2, we do not need to find all real solutions to quadratic equation (3.3) – any single solution is all that is needed to prove the required existence. It is perhaps even more important to understand that we don't need to consider the quadratic equation at all. We could instead prove the result by simply specifying a value in the domain and verifying that the open sentence is true for that value, without describing how we found that value. For example, the following is an alternative, and equally valid, proof of Proposition 2.

Alternative Proof of Proposition 2: Consider the real number $x = 3$. For $x = 3$, we have $x^2 + 3 = 3^2 + 3 = 12$. Also for $x = 3$, we have $4x = 4 \cdot 3 = 12$. Therefore, $x = 3$ is a real number for which $x^2 + 3 = 4x$ is true, and we have proved the result. $\square$

This alternative proof is certainly succinct, and it should convince almost any reader that the result is correct. However, it doesn't describe how the value $x = 3$ was discovered, so it may leave a reader wondering how to discover a proof of a different but similar statement.

Note that in both proofs of Proposition 2, we have continued to use sentences consisting largely of English words, together with some mathematical symbols. Moreover, instead of writing "$\exists\ x \in \mathbb{R}$" when the proposition itself was stated, we used some of the language introduced in Chapter 1, and wrote "For some real number x".

So far, we have considered proofs for quantified statements involving a single universal quantifier or a single existential quantifier. In the next example, we consider a pair of

quantified statements with one quantifier of each type, and recall from Section 1.5 that in this setting the quantifiers are said to be *nested*.

Example 5   Prove or disprove each of the following statements.

1. For all integers k, there is some real number x such that $x^2 + 2kx + k = 0$.

   Solution: We begin by expressing the quadratic as

   $$x^2 + 2kx + k = x^2 + 2kx + k^2 - k^2 + k = (x + k)^2 - k^2 + k = (x + k)^2 - k(k - 1),$$

   so the quadratric equation $x^2 + 2kx + k = 0$ can be rewritten in the form

   $$(x + k)^2 = k(k - 1). \tag{3.5}$$

   Let k be an arbitrary integer. To prove this result, we wish to choose some real number x that satisfies equation (3.5). Moreover, recall that with this nesting order of the quantifiers, we know the value of k before we choose the value of x, and hence the value of x can depend on the arbitrarily chosen value of k. But in Example 2 we proved that $k(k - 1) \geq 0$ for all integers k. Hence

   $$x = -k + \sqrt{k(k - 1)}$$

   is a real value of x, depending on k, that satisfies equation (3.5) (and therefore satisfies the given equation $x^2 + 2kx + k = 0$). Therefore, this statement is true, and we have proved it.

2. There is some real number x such that for all integers k, $x^2 + 2kx + k = 0$.

   Solution: To prove this result, we would need to choose a real number x that satisfies equation (3.5) for every integer k. However, with this nesting order of the quantifiers, we must choose the value of x without knowing the value of k, and that single choice of x must satisfy (3.5) for all integers k.

   Now, when $k = 1$, equation (3.5) becomes

   $$(x + 1)^2 = 0,$$

   and the only value of x satisfying this equation is $x = -1$. However, when $k = 0$, equation (3.5) becomes
   $$x^2 = 0,$$
   and the only value of x satisfying this equation is $x = 0$.

   Therefore, there is no single choice of x that satisfies equation (3.5) for both $k = 0$ and $k = 1$, so there certainly can be no choice of x satisfying (3.5) for *all* integer values of k. Therefore, this statement is false, and we have disproved it.

Note that the pair of quantified statements in Example 5 differ only in the order of the quantifiers. In particular, the statements can be written symbolically as

1. $\forall\, k \in \mathbb{Z},\ \exists\, x \in \mathbb{R},\ x^2 + 2kx + k = 0$,

2. $\exists\, x \in \mathbb{R},\ \forall\, k \in \mathbb{Z},\ x^2 + 2kx + k = 0$.

As we have seen in the above solutions for Example 5, the first of these quantified statements is true, and the second is false. The difference in truth values between these two quantified statements is similar to the pair of examples that were considered at the beginning of Section 1.5.1, and should be a good reminder that changing the order of nested quantifiers can change the truth value of a quantified statement.

Note also in the proof for part 1 of Example 5, that among the known facts about integers and real numbers that were used, was a result that was previously proved in Example 2. It is good to remember as the course proceeds, that every time we prove a new result, that result then becomes a "known fact" that we are able to use in future proofs if it is helpful.

Since not all existentially quantified statements are true, we'll finish this section by considering how to disprove an existentially quantified statement. In equation (1.3) on page 14 of Chapter 1 we have the logical equivalence

$$\neg\big(\exists\ x \in S,\ P(x)\big) \equiv \big(\forall\ x \in S,\ \neg P(x)\big).$$

Hence to disprove "$\exists\ x \in S,\ P(x)$" (i.e., to demonstrate that this existentially quantified statement is false), it is equivalent to prove its negation "$\forall\ x \in S,\ \neg P(x)$" (i.e., to demonstrate that this universally quantified statement is true). Hence we need to demonstrate that for all $x \in S$, "$\neg P(x)$" is true (i.e., $P(x)$ is false). This general procedure is stated as the proof method below.

Proof Method

To disprove the existentially quantified statement "$\exists\ x \in S,\ P(x)$":

Prove the *universally* quantified statement "$\forall\ x \in S,\ \neg P(x)$".

3.3   Proving Implications

Many results that mathematicians prove are implications, often with a universal quantifier. In this section, we will continue our development of the method of direct proof by considering statements that are implications.

Proof Method

For proving an implication:

1. To prove the implication "$A \implies B$", **assume** that the hypothesis A is true, and use this assumption to show that the conclusion B is true. The hypothesis A is what you start with. The conclusion B is where you must end up.

2. To prove the universally quantified implication "$\forall\ x \in S,\ P(x) \implies Q(x)$":

   Let x be an arbitrary element of S, assume that the hypothesis $P(x)$ is true, and use this assumption to show that the conclusion $Q(x)$ is true.

To see how this method of *assuming the hypothesis* works, consider the following proposition with three universally quantified variables.

Proposition 3   For all real numbers a, b and c, if $a \neq 0$ and $b^2 = 4ac$, then $x = -\frac{b}{2a}$ is a solution to $ax^2 + bx + c = 0$.

Proof of Proposition 3: Let a, b and c be arbitrary real numbers, and assume that $a \neq 0$ and $b^2 = 4ac$. Since $a \neq 0$, therefore we are allowed to divide by a, and hence the fraction $-\frac{b}{2a}$ is a real number. We know that products and sums of real numbers are also real numbers, so $ax^2 + bx + c$ is a real number for any real number x. Substituting the real number $x = -\frac{b}{2a}$ into the left side of $ax^2 + bx + c = 0$ and simplifying, we obtain

$$
\begin{aligned}
ax^2 + bx + c &= a\left(-\frac{b}{2a}\right)^2 + b\left(-\frac{b}{2a}\right) + c \\
&= a\left(\frac{b^2}{4a^2}\right) - \frac{b^2}{2a} + c \\
&= \frac{b^2}{4a} - \frac{b^2}{2a} + c \\
&= \frac{b^2 - 2b^2 + 4ac}{4a} \\
&= \frac{-b^2 + 4ac}{4a}, \\
&= 0,
\end{aligned}
$$

where for the last equality, we have used the assumption that $b^2 = 4ac$ to simplify the numerator, and the assumption that $a \neq 0$ to ensure that the denominator is not 0.

Therefore, $x = -\frac{b}{2a}$ is a solution to $ax^2 + bx + c = 0$.   □

Note that in the above proof we worked with algebraic symbols a, b and c, representing arbitrary real numbers that also satisfy the assumptions "$a \neq 0$" and "$b^2 = 4ac$". It is important to note that the required conclusion, which is the statement "$ax^2 + bx + c = 0$", did not appear until the very end of the above proof. However, the quantity $ax^2 + bx + c$, which is not itself a statement, was introduced earlier. Then, we used two types of information about the arbitrary real numbers a, b, c on the way to reaching our conclusion. First, we used known facts about all real numbers a, b, c. Second, we used ("assumed") the given hypothesis.

In the above proof we explicitly mentioned that "products and sums of real numbers are also real numbers". This is perhaps an example of a known fact that would be obvious to every student in the course, and could be omitted from the proof. Henceforth, we will omit explicit mention of many obvious facts of this type from our proofs.

We continue with a proposition that we have seen before, in Example 7 of Chapter 2.

Proposition 4   For all real numbers x, if $x > 4$ then $x^2 > 9$.

Proof: Let x be an arbitrary real number, and assume that $x > 4$. Then x is positive, so multiplying $x > 4$ on both sides by x gives the inequality $x^2 > 4x$. Also, multiplying $x > 4$ on both sides by 4 gives the inequality $4x > 4^2 = 16$. Combining these two inequalities, we obtain

$$ x^2 > 4x > 16 > 9, $$

and we conclude that $x^2 > 9$.   □

Note that the solution to Example 7 in the previous chapter, which also demonstrated that this universally quantified implication is true, was a lot longer and more involved than the above proof of Proposition 4. When we gave the solution to Example 7, it was early in our study of universally quantified implications, and the solution included additional material to help us understand implications of this type in detail. The proof of Proposition 4 above is the type of proof that you should be aiming for when you create your own proofs.

Now we consider a proposition where finding a proof might be more challenging, though checking that the proof is correct should be quite straightforward.

Proposition 5   For all real numbers s, t, x and y, if $x^2 + y^2 = 1$, then $(sx + ty)^2 \leq s^2 + t^2$.

Proof: Let s, t, x and y be arbitrary real numbers, and assume that $x^2 + y^2 = 1$. We now do some algebraic manipulation with s, t, x and y, to obtain

$$
\begin{aligned}
(sx + ty)^2 + (sy - tx)^2 &= s^2 x^2 + t^2 y^2 + 2stxy + s^2 y^2 + t^2 x^2 - 2stxy \\
&= s^2 x^2 + t^2 y^2 + s^2 y^2 + t^2 x^2 \\
&= \left(s^2 + t^2\right)\left(x^2 + y^2\right) \\
&= \left(s^2 + t^2\right)(1) \\
&= s^2 + t^2,
\end{aligned}
$$

where for the fourth equality we have used the hypothesis $x^2 + y^2 = 1$. Since $sy - tx$ is a real number, we have $(sy - tx)^2 \geq 0$, from which we get

$$(sx + ty)^2 = (sx + ty)^2 + 0 \leq (sx + ty)^2 + (sy - tx)^2 = s^2 + t^2,$$

and hence $(sx + ty)^2 \leq s^2 + t^2$.   □

So far, we have used various types of information in our proofs: known facts about elements of the given domain, previously proved results, and (for implications) a given hypothesis. Another type of information that mathematicians use in proofs is a **definition**. For example, here is a definition that you may already know, for an integer to be *even* or *odd*.

Definition 3.3.1   We say that an integer is **even** if it can be written in the form $2k$ where k is an integer.
even, odd   Otherwise, an integer can be written in the form $2k + 1$ where k is an integer and we say that the integer is **odd**.

Note that the English word "if" is used in the above definition. However, we should not confuse this usage with the "if" that appears in an implication. In a *definition*, the convention that mathematicians generally use is that "if" is used with the same meaning as "if and only if", or "means precisely". We will use this convention throughout the course.

In the definition of "even", for each even integer we are only requiring the *existence* of an integer k. This definition is an example of nested quantifiers that says: "For all even integers n, there exists an integer k such that $n = 2k$." For example, $n = 12 = 2 \cdot 6$ is even, with $k = 6$, and $n = 10 = 2 \cdot 5$ is even, with $k = 5$. Similarly, in the definition of "odd", for

each odd integer, we also have a nested quantifier requiring the existence of an integer k. For example $n = 17 = 2 \cdot 8 + 1$ is odd, with $k = 8$, and $n = -13 = 2 \cdot (-7) + 1$ is odd, with $k = -7$.

To understand the role of a definition in an implication and its proof, consider the following proposition.

Proposition 6   For all integers a, if a is odd, then $3a^2 + 4a - 1$ is even.

Before we prove this proposition, how does the definition of even and odd affect its statement? The statement is a short English sentence containing only a few symbols. This is because we have *defined* the English words "even" and "odd" to have a precise mathematical meaning involving other symbols, that therefore don't need to appear in the statement of the proposition.

Looking ahead, how will the definition of even and odd affect the proof of this proposition? Our **hypothesis** is that a is odd, so the definition of "odd" allows us to therefore assume that there exists an integer k such that $a = 2k + 1$. Our **conclusion** is that $3a^2 + 4a - 1$ is even, so the definition of "even" means that we therefore only need to prove that there exists an integer m such that $3a^2 + 4a - 1 = 2m$. Note that here m and k are different variable names, because the choice of k in the definition of even and odd depends on the even or odd integer being considered.

Proof of Proposition 6: Let a be an arbitrary integer, and assume that "a is odd" is true. This means that $a = 2k + 1$ for some integer k. Substituting $a = 2k + 1$ in the integer $3a^2 + 4a - 1$, we obtain

$$
\begin{aligned}
3a^2 + 4a - 1 &= 3(2k + 1)^2 + 4(2k + 1) - 1 \\
&= 3(4k^2 + 4k + 1) + 4(2k + 1) - 1 \\
&= 12k^2 + 12k + 3 + 8k + 4 - 1 \\
&= 12k^2 + 20k + 6 \\
&= 2(6k^2 + 10k + 3).
\end{aligned}
$$

Now $6k^2 + 10k + 3$ is an integer since k is an integer. Hence, from the definition, $3a^2 + 4a - 1$ is even.   □

Once again as in previous proofs, the conclusion "$3a^2 + 4a - 1$ is even" does not appear in the above proof until the very end. However, we consider the quantity $3a^2 + 4a - 1$ earlier. Then we use the hypothesis to prove that $3a^2 + 4a - 1 = 2m$ for the integer $m = 6k^2 + 10k + 3$, and hence deduce from the definition that $3a^2 + 4a - 1$ is even.

As we commented above, the definition of even and odd appears in both the hypothesis and conclusion of Proposition 6. But of course, because of the definition of even and odd, this means that implicitly an existential quantifier appears in the hypothesis. We used the algebraic symbol k in the proof, and worked with it using only properties that all integers satisfy.

This is typical of how we work with existential quantifiers in proofs of implications in general, whether they appear explicitly, or implicitly because of a definition as in the proof above. For example, suppose our hypothesis is an implication that is explicitly of the form

"$\exists\, x \in S,\ P(x)$". Then, to assume this hypothesis, we use a symbol, such as "x", to denote an element of the set S for which $P(x)$ is true. Since we do not know exactly which element of S satisfies $P(x)$, we cannot assign a specific value to x. Instead, we work with x as a variable. We then apply the rules of mathematics and other established results about the elements of S to this "x" in order to prove the desired conclusion.

REMARK

Especially since an important part of this course is for students to create their own proofs, it is also useful to comment at this stage on how proofs like the one above might actually be created or discovered. There is no precise recipe, but the process of creating a proof almost always involves **rough work** of some sort, and for this one does not need to only work forwards from the hypothesis to the conclusion. Instead, one might often also work *backwards* from the conclusion and meet somewhere in the middle. Then, when all steps of a complete proof have been discovered, a final version is written, in which it is important to proceed only *forwards*, from the hypothesis toward the conclusion.

From Proposition 6 and its proof, we have seen how definitions play an important part in the language of mathematics. In the next section, we will introduce new definitions that will give us considerable practice with implications and their proofs.

3.4   Divisibility of Integers

In this section we introduce *divisibility* of integers, starting with definitions, and then proving various propositions using those definitions. Here, this will give us considerable practice with proving implications. Later in the course, we develop the topic further to prove a number of important mathematical results.

For the time being, we will focus exclusively on integers. Division turns out to be a fairly complicated operation on integers. For example, if we try to divide 6 by 2, we get the integer 3 as a result; but if we divide 6 by 4, then the result $\frac{6}{4} = 1.5$ is no longer an integer. Since we want to deal solely with integers, we are interested in learning more about the cases where the result of a division is an integer.

Definition 3.4.1

divides, divisor, factor, multiple, divisible by

An integer m **divides** an integer n, and we write $m \mid n$, if there exists an integer k so that $n = km$.

If $m \mid n$, then we say that m is a **divisor** or a **factor** of n, and that n is a **multiple** of m or that n is **divisible by** m.

If m does not divide n, then we write $m \nmid n$.

Example 6

Consider the following examples.

- $3 \mid 6$ since $6 = 3 \times 2$. That is, there exists an integer k such that $6 = 3k$.

- $-3 \mid 6$ since $6 = (-3) \times (-2)$. That is, there exists an integer k such that $6 = (-3)k$.

- $5 \nmid 6$ since no integer k exists so that $6 = k \times 5$.

- For all integers a, $a \mid 0$ since $0 = 0 \times a$. In particular, $0 \mid 0$ is true.

- For all non-zero integers a, $0 \nmid a$ since there is no integer k so that $k \times 0 = a$.

- For all integers b, $1 \mid b$ since b can be written as $b = b \times 1$, and $-1 \mid b$ since b can be written as $b = (-b) \times (-1)$.

This definition is again an example of nested quantifiers that says: "for all integers n that are divisible by m, there exists an integer k such that $n = mk$."

3.4.1   Transitivity of Divisibility

We now consider our first proposition about divisibility, in which we prove that divisibility is *transitive*.

Proposition 7

(Transitivity of Divisibility (TD))

For all integers a, b and c, if $a \mid b$ and $b \mid c$, then $a \mid c$.

When one first encounters a proposition, as part of one's rough work, it often helps to work through some examples to understand the proof.

Example 7

Suppose $a = 3$, $b = 6$ and $c = 42$. Then "$a \mid b$ and $b \mid c$" becomes "$3 \mid 6$ and $6 \mid 42$", both of which are true, and hence Transitivity of Divisibility allows us to conclude that "$a \mid c$", which becomes "$3 \mid 42$".

Of course, you know that $3 \mid 42$. The strength of this proposition is that it works for all integers a, b and c that satisfy the condition "$a \mid b$ and $b \mid c$", not just for the particular integers of our example.

Proposition 7, which we refer to as "Transitivity of Divisibility", is a universally quantified implication featuring *divisibility* (i.e., $m \mid n$ for integers m and n) both in the hypothesis and in the conclusion. In the following proof of this proposition, notice how the definition of divisibility is used.

Proof of Proposition 7: Let a, b and c be arbitrary integers, and assume that $a \mid b$ and $b \mid c$. Since $a \mid b$, there exists an integer r so that $b = ra$. Since $b \mid c$, there exists an integer s so that $c = sb$. Substituting ra for b in the equation $c = sb$, we get

$$c = s(ra) = sra = (sr)a,$$

using properties of products of integers. But sr is an integer, so from the definition of divisibility, we conclude that $a \mid c$.   □

Note that we consider the integer c before the end of the above proof, but we do *not assume* that it is divisible by a. We are able to *conclude* that c is divisible by a by proving that $c = ka$ for some integer k, using only the assumptions, together with known facts about products of integers.

For more practice with proofs of implications involving divisibility in both the hypothesis and the conclusion, we consider the following proposition.

Proposition 8   For all integers a, b and c, if $a \mid b$ or $a \mid c$, then $a \mid bc$.

Before we give the proof, we will note something important about this result. It is a universally quantified statement in which the open sentence is an implication of the form $A \vee B \implies C$. Previously, in Propositions 3 and 7, we have encountered universally quantified statements in which the open sentence is an implication of the form $A \wedge B \implies C$. In our proofs of the latter results, when we assumed the hypothesis, we assumed that *both* A and B were true, and using these, we were able to deduce that C was true.

What does it mean to assume the hypothesis when it is of the form $A \vee B$? In equation (2.2) on page 29 of Chapter 2, we have the logical equivalence

$$\left(A \vee B \implies C\right) \equiv \left((A \implies C) \wedge (B \implies C)\right).$$

This means that proving the implication $A \vee B \implies C$ is equivalent to proving *both* of the implications $A \implies C$ and $B \implies C$.

Proof: Let a, b and c be arbitrary integers. From the discussion above, we will prove this result by proving *both* of the following implications:

1. If $a \mid b$, then $a \mid bc$,

2. If $a \mid c$, then $a \mid bc$.

Here are the two proofs:

1. Assume that $a \mid b$. Hence there exists an integer k so that $b = ka$. Using this equation to substitute for b in the product bc, we get

$$bc = (ka)c = kac = (kc)a,$$

using properties of products of integers. Now kc is an integer, so from the definition of divisibility, we conclude that $a \mid bc$.

2. Assume that $a \mid c$. Hence there exists an integer m so that $c = ma$. Substituting this equation, we get
$$bc = b(ma) = bma = (bm)a,$$
using properties of products of integers. Now bm is an integer, so from the definition of divisibility, we conclude that $a \mid bc$.

We have now proved the result.   □

Note that the proof of Proposition 8 consists of two separate proofs. In the first, we assume one portion of the hypothesis, and in the second, we assume the second portion of the hypothesis. This is another context in which mathematicians use a *case analysis*. Hence, another way of presenting the above proof would be to replace "1." by "**Case 1:**", and "2." by "**Case 2:**".

REMARK

Note also that the two implications 1 and 2 in the proof of Proposition 8 above are almost identical – they can be obtained from each other by interchanging variables b and c. In fact, with the same interchange of variables, the proofs of these implications are also almost identical word for word. Often for a pair of proofs like these, we won't repeat the details of the second proof. Instead, for example, we might replace the second proof above by:

2. The proof that if $a \mid c$, then $a \mid bc$ is similar, and is omitted.

3.4.2   Divisibility of Integer Combinations

In the previous section we used the proof techniques that we have learned so far to give a proof of *Transitivity of Divisibility*.

Divisibility is a mathematical concept that we will use extensively in this course, and it is therefore important to discover what general properties it satisfies. For example, we know that 5 divides both 10 and 15. Then, by *Transitivity of Divisibility* (*TD*), 5 divides all multiples of 10, and similarly, 5 divides all multiples of 15. If we add a multiple of 10 to a multiple of 15, does 5 divide the result? For instance, does 5 divide

$$(10 \times 3) + (15 \times 4) = 90?$$

Of course, in this case the answer is "yes"! It is not difficult to believe that 5 divides all possible *integer linear combinations* of 10 and 15 (i.e., any expression of the form $10x + 15y$, where x and y are integers). This is in fact true, and can be generalized to give the following proposition.

Proposition 9

(Divisibility of Integer Combinations (DIC))

For all integers a, b and c, if $a \mid b$ and $a \mid c$, then for all integers x and y, $a \mid (bx + cy)$.

Before we give the proof, we will note something important about this result. The conclusion uses the universal quantifier for integers x and y - these integers are not referred to in the hypothesis at all. What is the significance of the requirement that "x and y are integers"? For example, would the result still hold if this we changed this to the requirement that "x and y are rational numbers"? The answer is that the proposition would no longer be true, and has many counter-examples. As one counter-example, let $a = 3$, $b = 6$, $c = 27$, with $x = 1/3$ and $y = 2$. Then we have $bx + cy = 56$, which is an integer, but it is certainly not divisible by 3. As a second counter-example, let $a = 3$, $b = 6$, $c = 27$, with $x = 3/2$ and $y = 1/4$. Then we have $bx + cy = 99/2$, which is not even an integer (and hence "divisibility" does not make sense). Therefore, this result would be false if the variables x and y in the conclusion were arbitrary rational numbers, instead of simply integers.

Proof of Proposition 9: Let a, b and c be arbitrary integers, and assume that $a \mid b$ and $a \mid c$. Since $a \mid b$, there exists an integer r such that $b = ra$. Since $a \mid c$, there exists an integer s such that $c = sa$. Let x and y be arbitrary integers. Then $bx + cy$ is also an

integer, using known facts about products and sums of integers. Using the assumptions, we have

$$bx + cy = (ra)x + (sa)y = rax + say = (rx + sy)a.$$

Since $rx + sy$ is an integer, it follows from the definition of divisibility that $a \mid (bx + cy)$. □

Proposition 9 (which we will usually refer to simply as DIC) will be used repeatedly in this course. To help understand this important result, consider the following pair of examples, in which we have interchanged the hypothesis and conclusions of DIC in two different ways.

Example 8   Prove or disprove each of the following statements.

1. For all integers a, b and c, if $a \mid (bx + cy)$ for all integers x and y, then $a \mid b$ and $a \mid c$.

   Solution: This statement is true, and we give the following proof. We consider arbitrary integers a, b and c, and assume the hypothesis holds for those integers. That is, $a \mid (bx + cy)$ for all integers x and y. Therefore, the hypothesis holds for every choice of the integers x and y.

   For the choice $x = 1$ and $y = 0$, the hypothesis becomes $a \mid (b \cdot 1 + c \cdot 0)$. Now of course $b \cdot 1 + c \cdot 0 = b$, so for these choices of x and y, the hypothesis immediately gives $a \mid b$.

   For the choice $x = 0$ and $y = 1$, the hypothesis becomes $a \mid (b \cdot 0 + c \cdot 1)$. Now of course $b \cdot 0 + c \cdot 1 = c$, so for these choices of x and y, the hypothesis immediately gives $a \mid c$.

   We have thus proved the result, that $a \mid b$ and $a \mid c$.

2. For all integers a, b, c, x and y, if $a \mid (bx + cy)$, then $a \mid b$ and $a \mid c$.

   Solution: This statement is false, and we disprove it by finding a counter-example. Suppose we let $a = 3$, $b = 1$, $c = 2$, $x = 8$ and $y = 5$. Then we have

   $$bx + cy = 1 \cdot 8 + 2 \cdot 5 = 18,$$

   and 18 is indeed divisible by 3. However, 1 is *not* divisible by 3, and 2 is *not* divisible by 3. Therefore, for these choices of integers a, b, c, x and y, the hypothesis "$a \mid (bx + cy)$" is true, and the conclusion "$a \mid b$ and $a \mid c$" is false. Hence the implication is false for these choices of a, b, c, x and y, so we have found a counter-example to this universally quantified implication. We have thus disproved the result.

Note that the universal quantifier appears in the hypothesis of the implication in part 1 of Example 8, in the same way that it appears in the conclusion of Proposition 9. In part 1 of Example 8, we choose particular values of the variables x and y, and use the fact that the hypothesis is true for those values to prove that the conclusion is also true.

REMARK

This is typical of how we work with universal quantifiers in proofs of implications in general. For example, suppose our hypothesis is of the form "$\forall\ x \in S,\ P(x)$". Then, to assume this hypothesis, we can substitute one or more values of x chosen from the set S into the open sentence $P(x)$, and use the fact that $P(x)$ is true for these values to prove the desired conclusion. Of course, there is no recipe about how to choose appropriate values of x in any given situation, but after some practice, you will start to develop a feel for what will work.

We now turn to another pair of examples related to DIC.

Example 9   Prove or disprove each of the following statements.

1. For all integers a, b and c, if $a \mid (b + c)$ and $a \mid (2b + c)$, then $a \mid b$ and $a \mid c$.

   Solution: This statement is true, and we give the following proof. Let a, b and c be arbitrary integers, and assume that $a \mid (b+c)$ and $a \mid (2b+c)$. Then DIC implies that

   $$a \mid \big((b + c)x + (2b + c)y\big), \tag{3.6}$$

   for all integers x and y. Substituting the values $x = -1$ and $y = 1$, we obtain

   $$(b + c)x + (2b + c)y = (b + c) \cdot (-1) + (2b + c) \cdot 1 = -b - c + 2b + c = b,$$

   so for these values (3.6) becomes $a \mid b$. Hence we have proved that $a \mid b$.

   Also, substituting the values $x = 2$ and $y = -1$, we obtain

   $$(b + c)x + (2b + c)y = (b + c) \cdot 2 + (2b + c) \cdot (-1) = 2b + 2c - 2b - c = c,$$

   so for these values (3.6) becomes $a \mid c$. Hence we have proved that $a \mid c$.

   Since have proved that "$a \mid b$" is true, and that "$a \mid c$" is true, we have thus proved that "$a \mid b$ and $a \mid c$" is true.

2. For all integers a, b and c, if $a \mid (b + c)$ and $a \mid (3b + c)$, then $a \mid b$ and $a \mid c$.

   Solution: This statement is false, and we disprove it by finding a counter-example. Suppose we let $a = 2$, $b = 3$ and $c = 5$. Then we have $b + c = 8$ and $3b + c = 14$, both of which are divisible by 2, so for these choices of integers a, b and c, the hypothesis is true. But 3 is not divisible by 2, so component "$a \mid b$" of the conclusion is false, and therefore the conclusion is false (note that 5 is not divisible by 2, so component "$a \mid c$" of the conclusion is also false; but since the conclusion is a conjunction, we only need to demonstrate that one component is false in order to justify that the conclusion is false). Hence we have found a counter-example to this universally quantified implication, and we have disproved the result.

Note that in the proof for part 1 of Example 9, in addition to assuming the hypothesis, we have used DIC as a "known fact". This is typical of proofs that mathematicians give when they are creating a sequence of results in a single subject area - they use previously proved results in proofs of subsequent results. In this way, propositions often act as "building blocks" for knowledge in a subject area. In this course we will use exactly this process to create sequences of results in a few different but related subject areas of mathematics. Hopefully you will see that generally the later results will be more powerful than earlier results in each of these sequences of propositions.

So far, the only proofs we have considered are direct proofs. In the remaining sections of this chapter, we will consider other standard methods of proofs used by mathematicians. Before we do so, it is useful to note some of the pitfalls that can occur when trying to discover and create direct proofs. Therefore we end this section by giving two *incorrect* proofs.

Example 10   The following are examples of incorrect proofs.

1. **Statement:** For all integers a, b and c, if $a \mid b$ and $a \mid c$, then for all integers x and y, $a \mid (bx + cy)$.

   Incorrect proof: Let a, b and c be arbitrary integers, and assume that $a \mid b$ and $a \mid c$. Since $a \mid b$, there exists an integer k such that $b = ka$. Since $a \mid c$, there exists an integer k such that $c = ka$. Let x and y be arbitrary integers. Then $bx + cy$ is also an integer, using known facts about products and sums of integers. Using the assumptions, we have

   $$bx + cy = (ka)x + (ka)y = kax + kay = k(x + y)a.$$

   Since $k(x+y)$ is an integer, it follows from the definition of divisibility that $a \mid (bx+cy)$. $\square$

2. **Statement:** For all integers a, b and c, if $a \mid b$ and $a \mid c$, then $b \mid c$.

   Incorrect proof: Let a, b and c be arbitrary integers. From the definition of divisibility we have $b = ra$, $c = sa$ and $c = tb$ for some integers r, s, t. Then we have

   $$c = tb = t(ra) = tra = (tr)a.$$

   But we also have $c = sa$, so $sa = (tr)a$, which gives $s = tr$. Since t and r are integers, then their product tr is also an integer, which can then be equal to the integer s. Thus we have proved that $b \mid c$. $\square$

What is wrong with the proof in part 1 of Example 10? Here the statement is DIC itself, which we have previously proved to be true. Moreover, the incorrect proof above is almost identical to the correct proof that we gave on page 50. However, there is a difference: in the *correct* proof we used the hypothesis to get $b = ra$ and $c = sa$ for some integers r, s; in the *incorrect* proof, we use the hypothesis to get $b = ka$ and $c = ka$ for some integer k. But the latter then means that $b = c$ (since both are equal to ka). Hence the incorrect proof above is incorrect because it only works for choices of integers b, c with $b = c$, and not for *all* choices of integers b, c.

What is wrong with the proof in part 2 of Example 10? Here the statement is almost the same as TD, but the order of the three divisibilities $a \mid b$, $a \mid c$ and $b \mid c$ is different. In the incorrect proof we assume the conclusion $b \mid c$ at an early stage, and then write down some equations that are true. Then we state something about the integer s, and finish by claiming that we have proved the conclusion. But there is no logical basis for this claim. In fact, it is not possible to give a correct proof, since the statement given in part 2 is false, with many counter-examples. For example, for the choices of integers $a = 3$, $b = 15$ and $c = 21$, we have $a \mid b$ and $a \mid c$, but we certainly do not have $b \mid c$. This counter-example disproves the statement, so the statement is false and no claimed "proof" can be correct!

3.5   Proof by Contrapositive

As we have seen, many mathematical propositions are stated as implications, and we have now had some practice in using the method of direct proof for proving implications. The

procedure for a direct proof is simple: assume that the hypothesis is true and use it to prove the conclusion.

However, sometimes it is difficult to apply the method of direct proof, so mathematicians have developed other methods of proving implications that can be easier in some instances. The key to these methods is to prove a statement that is logically equivalent to the given implication. For example, equation (2.3) on page 31 of Chapter 2 gives the logical equivalence of an implication and its contrapositive. This is the basis of the *contrapositive method* for proving implications.

Proof Method

For proving an implication using the contrapositive:

1. To prove the implication "$A \implies B$", replace it with its contrapositive "$(\neg B) \implies (\neg A)$". Then prove this contrapositive, usually by a direct proof. That is, assume $\neg B$ is true and deduce that $\neg A$ must be true as well.

2. To prove the universally quantified implication "$\forall\, x \in S,\; P(x) \implies Q(x)$", replace it with its universally quantified contrapositive "$\forall\, x \in S,\; \big(\neg Q(x)\big) \implies \big(\neg P(x)\big)$". Then prove this universally quantified contrapositive.

As we have already seen in a number of situations, implications are usually encountered within a universally quantified statement. To see how proofs by contrapositive work, consider the following examples.

Example 11

Prove each of the following statements.

1. For all real numbers x, if $x^5 - 3x^4 + 2x^3 - x^2 + 4x - 1 \geq 0$, then $x \geq 0$.

   Proof: Let x be an arbitrary real number. We prove the contrapositive:

   > If $x < 0$, then $x^5 - 3x^4 + 2x^3 - x^2 + 4x - 1 < 0$.

   Hence assume that $x < 0$. Then $x^5 < 0$, $2x^3 < 0$ and $4x < 0$. In addition, $-3x^4 < 0$, $-x^2 < 0$ and $-1 < 0$. Adding these six inequalities together, we obtain

   $$x^5 - 3x^4 + 2x^3 - x^2 + 4x - 1 < 0.$$

   Since the contrapositive is true, the original implication must be true, and hence we have proved the universally quantified implication. □

2. For all integers a, if $3a^2 - 6a + 8$ is even, then a is even.

   Proof: Let a be an arbitrary integer. We prove the contrapositive:

   > If a is odd, then $3a^2 - 6a + 8$ is odd.

Hence assume that a is odd. This means that $a = 2k + 1$ for some integer k. Substituting $a = 2k + 1$ in the integer $3a^2 - 6a + 8$, we obtain

$$
\begin{aligned}
3a^2 + 5a - 2 &= 3(2k+1)^2 - 6(2k+1) + 8 \\
&= 3(4k^2 + 4k + 1) - 6(2k+1) + 8 \\
&= 12k^2 + 12k + 3 - 12k - 6 + 8 \\
&= 12k^2 + 5 \\
&= 2(6k^2 + 2) + 1.
\end{aligned}
$$

Hence by the definition, $3a^2 - 6a + 8$ is odd. Since the contrapositive is true, the original implication must be true, and hence we have proved the universally quantified implication. $\square$

Note that for each of the above examples, assuming the hypothesis in a direct proof would give us a fact about a polynomial in some variable. But analyzing a polynomial can be quite complicated. Instead, we prove the contrapositive: assuming the hypothesis in the contrapositive gives us a fact about the variable itself, and that makes the above proofs quite straightforward.

In general, when should we consider using the contrapositive method to prove an implication? There is no precise recipe to tell us when to use the contrapositive, but there are a few things we can say; after some practice, you should develop a feel for when it is a good technique to attempt. One situation in which we might use this method is when the hypothesis seems more complicated than the conclusion, as illustrated in the examples above. Another situation in which the contrapositive method can be useful is when the conclusion is a disjunction, so the implication is of the form $A \implies B \vee C$. Hence the contrapositive of this implication is

$$
(\neg B) \wedge (\neg C) \implies \neg A, \tag{3.7}
$$

where in the hypothesis we have applied De Morgan's Law to replace $\neg(B \vee C)$ by the logically equivalent $(\neg B) \wedge (\neg C)$.

The following example shows how the contrapositive method works in this situation.

Example 12   Prove the following statement: For all $x \in \mathbb{R}$, if $x^2 - 7x + 10 \geq 0$, then $x \leq 3$ or $x \geq 4$.

Proof: Let x be an arbitrary real number. Using the logical equivalence (3.7), we prove the contrapositive:

If $x > 3$ and $x < 4$, then $x^2 - 7x + 10 < 0$,

since the negation of "$x \leq 3$" is "$x > 3$", and the negation of "$x \geq 4$" is "$x < 4$". Hence assume that $x > 3$ and $x < 4$. Now, factoring gives

$$
x^2 - 7x + 10 = (x - 2)(x - 5).
$$

Since $x > 3$, we have $x - 3 > 0$, and therefore

$$
x - 2 = (x - 3) + 1 > 0 + 1 = 1 > 0.
$$

Also, since $x < 4$, we have $x - 4 < 0$, and therefore

$$x - 5 = (x - 4) - 1 < 0 - 1 = -1 < 0.$$

Hence, multiplying the inequalities $x - 2 > 0$ and $x - 5 < 0$, we obtain

$$x^2 - 7x + 10 = (x - 2)(x - 5) < 0.$$

Since the contrapositive is true, the original implication must be true, and hence we have proved the universally quantified implication.   □

Note that the hypothesis in the above contrapositive is a conjunction, so we are able to assume both parts of the hypothesis, and this leads to a simple proof.

Before we move on to other methods of proof, we will demonstrate in another way that there is no precise recipe when it comes to proofs. Even for proving an implication of the form $A \implies B \vee C$, the contrapositive method is not the only logical equivalence that we can use. Consider the following exercise.

EXERCISE

Prove the logical equivalence

$$\Big(A \implies (B \vee C)\Big) \equiv \Big((A \wedge (\neg B)) \implies C\Big). \tag{3.8}$$

REMARK

Logical equivalence (3.8) tells us that another way to prove the implication "$A \implies (B \vee C)$" is to prove the logically equivalent implication "$(A \wedge (\neg B)) \implies C$". This method of proof is sometimes called the method of **elimination**, because in the hypothesis we assume that $\neg B$ is true (i.e., so B is false), thus "eliminating" the possibility of component B in the conclusion of the original implication. In this situation, having eliminated the possibility that B is true, we must prove that component C of the conclusion is true.

For example, we now apply the method of elimination to give a different proof of the implication in Example 12.

Alternative proof for Example 12: Let x be an arbitrary real number. Using the logical equivalence (3.8), we prove the implication:

    If $x^2 - 7x + 10 \geq 0$ and $x > 3$, then $x \geq 4$,

since the negation of "$x \leq 3$" is "$x > 3$". Hence assume that $x^2 - 7x + 10 \geq 0$ and $x > 3$. Factoring gives $x^2 - 7x + 10 = (x - 2)(x - 5)$, so

$$(x - 2)(x - 5) \geq 0.$$

But since $x > 3$, we have $x - 2 > 0$. Dividing the above inequality by $x - 2$ gives $x - 5 \geq 0$. Adding 5 to both sides of this inequality gives $x \geq 5$, so

$$x \geq 5 > 4,$$

and hence we have $x \geq 4$. Since the given implication is true, the original implication must be true, and hence we have proved the universally quantified implication. $\square$

3.6   Proof by Contradiction

We'll now consider another standard method of proof that is often used by mathematicians, and is certainly not direct. It is based on a *contradiction*, which is defined as follows.

Definition 3.6.1
contradiction

Let A be a statement. Note that either A or $\neg A$ must be false, so the compound statement $A \wedge (\neg A)$ is always *false*. The statement "$A \wedge (\neg A)$ is true" is called a **contradiction**.

In other words, any time we come across an argument that claims both A and $\neg A$ are true for some statement A, we say that there must be a contradiction in the argument. This is the basis of *proof by contradiction*. It is difficult to give a precise recipe for how to use proof by contradiction, and we will not give a formal "Proof Method" for it. Instead, we will give a number of examples of its usage so that you can develop a feel for it.

The first example of a proof by contradiction is for a very well known proposition, which involves a simple statement with no quantifiers.

Proposition 10

Prove that $\sqrt{2}$ is irrational.

Proof: Assume, for the sake of contradiction, that $\sqrt{2}$ is rational, that is, we have $\sqrt{2} \in \mathbb{Q}$. Then we can write

$$\sqrt{2} = \frac{a}{b}, \tag{3.9}$$

where a and b are both integers with $b \neq 0$. Since $\sqrt{2}$ is positive, we can assume that a and b are both positive (if a and b are both negative, multiply both of them by -1 to make the numerator and denominator in equation (3.9) both positive; if $a = 0$, then this is a contradiction of the fact that $\sqrt{2}$ is positive).

We can also assume that a and b are *not both even*, for the following reason: If a and b are both even, then there exist positive integers c and d with $a = 2c$ and $b = 2d$. Now since c and d are positive, we have $c < 2c$ and $d < 2d$, so $c < a$ and $d < b$. Also, we have

$$\frac{a}{b} = \frac{2c}{2d} = \frac{c}{d},$$

so we can replace the ratio a/b by c/d in equation (3.9). If c and d are not both even, then we are done (after this replacement of a/b by c/d has been made). Otherwise, c and d are both even, in which case we repeat the above argument and make another replacement of the ratio in (3.9). Since the integer numerator in the ratio is decreased with each replacement, yet is always positive, we can only repeat this replacement process a finite number of times

until we reach the situation in which both numerator and denominator are not both even, which is what we wanted to prove.

Now multiply on both sides of equation (3.9) by b (which is non-zero), and square both sides, to obtain the equation

$$2b^2 = a^2. \tag{3.10}$$

By the definition of divisibility, we thus obtain $2 \mid a^2$, so a^2 is even. Assume, for the sake of contradiction, that a is odd, so there exists an integer k with $a = 2k + 1$. Then we have

$$a^2 = (2k+1)^2 = 4k^2 + 4k + 1 = 2(2k^2 + 2k) + 1,$$

and we hence conclude that a^2 is *odd*, which is a contradiction. As a result, we conclude that a is even.

Therefore there exists an integer m such that $a = 2m$, and substituting this into equation (3.10) and rearranging, we get

$$b^2 = 2m^2.$$

Analyzing this equation in a similar way as we analyzed equation (3.10), we conclude that b is even. Hence a and b are both even, which is a contradiction (of the fact that a and b are not both even).

We conclude that $\sqrt{2}$ is irrational, that is, we have $\sqrt{2} \notin \mathbb{Q}$.   □

We will now turn to other types of statements. To see how a proof by contradiction is used to prove an implication, consider the logical equivalence

$$\neg(A \implies B) \equiv (A \wedge (\neg B)),$$

given in equation (2.4) on page 32 of Chapter 2. To use a proof by contradiction to prove that "$A \implies B$ is true", we assume that "$A \implies B$ is false". Hence, from the logical equivalence above, our assumption becomes "A is true and B is false". If we can demonstrate that this assumption leads to a contradiction, we have then proved the given implication "$A \implies B$". We demonstrate this with two examples.

Example 13   Prove each of the following statements.

1. For all integers a, b and c, if $a \mid (b + c)$ and $a \nmid b$, then $a \nmid c$.

   Solution: Assume, for the sake of contradiction, that there exist integers a, b and c such that $a \mid (b + c)$ and $a \nmid b$ and $a \mid c$ (noting that the negation of "$a \nmid c$" is "$a \mid c$"). Since $a \mid (b + c)$ and $a \mid c$, then by DIC, we have

   $$a \mid [(1)(b + c) + (-1)(c)],$$

   or in other words, $a \mid b$.

   But among our initial assumptions was that $a \nmid b$, and now we have concluded that $a \mid b$. This is a contradiction.

   As a result, the implication "if $a \mid (b + c)$ and $a \nmid b$, then $a \nmid c$" must be true.

2. For all real numbers x and y, if $x + y < 10$, then $x < 5$ or $y < 5$.

   Solution: Assume, for the sake of contradiction, that there exist real numbers x and y such that $x + y < 10$ and $x \geq 5$ and $y \geq 5$ (noting that the negation of "$x < 5$ or $y < 5$" is "$x \geq 5$ and $y \geq 5$").

   Since $x \geq 5$ and $y \geq 5$, adding these inequalities together gives $x + y \geq 10$.

   But among our initial assumptions was that $x + y < 10$, and now we have concluded that $x + y \geq 10$. This is a contradiction.

   As a result, the implication "if $x + y < 10$ then $x < 5$ or $y < 5$" must be true.

As illustrated in this pair of examples, proof by contradiction gives us another method for proving implications when a direct proof seems difficult. When should we use a proof by contradiction, and what contradiction are we looking for? Again, there is no precise recipe, and you will have to practice with it before you can develop a feel for when it is a good technique to use instead of, say, the contrapositive.

Note that there is a strong logical connection between a proof by contrapositive and a proof by contradiction for an implication "$A \implies B$". In a proof by contradiction, we would start by assuming "$A \wedge (\neg B)$ is true", and one possible contradiction to deduce is that "A is false". However, this is similar to proving the contrapositive "$(\neg B) \implies (\neg A)$". To demonstrate this connection, consider the following alternative proof using the contrapositive method for the statement in part 2 of Example 13.

Contrapositive proof of Example 13, part 2: Let x and y be arbitrary real numbers. We prove the contrapositive:

$$\text{If } x \geq 5 \text{ and } y \geq 5, \text{ then } x + y \geq 10.$$

Hence assume that $x \geq 5$ and $y \geq 5$. Adding these inequalities together, we get $x + y \geq 10$.

Since the contrapositive is true, the original implication must be true, and hence we have proved the universally quantified implication.   □

Comparing these proofs, you will find that they have much in common. Both are correct, and which one you give is a matter of personal preference. You will see other types of situations in which proof by contradiction is used later in the course. You will also have a lot of opportunities to practice your own proofs throughout this course, and that will help you to find which methods, including proof by contradiction, seem to work best for you.

3.6.1   Proving Uniqueness

One common type of mathematical statement that we haven't considered yet asserts the existence of *exactly one* element in a given set with certain properties. This can be viewed as a variant of the existential quantifier, which asserts the existence of *at least one* element. We refer to this informally as a *uniqueness* statement. Two possible methods of proof are given below, one of which uses a proof by contradiction.

Proof Method

To prove the statement "There is a **unique** element $x \in S$ such that $P(x)$ is true":

Step 1("Existence"): Prove that there is at least one element $x \in S$ such that $P(x)$ is true (i.e., prove the existentially quantified statement "$\exists \ x \in S, \ P(x)$").

Step 2("Uniqueness"): Do either (a) or (b) below.

(a) Assume that $P(x)$ and $P(y)$ are true for $x, y \in S$, and prove that this assumption leads to the conclusion $x = y$,

(b) Assume that $P(x)$ and $P(y)$ are true for distinct $x, y \in S$ (so $x \neq y$), and prove that this assumption leads to a contradiction.

To give an idea of what is involved, consider the following example, in which we prove a simple uniqueness result twice, once using each of the two possible methods for Step 2.

Example 14

Prove the following statement using the two methods above.

For all odd integers a, there is a unique integer k such that $a = 2k + 1$.

First proof: From the definition of odd, we know that $a = 2k + 1$ for some integer k. Assume that $a = 2k + 1$ and $a = 2m + 1$ for integers k and m. Using these equations, we have $2k + 1 = a = 2m + 1$, which gives

$$2k + 1 = 2m + 1.$$

Now subtract 1 from both sides of this equation, and divide on both sides by 2, which gives $k = m$. This proves the result.   □

Second proof: From the definition of odd, we know that $a = 2k + 1$ for some integer k. Assume that $a = 2k + 1$ and $a = 2m + 1$ for integers k and m with $k \neq m$. Using these equations, we have $2k + 1 = a = 2m + 1$, which gives

$$2k + 1 = 2m + 1.$$

Now subtract 1 from both sides of this equation, and divide on both sides by 2, which gives $k = m$. This is a contradiction, and we have proved the result.

□

Both of the proofs in Example 14 are correct. After some practice with these methods, you will develop a feel for which you prefer, and for which is easier in a given circumstance.

3.7   Proving If and Only If Statements

Some of the most important results that mathematicians prove are if and only if statements, often universally quantified. In equation (2.5) on page 32 in Chapter 2, we have the logical

equivalence
$$(A \iff B) \equiv ((A \implies B) \land (B \implies A)).$$

This is the basis of the usual method of proof for an if and only if statement.

Proof Method

For proving an if and only if statement:

1. To prove the statement "$A \iff B$", it is equivalent to prove both the implication "$A \implies B$" and its converse "$B \implies A$".

2. To prove the universally quantified statement "$\forall\, x \in S,\ P(x) \iff Q(x)$", it is equivalent to do either (a) or (b) below.

    (a) Let x be an arbitrary element of S, and prove both the implication "$P(x) \implies Q(x)$" and its converse "$Q(x) \implies P(x)$",

    (b) Prove both the universally quantified implication "$\forall\, x \in S,\ P(x) \implies Q(x)$" and its universally quantified converse "$\forall\, x \in S,\ Q(x) \implies P(x)$".

The above proof method means that to prove an if and only if statement, we can simply prove two implications. Previously in this chapter, we have had a lot of practice in proving implications, so we can take advantage of that here.

To see how this works, consider the following proof of an if and only if statement about divisibility of integers.

Example 15

Prove the following statement:

For all integers a, b and c, $a \mid b$ and $a \mid c$ if and only if, for all integers x and y, $a \mid (bx + cy)$.

Solution: To prove that this statement is true, we could prove both of the universally quantified implications

(i) For all integers a, b and c, if $a \mid b$ and $a \mid c$, then for all integers x and y, $a \mid (bx + cy)$,

(ii) For all integers a, b and c, if $a \mid (bx + cy)$ for all integers x and y, then $a \mid b$ and $a \mid c$.

Now, we have previously proved both of these statements: we proved statement (i) as Proposition 9 (DIC), and we proved statement (ii) in part 1 of Example 8. Since (i) and (ii) are both true, the universally quantified if and only if statement must be true as well.

Of course, not all if and only if statements are true, so we finish with an example in which we're asked to prove or disprove a pair of if and only if statements.

Example 16   Prove or disprove each of the following statements.

1. For all integers a, a is odd if and only if $3a^2 + 4a - 1$ is even.

   Solution: Let a be an arbitrary integer. To prove this statement, it is equivalent to prove both of the implications

   (i) If a is odd, then $3a^2 + 4a - 1$ is even,

   (ii) If $3a^2 + 4a - 1$ is even, then a is odd.

   Now, we have previously proved implication (i) as Proposition 6 on page 46.

   For implication (ii), we prove the contrapositive

   (iii) If a is even, then $3a^2 + 4a - 1$ is odd.

   To prove (iii), assume that a is even, which means that $a = 2k$ for some integer k. Substituting $a = 2k$ into the integer $3a^2 + 4a - 1$, we obtain

   $$3a^2 + 4a - 1 = 3(2k)^2 + 4(2k) - 1$$
   $$= 3(4k^2) + 4(2k) - 1$$
   $$= 12k^2 + 8k - 1$$
   $$= 2(6k^2 + 4k - 1) + 1.$$

   Hence, by the definition, $3a^2 + 4a - 1$ is odd.

   Since implication (iii) is true, then implication (ii) is true. Then since (i) and (ii) are both true, the if and only if statement must be true as well.

2. For all $x \in \mathbb{R}$, $x^2 - 7x + 10 \geq 0$ if and only if $x \leq 3$ or $x \geq 4$.

   Solution: Let x be an arbitrary real number. To prove this statement, it is equivalent prove both of the implications

   (i) If $x^2 - 7x + 10 \geq 0$ then $x \leq 3$ or $x \geq 4$,

   ii) If $x \leq 3$ or $x \geq 4$ then $x^2 - 7x + 10 \geq 0$.

   Now, we have previously proved implication (i) in Example 12.

   Implication (ii) is false, and we disprove it by finding a counter-example. Suppose we let $x = 2.5$. Then the hypothesis is true, since $2.5 \leq 3$. Also, we have

   $$x^2 - 7x + 10 = (2.5)^2 - 7(2.5) + 10 = 6.25 - 17.5 + 10 = -1.25,$$

   and -1.25 is certainly not greater than or equal to 0. Therefore the conclusion is false for this choice of x, and we have found a counter-example.

   Since implication (ii) is false, the if and only if statement is false.

Note that proving an if and only if statement is equivalent to proving both an implication and its converse, but such proofs do not need to be direct. For example, in the proof for

part 1 of Example 16 above, the proof of the implication was direct, but the proof of the converse used the contrapositive.

Chapter 4

Mathematical Induction

4.1 Notation for Summations, Products and Recurrences

We begin this chapter by introducing some notation that you may not have seen before. Suppose we had ten squares with sides of lengths 1 to 10. Then their total area is given by the somewhat lengthy sum

$$1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2 + 7^2 + 8^2 + 9^2 + 10^2.$$

Mathematicians often write sums like this in a compact form by using *summation* notation.

Definition 4.1.1

summation notation

For integers n and m with $n \geq m$, the notation $\displaystyle\sum_{i=m}^{n} x_i$, also written as $\sum_{i=m}^{n} x_i$, is called **summation notation**. It represents the sum

$$x_m + x_{m+1} + \cdots + x_{n-1} + x_n.$$

The summation symbol $\sum$ is the upper case Greek letter *sigma*. The letter i is the **index of summation**; the letter m is the **lower bound of summation**, and the letter n is the **upper bound of summation**. The notation means that the index i begins with an initial value of m and increments by 1 stopping when $i = n$. The index of summation is a *dummy* variable and any letter could be used in its place.

Example 1

$$\sum_{i=1}^{10} i^2 = 1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2 + 7^2 + 8^2 + 9^2 + 10^2,$$

$$\sum_{k=0}^{3} \sin(k\pi) = \sin(0) + \sin(\pi) + \sin(2\pi) + \sin(3\pi),$$

$$\sum_{\ell=1}^{n} \frac{1}{\ell^2} = 1 + \frac{1}{4} + \frac{1}{9} + \cdots + \frac{1}{(n-1)^2} + \frac{1}{n^2}.$$

There are a number of rules that help us manipulate summations with a finite number of terms.

Proposition 1   **(Properties of Summation (PS))**

1. Multiplying by a constant

$$\sum_{i=m}^{n} cx_i = c \sum_{i=m}^{n} x_i, \quad \text{where } c \text{ is a constant}$$

2. Adding two sums and subtracting two sums

$$\sum_{i=m}^{n} x_i + \sum_{i=m}^{n} y_i = \sum_{i=m}^{n} (x_i + y_i)$$

$$\sum_{i=m}^{n} x_i - \sum_{i=m}^{n} y_i = \sum_{i=m}^{n} (x_i - y_i)$$

3. Changing the bounds of the index of summation

$$\sum_{i=m}^{n} x_i = \sum_{i=m+k}^{n+k} x_{i-k}$$

The first two properties tell us that summation is *linear*. They require indices with the same upper and lower bounds. The last property allows us to change the bounds of the index of summation, which is often useful when combining summation expressions.

Similarly, mathematicians often write products in a compact form using *product* notation.

Definition 4.1.2

product notation

For integers n and m with $n \geq m$, the notation $\prod_{i=m}^{n} x_i$, also written as $\prod_{i=m}^{n} x_i$, is called **product notation**. It represents the product

$$x_m x_{m+1} \cdots x_{n-1} x_n.$$

The product symbol $\prod$ is the upper case Greek letter *pi*. The index i and the upper and lower bounds m and n behave just as they do for sums.

Example 2

$$\prod_{i=2}^{n} \left(1 - \frac{1}{i^2}\right) = \left(1 - \frac{1}{4}\right)\left(1 - \frac{1}{9}\right)\left(1 - \frac{1}{16}\right)\cdots\left(1 - \frac{1}{(n-1)^2}\right)\left(1 - \frac{1}{n^2}\right)$$

Mathematicians also use terminology and notation to define sequences of objects iteratively or recursively.

Definition 4.1.3
recurrence relation

A **recurrence relation** defines a sequence of values by giving one or more initial terms, together with an equation expressing each subsequent term in terms of earlier ones.

Example 3

Let $s_n = \sum_{i=1}^{n} i$, for $n \geq 1$, so the first four terms are $s_1 = 1$, $s_2 = 3$, $s_3 = 6$ and $s_4 = 10$. Then we can also define these sums by the following recurrence relation:

The initial term is defined as $s_1 = 1$, and the subsequent terms are given by the equation $s_n = s_{n-1} + n$, for $n \geq 2$.

The famous Fibonacci sequence $f_1, f_2, f_3, \ldots$ is defined by a particularly simple recurrence relation.

Example 4

The initial two terms of the *Fibonacci sequence* are defined as $f_1 = 1$ and $f_2 = 1$. All subsequent terms are defined by the equation $f_n = f_{n-1} + f_{n-2}$, for $n \geq 3$. Thus the first eight terms of the sequence are $1, 1, 2, 3, 5, 8, 13, 21$.

4.2   Proof by Induction

Definition 4.2.1
axiom

An **axiom** of a mathematical system is a statement that is assumed to be true. No proof is given. From axioms we derive propositions and theorems.

Axioms are sometimes described as *self-evident*, though many are not, and are defining properties of mathematical systems. The *Principle of Mathematical Induction* is one such axiom, taken as a defining property of $\mathbb{N}$, the set of natural numbers.

Axiom 1

Principle of Mathematical Induction (POMI)

Let $P(n)$ be a statement that depends on $n \in \mathbb{N}$.

If statements 1 and 2 are both true:

1. $P(1)$

2. For all $k \in \mathbb{N}$, if $P(k)$, then $P(k+1)$.

then statement 3 is true:

3. For all $n \in \mathbb{N}$, $P(n)$.

This immediately gives rise to the following method for proving a universally quantified statement which is in the form of statement 3 in Axiom 1 above.

Proof Method   To prove the universally quantified statement "For all $n \in \mathbb{N}$, $P(n)$":

1. Prove "$P(1)$".

2. Prove the universally quantified implication "For all $k \in \mathbb{N}$, if $P(k)$, then $P(k+1)$."

A proof using this method is known as a proof by *mathematical induction*, or, more simply, by *induction*. The statement "$P(1)$" in part 1 of the proof method is called the *base case*. The implication "$P(k) \implies P(k+1)$" in part 2 of the proof method is called the *inductive step*, the hypothesis "$P(k)$" of this implication is called the *inductive hypothesis*, and the conclusion "$P(k+1)$" is called the *inductive conclusion*.

We will give many proofs by induction in these notes, and present our proofs in a standard format, using the language described above.

Format: To prove "For all $n \in \mathbb{N}$, $P(n)$." by induction on n.

Begin by stating that the proof is by induction on n, and identifying the statement $P(n)$.

Base Case Prove $P(1)$, which we may also write as "Verify $P(1)$".

Inductive Step Let k be an arbitrary natural number.

Assume $P(k)$, the inductive hypothesis. We usually explicitly write out the statement $P(k)$.

Next we prove $P(k+1)$, the inductive conclusion, using the assumption $P(k)$. We often explicitly write out the statement $P(k+1)$ before giving its proof.

Following the proof of $P(k+1)$, we usually end the proof by stating that the result is true for $n = k+1$, and hence holds for all $n \geq 1$ by the Principle of Mathematical Induction.

In our first example of induction using the standard format, we prove a closed formula for the sum of squares of the first n positive integers.

Proposition 2   For every integer $n \in \mathbb{N}$,
$$\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}.$$

Proof: The proof is by induction on n, where $P(n)$ is the statement
$$\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}.$$

Base Case The statement $P(1)$ is given by

$$\sum_{i=1}^{1} i^2 = \frac{1(1+1)(2 \times 1 + 1)}{6}.$$

The expression on the left hand side of this equation evaluates to

$$\sum_{i=1}^{1} i^2 = 1^2 = 1,$$

and the expression on the right hand side evaluates to

$$\frac{1(1+1)(2 \times 1 + 1)}{6} = 1.$$

Since both sides are equal to each other, $P(1)$ is true.

Inductive Step Let k be an arbitrary natural number.

Assume the inductive hypothesis, $P(k)$. That is, we assume that

$$\sum_{i=1}^{k} i^2 = \frac{k(k+1)(2k+1)}{6}.$$

We wish to prove $P(k+1)$, which is the statement

$$\sum_{i=1}^{k+1} i^2 = \frac{(k+1)((k+1)+1)(2(k+1)+1)}{6}.$$

Then, starting with the summation on the left hand side of $P(k+1)$, we obtain

$$
\begin{aligned}
\sum_{i=1}^{k+1} i^2 &= \left(\sum_{i=1}^{k} i^2 \right) + \left((k+1)^2 \right), && \text{by properties of summation notation} \\
&= \left(\frac{k(k+1)(2k+1)}{6} \right) + \left((k+1)^2 \right), && \text{by the inductive hypothesis} \\
&= \frac{k(k+1)(2k+1) + 6(k+1)^2}{6} \\
&= \frac{(k+1)\left(2k^2 + 7k + 6\right)}{6}, && \text{factoring out } k+1 \text{ and rearranging} \\
&= \frac{(k+1)(k+2)(2k+3)}{6}, && \text{by factoring} \\
&= \frac{(k+1)((k+1)+1)(2(k+1)+1)}{6}, && \text{by rearranging.}
\end{aligned}
$$

The result is true for $n = k + 1$, and hence holds for all $n \geq 1$ by the Principle of Mathematical Induction. $\qquad\square$

REMARK

Informally, the reason that a proof by induction works is that for any natural number n, we are proving $P(n)$ by the following chain of results: We establish that $P(1)$ is true as the base case. Then, having established that the implication $P(k) \implies P(k+1)$ is true for an arbitrary positive integer k, we know that $P(2)$ must also be true (since $P(1)$ is true). Similarly, this then tells us that $P(3)$ is true (since $P(2)$ is true). Continuing in this way, we can establish that $P(n)$ is true for all positive integers n. Visually, we might write this as the chain of implications

$$P(1) \implies P(2) \implies P(3) \implies P(4) \implies \cdots \implies P(n-1) \implies P(n). \qquad (4.1)$$

Now suppose that we wish to prove a universally quantified statement of the form "For all integers $n \geq b$, $P(n)$," where b is a fixed integer but $b \neq 1$. Since $b \neq 1$, we cannot apply proof by induction as it is stated above. However, note that when considering only integers n with $n \geq b$, we could prove $P(n)$ by the chain of implications

$$P(b) \implies P(b+1) \implies P(b+2) \implies \cdots \implies P(n-1) \implies P(n),$$

using the informal language of the remark above. From this point of view, we see that proof by induction can be changed very slightly to give a proof method for statements of the form

For every integer $n \geq b$, $P(n)$ is true.

What changes are needed in terms of the standard format?

- For the base case, prove $P(b)$.

- For the inductive step we assume that k is an arbitrary integer where $k \geq b$. We then proceed the same as before by assuming the inductive hypothesis, $P(k)$, and then proving that the inductive conclusion, $P(k+1)$, is true.

We often refer to this modified method of proof as *induction with base case b*, or we might also say *induction with starting point b*, since the chain of implications starts with $P(b)$.

As our first example of this, we give a proof by induction with starting point 3.

Proposition 3   For every integer $n \geq 3$, $n^2 > 2n + 1$.

Proof: We prove this result by induction on n, where $P(n)$ is the statement $n^2 > 2n + 1$.

Base Case The statement $P(3)$ is given by $3^2 > 2(3) + 1$. Now $3^2 = 9$, $2(3) + 1 = 7$, and $9 > 7$, which proves $P(3)$.

Inductive Step Let k be an arbitrary integer where $k \geq 3$.

Assume the inductive hypothesis, $P(k)$. That is, we assume that $k^2 > 2k + 1$.

We wish to prove $P(k+1)$, which is the statement

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

Starting with the expression on the left hand side of $P(k+1)$, we obtain

$$(k+1)^2 = k^2 + 2k + 1 > (2k+1) + (2k+1) \geq 2k + 1 + 7 > 2k + 3 = 2(k+1) + 1,$$

where the first ">" follows by the inductive hypothesis and the "$\geq$" uses the fact that $2k + 1 \geq 7$, since $k \geq 3$.

The result is true for $n = k + 1$, and hence holds for all $n \geq 3$ by the Principle of Mathematical Induction.   □

Next we prove the statement that appeared as Example 2 at the beginning of Chapter 1. Our proof is by induction with starting point 5.

Proposition 4   For every integer $n \geq 5$, $2^n > n^2$.

Proof: We prove this result by induction on n, where $P(n)$ is the statement $2^n > n^2$.

Base Case The statement $P(5)$ is given by $2^5 > 5^2$. Now $2^5 = 32$, $5^2 = 25$, and $32 > 25$, which proves $P(5)$.

Inductive Step Let k be an arbitrary integer where $k \geq 5$.

Assume the inductive hypothesis, $P(k)$. That is, we assume $2^k > k^2$.

We wish to prove $P(k+1)$, which is the statement

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

Starting with the expression on the left hand side of $P(k+1)$, we obtain

$$2^{k+1} = 2 \times 2^k > 2 \times k^2 = k^2 + k^2 > k^2 + 2k + 1 = (k+1)^2,$$

where the first inequality follows by the inductive hypothesis and the second inequality uses Proposition 3, which gives $k^2 > 2k + 1$ for $k \geq 3$, and hence for $k \geq 5$.

The result is true for $n = k + 1$, and hence holds for all $n \geq 5$ by the Principle of Mathematical Induction.   □

For our final example in this section, we give a proof by induction with the standard starting point of 1, but for a more complicated type of problem involving two dimensional objects. A **triomino** is a tile consisting of three unit squares, of the form . A covering of a squared grid means that every square is covered exactly once.

Proposition 5   For all $n \in \mathbb{N}$, all $2^n \times 2^n$ grid of squares with one square removed can be covered by triominoes.

Proof: We prove this result by induction on n, where $P(n)$ is the statement

All $2^n \times 2^n$ grids of squares with one square removed can be covered by triominoes.

Base Case The statement $P(1)$ is given by

All 2×2 grids of squares with one square removed can be covered by triominoes.

All 2×2 grids with one square removed are given by ▨ or ▨ or ▨ or ▨.

They can all be covered by a single triomino (rotated by some multiple of $\frac{\pi}{2}$), proving the statement $P(1)$.

Inductive Step Let k be an arbitrary natural number.

Assume the inductive hypothesis, $P(k)$. That is, we assume that all $2^k \times 2^k$ grids of squares with one square removed can be covered by triominoes. Note that this inductive hypothesis includes all possible positions for the empty square within the grid.

We wish to prove $P(k+1)$, which is the statement

All $2^{k+1} \times 2^{k+1}$ grids of squares with one square removed can be covered by triominoes.

Consider an arbitrary $2^{k+1} \times 2^{k+1}$ grid with one square removed (where the missing square is drawn in black):
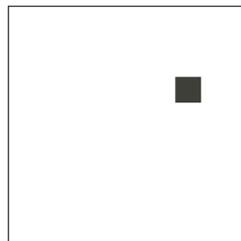
Split the $2^{k+1} \times 2^{k+1}$ grid in half vertically and horizontally by dotted lines, forming four $2^k \times 2^k$ subgrids:
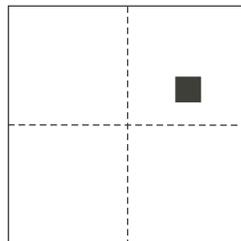
The missing square occurs in one of these four subgrids. Now, to start covering the grid by triominoes, we'll place one tile around the centre of the grid, covering a corner square in each of the three $2^k \times 2^k$ subgrids that do not contain the missing square:
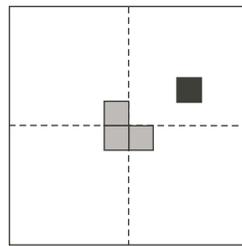
We can now view the grid as being made up of four $2^k \times 2^k$ subgrids, each with one square missing (that is, one square that cannot be covered by the additional triominoes). The Inductive Hypothesis tells us that each of these four grids can be covered by triominoes. Together with the initial triomino in the centre, this means that the $2^{k+1} \times 2^{k+1}$ grid (with the missing black square) can be covered. This proves that the result is true for $n = k + 1$, and hence holds for all $n \geq 1$ by the Principle of Mathematical Induction.   □

4.3   The Binomial Theorem

In this section we use induction to prove the binomial theorem, one of the most important results in mathematics. First, we need to introduce some notation and terminology that will be useful in stating the theorem.

Definition 4.3.1
factorial

For a positive integer m, we define
$$m! = \prod_{i=1}^{m} i,$$
called m **factorial**. We also define $m!$ for $m = 0$, to have the value $0! = 1$.

Example 5

$$1! = 1,$$
$$2! = 1 \cdot 2 = 2,$$
$$3! = 1 \cdot 2 \cdot 3 = 6,$$
$$4! = 1 \cdot 2 \cdot 3 \cdot 4 = 24.$$

Definition 4.3.2
binomial coefficient

For a non-negative integer n and non-negative integer $m \leq n$, we define
$$\binom{n}{m} = \frac{n!}{(n-m)!m!},$$
called n *choose* m, or (without needing to mention n or m) a **binomial coefficient**. If $m > n$, we define $\binom{n}{m} = 0$.

Example 6

$$\binom{n}{0} = \frac{n!}{n!0!} = 1,$$

$$\binom{n}{1} = \frac{n!}{(n-1)!1!} = n,$$

$$\binom{n}{2} = \frac{n!}{(n-2)!2!} = \frac{n(n-1)}{2},$$

$$\binom{7}{3} = \frac{7!}{4!3!} = \frac{7 \cdot 6 \cdot 5}{3!} = 35.$$

The following result is a recurrence for binomial coefficients that is needed for our inductive proof of the binomial theorem.

Proposition 6

(Pascal's Identity (PI))

For all positive integers n and m with $m < n$, we have

$$\binom{n}{m} = \binom{n-1}{m-1} + \binom{n-1}{m}.$$

Proof: Using the definition of binomial coefficient, we obtain

$$
\begin{aligned}
\binom{n-1}{m-1} + \binom{n-1}{m} &= \frac{(n-1)!}{(n-m)!(m-1)!} + \frac{(n-1)!}{(n-m-1)!m!} \\
&= \frac{(n-1)!m + (n-1)!(n-m)}{(n-m)!m!} \\
&= \frac{(n-1)!(m+n-m)}{(n-m)!m!} \\
&= \frac{(n-1)!n}{(n-m)!m!} \\
&= \frac{n!}{(n-m)!m!} \\
&= \binom{n}{m},
\end{aligned}
$$

where we have used the definition of binomial coefficient for the first and sixth equalities. □

Proposition 6 is a recurrence that allows us to calculate the value of the binomial coefficient $\binom{n}{m}$ for all non-negative integers n and m. The initial values are $\binom{n}{0} = 1$ for all non-negative integers n (see Example 6), and $\binom{0}{m} = 0$ for all positive integers m (since $\binom{n}{m} = 0$ for $m > n$). For example, the following table gives the values of $\binom{n}{m}$ for all n and m between 0 and 5.

	$m=0$	$m=1$	$m=2$	$m=3$	$m=4$	$m=5$
$n=0$	1	0	0	0	0	0
$n=1$	1	1	0	0	0	0
$n=2$	1	2	1	0	0	0
$n=3$	1	3	3	1	0	0
$n=4$	1	4	6	4	1	0
$n=5$	1	5	10	10	5	1

Note that all values in this table are non-negative integers; a fact that is not at all obvious from the definition of the binomial coefficient above. The portion of the table consisting of positive integer values is generally known as *Pascal's triangle*.

In fact, $\binom{n}{m}$ is a non-negative integer for all non-negative integers n and m, a fact that you are asked to prove by induction on n in the following Exercise. **Hint:** Here the statement $P(n)$, for $n \geq 0$, is

$\binom{n}{m}$ is a non-negative integer for all non-negative integers m.

EXERCISE

Use induction on n to prove that

For all non-negative integers n and m, $\dbinom{n}{m}$ is a non-negative integer.

Now we give the binomial theorem, proved by induction on the non-negative integer n.

Theorem 7   **(Binomial Theorem, Version 1 (BT1))**

For all integers $n \geq 0$ and all real numbers x,

$$(1+x)^n = \sum_{m=0}^{n} \binom{n}{m} x^m.$$

Proof: The proof is by induction on n, where $P(n)$ is the statement

$$\forall\, x \in \mathbb{R}, \quad (1+x)^n = \sum_{m=0}^{n} \binom{n}{m} x^m.$$

Base Case The statement $P(0)$ is given by

$$\forall\, x \in \mathbb{R}, \quad (1+x)^0 = \sum_{m=0}^{0} \binom{0}{m} x^m.$$

For the left hand side, we obtain $(1+x)^0 = 1$ for all real numbers x, and for the right hand side, we obtain $\binom{0}{0} x^0 = 1$ for all real numbers x. (Note: In this context, 0^0 is defined to be 1.) Since both sides are equal for all real numbers x, we have proved $P(0)$.

Inductive Step Let k be an arbitrary integer where $k \geq 0$.

Assume the inductive hypothesis, $P(k)$. That is, we assume

$$\forall\, x \in \mathbb{R}, \quad (1+x)^k = \sum_{m=0}^{k} \binom{k}{m} x^m.$$

We wish to prove $P(k+1)$, which is the statement

$$\forall\, x \in \mathbb{R}, \quad (1+x)^{k+1} = \sum_{m=0}^{k+1} \binom{k+1}{m} x^m.$$

We begin with the left hand side, and use the induction hypothesis to obtain

$$(1+x)^{k+1} = (1+x) \cdot (1+x)^k = (1+x) \sum_{m=0}^{k} \binom{k}{m} x^m.$$

We continue by expanding out the multiplication by $(1+x)$, and then rearranging the resulting summations, giving

$$(1+x)^{k+1} = \sum_{m=0}^{k} \binom{k}{m} x^m + \sum_{m=0}^{k} \binom{k}{m} x^{m+1}$$

$$= \sum_{m=0}^{k} \binom{k}{m} x^m + \sum_{i=1}^{k+1} \binom{k}{i-1} x^i$$

$$= \binom{k}{0} x^0 + \sum_{m=1}^{k} \left(\binom{k}{m} + \binom{k}{m-1} \right) x^m + \binom{k}{k} x^{k+1}.$$

Note how we have changed the bounds of summation, or changed the names of (dummy) summation variables, in some of the steps above. Finally, we use Pascal's Identity, and the facts that

$$\binom{k}{0} = \binom{k+1}{0} = \binom{k}{k} = \binom{k+1}{k+1} = 1,$$

for all integers $k \geq 0$, to obtain

$$(1+x)^{k+1} = x^0 + \sum_{m=1}^{k} \binom{k+1}{m} x^m + x^{k+1}$$

$$= \binom{k+1}{0} x^0 + \sum_{m=1}^{k} \binom{k+1}{m} x^m + \binom{k+1}{k+1} x^{k+1}$$

$$= \sum_{m=0}^{k+1} \binom{k+1}{m} x^m.$$

The result is true for $n = k + 1$, and hence holds for all $n \geq 0$ by the Principle of Mathematical Induction. $\square$

Example 7   When $n = 5$, using the values of $\binom{5}{m}$ calculated in the table above, the binomial theorem gives

$$(1 + x)^5 = 1 + 5x + 10x^2 + 10x^3 + 5x^4 + x^5.$$

We end the section by giving another version of the binomial theorem that is often convenient to use.

Corollary 8   **(Binomial Theorem, Version 2 (BT2))**

For all integers $n \geq 0$ and for all real numbers a and b,

$$(a + b)^n = \sum_{m=0}^{n} \binom{n}{m} a^{n-m} b^m.$$

Proof: First we consider the case $a = 0$. Then, substituting $a = 0$ in the right hand side of the result that we need to prove, we obtain

$$\sum_{m=0}^{n} \binom{n}{m} 0^{n-m} b^m = \binom{n}{n} b^n = b^n,$$

since $0^{n-m} = 0$ for all values of $m < n$, and $\binom{n}{n} = 1$ for all non-negative integers n. But substituting $a = 0$ in the left hand side of the result we need to prove, we also obtain $(0 + b)^n = b^n$, so the result is true when $a = 0$.

Otherwise, we have the case $a \neq 0$. Then we have

$$(a + b)^n = \left(a \cdot \left(1 + \frac{b}{a} \right) \right)^n = a^n \left(1 + \frac{b}{a} \right)^n.$$

Now, since $\frac{b}{a}$ is a real number when a and b is real and a is non-zero, we can use the Binomial Theorem, Version 1 with $x = \frac{b}{a}$, to obtain

$$(a + b)^n = a^n \sum_{m=0}^{n} \binom{n}{m} \left(\frac{b}{a} \right)^m$$

$$= a^n \sum_{m=0}^{n} \binom{n}{m} a^{-m} b^m$$

$$= \sum_{m=0}^{n} \binom{n}{m} a^{n-m} b^m,$$

so the result is true when $a \neq 0$.

Hence the result is true both for $a = 0$ and $a \neq 0$, so it is true for all a (and all real b and all non-negative integers n). $\square$

4.4   Proof by Strong Induction

Sometimes assuming $P(k)$ isn't enough to prove $P(k+1)$, and we need to assume some or all of the statements $P(k-1)$, $P(k-2)$, $\ldots$ $P(1)$ as well. In this situation we can't use POMI, but we can use a variation known as the *Principle of Strong Induction*.

Axiom 2

Principle of Strong Induction (POSI)

Let $P(n)$ be a statement that depends on $n \in \mathbb{N}$.

If statements 1 and 2 are both true:

1. $P(1)$

2. For all $k \in \mathbb{N}$, if $P(1) \wedge P(2) \wedge \cdots \wedge P(k)$, then $P(k+1)$.

then statement 3 is true:

3. For all $n \in \mathbb{N}$, $P(n)$.

The Principle of Mathematical Induction gave rise to "proof by induction", which we later adapted to allow for a different starting point. In a similar fashion the Principle of Strong Induction gives rise to a proof method, and has a variation that allows for a different starting point. However, as an additional variation, we can also adapt POSI by allowing for more than one base case. Both of these variations are incorporated in the following method of proof, that we refer to as *strong induction*.

Proof Method

To prove the universally quantified statement "For all integers $n \geq b$, $P(n)$":

1. Prove "$P(b) \wedge P(b+1) \wedge \cdots \wedge P(B)$", for some integer $B \geq b$.

2. Prove the universally quantified implication "For all integers $k \geq B$, if $P(b) \wedge P(b+1) \wedge \cdots \wedge P(k)$, then $P(k+1)$."

To implement the proof method of strong induction, what changes are needed in terms of the standard format?

- For the base case, prove all of $P(b), P(b+1), \ldots, P(B)$. When $b < B$ there is more than one case to prove, so we label them as **Base cases**, and refer to b as the smallest base case and B as the largest base case.

- For the inductive step, we assume that k is an arbitrary integer where $k \geq B$. We assume the inductive hypothesis, $P(b) \wedge P(b+1) \wedge \ldots \wedge P(k)$. That is, we assume $P(i)$, for all integers $i = b, b+1, \ldots, k$. We then prove $P(k+1)$ using the assumptions $P(i)$ for all integers $i = b, b+1, \ldots, k$.

For our first example of a proof by strong induction, we consider a recurrence relation with two initial terms, and in which each subsequent term is expressed in terms of the two previous terms. The result gives an explicit formula for the terms in the sequence, which we usually refer to as the "solution" to the recurrence relation.

Proposition 9

Let the sequence $x_1, x_2, \ldots$ be defined by the recurrence relation $x_1 = 0$, $x_2 = 30$, and $x_m = x_{m-1} + 6x_{m-2}$ for $m \geq 3$. For all integers $n \geq 1$,

$$x_n = 2 \cdot 3^n + 3 \cdot (-2)^n.$$

Proof: We prove this result by strong induction on n, where $P(n)$ is the statement

$$x_n = 2 \cdot 3^n + 3 \cdot (-2)^n.$$

Base Cases The statement $P(1)$ is given by

$$x_1 = 2 \cdot 3^1 + 3 \cdot (-2)^1 = 6 - 6 = 0.$$

But from the recurrence relation we are given $x_1 = 0$, and thus we have proved $P(1)$.

The statement $P(2)$ is given by

$$x_2 = 2 \cdot 3^2 + 3 \cdot (-2)^2 = 18 + 12 = 30.$$

But from the recurrence relation we are given $x_2 = 30$, thus we have proved $P(2)$.

Inductive Step Let k be an arbitrary integer where $k \geq 2$.

Assume the inductive hypothesis, $P(1) \wedge P(2) \wedge \ldots \wedge P(k)$. That is, we assume $x_i = 2 \cdot 3^i + 3 \cdot (-2)^i$, for all integers $i = 1, 2, \ldots, k$.

We wish to prove $P(k+1)$, which is the statement

$$x_{k+1} = 2 \cdot 3^{k+1} + 3 \cdot (-2)^{k+1}.$$

Starting with the left-hand side of $P(k+1)$, and noting that $k + 1 \geq 3$ since $k \geq 2$, we obtain

$$
\begin{aligned}
x_{k+1} &= x_k + 6x_{k-1}, && \text{by the recurrence relation} \\
&= \left(2 \cdot 3^k + 3(-2)^k\right) + 6\left(2 \cdot 3^{k-1} + 3(-2)^{k-1}\right), && \text{by the inductive hypothesis} \\
&= \left(2 \cdot 3 + 6 \cdot 2\right)3^{k-1} + \left(3(-2) + 6 \cdot 3\right)(-2)^{k-1}, && \text{expanding and factoring} \\
&= 18 \cdot 3^{k-1} + 12(-2)^{k-1} \\
&= 2 \cdot 3^2 \, 3^{k-1} + 3 \cdot (-2)^2 (-2)^{k-1} \\
&= 2 \cdot 3^{k+1} + 3 \cdot (-2)^{k+1},
\end{aligned}
$$

where, for the second equality, we have used the inductive hypothesis with $i = k$ and $i = k - 1$, which is valid since $k - 1 \geq 1$ when $k \geq 2$.

The result is true for $n = k + 1$, and so holds for all $n \geq 1$ by the Principle of Strong Induction.   □

Note that in the above proof there are two base cases, $n = 1$ and $n = 2$. The reason there are two base cases is that the smallest value of n for which the equation

$$x_n = x_{n-1} + 6x_{n-2}$$

can be used to prove the inductive hypothesis is $n = 3$. But to use this equation to determine the value of x_3, we need to already know the values of $x_{3-1} = x_2$, and $x_{3-2} = x_1$. Hence to start the induction, as base cases, we need to know the values of x_1 and x_2 (and thus to have proved the formula giving the values of x_n when $n = 1$ and $n = 2$).

For our second example of strong induction, the base cases are $n = 6$, $n = 7$ and $n = 8$, so in the notation used above, we have $b = 6$ and $B = 8$.

Proposition 10   For all integers $n \geq 6$, there exist non-negative integers x and y so that $3x + 4y = n$.

Proof: We prove this result by strong induction on n, where $P(n)$ is the statement

There exist non-negative integers x and y so that $3x + 4y = n$.

Base Cases Observe that

$$3(2) + 4(0) = 6, \qquad 3(1) + 4(1) = 7, \qquad 3(0) + 4(2) = 8,$$

so we have proved $P(6)$, $P(7)$ and $P(8)$.

Inductive Step Let k be an arbitrary integer where $k \geq 8$.

Assume the inductive hypothesis, $P(6) \wedge P(7) \wedge \ldots \wedge P(k)$. That is, we assume there exist non-negative integers x and y so that $3x + 4y = i$, for all integers $i = 6, 7, \ldots, k$.

We wish to prove $P(k + 1)$, which is the statement

There exist non-negative integers x and y so that $3x + 4y = k + 1$.

To prove $P(k + 1)$, note that

$$k + 1 = 3 + (k - 2). \tag{4.2}$$

Since $k \geq 8$, we have $k - 2 \geq 6$, and of course $k - 2 < k$, so from the inductive hypothesis there exist non-negative integers x_0 and y_0 such that $3x_0 + 4y_0 = k - 2$. Substituting this into equation (4.2), we obtain

$$k + 1 = 3 + \big(3x_0 + 4y_0\big) = 3 + 3x_0 + 4y_0 = 3x_1 + 4y_1,$$

where $x_1 = x_0 + 1 \geq 1$ and $y_1 = y_0 \geq 0$ are non-negative integers.

The result is true for $n = k + 1$, and hence holds for all $n \geq 6$ by the Principle of Strong Induction.   □

REMARK

Now is a good time to comment on the way in which proofs are actually created. Hopefully you understand the above proof of Proposition 10 after you have read it and worked through the details. However, you might be puzzled about the process of creating it - how would someone think of doing it this way?

In your rough work, even before thinking about whether you would try to use induction or not to prove the result, you would surely check the first few cases and hence quickly write down something like

$$3(2) + 4(0) = 6, \;\; 3(1) + 4(1) = 7, \;\; 3(0) + 4(2) = 8, \;\; 3(3) + 4(0) = 9, \;\; 3(2) + 4(1) = 10,$$

or even a few more cases. You might even write a computer program to check the result for many more cases than these.

Then, since the result is of the form "For all integers $n \geq b$, $P(n)$", with $b = 6$, you might think about trying a proof by induction, and hence you would wish to prove the result for $n = k + 1$. This would lead you to writing $k + 1 = 3 + (k - 2)$, which means that you could prove the result for $n = k + 1$ if you could prove it for $n = k - 2$. This requires strong induction, and you then only need to figure out how many of the first few cases (that you checked above) are needed as base cases to get the strong induction started. Finally, once you've worked out all the details in your rough work, you would write a final version like the one above, using the standard format for a strong induction proof.

EXERCISE

Give an alternate proof of Proposition 10 by strong induction, using the fact that $k + 1 = 4 + (k - 3)$ to prove the inductive conclusion. Be especially careful that you have enough base cases (**Hint:** You will need more than three base cases for this proof.)

For our final example of a proof by strong induction, the proof of the inductive conclusion uses a case analysis.

Proposition 11   Every positive integer n can be written as a sum of distinct non-negative integer powers of 2.

Proof: We prove this result by strong induction on n, where $P(n)$ is the statement

> The positive integer n can be written as a sum of distinct non-negative integer powers of 2.

Base Case When $n = 1$, we have $1 = 2^0$, which proves $P(1)$.

Inductive Step Let k be an arbitrary natural number.

> Assume the inductive hypothesis, $P(1) \wedge P(2) \wedge \ldots \wedge P(k)$. That is, we assume the positive integer i can be written as a sum of distinct non-negative integer powers of 2, for all integers $i = 1, 2, \ldots, k$.

> We wish to prove $P(k + 1)$, which is the statement

The positive integer $k + 1$ can be written as a sum of distinct non-negative integer powers of 2.

The integer $k+1$ is either odd or even, and we prove the inductive conclusion separately for these two cases..

- **Case 1:** Suppose $k + 1$ is odd. By the inductive hypothesis, k is the sum of distinct non-negative integer powers of 2. Now k is even, so this sum cannot include 2^0, since 2^0 is the only non-negative integer power of 2 that is odd. Thus by adding 2^0 to this sum, we obtain $k + 1$ as a sum of distinct non-negative integer powers of 2, proving $P(k + 1)$ in this case.

- **Case 2:** Suppose $k + 1$ is even. Then $(k + 1)/2$ is a positive integer less than $k+1$, so by the inductive hypothesis, $(k+1)/2$ is the sum of distinct non-negative integer powers of 2. If we multiply each term in the sum by 2, then the powers are all increased by 1, so they remain distinct (and positive, which is also a special type of non-negative). Now, the sum of these distinct positive integer powers of 2 is twice the original sum, or $2 \times (k+1)/2 = k+1$, which proves $P(k+1)$ in this case also.

The result holds for $n = k + 1$, and hence holds for all $n \geq 1$, by the Principle of Strong Induction. $\qquad\qquad\square$

REMARK

Students often ask when to use simple induction (POMI) and when to use strong induction (POSI) in a proof. As a general rule, if you can prove the inductive conclusion $P(k+1)$ by assuming only $P(k)$, then use POMI, but if you need to assume $P(i)$ for one or more i with $i < k$, then use POSI. Usually the details of how many base cases are needed is secondary, and can be worked out later.

In fact, it turns out that POMI and POSI are logically equivalent. That is, anything that can be proved by one of these Principles can also be proved by the other. Moreover, both POMI and POSI are equivalent to another famous Principle in mathematics known as the Well-Ordering Principle. These equivalences are studied in more advanced courses in logic, and won't be proved in this course.

Chapter 5

Sets

5.1 Introduction

Sets were introduced in Chapter 1, and so far they have played an important role as the domain of a quantified statement. We have made substantial use of the specially designated sets $\mathbb{N}, \mathbb{Z}, \mathbb{R}$, but have not dealt in any detail with sets in general. In this chapter, we will consider further properties of sets, and how they are used in mathematics.

We begin with an important new set.

Definition 5.1.1

empty set

The set $\{\ \}$ contains no elements and is known as the **empty set**. We often use $\emptyset$ as a symbol for the empty set, that is,

$$\emptyset = \{\ \}.$$

REMARK

A common mistake is to think that the set $\{\emptyset\}$ is the same as the empty set $\emptyset$. However, the set $\{\emptyset\}$ is actually not the empty set, since it contains the set $\emptyset$ as an element! Thus

$$\{\emptyset\} \neq \emptyset.$$

The number of elements in a finite set S is called the **cardinality** of S, denoted by $|S|$. The cardinality of the empty set is defined to be zero, that is, $|\emptyset| = 0$. However, consistent with the remark above, we also have $|\{\emptyset\}| = 1$.

Example 1

Consider the following examples of cardinality of finite sets.

- $|\{2, 4, 6, 8, 10\}| = 5$,

- $|\{0\}| = |\{1\}| = |\{2\}| = 1$,

- $|\{\clubsuit, \diamondsuit, \heartsuit, \spadesuit\}| = 4$,

- $|\{1, 2, \{1, 2, 3\}\}| = 3$.

5.2   Set-builder Notation

Although the elements of small sets can be explicitly listed, most sets that mathematicians work with are too large or complicated for listing to be convenient. In this section, we introduce *set-builder notation*, which provides a compact, flexible method for describing a wide variety of sets.

We will require two ingredients for set-builder notation. First, we assume the existence of a very large set, known as the **universe of discourse**, usually denoted by $\mathcal{U}$, that contains all the objects that we might encounter in a given situation. Usually we simply refer to the set $\mathcal{U}$ as the *universe*. The universe is often not explicitly stated, we just assume that it exists. For example, in our work on divisibility, we are primarily concerned with integers, so it is safe to assume that the set of integers $\mathbb{Z}$ is the universe, even when we don't explicitly state this. The second ingredient required for set-builder notation is an *open sentence $P(x)$*, such that $P(x)$ is true or false for each object x in $\mathcal{U}$.

Given a set $\mathcal{U}$ as the universe, and an open sentence $P(x)$ whose value is defined for every $x \in \mathcal{U}$, we now define the first type of set-builder notation.

Definition 5.2.1

set-builder notation, type 1

The notation

$$\{x \in \mathcal{U} : P(x)\}$$

describes the set consisting of all objects x such that

- x is an element of $\mathcal{U}$, and

- $P(x)$ is true.

Informally, we read $\{x \in \mathcal{U} : P(x)\}$ as "The set of all x in $\mathcal{U}$ such that $P(x)$ is true.", or "The set of all x in $\mathcal{U}$ with property P."

REMARK

Some mathematicians replace the colon in set-builder notation by a vertical bar, so the above set becomes "$\{x \in \mathcal{U} \mid P(x)\}$". We will avoid this usage of the vertical bar in this course, to avoid confusion with our use of "$\mid$" for *divides*.

Some mathematicians omit explicit mention of the universe of discourse in the above notation, and simply write $\{x : P(x)\}$, assuming that the universe is implied by the context. In this course, we will omit mention of the universe in this way from time to time.

Example 2

The following are examples of the use of set-builder notation of type 1.

- The set $\{\ldots, -4, -2, 0, 2, 4, \ldots\}$ of all even integers can be described by

$$\{n \in \mathbb{Z} : 2 \mid n\}.$$

- The set $\{1, 2, 3, 5, 6, 10, 15, 30\}$ of all positive divisors of 30 can be written as

$$\{n \in \mathbb{N} : n \mid 30\}.$$

- The closed interval $[a, b]$, for $a \leq b$ can be described by

$$\{x \in \mathbb{R} : a \leq x \leq b\}.$$

- The open interval (a, b), for $a \leq b$ can be written as

$$\{x \in \mathbb{R} : a < x < b\}.$$

- We have $\{x \in \mathbb{R} : x^2 = 2\} = \{-\sqrt{2}, \sqrt{2}\}$, and $\{x \in \mathbb{Z} : x^2 = 2\} = \emptyset$.

For a second type of set-builder notation, we introduce another ingredient, and consider a set whose elements are of the form $f(x)$ for all x in the universe $\mathcal{U}$, where f is some function.

Definition 5.2.2

set-builder notation, type 2

The notation

$$\{f(x) : x \in \mathcal{U}\}$$

describes the set consisting of all objects of the form $f(x)$ such that

- x is an element of $\mathcal{U}$.

Example 3

The following are examples of the use of set-builder notation of type 2.

- The set of all even integers can be described as

$$\{2k : k \in \mathbb{Z}\}.$$

Here we have $f(k) = 2k$.

- The set $\{5, 25, 125, \ldots\}$ of all integers that are positive integer powers of 5 is given by

$$\{5^n : n \in \mathbb{N}\}.$$

Here we have $f(n) = 5^n$.

- The set $\{0, 1, 4, 9, 16\}$ can be described as

$$\{x^2 : x \in \{0, 1, 2, 3, 4\}\}.$$

Here we have $f(x) = x^2$.

Finally, the third type of set-builder notation combines all three ingredients - a universe $\mathcal{U}$, an *open sentence* P, and a function f.

Definition 5.2.3

set-builder notation, type 3

The notations

$$\{f(x) : x \in \mathcal{U}, P(x)\} \qquad \text{or} \qquad \{f(x) : P(x), x \in \mathcal{U}\}$$

both describe the set consisting of all objects of the form $f(x)$ such that

- x is an element of $\mathcal{U}$, and

- $P(x)$ is true.

Example 4   The following are examples of the use of set-builder notation of type 3.

- The set $\{5, 25, 125, \ldots\}$ of all integers that are positive integer powers of 5 is given by

$$\{5^k : k \in \mathbb{Z}, k > 0\}.$$

- The set $\{\ldots, -12, -6, 0, 6, 12, \ldots\}$ of all integers that are divisible by 6 can be described by

$$\{3a : 2 \mid a, a \in \mathbb{Z}\}.$$

These are the types of set-builder notation that we will use most frequently in this course. All three involve a pair of brace brackets. Inside the brace brackets is a single colon, with expressions on its left and on its right.

- The expression on the *left* of the colon gives a typical element of the set, in terms of a single variable.

- The expression on the *right* of the colon gives a list of conditions that the variable must satisfy, with the convention that each comma separating conditions is interpreted as "AND".

In situations where it is useful, follow the above pattern to create even more complicated set-builder notation with multiple variables.

Example 5   The following is an example of the use of set-builder notation with two variables.

- The set of rational numbers $\mathbb{Q}$ can be written as

$$\mathbb{Q} = \left\{ \frac{a}{b} : a \in \mathbb{Z}, b \in \mathbb{Z}, b \neq 0 \right\}.$$

REMARK

One of the conventions that we have used in expressing $\mathbb{Q}$ in the set-builder notation in Example 5 concerns the fact that a set either contains an object, or it does not contain that object, for each object in the universe. Now, $\mathbb{Q}$ does contain the number $\frac{2}{3}$. Since we have

$$\frac{2}{3} = \frac{-2}{-3} = \frac{4}{6} = \frac{-4}{-6} = \frac{6}{9} = \frac{-6}{-9} = \ldots,$$

then these equalities of fractions demonstrate that the number $\frac{2}{3}$ is created in the form $\frac{a}{b}$ for more than one choice of the pair a, b: for $a = 2, b = 3$, for $a = -2, b = -3$, for $a = 4, b = 6$, etc.

Hence, when we create the elements in $\mathbb{Q}$ as objects of the form $\frac{a}{b}$, where a and b are integers, and b is non-zero, we do not worry about the fact that we are "creating" the number $\frac{2}{3}$ in more than one way (i.e., with different choices of a and b) - we simply have to create each number in $\mathbb{Q}$ at least once (and our convention is that the "extra" copies of this number are ignored).

5.3   Set Operations

In this section we consider some of the basic set operations that you need to be familiar with. In the definitions below, note how we use logical operators within set-builder notation, and that we sometimes omit mention of the universe $\mathcal{U}$, depending on the context.

Definition 5.3.1

union

The **union** of two sets S and T, written $S \cup T$, is the set of all elements belonging to either set S or set T (or both). Symbolically we write

$$S \cup T = \{x : x \in S \text{ OR } x \in T\} = \{x : (x \in S) \vee (x \in T)\}$$

Definition 5.3.2

intersection

The **intersection** of two sets S and T, written $S \cap T$, is the set of all elements belonging to both set S and set T. Symbolically we write

$$S \cap T = \{x : x \in S \text{ AND } x \in T\} = \{x : (x \in S) \wedge (x \in T)\}$$

Definition 5.3.3

set-difference

The **set-difference** of two sets S and T, written $S - T$ (or $S \setminus T$), is the set of all elements belonging to S but not T. Symbolically we write

$$S - T = \{x : x \in S \text{ AND } x \notin T\} = \{x : (x \in S) \wedge (x \notin T)\} = \{x : (x \in S) \wedge (\neg(x \in T))\}$$

Definition 5.3.4

complement

(*Without explicit mention of the universe $\mathcal{U}$*) The **complement** of a set S, written $\overline{S}$, is the set of all elements not in S. Symbolically, we write

$$\overline{S} = \{x : x \notin S\}.$$

(*With explicit mention of the universe $\mathcal{U}$*) The **complement** of a set S whose elements belong to $\mathcal{U}$, is written $\overline{S}$, and it is the set of all elements in $\mathcal{U}$ but not in S. Symbolically, we write

$$\overline{S} = \{x \in \mathcal{U} : x \notin S\}.$$

Note that in terms of set-difference notation, we have $\overline{S} = \mathcal{U} - S$.

Example 6   Let $\mathcal{U} = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$, $S = \{1, 3, 5, 7, 9\}$ and $T = \{0, 1, 2, 7, 8, 9\}$. Then we have

$$S \cup T = \{0, 1, 2, 3, 5, 7, 8, 9\},$$
$$S \cap T = \{1, 7, 9\},$$
$$S - T = \{3, 5\},$$
$$T - S = \{0, 2, 8\},$$
$$\overline{S} = \{0, 2, 4, 6, 8\},$$
$$\overline{T} = \{3, 4, 5, 6\}.$$

Example 7   Let S be an arbitrary subset of a universe $\mathcal{U}$.

- We have $S \cup \emptyset = S$, $S \cup \mathcal{U} = \mathcal{U}$ and $S \cup S = S$.

- We have $S \cap \emptyset = \emptyset$, $S \cap \mathcal{U} = S$ and $S \cap S = S$.

- We have $S - \emptyset = S$, $\emptyset - S = \emptyset$ and $S - S = \emptyset$.

- We have $\overline{\mathcal{U}} = \emptyset$, $\overline{\emptyset} = \mathcal{U}$, $S \cup \overline{S} = \mathcal{U}$ and $S \cap \overline{S} = \emptyset$.

5.4   Subsets of a Set

In the previous sections, we considered some basic set operations. In this section, we will consider how to compare two sets, say S and T.

Definition 5.4.1

disjoint

Two sets S and T are said to be **disjoint** when $S \cap T = \emptyset$.

In other words, S and T are disjoint when they have no elements in common. Note that $S \cap \emptyset = \emptyset \cap S = \emptyset$ for all choices of S, so S and the empty set $\emptyset$ are disjoint for all sets S. Also, we have $S \cap \overline{S} = \emptyset$ for all choices of S, so S and its complement are disjoint for all sets S.

Definition 5.4.2

subset, proper subset

A set S is called a **subset** of a set T, written symbolically as $S \subseteq T$, when every element of S belongs to T.

A set S is called a **proper subset** of a set T, written symbolically as $S \subsetneq T$, when S is a subset of T and there exists an element in T which does not belong to S.

When S is *not* a subset of T, we can write it symbolically as $S \nsubseteq T$. When we have $S \subseteq T$, we also say that T is a *superset* of S, and when we have $S \subsetneq T$, we also say that T is a *proper superset* of S.

Example 8   Consider the following examples of subsets and proper subsets.

- We have $\emptyset \subseteq S$ for all sets S.

- We have $S \subseteq S$ for all sets S, but S is never a *proper* subset of S.

- $\{1, 2, 3\} \subseteq \{1, 2, 3, 4\}$, and $\{1, 2, 3\} \subsetneq \{1, 2, 3, 4\}$.

- $\{2, 4, 6\} \not\subseteq \{1, 2, 4, 5\}$, and $\{2, 4, 6\} \not\subseteq \{3\}$.

- We have $\mathbb{N} \subseteq \mathbb{Z}$, $\mathbb{Z} \subseteq \mathbb{Q}$, and $\mathbb{Q} \subseteq \mathbb{R}$.

REMARK

Some mathematicians use the notation $S \subset T$ to mean "S is a proper subset of T". However, this is not universal, since $S \subset T$ is used by others to also mean "S is a subset of T". To avoid any potential confusion, we will not use the notation $S \subset T$ in this course. Instead, we will explicitly use $S \subseteq T$ or $S \subsetneq T$ as needed.

There is a nice formula for the number of subsets of a finite set, given in the next result. In the statement of this result, we use the notation $S_n = \{1, 2, 3, \ldots, n\}$, for a positive integer n. The proof is by Mathematical Induction on n.

Proposition 1   For all integers $n \geq 1$, S_n has 2^n subsets.

Proof: We prove this result by induction on n, where $P(n)$ is the statement

S_n has 2^n subsets.

Base Case When $n = 1$, we have $S_1 = \{1\}$, whose only subsets are $\emptyset$ and $\{1\}$. Hence S_1 has $2 = 2^1$ subsets, so we have proved $P(1)$.

Inductive Step Let k be an arbitrary natural number.

Assume the inductive hypothesis, $P(k)$. That is, we assume S_k has 2^k subsets.

We wish to prove $P(k + 1)$, which is the statement

S_{k+1} has 2^{k+1} subsets.

Every subset of S_{k+1} either includes the element $k + 1$ or it does not. So we divide the subsets of S_{k+1} into two disjoint sets. We will let A be the set of subsets of S_{k+1} which do not include the element $k + 1$ and we will let B be the set of subsets of S_{k+1} which do include the element $k + 1$. Since the sets A and B are disjoint, the number of subsets of S_{k+1} is equal to $|A| + |B|$.

The sets in A do not include the element $k + 1$ and therefore, they are all of the subsets of S_k. Therefore, by the inductive hypothesis, $|A| = 2^k$.

If we take any set from A and add the element $k+1$ to it, then we will have a subset of S_{k+1} which includes the element $k+1$ and therefore, this set is in B. Therefore, B as at least as many sets in it as A. Therefore, $|A| \le |B|$. (Convince yourself that adding $k+1$ to two distinct sets in A will produce two distinct sets in B.)

Any set in B includes the element $k+1$. If we take any set from B and remove the element $k+1$ from it, then we will have a subset of S_{k+1} which does not include the element $k+1$ and therefore, this set is in A. Therefore, $|B| \le |A|$. (Convince yourself that removing $k+1$ from two distinct sets in B will produce two distinct sets in A.)

Since $|A| \le |B|$ and $|B| \le |A|$, we have that $|A| = |B|$. Therefore, $|B| = 2^k$. Therefore, the number of subsets of S_{k+1} is $|A| + |B| = 2^k + 2^k = 2^{k+1}$.

The result is true for $n = k+1$, and hence holds for all $n \ge 1$ by the Principle of Mathematical Induction. □

Note that Proposition 1 also holds for $n = 0$, using the convention that $S_0 = \{\}$. That is, the only subset of the empty set is the empty set itself, so S_0 has $1 = 2^0$ subsets, in agreement with the result in Proposition 1 for the case $n = 0$.

5.5   Subsets, Set Equality and Implications

We now look at subsets of a set in a different way, in terms of the universally quantified statement

$$\forall\, x \in \mathcal{U},\ (x \in S) \implies (x \in T). \tag{5.1}$$

Now, statement (5.1) is true when for all objects x for which the hypothesis "$x \in S$" is true, the conclusion "$x \in T$" must also be true. This happens exactly when S is a subset of T. Moreover, statement (5.1) is false when there exists some $x \in S$ such that $x \notin T$, which happens exactly when S is not a subset of T. In summary, $S \subseteq T$ when (5.1) is true, and $S \nsubseteq T$ when (5.1) is false. This immediately gives us the following method for proving whether $S \subseteq T$ or not, for an arbitrary pair of sets S and T.

Proof Method

To prove that $S \subseteq T$, prove the universally quantified implication:

$$\forall\, x \in \mathcal{U},\ (x \in S) \implies (x \in T).$$

In the following examples of this proof method, we consider particular sets of numbers, in the first example integers, and in the second example real numbers. Note how set-builder notation is used to describe these sets in a compact way.

Example 9

Let $A = \{n \in \mathbb{Z} : 10 \mid n\}$ and $B = \{n \in \mathbb{Z} : 2 \mid n\}$.

(a) Prove that $A \subseteq B$.

(b) Prove that $B \nsubseteq A$.

Solution: For each part, we apply the proof method above.

(a) We prove "$\forall\ n \in \mathbb{Z},\ (n \in A) \implies (n \in B)$": Let n be an arbitrary integer, and assume that $n \in A$, so we have $10 \mid n$. Hence there exists an integer k such that $n = 10k$. We thus have $n = 2(5k)$ where $5k$ is an integer, and so we conclude that $2 \mid n$, or $n \in B$.

(b) We disprove $B \subseteq A$, by disproving "$\forall\ n \in \mathbb{Z},\ (n \in B) \implies (n \in A)$" using a counter-example: Suppose we let $n = 2$. Now 2 is divisible by 2, but 2 is *not* divisible by 10. Therefore, for the choice $n = 2$, the hypothesis "$n \in B$" is true, and the conclusion "$n \in A$" is false. Hence $n = 2$ is a counter-example for the given implication.

Example 10   Let $C = \{n\pi : n \in \mathbb{Z}\}$ and $D = \{x \in \mathbb{R} : f(x) = 0\}$, where $f(x) = (x^2 - 1)\sin x$.

(a) Prove that $C \subseteq D$.

(b) Prove that $D \nsubseteq C$.

Solution: We again apply the proof method above.

(a) We prove "$\forall\ x \in \mathbb{R},\ (x \in C) \implies (x \in D)$": Let x be an arbitrary element in C, so we have $x = n\pi$ for some $n \in \mathbb{Z}$. Now $f(n\pi) = ((n\pi)^2 - 1)\sin(n\pi) = 0$, since $\sin(n\pi) = 0$ for $n \in \mathbb{Z}$. Hence $x = n\pi \in D$.

(b) We disprove $D \subseteq C$, by disproving "$\forall\ x \in \mathbb{R},\ (x \in D) \implies (x \in C)$" using a counter-example: Suppose we let $x = 1$. Now $f(1) = 0$, but $1 \neq n\pi$ for any $n \in \mathbb{Z}$. Therefore, for the choice $x = 1$, the hypothesis "$x \in D$" is true, and the conclusion "$x \in C$" is false. Hence $x = 1$ is a counter-example for the given implication.

Now we consider a different type of example, in which the sets are arbitrary, within an arbitrary universe.

Example 11   Prove each of the following statements for an arbitrary universe $\mathcal{U}$.

(a) For all sets A and B, $(A \cap B) \subseteq A$.

(b) For all sets A and B, $\overline{A \cup B} \subseteq \overline{A}$.

Solution: We prove each part using the corresponding universally quantified implication.

(a) We prove "$\forall\ x \in \mathcal{U}, (x \in A \cap B \implies x \in A)$": Let x be an arbitrary element of $\mathcal{U}$, let A and B be arbitrary sets of elements in $\mathcal{U}$, and assume that $x \in A \cap B$. Thus we have $x \in A$ and $x \in B$, and hence $x \in A$.

(b) We prove "$\forall\ x \in \mathcal{U}, (x \in \overline{A \cup B} \implies x \in \overline{A})$": Let x be an arbitrary element of $\mathcal{U}$, let A and B be arbitrary sets of elements in $\mathcal{U}$, and assume that $x \in \overline{A \cup B}$. Now,

using the definitions of set union, intersection and complement, we have

$$\begin{aligned}
\overline{A \cup B} &= \{x \in \mathcal{U} : \neg(x \in A \cup B)\} \\
&= \{x \in \mathcal{U} : \neg((x \in A) \vee (x \in B))\} \\
&= \{x \in \mathcal{U} : (\neg(x \in A) \wedge (\neg(x \in B))\} \\
&= \overline{A} \cap \overline{B},
\end{aligned}$$

where we have used DeMorgan's Laws for the third equality. Therefore, since $x \in \overline{A \cup B}$, we have $x \in \overline{A} \cap \overline{B}$. But this gives $x \in \overline{A}$ and $x \in \overline{B}$, and hence $x \in \overline{A}$.

For our final topic on sets, we define what it means for two sets to be equal.

Definition 5.5.1

equal sets

We say that two sets S and T are **equal**, and write $S = T$, when $S \subseteq T$ and $T \subseteq S$.

This definition of set equality simply says that two sets S and T are equal when every element of S is in T *and* every element of T is in S (i.e., the sets S and T have exactly the same elements).

For example, the sets A and B defined in Example 10 are not equal (i.e., $A \neq B$), since as we proved in part (b), $B \not\subseteq A$. Similarly, the sets C and D defined in Example 9 are not equal, since as we proved in part (b), $D \not\subseteq C$.

Of course, from our proof method for proving $S \subseteq T$ and $T \subseteq S$ via universally quantified implications, we can prove that $S = T$ by proving the statement

$$\Big(\forall \, x \in \mathcal{U}, \; (x \in S) \implies (x \in T)\Big) \wedge \Big(\forall \, x \in \mathcal{U}, \; (x \in T) \implies (x \in S)\Big), \qquad (5.2)$$

or, equivalently,

$$\forall \, x \in \mathcal{U}, \; (x \in S) \iff (x \in T).$$

Consider the following example of applying this method to prove equality of sets.

Example 12

Define the three sets

$$\begin{aligned}
A &= \{(t, \tfrac{\sqrt{3}}{2}t) : t \in \mathbb{R}\}, \\
B &= \{(t, -\tfrac{\sqrt{3}}{2}t) : t \in \mathbb{R}\}, \\
C &= \{(x, y) : x, y \in \mathbb{R}, 16y^4 - 9x^4 = 0\}.
\end{aligned}$$

Prove that $A \cup B = C$.

Solution: We will prove $A \cup B = C$ by proving that $A \cup B \subseteq C$ and $C \subseteq A \cup B$, using universally quantified implications.

To prove $A \cup B \subseteq C$: We prove the implication

For all real numbers x and y, if $(x, y) \in A \cup B$, then $(x, y) \in C$.

Hence, let t be an arbitrary real number, and assume that (i) $(x, y) = (t, \frac{\sqrt{3}}{2}t) \in A$, or (ii) $(x, y) = (t, -\frac{\sqrt{3}}{2}t) \in B$. Then for (i) we obtain

$$16y^4 - 9x^4 = 16\left(\tfrac{\sqrt{3}}{2}t\right)^4 - 9t^4 = 16\tfrac{9}{16}t^4 - 9t^4 = 9t^4 - 9t^4 = 0,$$

and thus we conclude that $(x, y) \in C$. Similarly, for (ii) we obtain

$$16y^4 - 9x^4 = 16\left(-\tfrac{\sqrt{3}}{2}t\right)^4 - 9t^4 = 16\tfrac{9}{16}t^4 - 9t^4 = 9t^4 - 9t^4 = 0,$$

and again we conclude that $(x, y) \in C$.

To prove $C \subseteq A \cup B$: We prove the implication

For all real numbers x and y, if $(x, y) \in C$, then $(x, y) \in A \cup B$.

Hence, assume that $(x, y) \in C$, so we have

$$16y^4 - 9x^4 = 0.$$

Now the left hand side of the equation is a difference of squares, so we can factor it as $16y^4 - 9x^4 = (4y^2 - 3x^2)(4x^2 + 3y^2)$, and similarly we have $4y^2 - 3x^2 = (2y - \sqrt{3}x)(2y + \sqrt{3}x)$. Substituting these into the equation above, we obtain

$$\left(2y - \sqrt{3}x\right)\left(2y + \sqrt{3}x\right)\left(4y^2 + 3x^2\right) = 0. \tag{5.3}$$

On the left hand side of equation (5.3) we have a product of three real factors, so (x, y) is a solution to (5.3) if and only if it makes at least one of the factors equal to zero. Hence we consider three cases for (x, y).

Case 1: Suppose $2y - \sqrt{3}x = 0$. Then we have $y = \frac{\sqrt{3}}{2}x$ where $x \in \mathbb{R}$. This gives

$$(x, y) = (x, \tfrac{\sqrt{3}}{2}x) \in A,$$

and hence $(x, y) \in A \cup B$.

Case 2: Suppose $2y + \sqrt{3}x = 0$. Then we have $y = -\frac{\sqrt{3}}{2}x$ where $x \in \mathbb{R}$. This gives

$$(x, y) = (x, -\tfrac{\sqrt{3}}{2}x) \in B,$$

and hence $(x, y) \in A \cup B$.

Case 3: Suppose $4y^2 + 3x^2 = 0$. Since x and y are real, we have $x^2 \geq 0$ and $y^2 \geq 0$, so $4y^2 + 3x^2 \geq 0$, with equality only when $x = y = 0$. This gives $(x, y) = (0, 0) \in A \cup B$. (Note that $(0, 0)$ is contained in both A and B, and that $(0, 0)$ also occurs in both Cases 1 and 2, with $x = 0$.)

Note that the set C in the example above is the set of solutions to an equation, and the set $A \cup B$ is a set of values that we might regard as "candidates" as solutions to that equation. When we prove that $A \cup B = C$, we prove that all values in $A \cup B$ are solutions, *and* that none of them are extraneous. We will apply this proof method for solutions to an equation later in these notes, in Chapter 7 on page 118, and in Chapter 10 on page 174.

Chapter 6

The Greatest Common Divisor

6.1 The Division Algorithm

In this chapter, we return to *divisibility* of integers, which was defined on page 47 in Chapter 3. We begin with a preliminary result about the *absolute value* $|x|$ of the real number x.

Proposition 1

For all real numbers x, we have $x \leq |x|$.

Proof: Let x be an arbitrary real number. Then x is either negative, or non-negative, and we consider these as two cases.

Case 1: When x is negative, we have $x < 0 < |x|$. Hence $x < |x|$, so $x \leq |x|$ in this case.

Case 2: When x is non-negative, we have $x = |x|$, giving $x \leq |x|$ in this case. $\qquad\square$

REMARK

To clarify the above proof, note that Proposition 1 can be rewritten as

For all real numbers x, $\quad \big(x < |x|\big)$ or $\big(x = |x|\big)$.

That is, the "$\leq$" relationship can be regarded informally as ("$<$" or "$=$").

Now we use Proposition 1 to prove a very simple result about divisibility that will be helpful in proving other results later in the notes.

Proposition 2

(Bounds By Divisibility (BBD))

For all integers a and b, if $b \mid a$ and $a \neq 0$ then $b \leq |a|$.

Proof: Let a and b be arbitrary integers, and assume that $b \mid a$ and $a \neq 0$. Since $b \mid a$, from the definition of divisibility there exists an integer q so that $a = qb$. Since $a \neq 0$, this gives $qb \neq 0$, and hence $q \neq 0$. Therefore $q \in \{\ldots, -2, -1, 1, 2, \ldots\}$, so $|q| \geq 1$. Then, using properties of absolute value, we have $|a| = |qb| = |q| \cdot |b| \geq 1 \cdot |b| = |b|$, which gives $|b| \leq |a|$. Now, from Proposition 1 we also have the inequality $b \leq |b|$, and putting these two inequalities together, we obtain $b \leq |b| \leq |a|$, which gives $b \leq |a|$. $\square$

In the above proposition, one of the hypotheses is that b is divisible by a. In the proof, after assuming this, we used the definition of divisibility to then write $a = qb$, where q is an integer. We chose the variable name q for this integer because it is the *quotient* when a is divided by b in the situation when a is divisible by b.

As you learned when you studied arithmetic in school, you can divide a by b for any two integers with divisor b positive. In general, the result of this division is a quotient q, and a *remainder* which we will call r (which is equal to 0 if and only if a is divisible by b). For example, if we divide 94 by 13, then the quotient is 7, and the remainder is 3 (since $7 \times 13 = 91$ and $94 - 91 = 3$). In this case, we deduce that $94 = 7 \times 13 + 3$, and note that the remainder 3 is non-negative, and strictly smaller than the divisor 13. In general, when we divide an integer a by a positive integer b, we obtain a quotient q and remainder r, and these quantities are related by the equation

$$a = qb + r, \qquad 0 \leq r < b. \tag{6.1}$$

In this course we assume this result, and do not prove it. To be precise, we assume that for all integers a and positive integers b, there exist integers q and r satisfying equation (6.1). In the following proposition, we assume the existence of integers q and r as above, and prove that they are unique.

Proposition 3 **(Division Algorithm (DA))**

For all integers a and positive integers b, there exist unique integers q and r such that

$$a = qb + r, \qquad 0 \leq r < b.$$

Proof: Let a be an arbitrary integer, and b be an arbitrary positive integer. As discussed above, we assume that there exist integers q and r such that $a = qb + r$, with $0 \leq r < b$.

To prove that the integers q and r are unique, assume that q_1 and r_1 are integers such that $a = q_1 b + r_1$, with $0 \leq r_1 < b$, and that q_2 and r_2 are integers such that $a = q_2 b + r_2$, with $0 \leq r_2 < b$. Then we have the pair of inequalities

$$
\begin{aligned}
0 \leq\ \ r_1\ &< b, \\
-b < -r_2\ &\leq 0,
\end{aligned}
$$

where the second inequality is obtained from $0 \leq r_2 < b$ by multiplying through by -1 (which therefore reverses the inequalities). Now, adding this pair of inequalities, we obtain

$$-b < r_1 - r_2 < b. \tag{6.2}$$

However, we also have

$$0 = a - a = (q_1 b + r_1) - (q_2 b + r_2) = (q_1 - q_2)b + (r_1 - r_2),$$

and rearranging the equation $0 = (q_1 - q_2)b + (r_1 - r_2)$ gives $(q_2 - q_1)b = r_1 - r_2$. This means that $b \mid (r_1 - r_2)$, so $r_1 - r_2 = kb$ for some integer k. Now substituting for $r_1 - r_2$ in (6.2) gives $-b < kb < b$, and dividing through this inequality by b (which is positive) gives $-1 < k < 1$. Since k is an integer, we conclude that $k = 0$. Now $k = 0$ gives $r_1 - r_2 = kb = 0\,b = 0$, so we have $r_1 = r_2$. Finally, substituting $r_1 = r_2$ gives $(q_2 - q_1)b = r_1 - r_2 = 0$, so $(q_2 - q_1)b = 0$ and since $b \neq 0$, we have $q_2 - q_1 = 0$, and hence $q_1 = q_2$.

We have proved that $q_1 = q_2$ and $r_1 = r_2$, and therefore conclude that the integers q and r are unique. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Note that in the above proof, we have proceeded as in part (a) of Step 2 in the proof method for uniqueness given on page 59.

Example 1   As examples of the Division Algorithm with $b = 13$, we have

- $3142 = 241(13) + 9$

- $-3142 = -242(13) + 4$

- $5 = 0(13) + 5$

6.2   The Greatest Common Divisor

Now we consider a pair of integers, both of which are divisible by a given integer.

Definition 6.2.1

common divisor

Let a and b be integers. An integer c is called a **common divisor** of a and b if $c \mid a$ and $c \mid b$.

In other words, when c is a common divisor of a and b, then one of the properties that a and b have in *common* is that they both have c as a divisor. This English usage of the word "common" is the reason that we use it in the above definition.

Definition 6.2.2

greatest common divisor

Let a and b be integers.

1. If a and b are not both zero, an integer $d > 0$ is the **greatest common divisor** of a and b, written $d = \gcd(a, b)$, when

   - d is a common divisor of a and b, and
   - for all integers c, if c is a common divisor of a and b, then $c \leq d$.

2. If a and b are both zero, we define $\gcd(a, b) = \gcd(0, 0) = 0$.

In other words the greatest common divisor of integers a and b that are not both zero, is the largest integer that is both a divisor of a and a divisor of b.

In the following example, we state some simple properties of the gcd. Note that to prove some of these properties, we need to consider the cases $a = 0$ and $a \neq 0$.

Example 2   Let a be an arbitrary integer.

- We have $\gcd(a, a) = \gcd(a, -a) = |a|$ (note that when $a = 0$, this gives $\gcd(0, 0) = 0$).

- We have $\gcd(a, 1) = \gcd(a, -1) = 1$.

- We have $\gcd(a, 0) = |a|$ (note that when $a = 0$, this also gives $\gcd(0, 0) = 0$).

Now we consider a numerical example in which we exhaustively list all divisors of two integers, then list their common divisors and hence determine the gcd.

Example 3   Consider $a = 18$ and $b = 12$.

- The divisors of $a = 18$ are -18, -9, -6, -3, -2, -1, 1, 2, 3, 6, 9, 18.

- The divisors of $b = 12$ are -12, -6, -4, -3, -2, -1, 1, 2, 3, 4, 6, 12.

- The common divisors of $a = 18$ and $b = 12$ are -6, -3, -2, -1, 1, 2, 3, 6.

- The greatest common divisor of $a = 18$ and $b = 12$ is 6 (which is the largest integer in the list of common divisors above), so we have $\gcd(a, b) = \gcd(18, 12) = 6$.

In general, proceeding as in the above example is not an efficient way to determine the gcd of a pair of integers. The following proposition is introduced at this stage in order to develop an efficient way for determining greatest common divisors.

Proposition 4   **(GCD With Remainders (GCD WR))**

For all integers a, b, q and r, if $a = qb + r$, then $\gcd(a, b) = \gcd(b, r)$.

Proof: Let a, b, q and r be arbitrary integers, and assume that $a = qb + r$. Now either a and b are not both zero, or they are both zero, and we consider these possibilities as two cases.

Case 1: When a and b are not both zero, let $d = \gcd(a, b)$. Hence from the definition of gcd we have $d \mid a$ and $d \mid b$, and by the Divisibility of Integer Combinations (DIC), this implies that d divides $a(1) + b(-q) = a - qb$. But rearranging $a = qb + r$, we get $a - qb = r$, so d divides r, and hence d is a common divisor of b and r.

Now let c be an arbitrary common divisor of b and r. Since $c \mid b$ and $c \mid r$, then we have $c \mid (qb + r)$ by DIC. Now $a = qb + r$, so $c \mid a$. Since $d = \gcd(a, b)$ and $c \mid a$ and $c \mid b$, then from the definition of gcd we have $c \leq d$. Hence $\gcd(b, r) = d = \gcd(a, b)$ in this case.

Case 2: When a and b are both zero, then $a = qb + r$ becomes $0 = 0 + r$, so we have $r = 0$. We thus have $\gcd(a, b) = \gcd(0, 0) = 0$, and $\gcd(b, r) = \gcd(0, 0) = 0$, giving $\gcd(a, b) = \gcd(b, r)$ in this case. $\qquad\square$

Note that the hypothesis in GCD WR is the equation $a = qb + r$ relating the integers a, b, q and r. However, unlike in the statement of the Division Algorithm, the restriction $0 \leq r < b$ does *not* appear in GCD WR.

Note also the structure of the proof of the above proposition GCD WR. It is based on a case analysis to consider different possible values for the integers a and b, and it uses the definition of gcd in a careful way. The definition of gcd for a and b not both zero has two parts, that it is a common divisor, and that it is the largest integer among all common divisors, and indeed both of these parts arise within the proof for Case 1. In addition, the variable d is introduced in the above proof by use of the word "let", though d is not mentioned in the statement of the proposition. Of course, d has a precise meaning in terms of the variables a and b that do appear in the statement of the proposition.

REMARK

In creating a proof, you can always use "let" (or "define") to introduce a new variable, if it helps to make the reasoning clear and compact.

One way in which GCD With Remainders can be used is to determine the greatest common divisor of a pair of integers in an efficient way. The following example illustrates how to use GCD WR iteratively to determine $\gcd(1386, 322)$, avoiding the exhaustive listing of common divisors that appeared in Example 3 above.

Example 4   We use GCD WR and the Division Algorithm to determine $\gcd(1386, 322)$ as follows.

Stage 1:	Since	$1386 = 4 \times 322 + 98,$	$\gcd(1386, 322) = \gcd(322, 98).$
Stage 2:	Since	$322 = 3 \times 98 + 28,$	$\gcd(322, 98) = \gcd(98, 28).$
Stage 3:	Since	$98 = 3 \times 28 + 14,$	$\gcd(98, 28) = \gcd(28, 14).$
Stage 4:	Since	$28 = 2 \times 14 + 0,$	$\gcd(28, 14) = \gcd(14, 0).$

Since $\gcd(14, 0) = 14$, the chain of equalities in the right hand column above gives

$$\gcd(1386, 322) = \gcd(322, 98) = \gcd(98, 28) = \gcd(28, 14) = \gcd(14, 0) = 14,$$

so we have $\gcd(1386, 322) = 14$.

REMARK (Euclidean Algorithm)

Note how we use GCD With Remainders in the above example. Initially, we wish to determine the gcd of a pair of positive integers (a, b) with $a > b$. At every stage we use the Division Algorithm, writing $a = qb + r$, with $0 \leq r < b$, and deduce that $\gcd(a, b) = \gcd(b, r)$ from GCD WR. Hence we now have the gcd of a smaller pair (b, r) than the pair (a, b) that we started with (since $a > b > r$). We continue stage by stage until we have the gcd of a pair $(c, 0)$ for some positive integer c, and then use the fact that $\gcd(c, 0) = c$ to conclude that $\gcd(a, b) = c$. This process for determining the greatest common divisor is known as the *Euclidean Algorithm*.

We will return to the Euclidean Algorithm in the next section, and finish this section with two examples of how GCD With Remainders can also be used to determine greatest common divisors for expressions involving variables.

Example 5   Prove that for all integers a, we have $\gcd(22a + 7, 3a + 1) = 1$.

Proof: Let a be an arbitrary integer. Since $22a + 7 = 7 \cdot (3a + 1) + a$, GCD WR gives $\gcd(22a + 7, 3a + 1) = \gcd(3a + 1, a)$. Also, since $3a + 1 = 3 \cdot a + 1$, GCD WR (again) gives $\gcd(3a + 1, a) = \gcd(a, 1)$. Now $\gcd(a, 1) = 1$ for all integers a, so we have

$$\gcd(22a + 7, 3a + 1) = \gcd(3a + 1, a) = \gcd(a, 1) = 1,$$

and hence $\gcd(22a + 7, 3a + 1) = 1$.   □

Example 6   Prove that for all integers a, we have $\gcd(a^2, a + 1) = 1$.

Proof: Let a be an arbitrary integer. Since $a^2 = (a - 1) \cdot (a + 1) + 1$, GCD WR gives $\gcd(a^2, a + 1) = \gcd(a + 1, 1)$. Now $\gcd(a + 1, 1) = 1$ for all integers a, so we have

$$\gcd(a^2, a + 1) = \gcd(a + 1, 1) = 1,$$

and hence $\gcd(a^2, a + 1) = 1$.   □

6.3   Certificate of Correctness and Bézout's Lemma

We now return to the task of determining the gcd of a pair of integers a and b in an efficient way. The following result will turn out to make an important contribution to this task, though it may seem unmotivated at first glance.

Theorem 5   **(GCD Characterization Theorem (GCD CT))**

For all integers a and b, and non-negative integers d, if

- d is a common divisor of a and b, and

- there exist integers s and t such that $as + bt = d$,

then $d = \gcd(a, b)$.

Proof: Let a and b be arbitrary integers, and d be an arbitrary non-negative integer. Assume that d is a common divisor of a and b, and that there exist integers s and t such that $as + bt = d$. We consider the two cases a and b not both zero, and $a = b = 0$.

Case 1: Suppose that a and b are not both zero. Since d is a common divisor of a and b, this means that $d \neq 0$, so d is positive.

Now let c be an arbitrary common divisor of a and b. Hence $c \mid a$ and $c \mid b$, so by the Divisibility of Integer Combinations, we have $c \mid (ax+by)$ for all integers x and y. Therefore $c \mid (as + bt)$, and since $as + bt = d$, we obtain $c \mid d$. Then from Bounds by Divisibility we

obtain $c \leq |d|$, so by the positivity of d we get $c \leq d$. Hence from the definition of gcd, we have $d = \gcd(a, b)$.

Case 2: Suppose that $a = b = 0$. From the assumption that there exist integers s and t such that $as + bt = d$, we must have $d = 0$, since $0s + 0t = 0$ for *all* integers s and t. Also 0 is a divisor of 0, so $d = 0$ is a common divisor of $a = 0$ and $b = 0$. Since $\gcd(0, 0) = 0$, then $d = \gcd(a, b)$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Example 7

To see how to use the GCD Characterization Theorem, suppose we wish to prove that the greatest common divisor of $a = 1386$ and $b = 322$ is equal to $d = 14$, but we have lost the details of our calculations in Example 4 above. Suppose also that we do not wish to repeat our work in Example 4, nor do we want to find all divisors of 1386 and 322 (as in Example 3).

Instead, first let's check that $14 \mid 1386$ and $14 \mid 322$. This follows easily, using division by 14 to obtain $1386 = 99 \times 14$ and $322 = 23 \times 14$ as confirmation. Then, second, the GCD Characterization Theorem says that, *if we are able to find* a pair of integers s and t such that $1386s + 322t = 14$, this is all we need to complete a proof that $\gcd(1386, 322) = 14$.

It turns out that $s = 10$ and $t = -43$ are exactly such a pair of integers, which we check easily by the calculation $1386 \times 10 + 322 \times (-43) = 13860 - 13846 = 14$. This completes the proof that $\gcd(1386, 322) = 14$.

REMARK

Note the key role of the pair of integers s and t appearing in the GCD Characterization Theorem, as illustrated in Example 7 above. Given a pair of integers a and b, and a non-negative integer d, having first checked that d is a common divisor of a and b, then the integers s and t satisfying $as + bt = d$ are all we need to confirm (or "certify") that $d = \gcd(a, b)$. In fact, reflecting this language, we will refer to the integers s and t as providing a *certificate of correctness* for d as the greatest common divisor of a and b.

Of course, the big question is: "How do we find s and t?" We answer this question in the following example for the integers $a = 1386$ and $b = 322$ whose gcd we certified as $d = 14$ in Example 7.

Example 8

Consider the calculations in Example 4, where we used the Euclidean Algorithm to first determine the gcd of 1386 and 322 in four stages, each using the Division Algorithm. At each stage, we have an equation of the form $a = bq + r$. Rewriting this in the form $a - bq = r$ for the first three stages, we get the equations

$$1386 - 4 \times 322 = 98, \qquad\qquad\qquad (6.3)$$
$$322 - 3 \times 98 = 28, \qquad\qquad\qquad (6.4)$$
$$98 - 3 \times 28 = 14. \qquad\qquad\qquad (6.5)$$

Note that the greatest common divisor in this example is the positive integer $d = 14$ that appears on the right hand side of equation (6.5). Now substitute equation (6.4) into equation (6.5) to eliminate 28, giving

$$98 - 3 \times (322 - 3 \times 98) = 14,$$

and expand and rearrange the left hand side to obtain

$$-3 \times 322 + 10 \times 98 = 14.$$

Next, substitute equation (6.3) into the above equation to eliminate 98, giving

$$-3 \times 322 + 10 \times (1386 - 4 \times 322) = 14,$$

and expand and rearrange the left hand side to obtain

$$10 \times 1386 - 43 \times 322 = 14. \tag{6.6}$$

Now, recalling that $a = 1386$, $b = 322$ and $d = 14$ for this example, we see that equation (6.6) is of the form $as + bt = d$ (well, to be precise, it is written as $sa + tb = d$), which immediately gives $s = 10$ and $t = -43$ as a certificate of correctness for $\gcd(1386, 322) = 14$ (since we have previously checked that 14 is a common divisor of 1386 and 322).

Note that in Example 8 above, our method of finding s and t for $a = 1386$ and $b = 322$ was based on the calculations for the Euclidean Algorithm that we applied in Example 4 to determine $\gcd(1386, 322)$. All we did in Example 8 was to write out the equations of the form $a - qb = r$ for all stages of the Euclidean Algorithm except the final stage (when $r = 0$), and then use substitution to create an equation that was precisely of the form $as + bt = d$ for $d = \gcd(a, b)$.

REMARK

Though we won't prove it in these notes, the above procedure always works: Given a pair of integers $a \geq b > 0$, we can apply the Euclidean Algorithm followed by substitution of equations, to determine the greatest common divisor $d = \gcd(a, b)$ as well as a certificate of correctness s and t (i.e., the integers s and t are such that $as + bt = d$).

This leads to the following famous result, that we state without proof.

Lemma 6   **(Bézout's Lemma (BL))**

For all integers a and b, there exist integers s and t such that $as + bt = d$, where $d = \gcd(a, b)$.

Note that some mathematicians also refer to Proposition 6 as *Bézout's Identity*.

Note also that we have only described how to apply the Euclidean Algorithm to compute $\gcd(a, b)$ when $a \geq b > 0$, but the domain of Bézout's Lemma is not restricted to integers a and b with $a \geq b > 0$.

- To compute the gcd, what do we do for other choices of a and b? When $a < b$, then we simply interchange a and b, using the fact that $\gcd(a, b) = \gcd(b, a)$. When a or b or both are negative, then we simply replace them by their absolute value, using the fact that $\gcd(a, b) = \gcd(|a|, |b|)$. When one or both of them are equal to 0, we use the fact that $\gcd(a, 0) = a$ for all integers a.

- To find s and t, what do we do for other choices of a and b? In all cases, it is easy to find s and t, often by suitably adjusting signs, as illustrated in the following example.

Example 9   As we have shown above, for $a = 1386$ and $b = 322$, we have $\gcd(1386, 322) = 14$, and $s = 10$, $t = -43$, by equation (6.6). Consider the following variations for choices of a and b.

1. For $a = -1386$ and $b = 322$, we have

$$\gcd(-1386, 322) = \gcd(1386, 322) = 14.$$

We also have $s = -10$, $t = -43$, since we can rewrite (6.6) as

$$(-1386) \times (-10) + 322 \times (-43) = 14.$$

2. For $a = -322$ and $b = -1386$, we have

$$\gcd(-322, -1386) = \gcd(322, 1386) = \gcd(1386, 322) = 14.$$

We also have $s = 43$, $t = -10$, since we can rewrite (6.6) as

$$(-322) \times (43) + (-1386) \times (-10) = 14.$$

6.4   The Extended Euclidean Algorithm

In the previous section, we described how to use the Euclidean Algorithm to compute the greatest common divisor of a pair of integers, and to find a certificate of correctness for the gcd. We will not state that algorithm formally in these notes. Instead, in this section we will introduce and give a formal statement of the *Extended Euclidean Algorithm*.

Given a pair of integers $a \geq b > 0$, the Extended Euclidean Algorithm (EEA) is an efficient way to compute both $d = \gcd(a, b)$ and the pair of integers s and t that provide a certificate of correctness for d. The advantage of the EEA is that the computations for d, s and t are all carried out at the same time, with no substitution of equations required later.

We begin by defining some new notation for real numbers, that will be helpful in describing the EEA compactly.

Definition 6.4.1   The **floor** of a real number x, written $\lfloor x \rfloor$, is the largest integer less than or equal to x.

floor

Example 10   Consider the following examples of floor notation.

1. $\lfloor 9.713 \rfloor = \lfloor 9.001 \rfloor = \lfloor 9 \rfloor = 9$,

2. $\lfloor -9.713 \rfloor = \lfloor -9.001 \rfloor = \lfloor -10 \rfloor = -10$. For example, since the floor of x is the largest integer less than or equal to x, -9 cannot be the floor of -9.713 since $-9.713 < -9$.

3. $\left\lfloor \dfrac{7}{2} \right\rfloor = \left\lfloor \dfrac{18}{5} \right\rfloor = \left\lfloor \dfrac{18}{6} \right\rfloor = \left\lfloor \dfrac{1007}{329} \right\rfloor = 3.$

We first describe how to use the Extended Euclidean Algorithm to compute $\gcd(1386, 322)$ (which we have already computed by the Euclidean Algorithm in the previous section). Begin by creating four columns labelled x, y, r (for remainder) and q (for quotient). We construct a sequence of rows that will tell us the gcd and provide a certificate. For the i-th row we denote the entries by x_i, y_i, r_i and q_i. There is something very important to observe about the table. If we are computing $\gcd(a, b)$, then for each row of the table we have

$$ax_i + by_i = r_i,$$

which will turn out to be very useful for finding the certificate of correctness.

Assuming $a \geq b > 0$, the first two rows (with $i = 1, 2$) are always

x	y	r	q
1	0	a	0
0	1	b	0

so in our specific problem the first two rows are

x	y	r	q
1	0	1386	0
0	1	322	0

We construct each of the remaining rows by using the two preceding rows. To generate the i-th row we first compute the quotient q_i using the formula

$$q_i \leftarrow \left\lfloor \frac{r_{i-2}}{r_{i-1}} \right\rfloor.$$

Then we compute the other three entries x_i, y_i, r_i in the i-th row using the formula

$$\text{Row}_i = \text{Row}_{i-2} - q_i \text{Row}_{i-1},$$

where "Row$_i$" denotes the row vector (x_i, y_i, r_i).

For example, for the third row, using the above formulas with $i = 3$, we obtain

$$q_3 = \left\lfloor \frac{r_1}{r_2} \right\rfloor = \left\lfloor \frac{1386}{322} \right\rfloor = 4,$$

and

$$\text{Row}_3 = \text{Row}_1 - 4 \times \text{Row}_2 = (1, 0, 1386) - 4 \times (0, 1, 322) = (1, -4, 98).$$

Writing this in the table gives the first three rows as

	x	y	r	q
Row$_1$	1	0	1386	0
$-4 \times$ Row$_2$	0	1	322	0
$=$ Row$_3$	1	-4	98	4

In a similar fashion, for the fourth row, we use $i = 4$ to obtain

$$q_4 = \left\lfloor \frac{r_2}{r_3} \right\rfloor = \left\lfloor \frac{322}{98} \right\rfloor = 3,$$

and

$$\text{Row}_4 = \text{Row}_2 - 3 \times \text{Row}_3 = (0, 1, 322) - 3 \times (1, -4, 98) = (-3, 13, 28).$$

Writing this in the table gives the first four rows as

	x	y	r	q
	1	0	1386	0
Row_2	0	1	322	0
$-3 \times \text{Row}_3$	1	-4	98	4
$= \text{Row}_4$	-3	13	28	3

The completely worked out example follows.

x	y	r	q
1	0	1386	0
0	1	322	0
1	-4	98	4
-3	13	28	3
10	-43	14	3
-23	99	0	2

We stop when the remainder is 0. The second last row provides the desired d, s and t. The greatest common divisor d is the entry in the r column, s is the entry in the x column and t is the entry in the y column. Hence, $d = 14$, $s = 10$, $t = -43$ (as before), and we can check the conditions of the GCD Characterization Theorem to certify correctness: Since $14 \mid 1386$ and $14 \mid 322$ and $1386 \times 10 + 322 \times (-43) = 14$, we have confirmed that $14 = \gcd(1386, 322)$.

Here is a formal statement of the Extended Euclidean Algorithm.

Extended Euclidean Algorithm (EEA)

Input: Integers a, b with $a \geq b > 0$.

Initialize: Construct a table with four columns so that

- the columns are labelled x, y, r and q,

- the first row in the table is $(1, 0, a, 0)$,

- the second row in the table is $(0, 1, b, 0)$.

Repeat: For $i \geq 3$,

- $q_i \leftarrow \left\lfloor \dfrac{r_{i-2}}{r_{i-1}} \right\rfloor$

- $\text{Row}_i \leftarrow \text{Row}_{i-2} - q_i \text{Row}_{i-1}$

Stop: When $r_i = 0$.

Output: Set $n = i - 1$. Then $\gcd(a, b) = r_n$, and $s = x_n$ and $t = y_n$ are a certificate of correctness.

We do not prove that this algorithm works in these notes.

Example 11   Let $d = \gcd(2172, 423)$.

1. Apply EEA to compute d and give a certificate of correctness for d.

2. Determine $d_1 = \gcd(423, -2172)$ and give a certificate of correctness for d_1.

Solution:

1.

x	y	r	q
1	0	2172	0
0	1	423	0
1	-5	57	5
-7	36	24	7
15	-77	9	2
-37	190	6	2
52	-267	3	1
-141	724	0	2

From the table constructed by applying EEA above, we have determined that $n = 7$, and $d = \gcd(2172, 423) = r_7 = 3$. The certificate of correctness is $s = x_7 = 52$ and $t = y_7 = -267$, and indeed we check that

$$2172 \times (52) + 423 \times (-267) = 112,944 - 112,941 = 3. \tag{6.7}$$

2. We have $d_1 = \gcd(423, -2172) = \gcd(2172, 423) = 3$, from part 1 above. Our certificate of correctness is $s = -267$ and $t = -52$, since we can rewrite equation (6.7) as

$$423 \times (-267) + (-2172) \times (-52) = 3.$$

6.5   Further Properties of the Greatest Common Divisor

In this section, we prove various properties of the greatest common divisor. Repeated use is made of Bézout's Lemma, GCD Characterization Theorem and Divisibility of Integer Combinations.

Proposition 7   **(Common Divisor Divides GCD (CDD GCD))**

For all integers a, b and c, if $c \mid a$ and $c \mid b$, then $c \mid \gcd(a, b)$.

Proof: Let a, b and c be arbitrary integers, and assume that $c \mid a$ and $c \mid b$. By Bézout's Lemma, there exist integers s and t such that $as + bt = d$, where $d = \gcd(a, b)$. Now, since $c \mid a$ and $c \mid b$, by Divisibility of Integer Combinations we get $c \mid (as + bt)$, and so $c \mid d$.   □

Example 12   Prove the following statement. For all integers a and b, and non-negative integers c, we have $\gcd(ca, cb) = c \gcd(a, b)$.

Solution: Let a and b be arbitrary integers, and let c be an arbitrary non-negative integer.

Let $d = \gcd(a, b)$, so we have $d \mid a$ and $d \mid b$, and therefore there exist integers k and m such that $a = kd$ and $b = md$. Multiplying these equations by c, we obtain $ca = ckd = (cd)k$ and $cb = cmd = (cd)m$, and hence, by the definition of divisibility, we have $cd \mid ca$ and $cd \mid cb$.

Also, by Bézout's Lemma, there exist integers s and t such that $as + bt = d$, and multiplying this equation by c gives
$$(ca)s + (cb)t = cd.$$
Now, we have $d \geq 0$, and we have $c \geq 0$ by hypothesis, which gives $cd \geq 0$. We conclude from the GCD Characterization Theorem that $\gcd(ca, cb) = cd$.

We continue by introducing a special name for pairs of integers whose greatest common divisor is equal to 1.

Definition 6.5.1

coprime

Two integers a and b are **coprime** if $\gcd(a, b) = 1$.

Example 13   Note that $\gcd(5, 7) = 1$, so 5 and 7 are coprime. Also, $\gcd(11, 17) = 1$, so 11 and 17 are coprime. In these cases, the numbers 5, 7, 11 and 17 are all prime numbers (which we consider in detail starting in the next section). However, note also that $\gcd(18, 35) = 1$, so 18 and 35 are coprime, but of course neither 18 nor 35 is a prime number.

We start our study of coprime pairs of integers with a simple but important result.

Theorem 8   **(Coprimeness Characterization Theorem (CCT))**

For all integers a and b, $\gcd(a, b) = 1$ if and only if there exist integers s and t such that $as + bt = 1$.

Proof: Let a and b be arbitrary integers. We prove two implications.

Assume that $\gcd(a, b) = 1$. Then, by Bézout's Lemma, there exist integers s and t so that $as + bt = 1$.

Assume that there exist integers s and t such that $as + bt = 1$. Now, $1 \mid a$ and $1 \mid b$ for all integers a and b, so from the GCD Characterization Theorem, we have $\gcd(a, b) = 1$.   □

In the following example, we show how the Coprimeness Characterization Theorem might be used in a proof.

Example 14   Prove the following proposition. For all integers a and b, and all natural numbers n, if $\gcd(a, b) = 1$, then $\gcd(a, b^n) = 1$.

Solution: Let a and b be arbitrary integers. We prove this result by induction on n, where $P(n)$ is the statement

$$\text{If } \gcd(a, b) = 1, \text{ then } \gcd(a, b^n) = 1.$$

Base Case The statement

$$\text{If } \gcd(a, b) = 1, \text{ then } \gcd(a, b) = 1.$$

is immediately true, since the hypothesis and conclusion are identical, proving $P(1)$.

Inductive Step Let k be an arbitrary natural number. Assume the inductive hypothesis, $P(k)$. That is, we assume

$$\text{If } \gcd(a, b) = 1, \text{ then } \gcd(a, b^k) = 1.$$

We wish to prove $P(k+1)$, which is the statement

$$\text{If } \gcd(a, b) = 1, \text{ then } \gcd(a, b^{k+1}) = 1.$$

Hence assume $\gcd(a, b) = 1$, and then by the Coprimeness Characterization Theorem (or by Bézout's Lemma), there exist integers s and t such that

$$as + bt = 1.$$

Since $\gcd(a, b^k) = 1$ from the inductive hypothesis, we can also use the Coprimeness Characterization Theorem (or Bézout's Lemma) again, to give that there exist integers u and v such that

$$au + b^k v = 1.$$

Multiplying these two equations together gives

$$a^2 su + ab^k sv + abtu + b^{k+1} tv = 1,$$

which we can rewrite as

$$a\big(asu + b^k sv + btu\big) + b^{k+1}(tv) = 1.$$

Let $e = asu + b^k sv + btu$ and $f = tv$. Then from the previous equation e and f are integers such that $ae + b^{k+1}f = 1$, and hence we deduce from the Coprimeness Characterization Theorem that $\gcd(a, b^{k+1}) = 1$. The result is true for $n = k+1$, and hence holds for all integers $n \geq 1$ by the Principle of Mathematical Induction.

Now we prove a result that is closely related to the Coprimeness Characterization Theorem.

Proposition 9   **(Division by the GCD (DB GCD))**

For all integers a and b, not both zero, $\gcd\left(\dfrac{a}{d}, \dfrac{b}{d}\right) = 1$, where $d = \gcd(a, b)$.

Proof: Let a and b be arbitrary integers, not both zero. By Bézout's Lemma, there exist integers s and t so that $as + bt = d$. Now a and b are not both zero, so $d > 0$, and hence we can divide the equation $as + bt = d$ on both sides by d, to obtain

$$\frac{a}{d}s + \frac{b}{d}t = 1.$$

Since $d = \gcd(a, b)$, we have $d \mid a$ and $d \mid b$, so by the definition of divisibility, $\dfrac{a}{d}$ and $\dfrac{b}{d}$ are integers. Hence, from the CCT, we obtain $\gcd\left(\dfrac{a}{d}, \dfrac{b}{d}\right) = 1$.   □

In the following example, we show how Division by the GCD can be applied to give a concise proof of a classical result about rational numbers.

Example 15   Prove the following proposition. For all rational numbers r, there exist coprime integers p and q, with $q \neq 0$, such that $r = \dfrac{p}{q}$.

Solution: Let r be an arbitrary rational number. Then we can write r in the form $r = \dfrac{a}{b}$ where a and b are integers with $b \neq 0$. Let $d = \gcd(a, b)$. Since $b \neq 0$, then a and b are not both zero, and so $d > 0$. Also, we have $d \mid a$ and $d \mid b$, so by the definition of divisibility, $p = \dfrac{a}{d}$ and $q = \dfrac{b}{d}$ are both integers. Hence, from Division by the GCD we have $\gcd(p, q) = 1$, and of course $r = \dfrac{a}{b} = \dfrac{p}{q}$, where we have divided both numerator a and denominator b by the non-zero integer d.

REMARK

One application of the proposition in Example 15 above is to our previous proof that $\sqrt{2}$ is irrational (Proposition 10 on page 57). In equation (3.9) of that proof by contradiction, we expressed $\sqrt{2}$ as the ratio a/b. Then we proved that a and b are not both even, that is, a and b do not have a common factor of 2. The proposition above says that we can express $\sqrt{2}$ as the ratio a/b where a and b are *coprime*, so of course a and b are not both even. Hence the proposition in Example 15 would allow us to omit the portion of the proof of Proposition 10 where we proved that a and b are not both even.

To end this section, we now prove a simple result about coprime pairs that will be important for the study of *prime* numbers in the next section.

Proposition 10   **(Coprimeness and Divisibility (CAD))**

For all integers a, b and c, if $c \mid ab$ and $\gcd(a, c) = 1$, then $c \mid b$.

Proof: Let a, b and c be arbitrary integers, and assume that $c \mid ab$ and $\gcd(a, c) = 1$. Since $\gcd(a, c) = 1$, by the Coprimeness Characterization Theorem (or by Bézout's Lemma) there exist integers s and t such that $as + ct = 1$. Multiplying this equation by b gives

$$abs + cbt = b. \tag{6.8}$$

Now we have $c \mid ab$ from the hypothesis, and $c \mid c$ from the definition of divisibility, so by the Divisibility of Integer Combinations, we have $c \mid \big((ab)s + c(bt)\big)$. Hence from equation (6.8) we obtain $c \mid b$. $\qquad\square$

6.6   Prime Numbers

In this section, we consider a subset of the natural numbers called the prime numbers.

Definition 6.6.1
prime, composite

A natural number $p > 1$ is called **prime** (or a *prime number*) if its only positive divisors are 1 and p itself. Otherwise, p is called **composite**.

Example 16

The integers $2, 3, 5$ and 7 are primes. The integers $4 = 2 \times 2$, $6 = 2 \times 3$, $8 = 2 \times 2 \times 2$ and $9 = 3 \times 3$ are composite. Note, that by definition, 1 is not a prime and 1 is not composite.

We begin our study of prime numbers with a pair of famous theorems.

Proposition 11

(Prime Factorization (PF))

Every natural number $n > 1$ can be written as a product of primes.

Before proving this result, we consider some examples to make the meaning clear.

Example 17

The integers $2, 3, 5$ and 7 are primes and mathematicians use the convention that a number by itself is a product (with a single factor). The integers $4 = 2 \times 2$, $6 = 2 \times 3$, $8 = 2 \times 2 \times 2$ and $9 = 3 \times 3$ have been written as products of primes.

Proof: We prove this result by strong induction on n, where $P(n)$ is the statement

The natural number n can be written as a product of primes.

Base Case The statement $P(2)$ is true, since 2 is a prime.

Inductive Step Let k be an arbitrary integer where $k \geq 2$.

Assume the inductive hypothesis, $P(2) \wedge P(3) \wedge \ldots \wedge P(k)$. That is, we assume that the natural number i can be written as a product of primes, for all integers $i = 2, 3, \ldots, k$.

We wish to prove $P(k + 1)$, which says that the natural number $k + 1$ can be written as a product of primes. Now the integer $k + 1$ is either prime or composite, and we prove the inductive conclusion separately for these two cases.

Case 1: If $k+1$ is a prime, then $P(k+1)$ is immediately true.

Case 2: If $k+1$ is composite, then we can write $k+1 = rs$, where r and s are natural numbers such that $2 \le r < k+1$ and $2 \le s < k+1$. Therefore, by the inductive hypothesis, r and s can be written as products of primes, and hence $k+1 = rs$ can be written as a product of primes, proving $P(k+1)$ in this case also.

The result is true for $n = k+1$, and hence holds for all $n \ge 2$ by the Principle of Strong Induction. □

The second of our famous theorems was recorded by Euclid. The proof that we give is by contradiction.

Theorem 12 **(Euclid's Theorem (ET))**

The number of primes is infinite.

Proof: Assume, for the sake of contradiction, that the number of primes is finite, say n, for some $n \in \mathbb{N}$. Let the primes be denoted by $p_1, p_2, p_3, \ldots, p_n$, and consider the integer

$$N = p_1 p_2 p_3 \cdots p_n + 1.$$

Since $N > p_i$ for all i, then N cannot equal any of the p_i's, and hence N is not a prime. Now $N = q_i\, p_i + 1$ where q_i is an integer, for each of the primes p_i. From the Division Algorithm, this means that we have remainder 1 when dividing N by p_i. But $1 \ne 0$, and $1 < p_i$ since $p_i \ge 2$, and hence p_i doesn't divide N for all i. This implies that N cannot be written as a product of primes. This is a contradiction, since by the proposition Prime Factorization, N can be written as a product of primes. As a result, the statement "The number of primes is infinite" must be true. □

6.7   The Unique Factorization Theorem

In the previous section, we proved the proposition Prime Factorization, that every natural number $n \ge 2$ can be written as a product of primes. In this section, we consider products of primes in more detail.

The following implication will play an important role. We prove this result as a simple consequence of Coprimeness and Divisibility, using the "method of elimination" described on page 56.

Lemma 13 **(Euclid's Lemma (EL))**

For all integers a and b, and prime numbers p, if $p \mid ab$, then $p \mid a$ or $p \mid b$.

Proof: Let a and b be arbitrary integers, and p be an arbitrary prime number. Using the logical equivalence (3.8) on page 56, we prove the implication:

If $p \mid ab$ and $p \nmid a$, then $p \mid b$.

Hence assume that $p \mid ab$ and $p \nmid a$. Now since p is prime, the only positive divisors of p are 1 and p, so the only possible values of $\gcd(a,p)$ are 1 and p. However, $\gcd(a,p) \neq p$ since $p \nmid a$, so we conclude that $\gcd(a,p) = 1$.

Since $p \mid ab$ and $\gcd(a,p) = 1$, we conclude from Coprimeness and Divisibility (with $c = p$) that $p \mid b$. $\square$

We will not prove the following generalization of Euclid's Lemma, in which the product of two factors ab in the hypothesis is replaced by a product of n factors $a_1 a_2 \cdots a_n$ for an arbitrary natural number n. To prove this result, apply Mathematical Induction and Euclid's Lemma in a straightforward way.

Proposition 14   Let p be a prime number, n be a natural number, and a_1, a_2, ..., a_n be integers. If $p \mid (a_1 a_2 \cdots a_n)$, then $p \mid a_i$ for some $i = 1, 2, \ldots, n$.

Since multiplication of integers is commutative, the product of primes in a prime factorizations can be written in any order. For example $12 = 2 \times 2 \times 3 = 2 \times 3 \times 2 = 3 \times 2 \times 2$. However, up to the order of the factors, the factorization of integers into primes is unique. We prove this fact in the following result. We call it the Unique Factorization Theorem here, but mathematicians also refer to it as the *Fundamental Theorem of Arithmetic*.

Theorem 15   **(Unique Factorization Theorem (UFT))**

Every natural number $n > 1$ can be written as a product of prime factors uniquely, apart from the order of factors.

Proof: We have already proved, in Proposition 11 (Prime Factorization), that every natural number $n > 1$ can be written as a product of prime factors. All that remains to prove is that apart from the order of factors, this factorization is unique.

We prove this result by strong induction on n, where $P(n)$ is the statement

> The natural number n can be written as a product of prime factors uniquely, apart from the order of the factors.

Base Case The statement $P(2)$ is true, since 2 is a prime (recall that a single prime number is considered a product, with a single factor).

Inductive Step Let k be an arbitrary integer where $k \geq 2$.

Assume the inductive hypothesis, $P(2) \wedge P(3) \wedge \ldots \wedge P(k)$. That is, we assume that the natural number i can be written as a product of prime factors uniquely, apart from the order of the factors, for all integers
$i = 2, 3, \ldots, k$.

We wish to prove $P(k + 1)$, which says that the natural number $k + 1$ can be written as a product of prime factors uniquely, apart from the order of the factors. Now the integer $k + 1$ is either prime or composite, and we prove the inductive conclusion separately for these two cases.

Case 1: If $k + 1$ is a prime, then $P(k + 1)$ is immediately true.

Case 2: If $k+1$ is composite, then we already know from Prime Factorization that it can be written as a product of primes. Now suppose that $k+1$ can be written as a product of primes with factors $p_1, p_2, \ldots, p_j$, and as a product of primes with factors $q_1, q_2, \ldots, q_\ell$, so we have

$$k+1 = p_1 p_2 \cdots p_j = q_1 q_2 \cdots q_\ell. \tag{6.9}$$

Since $k+1$ is composite, we must have $j, \ell \geq 2$.

Now we have $p_1 \mid (k+1)$, so from (6.9) we obtain $p_1 \mid (q_1 q_2 \ldots q_\ell)$. Hence by Proposition 14, $p_1 \mid q_i$ for some $1 \leq i \leq \ell$. If necessary, reorder the q_i's so that $p_1 \mid q_1$. Since p_1 and q_1 are both primes, this means that $p_1 = q_1$, and thus dividing equation 6.9 by $p_1 = q_1$, we obtain

$$m = p_2 \cdots p_j = q_2 \cdots q_\ell, \tag{6.10}$$

where, for convenience, we have defined $m = (k+1)/p_1$. Now $p_1 \geq 2$, so $m < k+1$, and $p_2 \cdots p_j \geq 2$ since $j \geq 2$, so $m \geq 2$. Thus we have $2 \leq m \leq k$, so by the inductive hypothesis, m can be written as a product of prime factors uniquely, apart from the order of the factors. But this means that in (6.10), the two factorizations $p_2 \cdots p_j$ and $q_2 \cdots q_\ell$ of m must be identical, apart from the order of the factors (i.e., $j = \ell$, and the two lists of primes $p_2, \ldots, p_j$ and $q_2, \ldots, q_\ell$ are the same, except possibly for the order of the primes in the list). Then, since $p_1 = q_1$, this means that the two factorizations $p_1 p_2 \cdots p_j$ and $q_1 q_2 \cdots q_\ell$ of $k+1$ must be identical, hence proving $P(k+1)$ in this case also.

The result is true for $n = k+1$, and hence holds for all $n \geq 2$ by the Principle of Strong Induction. $\qquad \square$

The Unique Factorization Theorem guarantees that the prime factors of a positive integer n are unique, but does not provide an algorithm for finding them. The next proposition shows that to find the prime factors of n, instead of checking *all* of the prime numbers less than n, we only have to check those less than or equal to the square root of n.

Proposition 16   **(Finding a Prime Factor (FPF))**

Every natural number $n > 1$ is either prime or has a prime factor less than or equal to $\sqrt{n}$.

Proof: Suppose that $n > 1$ is not prime. Then by the above proposition UFT (or PF), we can write $n = p_1 p_2 \ldots p_k$, where $k \geq 2$, and $p_1, p_2, \ldots, p_k$ are primes. Let p be the smallest prime factor of n, so we have $p \leq p_i$ for $i = 1, 2, \ldots, k$. Hence we obtain

$$p^2 \leq p^k \leq p_1 p_2 \ldots p_k = n,$$

and therefore, taking positive square roots, we have $p \leq \sqrt{n}$. $\qquad \square$

Example 18   Determine whether 73 is a prime number or not.

Solution: From the above result Finding a Prime Factor, either 73 is a prime, or it has a prime factor less than or equal to $\sqrt{73}$. Now $\sqrt{73} < \sqrt{81} = 9$. So either 73 is prime or it has a prime factor less than or equal to 8. The only primes less than or equal to 8 are 2, 3, 5 and 7. Since none of these divide 73, we conclude that 73 must be a prime number.

6.8   Prime Factorizations and the Greatest Common Divisor

Suppose that $p_1, p_2, \ldots, p_k$, $k \geq 1$, are all the distinct prime divisors of some natural number $n \geq 2$. Then the Unique Factorization Theorem tells us that we can represent n uniquely as the product

$$n = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_k^{\alpha_k},$$

where $\alpha_1, \alpha_2, \ldots, \alpha_k$ are positive integers. For example, the prime divisors of 80262 are $2, 3, 7$ and 13, and we have $80262 = 2^1 3^2 7^3 13^1$.

More generally, suppose that $n \geq 1$, and the list of distinct primes $p_1, p_2, \ldots, p_m$, $m \geq 1$, includes all the prime divisors of n, but possibly also includes primes that don't divide n. Then we can represent n as the product

$$n = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_m^{\alpha_m}, \tag{6.11}$$

where $\alpha_1, \alpha_2, \ldots, \alpha_m$ are now *non-negative* integers. Using our previous example, we can write $80262 = 2^1 3^2 5^0 7^3 13^1 19^0$. Note that the product above gives a representation of $n = 1$ for any distinct primes $p_1, p_2 \ldots, p_m$, using $\alpha_i = 0$ for all $i = 1, 2, \ldots, m$.

The next proposition uses this representation to characterize the positive divisors of a natural number.

Proposition 17

(Divisors From Prime Factorization (DFPF))

Let n and c be positive integers, and let

$$n = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_k^{\alpha_k}$$

be a way to express n as a product of the distinct primes $p_1, p_2, \ldots, p_k$, where some or all of the exponents may be zero. The integer c is a positive divisor of n if and only if c can be represented as a product

$$c = p_1^{\beta_1} p_2^{\beta_2} \cdots p_k^{\beta_k}, \text{ where } 0 \leq \beta_i \leq \alpha_i \text{ for } i = 1, 2, \ldots, k.$$

Proof: Let n be an arbitrary positive integer, and assume that $n = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_k^{\alpha_k}$, where $p_1, p_2, \ldots, p_k$ are distinct primes, and $\alpha_1, \alpha_2, \ldots, \alpha_k$ are non-negative integers. Also let c be an arbitrary positive integer. To prove the if and only if statement, we prove both implications.

For one implication, assume that c can be represented as the product

$$c = p_1^{\beta_1} p_2^{\beta_2} \cdots p_k^{\beta_k}, \text{ where } 0 \leq \beta_i \leq \alpha_i \text{ for } i = 1, 2, \ldots, k.$$

Then we have

$$\frac{n}{c} = \frac{p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_k^{\alpha_k}}{p_1^{\beta_1} p_2^{\beta_2} \cdots p_k^{\beta_k}} = p_1^{\alpha_1 - \beta_1} p_2^{\alpha_2 - \beta_2} \cdots p_k^{\alpha_k - \beta_k},$$

is a positive integer, since $\alpha_i - \beta_i \geq 0$ for $i = 1, 2, \ldots, k$. Hence we conclude that $c \mid n$.

For the other implication, assume that c is a positive divisor of n. Let p be an arbitrary prime divisor of c. Since $p \mid c$ and $c \mid n$, by the proposition Transitivity of Divisibility, we obtain $p \mid n$, so p is a prime divisor of n, and therefore $p = p_i$ for some $i = 1, 2, \ldots, k$. Hence

we have proved that the list of distinct primes $p_1, p_2, \ldots, p_k$ includes all the prime divisors of c, but possibly also includes primes that don't divide c. Therefore, as in (6.11) above, we can represent c as the product

$$c = p_1^{\beta_1} p_2^{\beta_2} \cdots p_k^{\beta_k},$$

where $\beta_1, \beta_2, \ldots, \beta_k$ are non-negative integers.

We now prove that $\beta_i \le \alpha_i$ for all $i = 1, 2, \ldots, k$. Assume, for the sake of contradiction, that $\beta_j > \alpha_j$ for some $j = 1, 2, \ldots, k$, and let $m = n/p_j^{\alpha_j}$ and $f = c/p_j^{\alpha_j}$, so we have

$$m = p_1^{\alpha_1} \cdots p_{j-1}^{\alpha_{j-1}} p_{j+1}^{\alpha_{j+1}} \cdots p_k^{\alpha_k}, \qquad f = p_1^{\beta_1} \cdots p_{j-1}^{\beta_{j-1}} p_j^{\beta_j - \alpha_j} p_{j+1}^{\beta_{j+1}} \cdots p_k^{\beta_k}. \qquad (6.12)$$

Note that all exponents in the products above are non-negative integers, so m and f are both integers.

Since $c \mid n$, there exists an integer t so that $n = tc$. Dividing both sides of this equation by $p_j^{\alpha_j}$, we obtain $n/p_j^{\alpha_j} = tc/p_j^{\alpha_j}$. Using the notation above, this equation can be written as $m = tf$, which gives $f \mid m$. Now $\beta_j > \alpha_j$ means that $\beta_j - \alpha_j$ is a positive integer, so we have $p_j \mid p_j^{\beta_j - \alpha_j}$, and from (6.12) we also have $p_j^{\beta_j - \alpha_j} \mid f$. Hence, by the proposition Transitivity of Divisibility, we obtain $p_j \mid f$. Now $f \mid m$, so again by the proposition Transitivity of Divisibility, we obtain $p_j \mid m$. But from (6.12), the list of distinct primes that divide m does not contain p_j; in other words, $p_j \nmid m$, which is a contradiction. As a result, we conclude that $\beta_i \le \alpha_i$ for $i = 1, 2, \ldots, k$, and since $0 \le \beta_i$ for $i = 1, 2, \ldots, k$, we have $0 \le \beta_i \le \alpha_i$ for $i = 1, 2, \ldots, k$. $\qquad \square$

Example 19

Find all positive divisors of 72.

Solution: Observe that

$$72 = 2^3 3^2.$$

Therefore, by the proposition Divisors From Prime Factorization, the positive divisors of 72 are integers of the form

$$c = 2^{\beta_1} 3^{\beta_2} \text{ where } 0 \le \beta_1 \le 3 \text{ and } 0 \le \beta_2 \le 2.$$

Hence the positive divisors are given by

$$\begin{array}{llll}
2^0 3^0 = 1, & 2^1 3^0 = 2, & 2^2 3^0 = 4, & 2^3 3^0 = 8, \\
2^0 3^1 = 3, & 2^1 3^1 = 6, & 2^2 3^1 = 12, & 2^3 3^1 = 24, \\
2^0 3^2 = 9, & 2^1 3^2 = 18, & 2^2 3^2 = 36, & 2^3 3^2 = 72.
\end{array}$$

EXERCISE

Prove that the number of positive divisors of an integer n with unique prime factorization

$$n = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_k^{\alpha_k},$$

is given by the product $(\alpha_1 + 1)(\alpha_2 + 1) \cdots (\alpha_k + 1)$.

Now that we have a method to obtain all the positive divisors of any positive integer using its prime factorization, we consider how to obtain the GCD of a pair of positive integers, say a and b, using their prime factorizations. First, note that each of a and b has only a finite number of distinct prime divisors, so we can produce a large enough list of distinct primes, say $p_1, p_2, \ldots, p_k$, that contains all the distinct prime divisors of both a and b together. This provides us with ways to write both a and b, respectively, in terms of $p_1, p_2, \ldots, p_k$ as

$$a = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_k^{\alpha_k}, \quad \text{and} \quad b = p_1^{\beta_1} p_2^{\beta_2} \cdots p_k^{\beta_k},$$

where $\alpha_1, \alpha_2, \ldots, \alpha_k$ and $\beta_1, \beta_2, \ldots, \beta_k$ are non-negative integers.

In the following result we give a formula for the GCD of a and b, based on their prime factorizations. The notation "$\min\{x, y\}$" denotes the *minimum* of x and y.

Proposition 18   **(GCD From Prime Factorization (GCD PF))**

Let a and b be positive integers, and let

$$a = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_k^{\alpha_k}, \quad \text{and} \quad b = p_1^{\beta_1} p_2^{\beta_2} \cdots p_k^{\beta_k},$$

be ways to express a and b as products of the distinct primes $p_1, p_2, \ldots, p_k$, where some or all of the exponents may be zero. We have

$$\gcd(a, b) = p_1^{\gamma_1} p_2^{\gamma_2} \cdots p_k^{\gamma_k} \text{ where } \gamma_i = \min\{\alpha_i, \beta_i\} \text{ for } i = 1, 2, \ldots, k.$$

Proof: Let a and b be arbitrary positive integers, where

$$a = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_k^{\alpha_k}, \quad \text{and} \quad b = p_1^{\beta_1} p_2^{\beta_2} \cdots p_k^{\beta_k},$$

for distinct primes $p_1, p_2, \ldots, p_k$, with $\alpha_i \geq 0$, $\beta_i \geq 0$ for $i = 1, 2, \ldots, k$.

Let $d = p_1^{\gamma_1} p_2^{\gamma_2} \cdots p_k^{\gamma_k}$ where $\gamma_i = \min\{\alpha_i, \beta_i\}$ for $i = 1, 2, \ldots, k$. Then $d > 0$, and since $\min\{\alpha_i, \beta_i\} \leq \alpha_i$, as well as $\min\{\alpha_i, \beta_i\} \leq \beta_i$, for each $i = 1, 2, \ldots, k$, by the proposition Divisors From Prime Factorization (DFPF), we obtain $d \mid a$ and $d \mid b$.

Suppose that $c \in \mathbb{Z}$ is such that $c \mid a$ and $c \mid b$. Then by DFPF, c is of the form

$$c = p_1^{\nu_1} p_2^{\nu_2} \cdots p_k^{\nu_k}, \quad \text{or} \quad c = -p_1^{\nu_1} p_2^{\nu_2} \cdots p_k^{\nu_k},$$

where $0 \leq \nu_i \leq \alpha_i$ (since $c \mid a$) and $0 \leq \nu_i \leq \beta_i$ (since $c \mid b$), for $i = 1, 2, \ldots, k$. But this means that $0 \leq \nu_i \leq \min\{\alpha_i, \beta_i\} = d_i$, for $i = 1, 2, \ldots, k$, and hence by DFPF we have that $c \mid d$. By the proposition Bounds By Divisibility, we then obtain $c \leq |d| = d$.

We conclude, from the definition of GCD, that $d = \gcd(a, b)$.   □

Example 20   Determine $\gcd(24750, 434511)$ by using prime factorizations.

Solution: Observe that

$$24750 = 2^1 3^2 5^3 11^1 = 2^1 3^2 5^3 7^0 11^1 19^0, \quad \text{and} \quad 434511 = 3^3 7^1 11^2 19^1 = 2^0 3^3 5^0 7^1 11^2 19^1.$$

Then from the proposition GCD From Prime Factorization, we obtain

$$\gcd(24750, 434511) = 2^{\min\{1,0\}} 3^{\min\{2,3\}} 5^{\min\{3,0\}} 7^{\min\{0,1\}} 11^{\min\{1,2\}} 19^{\min\{0,1\}}$$
$$= 2^0 3^2 5^0 7^0 11^1 19^0 = 99.$$

Note that finding the GCD by using prime factorizations works well on small numbers, as in the example above. However, for large numbers, finding prime factors is a difficult computational problem, as we shall see in Chapter 9, where we consider RSA encryption. This means that for larger numbers, finding the GCD by using prime factorizations is much slower and less efficient than using the Extended Euclidean Algorithm.

Chapter 7

Linear Diophantine Equations

7.1 The Existence of Solutions in Two Variables

In high school, you studied how to find all real solutions to systems of linear equations. However, there are many applications where we only want to find *integer* solutions. For example, a variable might represent the number of trucks to use when shipping a product over large distances (of course, in the truck example, the variable would additionally be restricted to non-negative values only).

Definition 7.1.1

Diophantine equation, linear

An equation in which both the coefficients and variables are integers, is called a **Diophantine equation**, named after the Greek mathematician Diophantus of Alexandria. A Diophantine equation is called **linear** if each term in the equation is a constant or a constant times a single variable.

In one variable x, we have the simplest linear Diophantine equation

$$ax = b.$$

To emphasize, here a and b are given integers, and we want to find all integer values of x such that $ax = b$. Suppose that a is non-zero. By the definition of divisibility, we know that ax is divisible by a for all integers x, so this equation only has an integer solution x if a divides b. Also, if a divides b, then this equation does have an integer solution, given by $x = \frac{b}{a}$, and this is the only solution.

We now consider what happens for linear Diophantine equations in two variables x and y. As we shall see, divisibility will again play a key role in finding solutions.

Theorem 1

(Linear Diophantine Equation Theorem, Part 1 (LDET 1))

For all integers a, b and c, with a and b both not zero, the linear Diophantine equation

$$ax + by = c$$

(in variables x and y) has an integer solution if and only if $d \mid c$, where $d = \gcd(a, b)$.

Proof: Let a, b and c be arbitrary integers, with a and b both not zero. First, assume that the linear Diophantine equation $ax + by = c$ has an integer solution, say $x = x_0$, $y = y_0$. Therefore we have $ax_0 + by_0 = c$. Since $d = \gcd(a, b)$, we have $d \mid a$ and $d \mid b$. Hence by the Divisibility of Integer Combinations, $d \mid (ax_0 + by_0)$, which gives $d \mid c$.

Conversely, assume that $d \mid c$. Then by the definition of divisibility there exists an integer k such that $c = kd$. Now, by Bézout's Lemma, there exist integers s and t such that

$$as + bt = d.$$

Multiplying this equation on both sides by k gives

$$a(ks) + b(kt) = c,$$

since $kd = c$. Hence $x = ks$ and $y = kt$ is an integer solution to $ax + by = c$.   □

Example 1   Which of the following linear Diophantine equations has a solution?

1. $33x + 18y = 23$

2. $33x + 18y = 24$

Solution: Note that $\gcd(33, 18) = 3$.

1. Since 3 does not divide 23, this equation has no integer solution.

2. Since 3 divides 24, this equation has an integer solution.

REMARK

How do we find an integer solution to $ax + by = c$ when $d \mid c$, where $d = \gcd(a, b)$? Here are two simple steps that will allow us to find a solution.

1. Find a certificate of correctness s and t for $d = \gcd(a, b)$, which gives

$$as + bt = d. \tag{7.1}$$

2. Multiply Equation 7.1 by $k = \dfrac{c}{d}$ to get $a(ks) + b(kt) = kd = c$. This gives the solution $x = ks$ and $y = kt$.

Example 2   Determine whether the Diophantine equation $84x + 35y = 63$ has a solution, and, if so, find a solution.

Solution:

x	y	r	q
1	0	84	0
0	1	35	0
1	−2	14	2
−2	5	7	2
5	−12	0	2

From the table constructed by applying EEA above, we have determined that $n = 4$, and hence $d = \gcd(84, 35) = r_4 = 7$. Since 7 divides 63 (we have $63 = 9 \times 7$), then from the Linear Diophantine Equation Theorem, Part 1, we conclude that this equation has a solution.

Continuing the two steps described above, the certificate of correctness we obtain from the EEA table is $s = x_4 = -2$ and $t = y_4 = 5$, so we have

$$84 \times (-2) + 35 \times 5 = 7.$$

Multiplying this equation by $\dfrac{c}{d} = \dfrac{63}{7} = 9$ gives

$$84 \times (-18) + 35 \times (45) = 63,$$

so one integer solution is $x = -18$ and $y = 45$.

7.2   Finding All Solutions in Two Variables

By the Linear Diophantine Equation Theorem, Part 1, we know whether a linear Diophantine equation in two variables has any solutions. Moreover, if there are solutions, it tells us how to construct one particular solution.

The next result can be regarded as a companion result to LDET 1, since it tells us how to construct *all* solutions, given one particular solution.

Theorem 2   **(Linear Diophantine Equation Theorem, Part 2, (LDET 2))**

Let a, b and c be integers with a and b both not zero, and define $d = \gcd(a, b)$. If $x = x_0$ and $y = y_0$ is one particular integer solution to the linear Diophantine equation $ax + by = c$, then the set of all solutions is given by

$$\left\{ (x, y) : x = x_0 + \tfrac{b}{d}n, \ y = y_0 - \tfrac{a}{d}n, \ n \in \mathbb{Z} \right\}.$$

Proof: Let a, b and c be arbitrary integers, with a and b both not zero, and let $d = \gcd(a, b)$. Define the two sets

$$A = \left\{ (x, y) : x = x_0 + \tfrac{b}{d}n, \ y = y_0 - \tfrac{a}{d}n, \ n \in \mathbb{Z} \right\},$$
$$B = \{ (x, y) : x, y \in \mathbb{Z}, \ ax + by = c \}.$$

We will prove that $A = B$ by proving both that $A \subseteq B$ and $B \subseteq A$, using universally quantified implications as described in the proof method for $S \subseteq T$ on page 89.

To prove $A \subseteq B$: We prove the implication

For all integers x and y, if $(x, y) \in A$, then $(x, y) \in B$.

Hence, let n be an arbitrary integer, and assume that $(x, y) = \left(x_0 + \frac{b}{d}n, y_0 - \frac{a}{d}n\right) \in A$. Then we have

$$ax + by = a\left(x_0 + \frac{b}{d}n\right) + b\left(y_0 - \frac{a}{d}n\right)$$

$$= ax_0 + by_0 + \frac{ab}{d}n - \frac{ab}{d}n$$

$$= ax_0 + by_0$$

$$= c,$$

where the last equality follows since $(x, y) = (x_0, y_0)$ is a solution to the equation $ax + by = c$. Thus we conclude that

$$(x, y) = \left(x_0 + \frac{b}{d}n, y_0 - \frac{a}{d}n\right) \in B.$$

To prove $B \subseteq A$: We prove the implication

For all integers x and y, if $(x, y) \in B$, then $(x, y) \in A$.

Hence, assume that $(x, y) \in B$, so we have

$$ax + by = c. \tag{7.2}$$

Now we also have

$$ax_0 + by_0 = c, \tag{7.3}$$

and subtracting equation (7.3) from (7.2) and rearranging gives the equation

$$a(x - x_0) = -b(y - y_0).$$

Dividing this equation by d (we can do this since a and b are both not zero, so $d > 0$ and hence $d \neq 0$) gives

$$\frac{a}{d}(x - x_0) = -\frac{b}{d}(y - y_0), \tag{7.4}$$

where $\frac{a}{d}$ and $\frac{b}{d}$ are both integers since $d = \gcd(a, b)$. From Division by the GCD, we know that $\frac{a}{d}$ and $\frac{b}{d}$ are coprime, and from equation (7.4) we know that $\frac{b}{d}$ divides $\frac{a}{d}(x - x_0)$. Hence from Coprimeness and Divisibility it follows that

$$\frac{b}{d} \,\bigg|\, (x - x_0),$$

and therefore, by the definition of divisibility, there exists an integer k such that $x - x_0 = \frac{b}{d}k$. Thus we have $x = x_0 + \frac{b}{d}k$, and substituting this into equation (7.4) yields $y = y_0 - \frac{a}{d}k$ (since $b \neq 0$). We conclude that

$$(x, y) = \left(x_0 + \frac{b}{d}k, y_0 - \frac{a}{d}k\right)$$

for some integer k, and hence $(x, y) \in A$. $\square$

Example 3   Find all solutions to the Diophantine equation $84x + 35y = 63$.

Solution: In Example 2 we showed that $\gcd(84, 35) = 7$, and found the integer solution $x = -18$ and $y = 45$. Based on this particular solution, from the Linear Diophantine Equation Theorem, Part 2, the set of all solutions to this equation is given by

$$\{(x, y) : x = -18 + 5n, y = 45 - 12n, n \in \mathbb{Z}\}.$$

Example 4   Find all solutions to the Diophantine equation $9x + 5y = 137$, in which both x and y are non-negative.

Solution:

x	y	r	q
1	0	9	0
0	1	5	0
1	-1	4	1
-1	2	1	1
5	-9	0	4

From the table constructed by applying EEA above, we have determined that $n = 4$, and hence $d = \gcd(9, 5) = r_4 = 1$. Since 1 clearly divides 137, then from the Linear Diophantine Equation Theorem, Part 1, we conclude that this equation has a solution. The certificate of correctness we obtain from the EEA table is $s = x_4 = -1$ and $t = y_4 = 2$, so we have

$$9 \times (-1) + 5 \times 2 = 1$$

Multiplying this equation by 137 gives

$$9 \times (-137) + 5 \times (274) = 137,$$

so a particular solution is $x = -137$ and $y = 274$. Based on this particular solution, from the Linear Diophantine Equation Theorem, Part 2, the set of all solutions to this Diophantine equation is given by

$$\{(x, y) : x = -137 + 5n, y = 274 - 9n, n \in \mathbb{Z}\}.$$

We need to determine the pairs (x, y) in this set for which $x \geq 0$ and $y \geq 0$, and hence we need to determine all $n \in \mathbb{Z}$ such that

$$\text{(i)} \quad -137 + 5n \geq 0 \qquad \text{and} \qquad \text{(ii)} \quad 274 - 9n \geq 0.$$

To find all solutions to inequality (i), add 137 to both sides and divide by 5, which gives $n \geq \frac{137}{5} = 27\frac{2}{5}$, and to find all solutions to inequality (ii), add $9n$ to both sides and divide by 9, which gives $n \leq \frac{274}{9} = 30\frac{4}{9}$. The values of n satisfying both inequalities are precisely the values of n such that

$$27\frac{2}{5} \leq n \leq 30\frac{4}{9},$$

so we have $n = 28, 29, 30$. Substituting these values of n into the pair $(-137 + 5n, 274 - 9n)$, we obtain precisely three non-negative integer solutions:

$$(3, 22), \quad (8, 13) \quad \text{and} \quad (13, 4).$$

REMARK

It is useful to summarize what the Linear Diophantine Equation Theorems, Part 1 and 2, taken together, tell us about the number of solutions to the linear Diophantine equation

$$ax + by = c, \tag{7.5}$$

where a and b are both not zero. Let $d = \gcd(a, b)$.

1. By LDET 1, if d does not divide c, then equation (7.5) has no solutions.

2. By LDET 1 and 2, if d does divide c, then equation (7.5) has infinitely many solutions.

In particular, when $c = d$, item 2 above tells us that there are infinitely many certificates of correctness for $d = \gcd(a, b)$ whenever a and b are both not zero (see the GCD Characterization Theorem and Bézout's Lemma)

Chapter 8

Congruence and Modular Arithmetic

8.1   Congruence

One of the difficulties in working out properties of divisibility is that we don't have an "arithmetic" of divisibility. Wouldn't it be nice if we could solve problems about divisibility in much the same way that we usually do arithmetic: add, subtract, multiply and divide?

Carl Friedrich Gauss (1777 – 1855) is widely regarded as the greatest mathematician of the last two centuries. In his landmark work, *Disquisitiones Arithmeticae*, published when Gauss was 23, he introduced congruence notation and provided a mechanism to treat divisibility with arithmetic.

Definition 8.1.1

congruent,
congruence,
modulo, modulus

Let m be a fixed positive integer. For integers a and b, we say that a is **congruent** to b **modulo** m, and write

$$a \equiv b \pmod{m},$$

when $m \mid (a - b)$. For integers a and b such that $m \nmid (a - b)$, we write $a \not\equiv b \pmod{m}$. We refer to $\equiv$ as **congruence**, and m as its **modulus**.

REMARK

The three horizontal bars used for congruence is the same notation that is used for logical equivalence, even though there is no connection. When you encounter this three horizontal bars notation, you will generally be able to determine the meaning from the context – if logical expressions are involved, then the notation means logical equivalence, and if integers are involved and (usually) "(mod m)" is included for some positive integer m, then the notation means congruence.

We also note that this notation is only defined when m is a positive integer. Therefore, we will use the convention that when we write "(mod m)", that it is implied that m is a positive integer.

Example 1   Consider the following examples.

- $-41 \equiv 41 \pmod 2$

- $41 \equiv -41 \pmod 2$

- $-41 \not\equiv 41 \pmod 3$

- $41 \equiv 41 \pmod 3$

- $-20 \equiv 4 \pmod 6$

- $24 \not\equiv 34 \pmod 6$

8.2   Elementary Properties of Congruence

One convenient aspect of congruence is that it behaves a lot like equality. For example, equality is an *equivalence relation* on the integers. This means that it has the following three properties:

1. *Reflexivity*: For all integers a, $a = a$.

2. *Symmetry*: For all integers a and b, if $a = b$, then $b = a$.

3. *Transitivity*: For all integers a, b and c, if $a = b$ and $b = c$, then $a = c$.

Most relationships that you can think of do not have these three properties. The relation *divides* fails symmetry. The relation *greater than* fails both reflexivity and symmetry. The non-mathematical relation *is a parent of* fails reflexivity, symmetry and transitivity. However, logical equivalence of logical expressions and similarity of triangles satisfy all three properties. Here and throughout this chapter, we consider m to be a fixed positive integer.

Proposition 1   **(Congruence is an Equivalence Relation (CER))**

For all integers a, b and c, we have

1. $a \equiv a \pmod m$.

2. If $a \equiv b \pmod m$, then $b \equiv a \pmod m$.

3. If $a \equiv b \pmod m$ and $b \equiv c \pmod m$, then $a \equiv c \pmod m$.

Proof: We prove each part in turn.

1. Let a be an arbitrary integer. Since $a - a = 0$ and $m \mid 0$, the definition of congruence gives $a \equiv a \pmod m$.

2. Let a and b be arbitrary integers, and assume that $a \equiv b \pmod{m}$. Hence we have $m \mid (a - b)$, so $a - b = mk$ for some integer k. Now, $b - a = -(a - b) = -mk = m(-k)$ so we have $m \mid (b - a)$, and the definition of congruence gives $b \equiv a \pmod{m}$.

3. Let a, b and c be arbitrary integers, and assume that $a \equiv b \pmod{m}$ and $b \equiv c \pmod{m}$. Hence we have $m \mid (a - b)$ and $m \mid (b - c)$. Now, by the Divisibility of Integer Combinations, we have $m \mid ((a - b)(1) + (b - c)(1))$, so $m \mid (a - c)$. The definition of congruence then gives $a \equiv c \pmod{m}$. $\square$

Now suppose that we have integers a_1, a_2, b_1 and b_2 satisfying the equalities $a_1 = b_1$ and $a_2 = b_2$. Then we know that we can add, subtract and multiply these equations to obtain

$$a_1 + a_2 = b_1 + b_2. \qquad a_1 - a_2 = b_1 - b_2, \qquad \text{and} \qquad a_1 a_2 = b_1 b_2.$$

In the next result, we prove that congruence satisfies these same three properties.

Proposition 2

For all integers a_1, a_2, b_1 and b_2, if $a_1 \equiv b_1 \pmod{m}$ and $a_2 \equiv b_2 \pmod{m}$, then

1. $a_1 + a_2 \equiv b_1 + b_2 \pmod{m}$,

2. $a_1 - a_2 \equiv b_1 - b_2 \pmod{m}$,

3. $a_1 a_2 \equiv b_1 b_2 \pmod{m}$.

Proof: Let a_1, a_2, b_1 and b_2 be arbitrary integers. Assume that $a_1 \equiv b_1 \pmod{m}$ and $a_2 \equiv b_2 \pmod{m}$, so from the definitions of congruence and divisibility there exist integers k and ℓ such that $a_1 - b_1 = km$ and $a_2 - b_2 = \ell m$.

For part 1: Since m divides both $a_1 - b_1$ and $a_2 - b_2$, it follows from Divisibility of Integer Combinations that m divides $(a_1 - b_1) + (a_2 - b_2)$. That is, m divides $(a_1 + a_2) - (b_1 + b_2)$. From the definition of congruence we have $a_1 + a_2 \equiv b_1 + b_2 \pmod{m}$.

For part 2: Since m divides both $a_1 - b_1$ and $a_2 - b_2$, it follows from Divisibility of Integer Combinations that m divides $(a_1 - b_1) - (a_2 - b_2)$. That is, m divides $(a_1 - a_2) - (b_1 - b_2)$. From the definition of congruence we have $a_1 - a_2 \equiv b_1 - b_2 \pmod{m}$.

For part 3: We have

$$\begin{aligned} a_1 a_2 - b_1 b_2 &= (b_1 + km)(b_2 + \ell m) - b_1 b_2 \\ &= b_1 b_2 + b_1 \ell m + b_2 km + k\ell m^2 - b_1 b_2 \\ &= (b_1 \ell + b_2 k + k\ell m)m, \end{aligned}$$

which implies that m divides $a_1 a_2 - b_1 b_2$, and hence from the definition of congruence we have $a_1 a_2 \equiv b_1 b_2 \pmod{m}$. $\square$

We will not prove the following generalization of the above result, in which the two congruent pairs of integers are replaced by n congruent pairs for an arbitrary natural number n. To prove this result, apply Mathematical Induction in a straightforward way. We also note that that we have not explicitly mentioned subtraction because subtraction can be accomplished by adding the negative.

Proposition 3   **(Congruence Add and Multiply (CAM))**

For all positive integers n, for all integers $a_1, \ldots, a_n$, and $b_1, \ldots, b_n$, if $a_i \equiv b_i \pmod{m}$ for all $1 \le i \le n$, then

1. $a_1 + a_2 + \cdots + a_n \equiv b_1 + b_2 + \cdots + b_n \pmod{m}$,

2. $a_1 a_2 \cdots a_n \equiv b_1 b_2 \cdots b_n \pmod{m}$.

The typical way in which we will apply part 2 of Proposition 3 above is for the special case when $a_i = a$ and $b_i = b$ for all $i = 1, \ldots, n$. This special case is important enough that we record it separately.

Proposition 4   **(Congruence Power (CP))**

For all positive integers n and integers a and b, if $a \equiv b \pmod{m}$, then $a^n \equiv b^n \pmod{m}$.

So far, we have considered how addition, subtraction and multiplication work for congruences. In the next result we consider how division works for congruences, and prove that a condition involving the greatest common divisor is key.

Proposition 5   **(Congruence Divide (CD))**

For all integers a, b and c, if $ac \equiv bc \pmod{m}$ and $\gcd(c, m) = 1$, then $a \equiv b \pmod{m}$.

Proof: Let a, b and c be arbitrary integers, and assume that $ac \equiv bc \pmod{m}$ and $\gcd(c, m) = 1$. Since $ac \equiv bc \pmod{m}$, we obtain $m \mid (ac - bc)$ by the definition of congruence, so $m \mid c(a - b)$. Since $\gcd(c, m) = 1$, we can apply the proposition Coprimeness and Divisibility, which gives $m \mid (a - b)$. Hence, by the definition of congruence we conclude that $a \equiv b \pmod{m}$. □

The importance of the assumption that $\gcd(c, m) = 1$ in the above result is illustrated in the following example, in which we choose $m = 6$.

Example 2   As an example of the hypothesis $ac \equiv bc \pmod{m}$ in the proposition Congruence Divide, the statement

$$180 \equiv 120 \pmod{6} \tag{8.1}$$

is certainly true. Now consider dividing on both sides of this congruence for three different choices of c.

- $c = 5$: Here the congruence (8.1) in the hypothesis becomes $36 \times 5 \equiv 24 \times 5 \pmod{6}$, and the other part of the hypothesis, $\gcd(c, m) = \gcd(5, 6) = 1$, is also true. Dividing on both sides of (8.1) by 5, we obtain the conclusion $36 \equiv 24 \pmod{6}$ of the proposition Congruence Divide. It is easy to check that this conclusion is a true statement, as guaranteed by the proof of Congruence Divide in the result above.

- $c = 3$: Here the congruence (8.1) in the hypothesis becomes $60 \times 3 \equiv 40 \times 3 \pmod 6$, but the other part of the hypothesis, $\gcd(c, m) = \gcd(3, 6) = 1$, is *false* (of course, $\gcd(3, 6) = 3$). If we were to divide on both sides of (8.1) by 3, we would obtain the conclusion $60 \equiv 40 \pmod 6$ of the proposition Congruence Divide. It is easy to check that this conclusion is a *false* statement. However, for this choice of c, the fact that the conclusion is false does not contradict the proposition, since in this case it is an implication in which one of the hypotheses is false.

- $c = 2$: Here the congruence (8.1) in the hypothesis becomes $90 \times 2 \equiv 60 \times 2 \pmod 6$, but the other part of the hypothesis, $\gcd(c, m) = \gcd(2, 6) = 1$, is *false* (of course, $\gcd(2, 6) = 2$). If we were to divide on both sides of (8.1) by 2, we would obtain the conclusion $90 \equiv 60 \pmod 6$ of the proposition Congruence Divide. It is easy to check that this conclusion is a *true* statement. For this choice of c, the fact that the conclusion is true is not supported by the proposition, since it is an implication in which one of the hypotheses is false.

8.3   Congruence and Remainders

In this section we consider the relationship between congruence modulo m and the remainder when an integer is divided by m.

Proposition 6   **(Congruent Iff Same Remainder (CISR))**

For all integers a and b, $a \equiv b \pmod m$ if and only if a and b have the same remainder when divided by m.

Proof: Let a and b be arbitrary integers. From the Division Algorithm, dividing a by m and b by m, we obtain

$$a = q_1 m + r_1, \qquad \text{and} \qquad b = q_2 m + r_2,$$

for unique integers q_1, r_1 and q_2, r_2 with $0 \le r_1 < m$ and $0 \le r_2 < m$. Subtracting the second equation from the first gives

$$a - b = (q_1 - q_2)m + (r_1 - r_2), \qquad -m < r_1 - r_2 < m, \tag{8.2}$$

where for a detailed proof of the inequality for $r_1 - r_2$, see the proof of inequality (6.2) that we gave on page 94 of Chapter 6 as part of the proof of the Division Algorithm.

We now prove the two implications that make up the if and only if statement of the result.

For one implication, assume that $a \equiv b \pmod m$. Then, by the definitions of congruence and divisibility, there exists an integer k such that $a - b = km$. Substituting this into equation (8.2), we obtain

$$km = (q_1 - q_2)m + (r_1 - r_2).$$

Rearranging, we get $r_1 - r_2 = (k - q_1 + q_2)m$, which implies that $m \mid (r_1 - r_2)$. But $-m < r_1 - r_2 < m$, so we have $r_1 - r_2 = 0$. That is, $r_1 = r_2$, and hence a and b have the same remainder when divided by m.

For the other implication, assume that a and b have the same remainder when divided by m, so we have $r_1 = r_2$. Substituting this in equation (8.2), we obtain $a - b = (q_1 - q_2)m$, so m divides $a - b$, and hence from the definition of congruence we have $a \equiv b \pmod{m}$. $\square$

Note what the proposition Congruent Iff Same Remainder says in the special case that b satisfies $0 \leq b < m$. In this case, when we divide b by m, we obtain $b = 0 \cdot m + b$, so the quotient is 0 and the remainder is b itself. This means that the condition that a and b have the same remainder can be restated as the condition that a has remainder b when divided by m. We record this special case as the next result.

Proposition 7   **(Congruent To Remainder (CTR))**

For all integers a and b with $0 \leq b < m$, $a \equiv b \pmod{m}$ if and only if a has remainder b when divided by m.

We now give some examples to demonstrate how our propositions on congruence can be applied to determine remainders in an elegant and surprisingly powerful way.

Example 3   Determine the remainder when 3^{47} is divided by 7.

Solution: Observe that $3^3 \equiv 27 \equiv -1 \pmod{7}$, so from the propositions Congruence Add and Multiply, and Congruence Power, we obtain

$$
\begin{aligned}
3^{47} &\equiv 3^{2+45} \pmod{7} \\
&\equiv 3^2 3^{45} \pmod{7} \\
&\equiv 9(3^3)^{15} \pmod{7} \\
&\equiv 2(-1)^{15} \pmod{7} \\
&\equiv 2(-1) \pmod{7} \\
&\equiv -2 \pmod{7} \\
&\equiv 5 \pmod{7}.
\end{aligned}
$$

Since $0 \leq 5 < 7$, we conclude from the proposition Congruent To Remainder that the remainder when 3^{47} is divided by 7 is equal to 5.

Example 4   What is the remainder when $2^{22}3^{33}5^{55}$ is divided by 11?

Solution: Observe that (for reasons that will become clear in the computations that follow)

$$2^5 \equiv 32 \equiv -1 \pmod{11}, \qquad 3^2 \equiv 9 \equiv -2 \pmod{11}, \qquad \text{and} \qquad 5^2 \equiv 25 \equiv 3 \pmod{11}.$$

Then, using propositions Congruence Add and Multiply, and Congruence Power, we get

$$
\begin{aligned}
2^{22}3^{33}5^{55} &\equiv 2^{2+20}3^{1+32}5^{1+54} \pmod{11} \\
&\equiv 2^2(2^5)^4(3)(3^2)^{16}(5)(5^2)^{27} \pmod{11} \\
&\equiv 2^2(-1)^4(3)(-2)^{16}(5)3^{27} \pmod{11} \\
&\equiv (5)2^{18}3^{28} \pmod{11} \\
&\equiv (5)2^3(2^5)^3(3^2)^{14} \pmod{11} \\
&\equiv (5)2^3(-1)^3(-2)^{14} \pmod{11} \\
&\equiv (-5)2^{17} \pmod{11} \\
&\equiv (-5)2^2(2^5)^3 \pmod{11} \\
&\equiv (-20)(-1)^3 \pmod{11} \\
&\equiv 20 \pmod{11} \\
&\equiv 9 \pmod{11}.
\end{aligned}
$$

Since $0 \leq 9 < 11$, we conclude from the proposition Congruent To Remainder that the remainder when $2^{22}3^{33}5^{55}$ is divided by 11 is equal to 9.

Example 5   What is the last decimal digit of 7^{3333}?

Solution: The last decimal digit of any non-negative integer a is precisely equal to the remainder when a is divided by 10. Therefore, we will work modulo 10, and first observe that $7^2 \equiv 49 \equiv -1 \pmod{10}$. Then, using propositions Congruence Add and Multiply, and Congruence Power, we get

$$
7^{3333} \equiv 7^{1+3332} \equiv 7(7^2)^{1666} \equiv 7(-1)^{1666} \equiv 7 \pmod{10}.
$$

Since $0 \leq 7 < 10$, we conclude from the proposition Congruent To Remainder that the remainder when 7^{3333} is divided by 10 is 7, and hence that the last decimal digit is 7.

REMARK

In the above examples, we have written all relations between integers as congruences, and not equalities. For example, we have written $2^5 \equiv 32 \pmod{11}$ even though, of course, it is also true that $2^5 = 32$.

The reason we have chosen to use only congruences is that care has to be taken if we use a mixture of congruences and equalities: For integers x and y, if $x = y$ then it is always true that $x \equiv y \pmod{m}$ for any positive integer m. However, this is not reversible: if $x \equiv y \pmod{m}$ for some positive integer m then it is not necessarily true that $x = y$. For example, $32 \equiv 10 \pmod{11}$, but of course $32 \neq 10$.

Another way in which our propositions on congruence can be applied is to prove classical results about divisibility of integers. For example, the next result gives a simple rule for divisibility by 3, stated in terms of the digits in the decimal representation.

Proposition 8   For all non-negative integers a, a is divisible by 3 if and only if the sum of the digits in the decimal representation of a is divisible by 3.

Proof: Suppose that a has the decimal representation $d_k d_{k-1} \cdots d_1 d_0$, where $0 \leq d_i \leq 9$ for $i = 0, 1, \ldots, k$, $k \geq 0$. Then we have the equation

$$a = d_k 10^k + d_{k-1} 10^{k-1} + \cdots + d_1 10 + d_0$$

that relates a and its digits $d_k, d_{k-1}, \ldots, d_1, d_0$. Now observe that $10 \equiv 1 \pmod{3}$. From the above equation for a, using propositions Congruence Add and Multiply, and Congruence Power, we get

$$\begin{aligned}
a &\equiv d_k 10^k + d_{k-1} 10^{k-1} + \cdots + d_1 10 + d_0 \pmod{3} \\
&\equiv d_k (1)^k + d_{k-1} (1)^{k-1} + \cdots + d_1 (1) + d_0 \pmod{3} \\
&\equiv d_k + d_{k-1} + \cdots + d_1 + d_0 \pmod{3}.
\end{aligned}$$

Now let $S = d_k + d_{k-1} + \cdots + d_1 + d_0$ represent the sum of the digits in the decimal representation of a. Using this notation, what we have proved above is that $a \equiv S \pmod{3}$.

But an integer is divisible by 3 if and only if the remainder when it is divided by 3 is 0. Hence from the proposition Congruent To Remainder, an integer is divisible by 3 if and only if it is congruent to 0 modulo 3. Since $a \equiv S \pmod{3}$, then $a \equiv 0 \pmod{3}$ if and only if $S \equiv 0 \pmod{3}$, proving the result. $\qquad \square$

Example 6   The sum of the digits of 6455874532635 is

$$6 + 4 + 5 + 5 + 8 + 7 + 4 + 5 + 3 + 2 + 6 + 3 + 5 = 63,$$

which is divisible by 3. This implies that the integer 6455874532635 is also divisible by 3.

The sum of the digits of 5748562331869 is

$$5 + 7 + 4 + 8 + 5 + 6 + 2 + 3 + 3 + 1 + 8 + 6 + 9 = 67,$$

which is not divisible by 3. This implies that the integer 5748562331869 is also not divisible by 3.

We continue with a similar rule for divisibility by 11.

Proposition 9   For all non-negative integers a, a is divisible by 11 if and only if $S_e - S_o$ is divisible by 11, where

- S_e is the sum of the digits of even powers (of 10) in the decimal representation of a,

- S_o is the sum of the digits of odd powers (of 10) in the decimal representation of a.

Proof: Suppose that a has the decimal representation $d_k d_{k-1} \cdots d_1 d_0$, where $0 \leq d_i \leq 9$ for $i = 0, 1, \ldots, k$, $k \geq 0$, so we have the equation

$$a = d_k 10^k + d_{k-1} 10^{k-1} + \cdots + d_1 10 + d_0.$$

Now observe that $10 \equiv -1 \pmod{11}$. From the above equation for a, using propositions Congruence Add and Multiply, and Congruence Power, we get

$$\begin{aligned} a &\equiv d_k 10^k + d_{k-1} 10^{k-1} + \cdots + d_1 10 + d_0 \pmod{11} \\ &\equiv d_k(-1)^k + d_{k-1}(-1)^{k-1} + \cdots + d_1(-1) + d_0 \pmod{11} \\ &\equiv S_e - S_o \pmod{11}, \end{aligned}$$

where, for k even, we have

$$S_e = d_k + d_{k-2} + \cdots + d_2 + d_0, \qquad S_o = d_{k-1} + \cdots + d_3 + d_1,$$

and, for k odd, we have

$$S_e = d_{k-1} + \cdots + d_2 + d_0, \qquad S_o = d_k + d_{k-2} + \cdots + d_3 + d_1.$$

Hence, for both of the cases k even and k odd, we have

- S_e is the sum of the digits of even powers in the decimal representation of a,

- S_o is the sum of the digits of odd powers in the decimal representation of a.

But an integer is divisible by 11 if and only if the remainder when it is divided by 11 is 0. Hence from the proposition Congruent To Remainder, an integer is divisible by 11 if and only if it is congruent to 0 modulo 11. Since $a \equiv S_e - S_o \pmod{11}$, then $a \equiv 0 \pmod{11}$ if and only if $S_e - S_o \equiv 0 \pmod{11}$, proving the result.   □

Example 7   For the integer 6455874532635, we have

$$S_e - S_o = (6 + 5 + 8 + 4 + 3 + 6 + 5) - (4 + 5 + 7 + 5 + 2 + 3) = 37 - 26 = 11,$$

which is divisible by 11. This implies that the integer 6455874532635 is also divisible by 11.

For the integer 5748562331869, we have

$$S_e - S_o = (5 + 4 + 5 + 2 + 3 + 8 + 9) - (7 + 8 + 6 + 3 + 1 + 6) = 36 - 31 = 5,$$

which is not divisible by 11. This implies that the integer 5748562331869 is also not divisible by 11.

EXERCISE

Find and prove a rule for dividing a non-negative integer by 9 in terms of the digits in its decimal representation.

8.4   Linear Congruences

One of the advantages of congruence over divisibility is that we have an "arithmetic" of congruence. This allows us to solve new kinds of "equations".

Definition 8.4.1

linear congruence

A relation of the form

$$ax \equiv c \pmod{m}$$

is called a **linear congruence** in the variable x. A **solution** to such a linear congruence is an integer x_0 such that

$$ax_0 \equiv c \pmod{m}.$$

In the following result we determine whether a linear congruence has a solution or not, and, if so, what the solutions are.

Theorem 10

(Linear Congruence Theorem (LCT))

For all integers a and c, with a non-zero, the linear congruence

$$ax \equiv c \pmod{m}$$

has a solution if and only if $d \mid c$, where $d = \gcd(a, m)$. Moreover, if $x = x_0$ is one particular solution to this congruence, then the set of all solutions is given by

$$\left\{ x \in \mathbb{Z} : x \equiv x_0 \pmod{\frac{m}{d}} \right\},$$

or, equivalently,

$$\left\{ x \in \mathbb{Z} : x \equiv x_0, x_0 + \frac{m}{d}, x_0 + 2\frac{m}{d}, \ldots, x_0 + (d-1)\frac{m}{d} \pmod{m} \right\}.$$

Proof: From the definitions of congruence and divisibility, $x = x_0$ is a solution to the congruence $ax \equiv c \pmod{m}$ precisely when there exists an integer k such that $ax_0 - c = mk$. Rearranging slightly, this means that the linear Diophantine equation

$$ax + my = c$$

has the solution $x = x_0$ and $y = -k$. Therefore $x = x_0$ is a solution to the linear congruence $ax \equiv c \pmod{m}$ if and only if there exists an integer k such that $x = x_0$ and $y = -k$ is a solution to the linear Diophantine equation $ax + my = c$. Then, from the Linear Diophantine Equation Theorem, Part 1 with $b = m$, the congruence $ax \equiv c \pmod{m}$ has a solution if and only if $d \mid c$.

Moreover, if $x = x_0$ is one particular solution to this congruence, the Linear Diophantine Equation Theorem, Part 2 with $b = m$ tells us that the set of all solutions to the congruence is given by

$$S = \left\{ x : x = x_0 + \frac{m}{d}n, n \in \mathbb{Z} \right\} = \left\{ x \in \mathbb{Z} : x \equiv x_0 \pmod{\frac{m}{d}} \right\},$$

where we have omitted the y values from the solutions to the linear Diophantine equation.

We can also express the elements of S in terms of congruence relations modulo m, as follows: each element $x \in S$ can be written uniquely in the form $x = x_0 + \frac{m}{d}n$, for some integer n. Now, dividing n by d, by the Division Algorithm we can write n uniquely in the form $n = dq + r$, for some integer quotient q and remainder r, where $0 \le r \le d - 1$. Hence we can write x uniquely in the form

$$x = x_0 + \frac{m}{d}(dq + r) = x_0 + r\frac{m}{d} + qm,$$

where $q \in \mathbb{Z}$ and $0 \le r \le d - 1$. For each fixed r, with $0 \le r \le d - 1$, the set of x of the form $x = x_0 + r\frac{m}{d} + qm$, where $q \in \mathbb{Z}$, is precisely the set of x such that $x \equiv x_0 + r\frac{m}{d} \pmod{m}$, and so we obtain

$$S = \left\{ x \in \mathbb{Z} : x \equiv x_0, x_0 + \frac{m}{d}, x_0 + 2\frac{m}{d}, \ldots, x_0 + (d-1)\frac{m}{d} \pmod{m} \right\}.$$

$\square$

REMARK

Note that the proof of the proposition Linear Congruence Theorem above tells us that to find solutions to the linear congruence relation $ax \equiv c \pmod{m}$, we should consider the corresponding linear Diophantine equation $ax + my = c$.

Example 8   If possible, solve the linear congruence

$$3x \equiv 5 \pmod{6}.$$

Solution: Since $\gcd(3, 6) = 3$ and $3 \nmid 5$, there is no solution to $3x \equiv 5 \pmod 6$, by the Linear Congruence Theorem.

Example 9   If possible, solve the linear congruence

$$3x \equiv 5 \pmod{76}.$$

Solution: Since $\gcd(3, 76) = 1$ and $1 \mid 5$, there is a solution, by the Linear Congruence Theorem. Moreover, since $\frac{m}{d} = \frac{76}{1} = 76$, the set of all solutions is given by the set of all integers x such that $x \equiv x_0 \pmod{76}$, where x_0 is some particular solution to the congruence. To find a particular solution, we consider the corresponding linear Diophantine equation $3x + 76y = 5$. Applying the Extended Euclidean Algorithm we obtain the table

y	x	r	q
1	0	76	0
0	1	3	0
1	-25	1	25
-3	76	0	3

From the second last row we obtain $76(1) + 3(-25) = 1$, or to match up with the order of the original equation, $3(-25) + 76(1) = 1$. Multiplying this equation by 5 gives

$$3(-125) + 76(5) = 5,$$

so one particular solution is given by $x = -125$. But $-125 \equiv -125 + 2(76) \equiv 27 \pmod{76}$, and hence the set of solutions to the linear congruence is given by all integers x such that

$$x \equiv 27 \pmod{76}.$$

Example 10   If possible, solve the linear congruence

$$4x \equiv 6 \pmod{10}.$$

Solution: Since $\gcd(4, 10) = 2$ and $2 \mid 6$, there is a solution, by the Linear Congruence Theorem. Moreover, since $\frac{m}{d} = \frac{10}{2} = 5$, the set of all solutions is given by the set of all integers x such that $x \equiv x_0$ or $x_0 + 5 \pmod{10}$, where x_0 is some particular solution to the congruence. Of course, we could solve the corresponding linear Diophantine equation to obtain a particular solution, but since 10 is a small modulus, we can also simply check all possibilities modulo 10 to find a particular solution:

$x \pmod{10}$	0	1	2	3	4	5	6	7	8	9
$4x \pmod{10}$	0	4	8	2	6	0	4	8	2	6

Hence the solutions are all integers x such that $x \equiv 4$ or $9 \pmod{10}$.

8.5   Non-Linear Congruences

By applying the Euclidean Algorithm or Extended Euclidean Algorithm, we have an efficient way to solve linear congruences, but we have no equivalent way to solve congruences involving higher powers of the variable. However, for higher powers, we can solve a congruence relation modulo m by checking all m values, which works quite well when m is small.

Example 11   Solve the congruence relation $x^2 + x \equiv 2 \pmod{8}$.

Solution: We use a table to check the 8 possible values of x.

$x \pmod 8$	0	1	2	3	4	5	6	7
$x^2 \pmod 8$	0	1	4	1	0	1	4	1
$x^2 + x \pmod 8$	0	2	6	4	4	6	2	0

Hence, the solution is given by all integers x such that $x \equiv 1$ or $6 \pmod 8$.

8.6   Congruence Classes and Modular Arithmetic

In previous sections, we have seen that the solutions to a congruence relation can be expressed in terms of sets consisting of all integers congruent to a given integer modulo m.

In this section, we study such sets, and an "arithmetic" for these sets that will be new for many students.

Definition 8.6.1
congruence class

The **congruence class** modulo m of the integer a is the set of integers

$$[a] = \{x \in \mathbb{Z} : x \equiv a \pmod{m}\}.$$

Note that, by the proposition Congruent Iff Same Remainder, there are exactly m different congruence classes modulo m, since there are m choices $0, 1, \ldots, m-1$ of remainder when dividing by m.

Example 12

For $m = 4$, the four congruence classes are given by

$$
\begin{array}{rclclcl}
[0] & = & \{x \in \mathbb{Z} : x \equiv 0 \pmod{4}\} & = & \{\ldots, -8, -4, 0, 4, 8, \ldots\} & = & \{4k : k \in \mathbb{Z}\}, \\
[1] & = & \{x \in \mathbb{Z} : x \equiv 1 \pmod{4}\} & = & \{\ldots, -7, -3, 1, 5, 9, \ldots\} & = & \{4k + 1 : k \in \mathbb{Z}\}, \\
[2] & = & \{x \in \mathbb{Z} : x \equiv 2 \pmod{4}\} & = & \{\ldots, -6, -2, 2, 6, 10, \ldots\} & = & \{4k + 2 : k \in \mathbb{Z}\}, \\
[3] & = & \{x \in \mathbb{Z} : x \equiv 3 \pmod{4}\} & = & \{\ldots, -5, -1, 3, 7, 11, \ldots\} & = & \{4k + 3 : k \in \mathbb{Z}\}.
\end{array}
$$

REMARK

Note that congruence classes have more than one representation. In the example above with $m = 4$, we have $[-4] = [0] = [4] = [8]$ and, in fact the congruence class $[0]$ has infinitely many representations of the form $[a]$ for different integers a. We can use any integer a where $4 \mid a$. This might seem strange at first, but remember that fractions are another example in which one number has infinitely many representations. For example $1/2 = 2/4 = 3/6 = \cdots$.

Now we define the arithmetic of congruence classes modulo m.

Definition 8.6.2
$\mathbb{Z}_m$, **addition and multiplication in** $\mathbb{Z}_m$, **modular arithmetic**

We define $\mathbb{Z}_m$ to be the set of m congruence classes

$$\mathbb{Z}_m = \{[0], [1], [2], \ldots, [m-1]\},$$

and we define two operations on $\mathbb{Z}_m$, **addition** and **multiplication**, as follows:

$$[a] + [b] = [a + b],$$
$$[a][b] = [ab].$$

When we apply these operations on the set $\mathbb{Z}_m$, we are doing what is known as **modular arithmetic**.

REMARK

Though the definition of the operations of modular arithmetic in the definition above may seem obvious, there is a fair amount going on here.

1. Sets are being treated as individual "numbers". Modular addition and multiplication are being performed on congruence classes, which are sets.

2. The addition and multiplication symbols on the left of the equals signs are in $\mathbb{Z}_m$ and those on the right are operations in the integers.

3. We are assuming that the operations are *well-defined*. That is, we are assuming that these operations make sense even when there are multiple representatives of a congruence class. For example, in $\mathbb{Z}_6$, $[2] = [8]$ and $[3] = [-9]$. Further, we can verify that $[2] + [3] = [8] + [-9]$.

Just as we have addition and multiplication tables for the integers, we can also create addition and multiplication tables for $\mathbb{Z}_m$. For example, addition and multiplication tables for $\mathbb{Z}_4$ are given by

+	[0]	[1]	[2]	[3]
[0]	[0]	[1]	[2]	[3]
[1]	[1]	[2]	[3]	[0]
[2]	[2]	[3]	[0]	[1]
[3]	[3]	[0]	[1]	[2]

×	[0]	[1]	[2]	[3]
[0]	[0]	[0]	[0]	[0]
[1]	[0]	[1]	[2]	[3]
[2]	[0]	[2]	[0]	[2]
[3]	[0]	[3]	[2]	[1]

How do we read these tables? In each case, let the elements in the left-most column be denoted by $[i]$, for $i = 0, 1, 2, 3$, and the elements in the top row be denoted by $[j]$, for $j = 0, 1, 2, 3$. Then in the addition table (which has "+" at the top left), the table entry for row $[i]$ and column $[j]$ is given by $[i + j]$. Note however that all table entries have been adjusted where necessary, to have representatives between 0 and 3. For example, the entry in row $[2]$ and column $[3]$ of the addition table for $\mathbb{Z}_4$ above is given as $[1]$, and not $[5]$, since $2 + 3 \equiv 5 \equiv 1 \pmod 4$.

Example 13

In modular arithmetic, the following properties hold for all $[a] \in \mathbb{Z}_m$:

1. $[a] + [0] = [0] + [a] = [a]$,

2. $[a][0] = [0][a] = [0]$,

3. $[a] + [-a] = [-a] + [a] = [0]$,

4. $[a][1] = [1][a] = [a]$.

The proofs of these properties are straightforward and follow directly from the definitions of addition and multiplication of congruence classes . For example, to prove Property 1, we have $[a] + [0] = [a + 0] = [a]$, and $[0] + [a] = [0 + a] = [a]$.

From Property 1 in Example 13, we say that $[0]$ is the *additive identity* in $\mathbb{Z}_m$, and from Property 3, we say that $[-a]$ is the *additive inverse* of $[a]$ in $\mathbb{Z}_m$, for all $[a] \in \mathbb{Z}_m$. Similarly, from Property 4, we say that $[1]$ is the *multiplicative identity* in $\mathbb{Z}_m$.

For any $[a] \in \mathbb{Z}_m$, if there exists $[b] \in \mathbb{Z}_m$ such that

$$[a][b] = [b][a] = [1], \tag{8.3}$$

then we say that $[b]$ is the *multiplicative inverse* of $[a]$, and in this situation we use the notation $[b] = [a]^{-1}$. Now, the situation for the multiplicative inverse is different than for the additive inverse. As we noted above, every element $[a]$ in $\mathbb{Z}_m$ has an additive inverse, but it is not the case that every element has a multiplicative inverse. For example, checking the multiplication table for $\mathbb{Z}_4$ on the previous page, we see that there only two occurrences of the element $[1]$ as a table entry – in row $[1]$ and column $[1]$, and in row $[3]$ and column $[3]$. This means that in $\mathbb{Z}_4$, the only choices of $[a]$, $[b]$ for which $[a][b] = [1]$ are $[a] = [b] = [1]$ and $[a] = [b] = [3]$. Hence we conclude that in $\mathbb{Z}_4$, $[1]^{-1} = [1]$ and $[3]^{-1} = [3]$, but $[0]$ and $[2]$ do not have a multiplicative inverse.

In general, for an arbitrary positive integer m and $a = 0, 1, \ldots, m-1$, when does $[a]^{-1}$ exist in $\mathbb{Z}_m$? To answer this question, note from (8.3) that $[x] = [a]^{-1}$ is exactly a solution to the linear equation $[a][x] = [1]$ in $\mathbb{Z}_m$. The following result specifies when solutions to linear equations in $\mathbb{Z}_m$ exist.

Theorem 11

(Modular Arithmetic Theorem (MAT))

For all integers a and c, with a non-zero, the equation

$$[a][x] = [c]$$

in $\mathbb{Z}_m$ has a solution if and only if $d \mid c$, where $d = \gcd(a, m)$. Moreover, when $d \mid c$, there are d solutions, given by

$$[x_0] , \left[x_0 + \frac{m}{d}\right] , \left[x_0 + 2\frac{m}{d}\right] , \ldots, \left[x_0 + (d-1)\frac{m}{d}\right] ,$$

where $[x] = [x_0]$ is one particular solution.

Proof: From the proposition Congruence Add and Multiply,

$$[a][x_0] = [c],$$

in $\mathbb{Z}_m$, if and only if

$$ax_0 \equiv c \pmod{m}.$$

Hence the solutions to the equation $[a][x] = [c]$ in $\mathbb{Z}_m$ can be determined precisely from the solutions to the congruence relation $ax \equiv c \pmod{m}$. The result now follows directly from the Linear Congruence Theorem. $\square$

REMARK

By comparing the proofs of the Linear Congruence Theorem and Modular Arithmetic Theorem, we see that to find solutions to the equation $[a][x] = [c]$ in $\mathbb{Z}_m$, we should consider the corresponding linear Diophantine equation $ax + my = c$.

Note that for the special case $c = 1$, the above Modular Arithmetic Theorem describes the solutions to the equation $[a][x] = [1]$ in $\mathbb{Z}_m$. Hence it answers exactly the question about whether the multiplicative inverse $[a]^{-1}$ exists in $\mathbb{Z}_m$. This case of the above proposition can be stated in a particularly simple way, since the only positive divisor of $c = 1$ is $d = 1$. We record this special case as the next result.

Corollary 12   **(Inverses in $\mathbb{Z}_m$ (INV $\mathbb{Z}_m$))**

Let a be an integer with $1 \leq a \leq m - 1$. The element $[a]$ in $\mathbb{Z}_m$ has a multiplicative inverse if and only if $\gcd(a, m) = 1$. Moreover, when $\gcd(a, m) = 1$, the multiplicative inverse is unique.

Consider the situation for multiplicative inverses when m is a prime, say $m = p$. Then in the above corollary Inverses in $\mathbb{Z}_m$, the condition $\gcd(a, p) = 1$ is true for all a not divisible by p. We record the result in this case separately, below.

Corollary 13   **(Inverses in $\mathbb{Z}_p$ (INV $\mathbb{Z}_p$))**

For all prime numbers p and non-zero elements $[a]$ in $\mathbb{Z}_p$, the multiplicative inverse $[a]^{-1}$ exists and is unique.

Example 14   If possible, find the multiplicative inverse of $[5]$ in $\mathbb{Z}_{10}$.

Solution: Observe that $\gcd(5, 10) = 5 \neq 1$. Hence, by the corollary Inverses in $\mathbb{Z}_m$, $[5]$ has no multiplicative inverse in $\mathbb{Z}_{10}$.

Example 15   If possible, find the multiplicative inverse of $[5]$ in $\mathbb{Z}_{42}$.

Solution: Observe that $\gcd(5, 42) = 1$. Hence, by the corollary Inverses in $\mathbb{Z}_m$, $[5]$ has a unique multiplicative inverse in $\mathbb{Z}_{42}$. To find the inverse, we consider the corresponding linear Diophantine equation $5x + 42y = 1$. Applying the Extended Euclidean Algorithm we obtain the table

y	x	r	q
1	0	42	0
0	1	5	0
1	-8	2	8
-2	17	1	2
5	-42	0	2

From the second last row we obtain $42(-2) + 5(17) = 1$, or to match up with the order of the original equation, $5(17) + 42(-2) = 1$. We conclude that $[5]^{-1} = [17]$ in $\mathbb{Z}_{42}$.

Example 16   Solve the following system of equations in $\mathbb{Z}_{11}$,

$$[2][x] + [7][y] = [4], \tag{8.4}$$
$$[3][x] + [2][y] = [9]. \tag{8.5}$$

Solution: This is a system of two linear equations in $\mathbb{Z}_{11}$ for the two unknowns $[x]$ and $[y]$. Our method of solution will be similar to the method we use to solve two linear equations in the reals for two unknowns. First, subtract $[2]$ times equation (8.5) from $[3]$ times equation (8.4) to obtain $[17][y] = [-6]$, and note that $[17] = [6]$, so we have the equation

$$[6][y] = [-6].$$

Now 11 is prime, and $[6] \neq [0]$, so by the corollary Inverses in $\mathbb{Z}_p$, the element $[6]$ has a multiplicative inverse in $\mathbb{Z}_{11}$. Multiplying on both sides of the above equation by $[6]^{-1}$, we obtain

$$[6]^{-1}[6][y] = [6]^{-1}[-6],$$

and using the fact that

$$[6]^{-1}[6][y] = [1][y] = [y],$$

and

$$[6]^{-1}[-6] = [6]^{-1}[6][-1] = [1][-1] = [-1] = [10]$$

in $\mathbb{Z}_{11}$, we get $[y] = [10]$. Substituting $[y] = [10]$ into equation (8.4) gives

$$[2][x] = [4] - [7][10] = [-66] = [0].$$

But $[2]$ has a multiplicative inverse in $\mathbb{Z}_{11}$, and multiplying on both sides of the equation $[2][x] = [0]$ by $[2]^{-1}$, we obtain $[x] = [0]$. Hence we obtain $[x] = [0], [y] = [10]$.

Checking to make sure that this is not an extraneous solution, we substitute these values for $[x]$ and $[y]$ into equations (8.4) and (8.5), to obtain

$$[2][x] + [7][y] = [2][0] + [7][10] = [70] = [4],$$

$$[3][x] + [2][y] = [3][0] + [2][10] = [20] = [9],$$

which confirms that $[x] = [0], [y] = [10]$ is indeed a solution, and hence the only solution.

8.7   Fermat's Little Theorem

In this section we consider Fermat's Little Theorem, a useful result for powers of an integer modulo a prime, named after Pierre de Fermat, a French mathematician in the 1600's. This result should not be confused with *Fermat's Last Theorem*, conjectured by Fermat, but only proved in the 1990's by the English mathematician Andrew Wiles.

Theorem 14

(Fermat's Little Theorem (FℓT))

For all prime numbers p and integers a not divisible by p, we have

$$a^{p-1} \equiv 1 \pmod{p}.$$

Proof: Let p be an arbitrary prime number, and a be an arbitrary integer not divisible by p. Since $p \nmid a$, we have $[a] \neq [0]$. We prove the equivalent result that, for all prime numbers

p and non-zero elements $[a]$ in $\mathbb{Z}_p$, we have

$$[a]^{p-1} = [1].$$

Now consider the list of $p - 1$ elements

$$[a], [2a], \ldots, [(p-1)a] \tag{8.6}$$

in $\mathbb{Z}_p$. Next we prove two facts about this list:

- Fact 1 – the element $[0]$ does not appear in the list;

- Fact 2 – the elements in the list are all different.

To prove Fact 1, assume, for the sake of contradiction, that $[0]$ does appear in list (8.6), and hence that

$$[ia] = [0]$$

for some i with $1 \le i \le p-1$. By the corollary Inverses in $\mathbb{Z}_p$, $[a]$ has a multiplicative inverse, and multiplying on both sides of the above equation by $[a]^{-1}$, we obtain $[a]^{-1}[ia] = [a]^{-1}[0]$. But on the left hand side, we have $[a]^{-1}[ia] = [a]^{-1}[a][i] = [1][i] = [i]$, and on the right hand side we have $[a]^{-1}[0] = [0]$, so we get $[i] = [0]$, a contradiction. Hence we conclude that $[0]$ does not appear in the list (8.6).

To prove Fact 2, assume, for the sake of contradiction, that two elements in the list are the same, and hence that

$$[ia] = [ja]$$

for some i and j with $i \ne j$ and $1 \le i, j \le p - 1$. Multiplying on both sides of the above equation by $[a]^{-1}$, we obtain $[a]^{-1}[ia] = [a]^{-1}[ja]$. But on the left hand side, we have $[a]^{-1}[ia] = [a]^{-1}[a][i] = [1][i] = [i]$, and on the right hand side we similarly have $[a]^{-1}[ja] = [a]^{-1}[a][j] = [1][j] = [j]$, so we get $[i] = [j]$, a contradiction. Hence we conclude that the elements in the list (8.6) are all different.

Facts 1 and 2 together mean that the elements in list (8.6) must be

$$[1], [2], \ldots, [p - 1], \tag{8.7}$$

the non-zero elements in $\mathbb{Z}_p$, in some order. Since lists (8.6) and (8.7) consist of the same elements in possibly different orders, the products of the elements in lists (8.6) and (8.7) must be equal. Hence we obtain

$$[a][2a] \cdots [(p-1)a] = [1][2] \cdots [p - 1],$$

and factoring each $[ia] = [i][a]$, for $i = 1, 2, \ldots, p - 1$, on the left hand side, this gives

$$[1][2] \cdots [p - 1][a]^{p-1} = [1][2] \cdots [p - 1].$$

Now from the corollary Inverses in $\mathbb{Z}_p$, $[i]^{-1}$ exists for all $i = 1, 2, \ldots, p-1$, and multiplying on both sides of the above equation by $[1]^{-1}[2]^{-1} \cdots [p - 1]^{-1}$, we obtain

$$[a]^{p-1} = [1],$$

which is equivalent to the given result, as noted above. $\square$

REMARK

Recall that from the corollary Inverses in $\mathbb{Z}_p$, the multiplicative inverse $[a]^{-1}$ exists for all primes p and non-zero elements $[a]$ in $\mathbb{Z}_p$. Note that under these same conditions on p and a, Fermat's Little Theorem gives $[a][a]^{p-2} = [1]$ in $\mathbb{Z}_p$, which means that

$$[a]^{-1} = [a^{p-2}],$$

giving a formula for the multiplicative inverse when $[a]$ is a non-zero element in $\mathbb{Z}_p$ and p is a prime.

The following corollary of Fermat's Little Theorem is particularly convenient, since unlike Fermat's Little Theorem, it holds for all integers a.

Corollary 15

For all prime numbers p and integers a, we have

$$a^p \equiv a \pmod{p}.$$

Proof: Let p be an arbitrary prime number, and a be an arbitrary integer. Note that a is either divisible by p, or it is not divisible by p.

In the case that a is divisible by p, we have $a \equiv 0 \pmod{p}$. Hence $a^p \equiv 0^p \equiv 0 \pmod{p}$, so we have $a^p \equiv a \pmod{p}$ in this case.

In the case that $p \nmid a$, then by Fermat's Little Theorem we have $a^{p-1} \equiv 1 \pmod{p}$. Multiplying on both sides by a gives $a^p \equiv a \pmod{p}$ in this case also. $\qquad\square$

Example 17

What is the remainder when 3167^{2531} is divided by 17?

Solution: Observe that
$$3167 \equiv 5 \pmod{17}.$$

Also, since $17 \nmid 5$, by Fermat's Little Theorem we have

$$5^{16} \equiv 1 \pmod{17}.$$

Then, using propositions Congruence Add and Multiply, and Congruence Power, we obtain

$$3167^{2531} \equiv 5^{2531} \equiv 5^{16 \cdot 158 + 3} \equiv (5^{16})^{158}(5^3) \equiv (1)^{158}(125) \equiv 6 \pmod{17}.$$

Since $0 \le 6 < 17$, we conclude from the proposition Congruent To Remainder that the remainder is equal to 6.

Example 18

Solve the congruence relation

$$38x^{47} + 26x^{26} + 5x^9 + 4x^3 + 3x^2 + 2x + 1 \equiv 2 \pmod{5}.$$

Solution: For notational convenience, let $f(x) = 38x^{47} + 26x^{26} + 5x^9 + 4x^3 + 3x^2 + 2x + 1$. Now we have that $x^5 \equiv x \pmod{5}$ for all integers x, from Corollary 15. Observing that

$38 \equiv 3 \pmod{5}$, from the propositions Congruence Add and Multiply and Congruence Power, we obtain

$$38x^{47} \equiv 3x^{47} \equiv 3x^2(x^5)^9 \equiv 3x^2x^9 \equiv 3x^1(x^5)^2 \equiv 3x^1x^2 \equiv 3x^3 \pmod{5},$$

for all integers x. Similarly, since $26 \equiv 1 \pmod{5}$, we obtain

$$26x^{26} \equiv x^{26} \equiv x^1(x^5)^5 \equiv x^1x^5 \equiv x^1x^1 \equiv x^2 \pmod{5},$$

for all integers x, and since $5 \equiv 0 \pmod{5}$, we obtain $5x^9 \equiv 0 \pmod{5}$ for all integers x. Putting these congruence relations together using the proposition Congruence Add and Multiply, we get

$$f(x) \equiv 3x^3 + x^2 + 0 + 4x^3 + 3x^2 + 2x + 1 \equiv 2x^3 + 4x^2 + 2x + 1 \pmod{5},$$

for all integers x. Now we can use a table to check the 5 possible values of x.

$x \pmod 5$	0	1	2	3	4
$2x^3 + 4x^2 + 2x + 1 \pmod 5$	1	4	2	2	1

Therefore, the solutions to this congruence relation are given by all integers x such that $x \equiv 2 \pmod{5}$ or $x \equiv 3 \pmod{5}$.

REMARK

There is no hard and fast rule about when to use Fermat's Little Theorem, and when to use the closely related result Corollary 15, when evaluating a large power modulo a prime p. In general, Fermat's Little Theorem is more convenient when considering a large power of a specific integer that is not divisible by p, as in Example 17. However, Corollary 15 is often preferable when considering a large power of a variable, as in Example 18.

8.8   The Chinese Remainder Theorem

The following problem was posed, likely in the third century CE, by Sun Zi in his *Mathematical Manual* and republished in 1247 by Qin Jiushao in the *Mathematical Treatise in Nine Sections*.

> There are certain things whose number is unknown. Repeatedly divided by 3, the remainder is 2; by 5 the remainder is 3; and by 7 the remainder is 2. What will be the number?

The word problem asks us to find an integer n that simultaneously satisfies the following three congruences:

$$n \equiv 2 \pmod{3}$$
$$n \equiv 3 \pmod{5}$$
$$n \equiv 2 \pmod{7}.$$

In this section, we will consider the general problem of solving a set of simultaneous congruences like those above. We start with a famous theorem for solving two simultaneous congruences whose moduli are coprime. While many mathematical theorems attributed to western mathematicians are named for their author(s), this theorem has been dubbed in the west as "The Chinese Remainder Theorem". While Sun Zi articulated the problem that the theorem solves, the technique for solving simultaneous congruences and the theorem characterizing their solution appear to have been developed over time by multiple mathematicians. Two factors make it different to pinpoint a single author for whom to name the theorem: limited historical documents, and the fact that existing documents generously frame the contributions of previous mathematicians.

Theorem 16   **(Chinese Remainder Theorem (CRT))**

For all integers a_1 and a_2, and positive integers m_1 and m_2, if $\gcd(m_1, m_2) = 1$, then the simultaneous linear congruences

$$n \equiv a_1 \pmod{m_1}$$
$$n \equiv a_2 \pmod{m_2}$$

have a unique solution modulo $m_1 m_2$. Thus, if $n = n_0$ is one particular solution, then the solutions are given by the set of all integers n such that

$$n \equiv n_0 \pmod{m_1 m_2}.$$

Proof: Let a_1 and a_2 be arbitrary integers, and m_1 and m_2 be arbitrary positive integers. Assume that $\gcd(m_1, m_2) = 1$. From the definitions of congruence and divisibility, the set of solutions to the congruence $n \equiv a_1 \pmod{m_1}$ is given by

$$\{m_1 x + a_1 : x \in \mathbb{Z}\}, \tag{8.8}$$

An element of this set satisfies the congruence $n \equiv a_2 \pmod{m_2}$ if and only if the integer x satisfies the linear congruence

$$m_1 x \equiv a_2 - a_1 \pmod{m_2}.$$

Now we have $\gcd(m_1, m_2) = 1$, and hence from the Linear Congruence Theorem (with $d = 1$) and the definitions of congruence and divisibility, the set of solutions to the above linear congruence is given by

$$\{m_2 y + x_0 : y \in \mathbb{Z}\},$$

where x_0 is one particular solution (which must exist). Hence, replacing x by $m_2 y + x_0$ in (8.8), the set of solutions to the simultaneous congruences is given by

$$\{m_1(m_2 y + x_0) + a_1 : y \in \mathbb{Z}\} = \{m_1 m_2 y + (m_1 x_0 + a_1) : y \in \mathbb{Z}\},$$

which is simply the congruence class $[n_0]$ in $\mathbb{Z}_{m_1 m_2}$, where $n_0 = m_1 x_0 + a_1$ is one particular solution. $\qquad\square$

In our first example of the use of the Chinese Remainder Theorem, we consider a pair of congruences whose moduli are very small.

Example 19   Solve the simultaneous congruences

$$n \equiv 2 \pmod{3}$$
$$n \equiv 3 \pmod{5}.$$

Solution: Observe that $\gcd(3, 5) = 1$. Therefore we can apply the Chinese Remainder Theorem, which says that the solution to these simultaneous congruences is the set of all integers n such that $n \equiv n_0 \pmod{15}$, where n_0 is one particular solution.

To find a suitable value for n_0, note that $n_0 = 3, 8, 13$ are the choices of integers between 0 and 14 that are congruent to 3 (mod 5). Now observe that $3 \equiv 0 \pmod{3}$, $8 \equiv 2 \pmod{3}$, and $13 \equiv 1 \pmod{3}$, so $n_0 = 8$ is a solution to both congruences simultaneously. We conclude that the set of solutions to the simultaneous congruences consists of all integers n such that

$$n \equiv 8 \pmod{15}.$$

We continue with another example of the use of the Chinese Remainder Theorem, in which the pair of moduli are somewhat larger.

Example 20   Solve the simultaneous congruences

$$n \equiv 237 \pmod{1000}$$
$$n \equiv 100 \pmod{343}.$$

Solution: Observe that the prime factorizations of the moduli are given by $1000 = 2^3 5^3$ and $343 = 7^3$, so by the proposition GCD From Prime Factorization, we obtain that $\gcd(1000, 343) = 1$. Therefore we can apply the Chinese Remainder Theorem, which says that the solution to these simultaneous congruences is the set of all integers n such that $n \equiv n_0 \pmod{343000}$, where n_0 is one particular solution.

To find a suitable value for n_0, we could exhaustively check modulo 343000, but given the large integers involved, we choose to proceed much like in the proof of the Chinese Remainder Theorem: An integer n satisfies the congruence $n \equiv 237 \pmod{1000}$ if and only if it is of the form

$$n = 1000x + 237, \tag{8.9}$$

for some integer x. An integer n of this form also satisfies the congruence $n \equiv 100 \pmod{343}$ if and only if the integer x satisfies the linear congruence

$$1000x \equiv -137 \pmod{343}. \tag{8.10}$$

Since $\gcd(1000, 343) = 1$, the Linear Congruence Theorem (with $d = 1$) says that this linear congruence has solutions. To find a solution, we consider the corresponding linear Diophantine equation $1000x + 343y = -137$.

Applying the Extended Euclidean Algorithm we obtain the table

y	x	r	q
1	0	1000	0
0	1	343	0
1	-2	314	2
-1	3	29	1
11	-32	24	10
-12	35	5	1
59	-172	4	4
-71	207	1	1
343	-1000	0	4

From the second last row we obtain $1000(-71)+343(207) = 1$, and multiplying this equation by -137 gives

$$1000(9727) + 343(-28359) = -137.$$

Hence one particular solution to linear congruence (8.10) is given by $x_0 = 9727$. Substituting $x = x_0$ in (8.9), we obtain the value $n_0 = 1000(9727) + 237 = 9727237$ as one particular solution to the simultaneous congruences. Now observe that $9727237 \equiv 123237$ (mod 343000), and we conclude that the set of solutions to the simultaneous congruences consists of all integers n such that

$$n \equiv 123237 \pmod{343000}.$$

As a check, dividing 123237 by 1000, the Division Algorithm gives $123237 = 123(1000)+237$, and dividing 123237 by 343, the Division Algorithm gives $123237 = 359(343) + 100$.

REMARK

Suppose we wish to avoid using an exhaustive computational search to obtain a particular solution to a pair of simultaneous congruences when applying the Chinese Remainder Theorem. Then, as an alternative, we can always proceed by determining a particular solution to the corresponding linear Diophantine equation using the Extended Euclidean Algorithm.

In the next example, we consider a set of three simultaneous congruences.

Example 21   Solve the simultaneous congruences

$$n \equiv 2 \pmod 3$$
$$n \equiv 3 \pmod 5$$
$$n \equiv 5 \pmod 7.$$

Solution: In Example 19, we proved that the set of integers n simultaneously satisfying the first two congruences above, $n \equiv 2$ (mod 3) and $n \equiv 3$ (mod 5), is precisely the set of integers n satisfying the single congruence $n \equiv 8$ (mod 15).

Hence solving the three simultaneous congruences given in this example is equivalent to solving the pair of simultaneous congruences

$$n \equiv 8 \pmod{15}$$
$$n \equiv 5 \pmod{7}.$$

Since $\gcd(15, 7) = 1$, we can now apply the Chinese Remainder Theorem to this pair of congruences. The solution turns out to be the set of all n such that

$$n \equiv 68 \pmod{105},$$

and we leave the details of verification to the reader.

We now turn to the general problem of solving a system of simultaneous congruences with k congruences, for an arbitrary positive integer k. We will not prove the following generalization of the Chinese Remainder Theorem, in which all pairs of moduli for the k congruences are coprime. To prove this result, generalize the approach used in the above example with three simultaneous congruences, and apply Mathematical Induction on k, together with the Chinese Remainder Theorem.

Theorem 17   **(Generalized Chinese Remainder Theorem (GCRT))**

For all positive integers k and $m_1, m_2, \ldots, m_k$, and integers $a_1, a_2, \ldots, a_k$, if $\gcd(m_i, m_j) = 1$ for all $i \neq j$, then the simultaneous congruences

$$n \equiv a_1 \pmod{m_1}$$
$$n \equiv a_2 \pmod{m_2}$$
$$\vdots$$
$$n \equiv a_k \pmod{m_k}$$

have a unique solution modulo $m_1 m_2 \cdots m_k$. Thus, if $n = n_0$ is one particular solution, then the solutions are given by the set of all integers n such that

$$n \equiv n_0 \pmod{m_1 m_2 \ldots m_k}.$$

EXERCISE

Solve the problem posed by Sun Zi that was discussed at the beginning of this section.

You should ask yourself "What happens if the moduli are *not* coprime?" That is a very interesting problem, but it is not covered in these notes.

8.9   Splitting a Modulus

In this section, we consider a surprisingly powerful result for a pair of simultaneous congruences in a special case.

Theorem 18 **(Splitting Modulus Theorem (SMT))**

For all integers a and positive integers m_1 and m_2, if $\gcd(m_1, m_2) = 1$, then the simultaneous congruences

$$n \equiv a \pmod{m_1}$$
$$n \equiv a \pmod{m_2}$$

have exactly the same solutions as the single congruence $n \equiv a \pmod{m_1 m_2}$.

Proof: Let a be an arbitrary integer, and m_1 and m_2 be arbitrary positive integers. Assume that $\gcd(m_1, m_2) = 1$. Therefore we can apply the Chinese Remainder Theorem and thus we obtain that the solutions to the simultaneous congruences are all integers n such that $n \equiv n_0 \pmod{m_1 m_2}$, where n_0 is one particular solution.

Now observe that $n_0 = a$ is a suitable choice of n_0, since of course $a \equiv a \pmod{m_1}$ and $a \equiv a \pmod{m_2}$. Hence these simultaneous congruences have exactly the same solutions as the single congruence $n \equiv a \pmod{m_1 m_2}$. □

Example 22   Find all integers x such that $x^3 + x^2 \equiv 26 \pmod{35}$.

Solution: This is a non-linear congruence in the variable x, and one method to solve it would be to check the 35 possible values of x modulo 35. However, carrying out these computations by hand would be quite involved, so we will use an alternative method of solution.

Observe that $35 = 5 \times 7$ and $\gcd(5, 7) = 1$, so we can apply the Splitting Modulus Theorem (with $m_1 = 5$ and $m_2 = 7$). This theorem says that the single non-linear congruence $x^3 + x^2 \equiv 26 \pmod{35}$ has exactly the same solutions as the simultaneous *non-linear* congruences

$$x^3 + x^2 \equiv 26 \equiv 1 \pmod{5}$$
$$x^3 + x^2 \equiv 26 \equiv 5 \pmod{7}.$$

Now we solve these two congruences separately. For the first congruence we use the table

$x \pmod 5$	0	1	2	3	4
$x^2 \pmod 5$	0	1	4	4	1
$x^3 \pmod 5$	0	1	3	2	4
$x^3 + x^2 \pmod 5$	0	2	2	1	0

We see from the last row that the solution to the congruence $x^3 + x^2 \equiv 1 \pmod{5}$ is the set of integers x such that $x \equiv 3 \pmod{5}$.

For the second congruence we use the table

$x \pmod 7$	0	1	2	3	4	5	6
$x^2 \pmod 7$	0	1	4	2	2	4	1
$x^3 \pmod 7$	0	1	1	6	1	6	6
$x^3 + x^2 \pmod 7$	0	2	5	1	3	3	0

We see from the last row that the solution to the congruence $x^3 + x^2 \equiv 5 \pmod 7$ is the set of integers x such that $x \equiv 2 \pmod 7$.

Putting the solutions to these two non-linear congruences together, we need to solve the simultaneous *linear* congruences

$$x \equiv 3 \pmod 5$$
$$x \equiv 2 \pmod 7.$$

Since $\gcd(5, 7) = 1$, we can apply the Chinese Remainder Theorem and thus we obtain that the solution to these simultaneous linear congruences is the set of all integers x such that $x \equiv n_0 \pmod{35}$, where n_0 is one particular solution.

To find a suitable value for n_0, note that $n_0 = 2, 9, 16, 23, 30$ are the choices of integers between 0 and 34 that are congruent to 2 $\pmod 7$. Now we observe that $2 \equiv 2 \pmod 5$, $9 \equiv 4 \pmod 5$, $16 \equiv 1 \pmod 5$, $23 \equiv 3 \pmod 5$ and $30 \equiv 0 \pmod 5$, so $n_0 = 23$ is a solution to both linear congruences simultaneously. We conclude that the set of solutions for the simultaneous linear congruences, and therefore for the original set of non-linear congruences, consists of all integers x such that

$$x \equiv 23 \pmod{35}.$$

REMARK

Note how we use the Splitting Modulus Theorem in the above example. We are given a problem involving a single congruence modulo $m_1 m_2$, where m_1 and m_2 are coprime, and apply the Splitting Modulus Theorem to replace the problem by an equivalent pair of problems involving congruences modulo m_1 and m_2. We then solve both of these problems, and combine the results back into a single congruence modulo $m_1 m_2$ via the Chinese Remainder Theorem.

In the process described above, the fact that we *split* the problem modulo $m_1 m_2$ into a pair of problems modulo m_1 and m_2 is referred to as "splitting" the modulus $m_1 m_2$ into the pair m_1 and m_2. This is the origin of the name "Splitting Modulus Theorem".

In the previous example of the Splitting Modulus Theorem, where we use a purely computational approach, the advantage of splitting the modulus 35 into the pair 5 and 7 is that we can replace 35 computations by $5 + 7 = 12$ computations, with smaller numbers.

In the following example of the Splitting Modulus Theorem, the advantage is quite different. In this example, we take a theoretical approach, and the advantage of splitting the modulus 77 into the pair 7 and 11 is that the latter are both prime, so we can apply Fermat's Little Theorem.

Example 23   Prove the following statement.

Let a be an integer. If $\gcd(a, 77) = 1$, then $a^{30} \equiv 1 \pmod{77}$.

Solution: Let a be an arbitrary integer, and assume that $\gcd(a, 77) = 1$. Observe that $77 = 7 \times 11$ and $\gcd(7, 11) = 1$, so we can apply the Splitting Modulus Theorem (with $m_1 = 7$ and $m_2 = 11$). This says that the single congruence $a^{30} \equiv 1 \pmod{77}$ holds if and only if the simultaneous congruences

$$a^{30} \equiv 1 \pmod 7$$
$$a^{30} \equiv 1 \pmod{11}$$

hold. Hence we can prove the single congruence $a^{30} \equiv 1 \pmod{77}$ by proving *both* of the congruences $a^{30} \equiv 1 \pmod 7$ and $a^{30} \equiv 1 \pmod{11}$, and that is how we will proceed.

Since $\gcd(a, 77) = 1$, by Bézout's Lemma, there exist integers s and t such that

$$as + 77t = 1.$$

Hence $as + 7(11t) = 1$, so from the proposition Comprimeness Characterization Theorem (CCT), we have $\gcd(a, 7) = 1$. Similarly, $as + 11(7t) = 1$, so again from CCT we have $\gcd(a, 11) = 1$.

Since $\gcd(a, 7) = 1$ and 7 is prime, by Fermat's Little Theorem we obtain $a^6 \equiv 1 \pmod 7$. Then by the proposition Congruence Power, taking the 5th power on both sides of the congruence, we obtain
$$a^{30} \equiv 1 \pmod 7.$$

Similarly, since $\gcd(a, 11) = 1$ and 11 is prime, by Fermat's Little Theorem we obtain $a^{10} \equiv 1 \pmod{11}$. Then by the proposition Congruence Power, cubing both sides of the congruence, we obtain
$$a^{30} \equiv 1 \pmod{11}.$$

REMARK

Note that in Examples 22 and 23 we have used the Splitting Modulus Theorem (SMT) with $x^3 + x^2$ and a^{30}, respectively, on the left hand side of the congruences. This may seem unjustified, since the SMT is stated in a way that looks quite different, with a single variable called n on the left hand side of the congruences.

However, the key fact about the SMT is that the *same* integer appears on the left hand sides of all three congruences - the simultaneous congruences modulo m_1 and m_2, and the single congruence modulo $m_1 m_2$. It doesn't matter what that integer is called, or how it is formed. Hence, in Examples 22 and 23, observe that $x^3 + x^2$ and a^{30} are integers for any integer choices of x and a, respectively, and hence our use of the SMT is indeed justified.

Chapter 9

The RSA Public-Key Encryption Scheme

9.1   Public-Key Cryptography

The need for secret communications has been appreciated for millenia. In the modern world, the need for secret communications is much greater than it was even in the recent past. Certainly the traditional requirements of military and diplomatic secrecy continue, but the credit card, debit card and web transactions of modern e-commerce, as well as privacy concerns for health, citizenship and other electronic records, have dramatically raised the need for secure communications and storage.

In its most elemental form, the objective of any secret communications scheme is to allow two parties, usually referred to as **Alice** (for person A) and **Bob** (for person B), to achieve **confidentiality** when they communicate over an insecure channel such as the internet. In other words, an eavesdropper **Eve** who observes the messages that Alice and Bob exchange should be unable to understand what is actually being communicated. The information that Alice wishes to communicate is called the **plaintext**. The act of transforming the plaintext into unintelligible **ciphertext** is called **encryption**. The encryption process uses an **encryption key**. The act of transforming the ciphertext to plaintext is called **decryption**. The decryption process uses a **decryption key**.

In traditional **symmetric-key encryption schemes**, the encryption key and decryption key are identical; that is, the same key k is used for encryption and decryption. Clearly k must be kept secret by the two communicating parties, Alice and Bob. This raises the problem of how Alice and Bob are to securely agree upon k in the first place.

Consider the example where Alice is an Outlook user and Bob is the Outlook web site (`https://outlook.live.com`) that is managed by Microsoft. When Alice wishes to read her email, she identifies herself to the Outlook web site by providing her username and password. Since Alice's password is sent over the internet to Bob, the password needs to be encrypted before it is transmitted. But if Alice uses a symmetric-key encryption scheme to encrypt her password with a secret key k, how can she securely transport k over the internet to Bob so that Bob can subsequently decrypt Alice's ciphertext and recover her password? This is known as the **key distribution problem**.

Imagine also the complexity of the **key management problem** that Microsoft might face of managing a huge number of secret keys, i.e., a different key k for each of its several hundred millions of Outlook users.

The revolutionary idea of **public-key cryptography** is to separate encryption keys from decryption keys. Each participant B has a decryption key which they hold secretly, and a related encryption key which they share in a public repository of some sort. Even though B's decryption key and encryption key are mathematically related, it should be computationally infeasible for an eavesdropping adversary to determine B's decryption key from B's encryption key. Now, if any user A wishes to send a confidential message M to user B, A would somehow obtain an authentic copy of B's public encryption key, encrypt M with this public key, and send the resulting ciphertext C to B. Since B is the only person who possesses the private decryption key, only B can decrypt C to recover M.

Such an arrangement alleviates the key distribution problem described above. The public keys need to be authenticated, but do not need to be kept secret. Moreover, each participant needs only one public key, thus solving the key management problem.

REMARK

In the Outlook example, Microsoft would send its *certificate* over the internet to the Outlook user. The certificate contains Microsoft's public key, and is signed by a Certification Authority using the RSA signature scheme (which is not covered in these notes). The Outlook user validates the signature in the certificate using the Certification Authority's public verification key which is preinstalled in the user's web browser. If the signature is valid, then the Outlook user is assured that it has an authentic copy of Microsoft's public key.

The notion of public-key cryptography was first proposed in 1975 by Whitfield Diffie, Martin Hellman and Ralph Merkle. The RSA scheme, proposed in 1977 and named after its inventors Ron Rivest, Adi Shamir and Len Adleman, is an example of a commercially implemented public-key encryption scheme.

RSA is now widely deployed. The following protocols and products, which embed RSA, are used by many of us daily. SSL (Secure Sockets Layer) is the most commonly used protocol for secure communication over the web. It is frequently used to encrypt user passwords before sending that data to a server. PGP (Pretty Good Privacy) is used by individuals and businesses to encrypt and authenticate messages. It was originally intended for email messages and attachments but is now also used for encrypting files, folders and entire hard drives. EMV (Europay, MasterCard and VISA) is a global standard for authenticating credit and debit card transactions at point of sale (POS) or automated teller machines (ATM).

In the next section, we will introduce the RSA public-key encryption scheme and then prove that it works. The security of RSA is a widely studied subject, but we will not address it here.

9.2   Implementing the RSA Scheme

In RSA, plaintexts are integers. How does one get an integer from an English message? We assign a number to each letter of the alphabet and then concatenate the digits together. If the message is too long, it may need to be broken into parts (so that each part M is less than n; see *RSA Encryption* below).

We now describe the implementation of the RSA public-key encryption scheme, in which Alice sends a message to Bob. There are three stages to RSA implementation, referred to as follows:

(a) Setting up RSA

(b) RSA Encryption

(c) RSA Decryption

The three stages are described below.

(a) Setting up RSA: To set up the RSA encryption scheme, Bob does the following.

1. Randomly choose two large, distinct primes p and q and let $n = pq$.

2. Select an arbitrary integer e so that $\gcd(e, (p-1)(q-1)) = 1$ and $1 < e < (p-1)(q-1)$.

3. Solve the congruence
$$ed \equiv 1 \pmod{(p-1)(q-1)}$$
   for an integer d where $1 < d < (p-1)(q-1)$.

4. Publish the public key (e, n).

5. Keep secret the private key (d, n), and the primes p and q.

(b) RSA Encryption: To encrypt a message as ciphertext and send securely to Bob, Alice does the following.

1. Obtain an authentic copy of Bob's public key (e, n).

2. Construct the plaintext message M, an integer such that $0 \leq M < n$.

3. Encrypt M as the ciphertext C, given by

$$C \equiv M^e \pmod{n} \text{ where } 0 \leq C < n.$$

4. Send C to Bob.

(c) RSA Decryption: To decrypt the ciphertext received from Alice, Bob does the following.

1. Use the private key (d, n) to decrypt the ciphertext C as the received message R, given by
$$R \equiv C^d \pmod{n} \text{ where } 0 \leq R < n.$$

2. *Claim:* The received message R equals the original plaintext message M, i.e., $R = M$.

REMARK

In order for RSA to be considered secure, it must be the case that it is computationally infeasible for an adversary to compute Bob's private key (d, n) from his public key (e, n). Hence it is vital that the adversary should be incapable of factoring n (since otherwise the adversary would find p and q and thereafter compute d in the same way as Bob). As of 2018, it is recommended that the primes p and q each be chosen to be at least 300 decimal digits long, so that n is at least 600 decimal digits long. This is because factoring such 600-decimal digit numbers n is well beyond the reach of the fastest known factoring algorithms and the world's fastest computers.

For our first example of the computations involved in carrying the RSA scheme, we consider a very small pair of distinct primes p and q, in order to give all details "by hand".

Example 1   Carry out the following calculations for the RSA scheme with $p = 5$, $q = 11$ and $e = 3$.

1. Determine the private key (d, n).

   Solution: In this case, $n = 5 \times 11 = 55$ and $(p - 1)(q - 1) = 4 \times 10 = 40$. To find d, we solve
   $$3d \equiv 1 \pmod{40}.$$

   To do so, we set up the Linear Diophantine Equation
   $$40x + 3d = 1$$

   and use the Extended Euclidean Algorithm

x	d	r	q
1	0	40	0
0	1	3	0
1	-13	1	13
-3	40	0	3

   Hence our solution for d is
   $$d \equiv -13 \pmod{40}.$$

   Since d must satisfy $1 < d < 40$, we obtain $d = 40 - 13 = 27$, so the private key is the pair $(d, n) = (27, 55)$.

2. Suppose Bob receives the ciphertext $C = 47$. Decrypt C to determine the message M that was encrypted by Alice.

   Solution: We wish to compute $R = 47^{27} \pmod{55}$. To simplify this computation, note that 5 and 11 are coprime, so by the Splitting Modulus Theorem, we can obtain R as the unique solution to the simultaneous congruences

   $$R \equiv 47^{27} \pmod{5}$$
   $$\text{and} \quad R \equiv 47^{27} \pmod{11}.$$

   Now $47 \equiv 2 \pmod{5}$ and $47 \equiv 3 \pmod{11}$, therefore we have $R \equiv 2^{27} \pmod{5}$ and $R \equiv 3^{27} \pmod{11}$. Since 5 and 11 are both prime numbers, we can apply Fermat's

Little Theorem (FℓT), which gives $2^4 \equiv 1 \pmod 5$ and $3^{10} \equiv 1 \pmod{11}$. Hence, from $27 = (6)(4) + 3$, we obtain

$$R \equiv 2^{27} \equiv (2^4)^6 2^3 \equiv (1)^6 2^3 \equiv 2^3 \equiv 8 \equiv 3 \pmod 5.$$

Similarly, from $27 = 2(10) + 7$, we obtain

$$R \equiv 3^{27} \equiv (3^{10})^2 (3)^7 \equiv (1)^2 3^7 \equiv 3^7 \equiv (9)^3 3 \equiv (-2)^3 3 \equiv 9 \pmod{11}.$$

Therefore, we have to solve the simultaneous congruences

$$R \equiv 3 \pmod 5$$
$$\text{and} \quad R \equiv 9 \pmod{11}.$$

Again note that 5 and 11 are coprime, and that these simultaneous congruences are linear, so by the Chinese Remainder Theorem, there is a unique solution modulo $5 \times 11 = 55$. To shorten our work, a quick check shows that $53 \equiv 3 \pmod 5$ and $53 \equiv 9 \pmod{11}$, and so the unique solution to these simultaneous congruences is given by $R \equiv 53 \pmod{55}$. Hence we have $M = 53$.

REMARK

Note in the above example that Bob knows the values of the primes p and q individually when he is carrying out the decryption. Hence he is able to apply the Splitting Modulus Theorem and Fermat's Little Theorem to make calculations modulo the primes p and q, and then combine the results via the Chinese Remainder Theorem to determine the final answer modulo n, where $n = pq$. Performing these computations modulo p and q is significantly faster than working modulo n.

Note also that Bob is the only one who knows the values of p and q individually. The value of n is known to everyone as part of the public key, and therefore to any eavesdropping adversary who might want to decrypt messages that are sent by others using RSA. This explains why the infeasibility of factoring a known n into its unknown prime factors p and q is critical to the security of the RSA scheme. Without knowing the values of p and q, an adversary is unable to take advantage of the computational simplifications of working modulo p and q, and in particular is unable to determine the private key (d, n).

We now give another example of RSA computations, in this case for 50 decimal digit primes p and q. Of course, the computations are not feasible by hand, and we have used a computer to carry them out, using Maple. You may wish to check these computations using your own preferred software. Note however, that though the numbers involved in this example are large, they are still much smaller than those used for current secure applications, where the primes are at least 300 decimal digits.

Example 2   Let the prime p be
9026694843 0929817462 4847943076 6619417461 5791443937,
and the prime q be

```
7138718791 1693596343 0802517103 2405888327 6844736583,
```
so $n = pq$ is equal to
```
6443903609 8539423089 8003779070 0502485677 1034536315
4526254586 6290164606 1990955188 1922989980 3977447271,
```
and $(p-1)(q-1)$ is equal to
```
6443903609 8539423089 8003779070 0502485677 1034536313
8360840952 3666750800 6340495008 2897684191 1341266752.
```
Choose e with $\gcd(e, (p-1)(q-1)) = 1$ and $1 < e < (p-1)(q-1)$ to be
```
9573596212 0300597326 2950869579 7174556955 8757345310 2344121731.
```
Solving the LDE

$$ed \equiv 1 \pmod{(p-1)(q-1)}$$

for d with $1 < d < (p-1)(q-1)$, we obtain that d is equal to
```
5587652122 6351022927 9795248536 5522717791 7285682675
6100082011 1849030646 3274981250 2583120946 4072548779.
```
Choose the integer message M with $0 \le M < n$ to be

$$M = 3141592653.$$

Computing the ciphertext $C \equiv M^e \pmod{n}$, with $0 \le C < n$, we obtain that C is equal to
```
4006696554 3080815610 2814019838 8509626485 8151054441
5245547382 5506759308 1333888622 4491394825 3742205367.
```

Finally, computing $R \equiv C^d \pmod{n}$, with $0 \le R < n$, we obtain

$$R = 3141592653.$$

Checking, we confirm that indeed $R = M$.

9.3 Proving that the RSA Scheme Works

Now that we have seen two examples of RSA and the associated computations, we prove in the following result that the RSA scheme always works. What we mean by this is that we prove the *Claim: $R = M$*, that the plaintext message M and the decrypted message received R are identical.

Theorem 1 **(RSA Works (RSA))**

For all integers p, q, n, e, d, M, C and R, if

1. p and q are distinct primes,

2. $n = pq$,

3. e and d are positive integers such that $ed \equiv 1 \pmod{(p-1)(q-1)}$ and $1 < e, d < (p-1)(q-1)$,

4. $0 \le M < n$,

5. $M^e \equiv C \pmod{n}$ where $0 \le C < n$,

6. $C^d \equiv R \pmod{n}$ where $0 \le R < n$,

then $R = M$.

Proof: Let p, q, n, e, d, M, C and R be arbitrary integers, and assume that they satisfy parts $1 - 6$ of the hypothesis. Now, from parts 5 and 6 of the hypothesis, we have

$$R \equiv C^d \equiv \left(M^e\right)^d \equiv M^{ed} \pmod{n}.$$

Since p and q are distinct primes, they must be coprime. Therefore, since $n = pq$, we can apply the Splitting Modulus Theorem to obtain that

$$R \equiv M^{ed} \pmod{p}$$

and

$$R \equiv M^{ed} \pmod{q}.$$

Now, we prove that $R \equiv M \pmod{p}$, by considering the two cases $p \mid M$ and $p \nmid M$.

- **Case 1:** If $p \mid M$, then we have $M \equiv 0 \pmod{p}$, and therefore,

$$R \equiv 0^{ed} \equiv 0 \pmod{p}.$$

Hence in this case both R and M are congruent to 0 modulo p, giving $R \equiv M \pmod{p}$.

- **Case 2:** If $p \nmid M$, then p and M are coprime, so by Fermat's Little Theorem, we have

$$M^{p-1} \equiv 1 \pmod{p}. \tag{9.1}$$

From part 3 of the hypothesis and the definitions of congruence and divisibility, there exists an integer k such that

$$ed = 1 + k(p-1)(q-1).$$

Moreover, since $ed > 1$ and $p - 1$ and $q - 1$ are positive integers, it must be the case that k is a positive integer. Putting these together, we obtain

$$R \equiv M^{1+k(p-1)(q-1)} \pmod{n},$$

for some positive integer k.

Substituting (9.1) gives

$$R \equiv M\left(M^{p-1}\right)^{k(q-1)} \equiv M(1)^{k(q-1)} \equiv M \pmod{p},$$

and hence in this case we also have $R \equiv M \pmod{p}$.

Since it is true in both cases, we conclude that $R \equiv M \pmod{p}$.

It is straightforward to modify the above proof (simply by interchanging p and q) to prove that $R \equiv M \pmod{q}$. We omit this proof, and leave it as an exercise for the reader.

Summarizing, we have established that R and M satisfy the following simultaneous congruences:

$$R \equiv M \pmod{p},$$
$$R \equiv M \pmod{q}.$$

Applying the Splitting Modulus Theorem again we obtain

$$R \equiv M \pmod{n}.$$

Finally, from parts 4 and 6 of the hypothesis, we know that R and M both lie between 0 and $n - 1$, which gives the required conclusion $R = M$.   □

REMARK

In our description of RSA implementation and in the above theorem RSA Works, one of the restrictions is $e > 1$. This restriction is not needed for the mathematics to hold – certainly the above proof for the conclusion $R = M$ of RSA Works continues to be valid if the hypotheses were to allow the choice $e = 1$ (in which case $d = 1$ also).

However, the reason that the choice $e = 1$ is excluded in the RSA implementation is that it is completely insecure. Bob can't publish $e = 1$ as part of the public key, because that would mean that the plaintext M would be "encrypted" as the ciphertext $C = M$, for all M. That is, there would be no actual encryption for any ciphertext M!

Chapter 10

Complex Numbers

10.1 Standard Form

When humans first counted, we tallied, literally making notches on sticks, stones and bones. Thus the natural numbers $\mathbb{N}$ were born. But it wouldn't be long before the necessity for fractions became obvious. For example, sharing one animal between four people (we will assume fairly), naturally involves the notion of $1/4$, and hence the equation

$$4x = 1,$$

though it would not have been expressed in these terms originally. For these reasons, we would have had to extend our notion of numbers to include fractions, leading to the rationals

$$\mathbb{Q} = \left\{ \frac{a}{b} : a, b \in \mathbb{Z}, b \neq 0 \right\}.$$

(This is an overstatement historically, because recognition of zero and negative numbers, which are permitted in $\mathbb{Q}$, was very slow to come.) But even these new numbers would not help solve equations of the form

$$x^2 = 2,$$

which would arise naturally for the length of the hypotenuse of a right angled triangle with two sides of length 1. Thus, the notion of number had to be extended to include irrational numbers, which, combined with the rationals, give us the set of real numbers $\mathbb{R}$.

Eventually, via Hindu and Islamic scholars, western mathematics began to recognize and accept both zero and negative numbers. Otherwise, equations like

$$3x = 5x, \qquad \text{or} \qquad 2x + 4 = 0,$$

would have no solution. Hence, mathematicians recognized that

$$\mathbb{N} \subsetneq \mathbb{Z} \subsetneq \mathbb{Q} \subsetneq \mathbb{R},$$

but even $\mathbb{R}$ was inadequate because equations of the form

$$x^2 + 1 = 0$$

have no real solutions. So, once again, our number system was extended to the set of *complex numbers* $\mathbb{C}$, with $\mathbb{R} \subsetneq \mathbb{C}$, defined below.

Definition 10.1.1

complex number,
real part, imaginary
part, equality

A **complex number** z in **standard form** is an expression of the form $z = x + yi$ where $x, y \in \mathbb{R}$. The real number x is called the **real part** of z, and is written $\mathrm{Re}(z)$. The real number y is called the **imaginary part** of z, and is written $\mathrm{Im}(z)$. The set of all complex numbers is denoted by

$$\mathbb{C} = \{x + yi : x, y \in \mathbb{R}\}.$$

The complex numbers $z = x + yi$ and $w = u + vi$ are **equal** (written $z = w$), if and only if $x = u$ and $y = v$.

Example 1

We have the following examples for complex numbers in standard form.

- For $z_1 = \sqrt{2} + \frac{3}{4}i$, we have $\mathrm{Re}(z_1) = \sqrt{2}$ and $\mathrm{Im}(z_1) = \frac{3}{4}$.

- For $z_2 = 3 + 0i$, we have $\mathrm{Re}(z_2) = 3$ and $\mathrm{Im}(z_2) = 0$. Complex numbers with imaginary part equal to 0 are said to be *purely real*, and are often written without the "$0i$". Hence z_2 is purely real, and we can also write it as $z_2 = 3$.

- For $z_3 = 0 + 5i$, we have $\mathrm{Re}(z_3) = 0$ and $\mathrm{Im}(z_3) = 5$. Complex numbers with real part equal to 0 are said to be *purely imaginary*, and are often written without the "0". Hence z_3 is purely imaginary, and we can also write it as $z_3 = 5i$.

- The complex number $1 = 1 + 0i$ is purely real.

- The complex number $i = 0 + 1i = 1i$ is purely imaginary.

- The complex number $0 = 0 + 0i = 0i$ is both purely real and purely imaginary!

We now formally define rules for the addition and multiplication of complex numbers, so we can do *complex arithmetic*.

Definition 10.1.2

addition and
multiplication of
complex numbers

Let $z = a + bi$ and $w = c + di$ be complex numbers.

- We define **addition** by

$$z + w = (a + c) + (b + d)i$$

- We define **multiplication** by

$$z\,w = (ac - bd) + (ad + bc)i$$

Example 2

We have the following examples of adding and multiplying complex numbers.

(a) $(1 + 7i) + (2 - 3i) = (1 + 2) + (7 - 3)i = 3 + 4i$

(b) $(1 + 7i)(2 - 3i) = (2 - 7(-3)) + (-3 + 7(2))i = 23 + 11i$

(c) $i^2 = i\,i = (0 + 1i)(0 + 1i) = (0 - 1) + (0 + 0)i = -1$

REMARK

In part (c) of Example 2, we showed that $i^2 = -1$ using the formal definition of multiplication for complex numbers. In fact, an informal but equivalent definition of multiplication for complex numbers is to expand products of sums in the usual way, with the special rule that $i^2 = -1$. For example, using this approach, we obtain

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

in agreement with the formal rule given in Definition 10.1.2.

Example 3   In complex arithmetic, the following properties hold for all $z \in \mathbb{C}$:

1. $z + 0 = 0 + z = z$,

2. $z\,0 = 0\,z = 0$,

3. $z + (-1)z = (-1)z + z = 0$,

4. $z\,1 = 1\,z = z$.

The proofs of these properties are straightforward from the definitions of addition and multiplication of complex numbers. For example, to prove the $z + 0 = z$ part of Property 1, letting $z = a + bi$ with $a, b \in \mathbb{R}$, we have

$$z + 0 = (a + bi) + (0 + 0i) = (a + 0) + (b + 0)i = a + bi = z.$$

Property 1 in Example 3 says that 0 is the *additive identity* in $\mathbb{C}$, and Property 3 says that $(-1)z$ is the *additive inverse* of z, for all $z \in \mathbb{C}$. For this reason, we usually write $(-1)z$ as $-z$. Property 4 says that 1 is the *multiplicative identity* in $\mathbb{C}$.

We know from our earlier work with congruences that the *multiplicative inverse* of a complex number z is a complex number, that we write as z^{-1}, with the property that $z\,z^{-1} = 1$ (where 1 appears on the right hand side of this equation because it is the multiplicative identity for the complex numbers). In the next result we determine when the multiplicative inverse exists, and if so, what it is.

Proposition 1   For all complex numbers z, the multiplicative inverse of z exists if and only if $z \neq 0$. Moreover, for $z = a + bi \neq 0$, the multiplicative inverse is unique, and is given by

$$z^{-1} = \frac{a}{a^2 + b^2} - \frac{b}{a^2 + b^2}i = \frac{a - bi}{a^2 + b^2}.$$

Proof: Let $z = a + bi$ be an arbitrary complex number, and suppose that the multiplicative inverse $z^{-1} = c + di$ exists. Then we have $(a + bi)(c + di) = 1$, and expanding out the product on the left hand side, we obtain

$$(ac - bd) + (ad + bc)i = 1 + 0i.$$

Hence, for each pair a and b, the real numbers c and d satisfy the simultaneous equations (one equation for the real part, one for the imaginary part)

$$ac - bd = 1, \tag{10.1}$$

$$ad + bc = 0. \tag{10.2}$$

Now, assume that $z = 0$, so $a = b = 0$. Then, for all real numbers c and d, equation (10.1) becomes $0 = 1$ which is certainly false. Hence there is no pair of real numbers c and d that satisfies equation (10.1), so we conclude that $z = 0$ has no multiplicative inverse.

We continue the proof by assuming that $z \neq 0$, so a and b are not both zero. Then, multiplying equation (10.1) on both sides by a, and multiplying equation (10.2) on both sides by b, and then adding the resulting equations and rearranging, we obtain

$$(a^2 + b^2)c = a. \tag{10.3}$$

Since a and b are real numbers, we have $a^2 \geq 0$, with equality only when $a = 0$, and $b^2 \geq 0$, with equality only when $b = 0$. Adding these inequalities, we get $a^2 + b^2 \geq 0$, with equality only when $a = b = 0$. But we have assumed that a and b are not both zero, so we have $a^2 + b^2 \neq 0$. Hence we can divide on both sides of equation (10.3), which gives $c = \frac{a}{a^2+b^2}$.

To obtain d, multiply equation (10.1) on both sides by $-b$, and multiply equation (10.2) on both sides by a, and then add the resulting equations and rearrange, to get

$$(a^2 + b^2)d = -b.$$

Now, dividing this equation on both sides by $a^2 + b^2$, we obtain $d = \frac{-b}{a^2+b^2}$.

Hence we have proved that, if z^{-1} exists, it must be $c + di = \frac{a-bi}{a^2+b^2}$. Checking to make sure this is not an extraneous solution, we confirm that $(a + bi)\frac{a-bi}{a^2+b^2} = \frac{a^2-b^2i^2}{a^2+b^2} = 1$, completing the proof. $\qquad\square$

The above result tells us that the only complex number without a multiplicative inverse is 0. This is the same as for the real numbers, where it is a standard fact that we can divide by any real number as long as it isn't equal to 0. In fact, complex arithmetic has many properties in common with real arithmetic (and rational arithmetic), collected below. The proofs of the parts that we haven't given already follow straightforwardly from the definitions of addition and multiplication, and are therefore omitted.

Proposition 2 **(Properties of Complex Arithmetic (PCA))**

In complex arithmetic, the following properties hold for all $u, v, z \in \mathbb{C}$:

1. Associativity of addition: $(u + v) + z = u + (v + z)$

2. Commutativity of addition: $u + v = v + u$

3. Additive identity: $0 = 0 + 0i$ has the property that $z + 0 = z$

4. Additive inverses: If $z = a + bi$ then there exists an *additive inverse* of z, written $-z$ with the property that $z + (-z) = 0$. The additive inverse of $z = a + bi$ is $-z = -a - bi$.

5. Associativity of multiplication: $(u\,v)\,z = u\,(v\,z)$

6. Commutativity of multiplication: $u\,v = v\,u$

7. Multiplicative identity: $1 = 1 + 0i$ has the property that $z\,1 = z$.

8. Multiplicative inverses: If $z = a + bi \neq 0$ then there exists a *multiplicative inverse* of z, written z^{-1}, with the property that $z\,z^{-1} = 1$. The multiplicative inverse of $z = a + bi \neq 0$ is $z^{-1} = \frac{a-bi}{a^2+b^2}$.

9. Distributivity: $z\,(u + v) = z\,u + z\,v$

REMARK

In chapter 4 we introduced the Binomial Theorem. Version 1 was stated in terms of the real variable x. Version 2 was stated in terms of the real variables a and b. It is easy to verify that our proofs of these theorems still hold for complex variables x, a and b. Therefore, we state a more general version of the Binomial Theorem.

Proposition 3   **(Binomial Theorem (BT))**

For all integers $n \geq 0$ and for all complex numbers a and b,

$$(a + b)^n = \sum_{m=0}^{n} \binom{n}{m} a^{n-m} b^m.$$

10.2   Conjugate and Modulus

Given a complex number, we associate with it a complex number called the complex conjugate, which we now define.

Definition 10.2.1   The complex **conjugate** of a complex number $z = x + yi$, written $\bar{z}$, is the complex number

conjugate

$$\bar{z} = x - yi.$$

Often we refer to the complex conjugate as simply the *conjugate*. For example, the conjugate of $z = 2 + 3i$ is $\bar{z} = 2 - 3i$, and the conjugate of $w = -\sqrt{5}\,i$ is $\bar{w} = \sqrt{5}\,i$.

Proposition 4   **(Properties of Conjugate (PCJ))**

For the complex conjugate, the following properties hold for all $z, w \in \mathbb{C}$:

1. $\overline{(\overline{z})} = z$

2. $\overline{z + w} = \overline{z} + \overline{w}$

3. $z + \overline{z} = 2\operatorname{Re}(z)$ and $z - \overline{z} = 2\operatorname{Im}(z)\,i$

4. $\overline{z\,w} = \overline{z}\,\overline{w}$

5. If $z \neq 0$, then $\overline{(z^{-1})} = (\overline{z})^{-1}$.

Proof: Let $z = a + bi$ and $w = c + di$ be arbitrary complex numbers.

1. We have $\overline{z} = a - bi$, and hence $\overline{(\overline{z})} = \overline{(a - bi)} = a - (-b)i = a + bi = z$.

2. We have $z + w = (a + c) + (b + d)i$, and hence

$$\overline{z + w} = (a + c) - (b + d)i = a + c - bi - di = (a - bi) + (c - di) = \overline{z} + \overline{w}.$$

3. From $\overline{z} = a - bi$ we obtain

$$z + \overline{z} = (a + bi) + (a - bi) = 2a = 2\operatorname{Re}(z),$$
$$z - \overline{z} = (a + bi) - (a - bi) = 2bi = 2\operatorname{Im}(z)\,i.$$

4. We have $z\,w = (ac - bd) + (ad + bc)i$, and hence

$$\overline{z\,w} = (ac - bd) - (ad + bc)i = (a - bi)(c - di) = \overline{z}\,\overline{w}.$$

5. Assume that $z \neq 0$, so $\overline{z} \neq 0$, and from Proposition 1, z^{-1} exists. Let $w = z^{-1}$, so $z\,w = 1$, and from part 4 of this result we obtain

$$\overline{z}\,\overline{w} = \overline{z\,w} = \overline{1} = \overline{1 + 0i} = 1 - 0i = 1.$$

Dividing on both sides of the equation $\overline{z}\,\overline{w} = 1$ by $\overline{z} \neq 0$, we have $\overline{w} = (\overline{z})^{-1}$. $\square$

REMARK

Consider what part 3 of the proposition Properties of Conjugate says about when a complex number is purely real $(\operatorname{Im}(z) = 0)$ or purely imaginary $(\operatorname{Re}(z) = 0)$. Part 3 says that z is *purely real* if and only if $z - \overline{z} = 0$, or, equivalently, $z = \overline{z}$. Also, part 3 says that z is *purely imaginary* if and only if $z + \overline{z} = 0$, or, equivalently, $z = -\overline{z}$.

Given a complex number, we also associate with it a non-negative real number called the modulus, which we now define.

Definition 10.2.2

modulus

The **modulus** of the complex number $z = x + yi$, written $|z|$, is the non-negative real number

$$|z| = \sqrt{x^2 + y^2}.$$

Note the important role of non-negativity is in the definition of modulus. First, both x and y are real numbers, so $x^2 \geq 0$ and $y^2 \geq 0$, and we can add these inequalities to obtain $x^2 + y^2 \geq 0$. This means that $|z|$ for $z = x + yi$ is *well-defined*, since we can always take the square root of a non-negative quantity. Second, when we take the square root, we use the standard mathematical meaning, that $\sqrt{0} = 0$, and that otherwise we take the *positive* root. Thus $|z|$ evaluates to a unique non-negative real number for every complex number z.

For example, the modulus of $z = 2 - 5i$ is $|z| = \sqrt{(2^2) + (-5)^2} = \sqrt{29}$, the modulus of $w = -3$ is $|w| = \sqrt{(-3)^2 + 0^2} = \sqrt{9} = 3$, and the modulus of $s = i$ is $|s| = \sqrt{0^2 + 1^2} = 1$.

REMARK

The two vertical bars used for modulus is the same notation that is used for absolute value. When you encounter this two vertical bars notation, you will generally be able to determine the meaning from the context – for a real number x, $|x|$ means the *absolute value* of x, and for a complex number z, $|z|$ means the *modulus* of z.

However, there is a close relationship between the two meanings: Suppose that $z = x + 0i = x$ is a purely real complex number. Then the modulus of z is given by $|z| = \sqrt{x^2 + 0^2} = \sqrt{x^2}$. At this stage, you have to be careful! It is tempting to simply write $\sqrt{x^2} = x$, but this is certainly not correct when x is negative. Instead, we must evaluate $\sqrt{x^2}$ as the absolute value of x. Hence for purely real complex numbers $z = x + 0i$, we have

$$|z| = |x|,$$

where on the left hand side we mean the modulus of z, and on the right hand side we mean the absolute value of x. In summary, the modulus and absolute value of a purely real complex number have exactly the same value!

We now give some properties of modulus. Note that some of these properties also involve the conjugate, thus giving us relationships between these two quantities.

Proposition 5   **(Properties of Modulus (PM))**

For the modulus, the following properties hold for all $z, w \in \mathbb{C}$:

1. $|z| = 0$ if and only if $z = 0$

2. $|\bar{z}| = |z|$

3. $\bar{z} \, z = |z|^2$

4. $|z \, w| = |z| \, |w|$

5. If $z \neq 0$, then $|z^{-1}| = |z|^{-1}$.

Proof: Let $z = a + bi$ and w be arbitrary complex numbers.

1. Since a and b are real numbers, we have $a^2 \geq 0$, with equality only when $a = 0$, and $b^2 \geq 0$, with equality only when $b = 0$. Adding these two inequalities, we get $a^2 + b^2 \geq 0$, with equality only when $a = b = 0$. Therefore $|z| = \sqrt{a^2 + b^2}$ is equal to 0 if and only if $a = b = 0$, or, equivalently, $z = 0$.

2. We have $|z| = \sqrt{a^2 + b^2}$ and $\overline{z} = a - bi$, so $|\overline{z}| = \sqrt{a^2 + (-b)^2} = \sqrt{a^2 + b^2} = |z|$.

3. We have $\overline{z}\, z = (a - bi)(a + bi) = a^2 - b^2(-1) = a^2 + b^2 = |z|^2$.

4. From part 3 of this result, we get

$$|z\, w|^2 = (z\, w)\, \overline{(z\, w)} = (z\, w)\, (\overline{z}\, \overline{w}) = z\, \overline{z}\, w\, \overline{w} = |z|^2 |w|^2,$$

   where we have used $\overline{(z\, w)} = \overline{z}\, \overline{w}$ for the second equality. Taking non-negative square roots on both sides of $|z\, w|^2 = |z|^2 |w|^2$, we obtain $|z\, w| = |z|\, |w|$.

5. Assume that $z \neq 0$, so from Proposition 1, z^{-1} exists. Let $w = z^{-1}$, so $z\, w = 1$, and from part 4 of this result, we obtain

$$|z|\, |w| = |z\, w| = |1| = \sqrt{1^2 + 0^2} = \sqrt{1} = 1.$$

   We can divide on both sides of the equation $|z|\, |w| = 1$ by $|z|$, since $z \neq 0$, so by part 1 of this result, we have $|z| \neq 0$. Hence we obtain $|w| = |z|^{-1}$. $\square$

Part 5 of Properties of Modulus tells us about the modulus of the multiplicative inverse. Note that we can also rearrange part 3 to tell us about the multiplicative inverse itself: Assume $z \neq 0$, so by part 1, $|z| \neq 0$. Then dividing both sides of $\overline{z}\, z = |z|^2$ by the complex number z and by the real number $|z|^2$, we obtain

$$\frac{1}{z} = \frac{\overline{z}}{|z|^2}.$$

Note that this is exactly the same formula for z^{-1} as given in Proposition 1, except that it is rewritten using the notation for conjugate and modulus. Note also that we can replace the denominator $|z|^2$ on the right hand side of this formula by $\overline{z}\, z$, and that the resulting formula could also be simply obtained by multiplying $\frac{1}{z}$ on the top and bottom by $\overline{z}$. This provides a convenient way to "rationalize the denominator" when dividing by a complex number z, as illustrated in the following example.

Example 4   Write the complex number $z = \dfrac{3 + 4i}{2 + 5i}$ in standard form.

Solution: Multiplying on the top and bottom by $2 - 5i$, which is the conjugate of the denominator, we obtain

$$z = \frac{3 + 4i}{2 + 5i} = \frac{(3 + 4i)(2 - 5i)}{(2 + 5i)(2 - 5i)} = \frac{(6 + 20) + (-15 + 8)i}{2^2 + 5^2} = \frac{26 - 7i}{29} = \frac{26}{29} - \frac{7}{29}i.$$

Part 3 of Properties of Modulus gives a surprisingly powerful way to prove other properties of modulus. This has already been illustrated in the proof of part 4 of Properties of Modulus. Next we give another example of this, for a more complicated property of the modulus.

Example 5   Prove the following statement:

For all complex numbers z and w, we have $|z + w|^2 + |z - w|^2 = 2\left(|z|^2 + |w|^2\right)$.

Solution: Let z and w be arbitrary complex numbers. In the proof we'll make repeated use of part 2 of Properties of Conjugates and part 3 of Properties of Modulus. Thus, starting with the expression on the left hand side, we get

$$
\begin{aligned}
|z + w|^2 + |z - w|^2 &= \overline{(z + w)}(z + w) + \overline{(z - w)}(z - w) \\
&= (\overline{z} + \overline{w})(z + w) + (\overline{z} - \overline{w})(z - w) \\
&= \overline{z}\,z + \overline{z}\,w + \overline{w}\,z + \overline{w}\,w + \overline{z}\,z - \overline{z}\,w - \overline{w}\,z + \overline{w}\,w \\
&= 2\,(\overline{z}\,z + \overline{w}\,w) \\
&= 2\left(|z|^2 + |w|^2\right).
\end{aligned}
$$

We have proved a number of results about the conjugate and modulus of the product or sum of a pair of complex numbers. We will not prove the following result, that extends the results for pairs to an arbitrary natural number n of complex numbers. To prove it, apply Mathematical Induction in a straightforward way.

Corollary 6   For all positive integers n and complex numbers $z_1, z_2, \ldots, z_n$, we have

1. $\overline{z_1 + z_2 + \cdots + z_n} = \overline{z_1} + \overline{z_2} + \cdots \overline{z_n}$,

2. $\overline{z_1 z_2 \cdots z_n} = \overline{z_1}\,\overline{z_2} \cdots \overline{z_n}$,

3. $|z_1 z_2 \cdots z_n| = |z_1|\,|z_2| \cdots |z_n|$.

So far, we have given various equalities for conjugate and modulus. We end with an *inequality* for modulus, but the proof of this important result is delayed until the next section.

Proposition 7   **(Triangle Inequality (TIQ))**

For all $z, w \in \mathbb{C}$, we have $|z + w| \leq |z| + |w|$.

10.3   The Complex Plane and Polar Form

In this section we introduce a geometric representation for complex numbers, that will give us elegant and powerful methods for complex arithmetic.

Definition 10.3.1
complex plane

The complex number $z = x + yi$ can be represented by the point (x, y) in a plane whose axes are called the **real axis** (for x values) and the **imaginary axis** (for y values). With this interpretation of axes, the plane is called the **complex plane** or the **Argand plane**.

For example, the complex number $z = 3 + 2i$ corresponds to the point $(3, 2)$ in the complex plane, as illustrated in Figure 10.1.

We now consider the geometric interpretation for the conjugate and modulus of the complex number $z = x + yi$, which is represented by the point (x, y) in the complex plane.
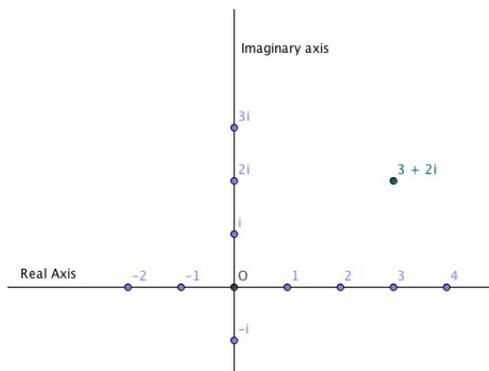
Figure 10.1: The Complex Plane

- The conjugate $\overline{z} = x - yi$ of z is represented by the point $(x, -y)$. Hence, in geometric terms, taking the *conjugate* of a complex number corresponds to *reflection* about the real axis in the complex plane.

- The modulus of z is given by $|z| = \sqrt{x^2 + y^2}$. By the Pythagorean Theorem, this gives the distance between the point (x, y) and the origin $(0, 0)$ in the complex plane. Hence, in geometric terms, the *modulus* of a complex number corresponds to the distance from the origin in the complex plane.

The above geometric interpretation of modulus leads to the following elegant proof of the Triangle Inequality for the modulus of complex numbers, which was stated but not proved at the end of the previous section.

Proof of the Triangle Inequality: Let $z = x + yi$ and $w = u + vi$ be arbitrary complex numbers. Note that $-w = -u - vi$ and $z + w = z - (-w) = (x - (-u)) + (y - (-v))i$, so

$$|z + w| = |z - (-w)| = \sqrt{(x - (-u))^2 + (y - (-v))^2}.$$

Now define three points in the complex plane: let A be the origin $(0, 0)$, let B be the point (x, y) corresponding to z, and let C be the point $(-u, -v)$ corresponding to $-w$. In the triangle ABC, let ℓ_{AB} be the length of side AB, ℓ_{BC} be the length of side BC, and let ℓ_{AC} be the length of side AC. A standard fact from geometry is that the length of any side of a triangle is less than or equal to the sum of the lengths of the other two sides of the triangle. One of the three inequalities that this gives for the triangle ABC is

$$\ell_{BC} \leq \ell_{AB} + \ell_{AC}.$$

As described above, by the Pythagorean Theorem, we have $\ell_{AB} = |z|$, $\ell_{AC} = |-w| = |w|$, and $\ell_{BC} = |z - (-w)| = |z + w|$. Substituting these into the inequality above gives

$$|z + w| \leq |z| + |w|. \qquad \square$$

REMARK

The proof of the Triangle Inequality above involves a triangle ABC in the complex plane, defined in terms of the complex numbers z and $-w$. Depending on the choices of z and w, it is possible that the three points A, B and C are collinear (i.e., they all occur on a single

line). In the above proof for this case, we still consider ABC to be a triangle, even though it has zero area (some mathematicians call this a *degenerate* triangle). In fact, this is precisely when the inequality becomes an equality. For example, if A, B and C are collinear, with A between B and C, then we immediately obtain $\ell_{BC} = \ell_{AB} + \ell_{AC}$.

When the complex number $z = x + yi$ is represented as a point in the complex plane, we refer to the ordered pair (x, y) as the *Cartesian coordinates* for the point. These coordinates allow us to specify the point by giving its horizontal position x, and its vertical position y.

However, there is another type of coordinate system that can be used to specify each point in the complex plane, and is especially useful when doing complex multiplication. Denote the origin by O, and the point whose Cartesian coordinates are (x, y) by P. Define the *polar axis* to be the positive half of the real axis, starting at O, and extending out to infinity at the right, as illustrated in Figure 10.2. Suppose that the length of OP is r, and that the counterclockwise angle of rotation from the polar axis, measured in radians, is θ. We define the pair (r, θ) to be the *polar coordinates* for the point P, also as illustrated in Figure 10.2.
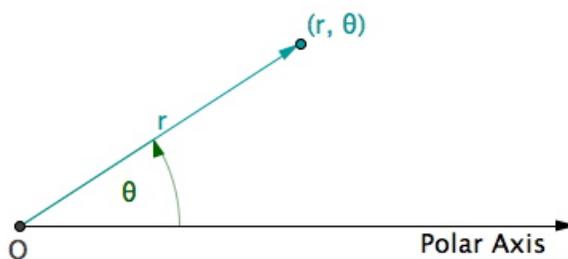
Figure 10.2: The polar axis and the point with polar coordinates (r, θ)

From the description above, note that r is a non-negative real number, and θ is a real number in the interval $[0, 2\pi)$ (when $r > 0$). However, for convenience, we will allow θ to be any real number. Therefore each point can be specified by polar coordinates in an infinite number of ways, since (r, θ) identifies the same point as $(r, \theta + 2\pi k)$ for any integer k.
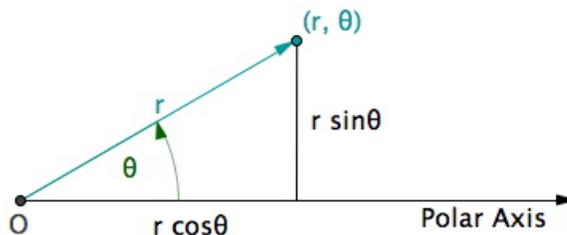
Figure 10.3: Relating polar and Cartesian coordinates

Given a point P representing the complex number z, what is the relationship between its Cartesian coordinates (x, y) and polar coordinates (r, θ)? Consider the right triangle with vertices O, P and the foot of the perpendicular from P to the real axis, illustrated in Figure 10.3 for the case when P with polar coordinates (r, θ) is in the first quadrant.

Since r is defined to be the length of the hypotenuse OP in our triangle, we immediately obtain $r = \sqrt{x^2 + y^2}$ by the Pythagorean Theorem, giving $r = |z|$, the modulus of z. Also, using simple trigonometry on this right-angled triangle, we get

$$x = r\cos\theta, \qquad\qquad y = r\sin\theta,$$

and, by a case analysis, it is straightforward to prove that these relationships hold for P in all 4 quadrants. This means that the complex number represented by the point P can be written in terms of the polar coordinates (r, θ) by

$$z = x + yi = r\cos\theta + r\sin\theta\, i = r(\cos\theta + i\sin\theta).$$

We give a special name to this way of expressing complex numbers in the following definition.

Definition 10.3.2

**polar form,
argument**

A **polar form** for the complex number z is

$$z = r(\cos\theta + i\sin\theta),$$

where $r \geq 0$ is the modulus of z. The angle $\theta \in \mathbb{R}$ is called an **argument** of z.

Given a complex number in standard form, determining its modulus is the first step to writing it in polar form, as illustrated in the following examples.

Example 6

Write the following complex numbers in polar form:

   (a) $z_1 = 5 - 5i$

   (b) $z_2 = -3\sqrt{3} - 9i$

   (c) $z_3 = 7i$

   (d) $z_4 = -\frac{3}{4}$

Solution:

   (a) We have $|z_1| = \sqrt{25 + 25} = \sqrt{50} = 5\sqrt{2}$, so

$$z_1 = 5\sqrt{2}\left(\frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}}i\right) = 5\sqrt{2}\left(\cos\frac{7\pi}{4} + i\sin\frac{7\pi}{4}\right).$$

      Note that $\frac{7\pi}{4} - 2\pi = -\frac{\pi}{4}$, so another polar form for z_1 is given by

$$z_1 = 5\sqrt{2}\left(\cos\left(-\frac{\pi}{4}\right) + i\sin\left(-\frac{\pi}{4}\right)\right).$$

   (b) We have $|z_2| = \sqrt{27 + 81} = \sqrt{108} = 6\sqrt{3}$, so

$$z_2 = 6\sqrt{3}\left(-\frac{1}{2} - \frac{\sqrt{3}}{2}i\right) = 6\sqrt{3}\left(\cos\frac{4\pi}{3} + i\sin\frac{4\pi}{3}\right).$$
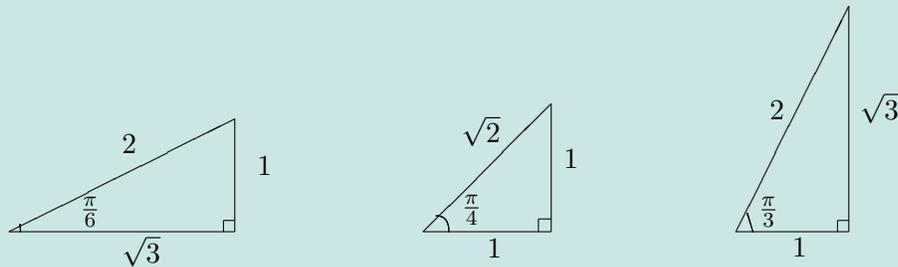
(c) We have $|z_3| = \sqrt{0 + 49} = \sqrt{49} = 7$, so

$$z_3 = 7\,(0 + i) = 7\left(\cos\frac{\pi}{2} + i\sin\frac{\pi}{2}\right).$$

(d) We have $|z_4| = \sqrt{\frac{9}{16} + 0} = \sqrt{\frac{9}{16}} = \frac{3}{4}$, so

$$z_4 = \frac{3}{4}\,(-1 + 0\,i) = \frac{3}{4}\,(\cos\pi + i\sin\pi).$$

REMARK

For θ in the first quadrant, so $0 \le \theta \le \frac{\pi}{2}$, the angles $\theta = 0, \frac{\pi}{6}, \frac{\pi}{4}, \frac{\pi}{3}, \frac{\pi}{2}$ show up repeatedly in examples of transforming a complex number into polar form. In order for you to more easily remember the sine and cosine values for these angles, it might be helpful to remember the special right triangles below:

From these triangles, we immediately obtain

$$\sin\frac{\pi}{6} = \cos\frac{\pi}{3} = \frac{1}{2}, \qquad \sin\frac{\pi}{4} = \cos\frac{\pi}{4} = \frac{1}{\sqrt{2}}, \qquad \sin\frac{\pi}{3} = \cos\frac{\pi}{6} = \frac{\sqrt{3}}{2}.$$

For θ in other quadrants, you will need to adapt the sine and cosine values for the first quadrant by appropriate changes of sign.

In the next result we show that multiplication of complex numbers can be expressed very compactly in terms of polar form. For the proof, it will be helpful to recall the following *sum of angle formulas* for sine and cosine.

$$\sin(\alpha + \beta) = \sin\alpha\cos\beta + \cos\alpha\sin\beta \tag{10.4}$$
$$\cos(\alpha + \beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta. \tag{10.5}$$

Proposition 8   **(Polar Multiplication in $\mathbb{C}$ (PM$\mathbb{C}$))**

For all complex numbers $z_1 = r_1(\cos\theta_1 + i\sin\theta_1)$ and $z_2 = r_2(\cos\theta_2 + i\sin\theta_2)$, we have

$$z_1 z_2 = r_1 r_2\big(\cos(\theta_1 + \theta_2) + i\sin(\theta_1 + \theta_2)\big).$$

Proof: Let r_1, r_2, θ_1 and θ_2 be arbitrary real numbers, where r_1 and r_2 are non-negative, $z_1 = r_1(\cos\theta_1 + i\sin\theta_1)$, and $z_2 = r_2(\cos\theta_2 + i\sin\theta_2)$. Then from the definition of multiplication for complex numbers, we have

$$
\begin{aligned}
z_1 z_2 &= r_1(\cos\theta_1 + i\sin\theta_1)\, r_2(\cos\theta_2 + i\sin\theta_2) \\
&= r_1 r_2\big((\cos\theta_1\cos\theta_2 - \sin\theta_1\sin\theta_2) + i(\cos\theta_1\sin\theta_2 + \sin\theta_1\cos\theta_2)\big) \\
&= r_1 r_2\big(\cos(\theta_1 + \theta_2) + i\sin(\theta_1 + \theta_2)\big),
\end{aligned}
$$

where we have used (10.4) and (10.5) for the third equality (with $\alpha = \theta_1$ and $\beta = \theta_2$). $\square$

10.4   De Moivre's Theorem

The next result, named after the French mathematician Abraham de Moivre, shows how the use of polar form dramatically simplifies taking powers of complex numbers.

Theorem 9

(De Moivre's Theorem (DMT))

For all real numbers θ and integers n, we have

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

Proof: For the integer n, we either have $n \geq 0$ or $n < 0$, and we consider these ranges of values for n separately.

For $n \geq 0$, we prove the result by induction on n, where $P(n)$ is the statement

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

Base Case For $n = 0$, the statement $P(0)$ is

$$(\cos\theta + i\sin\theta)^0 = \cos 0 + i\sin 0.$$

By convention, $z^0 = 1$ and so the left hand side of this equation is 1. Since $\cos 0 = 1$ and $\sin 0 = 0$, the right hand side also evaluates to 1, so $P(0)$ is true.

Inductive Step Let k be an arbitrary integer, where $k \geq 0$.

Assume the inductive hypothesis, $P(k)$. That is, we assume

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

We wish to prove $P(k + 1)$, which is the statement

$$(\cos\theta + i\sin\theta)^{k+1} = \cos(k + 1)\theta + i\sin(k + 1)\theta.$$

Starting with the left hand side of $P(k + 1)$, we have

$$
\begin{aligned}
(\cos\theta + i\sin\theta)^{k+1} &= (\cos\theta + i\sin\theta)^k(\cos\theta + i\sin\theta) \\
&= (\cos k\theta + i\sin k\theta)(\cos\theta + i\sin\theta), \quad \text{by the inductive hypothesis} \\
&= \cos(k + 1)\theta + i\sin(k + 1)\theta. \qquad\qquad \text{using Polar Multiplication}
\end{aligned}
$$

The result is true for $n = k + 1$, and hence $P(n)$ holds for all $n \geq 0$ by the Principle of Mathematical Induction.

For $n < 0$, we have $n = -m$ for some positive integer m. Then we obtain

$$
\begin{aligned}
(\cos\theta + i\sin\theta)^n &= \frac{1}{(\cos\theta + i\sin\theta)^m} \\
&= \frac{1}{\cos m\theta + i\sin m\theta} \\
&= \frac{1}{\cos m\theta + i\sin m\theta} \times \frac{\cos m\theta - i\sin m\theta}{\cos m\theta - i\sin m\theta} \\
&= \frac{\cos m\theta - i\sin m\theta}{\cos^2 m\theta + \sin^2 m\theta} \\
&= \cos m\theta - i\sin m\theta \\
&= \cos(-m\theta) + i\sin(-m\theta) \\
&= \cos n\theta + i\sin n\theta,
\end{aligned}
$$

where for the second equality we have used De Moivre's Theorem for the positive integer m, proved in the first case above. For the remaining equalities we used standard results from complex numbers and trigonometry.   □

The following corollary is an immediate result of De Moivre's Theorem.

Corollary 10   For all complex numbers $z = r(\cos\theta + i\sin\theta)$ and integers n, except when $|z| = r = 0$ and n is negative, we have
$$z^n = r^n(\cos n\theta + i\sin n\theta).$$

Example 7   Determine $(-1 + \sqrt{3}i)^{17}$.

Solution: In polar form, we have $-1 + \sqrt{3}i = 2\big(\cos 2\pi/3 + i\sin 2\pi/3\big)$. Then by De Moivre's Theorem, we obtain

$$
\begin{aligned}
(-1 + \sqrt{3}i)^{17} &= 2^{17}\big(\cos 2\pi/3 + i\sin 2\pi/3\big)^{17} \\
&= 2^{17}\big(\cos 34\pi/3 + i\sin 34\pi/3\big) \\
&= 2^{17}\big(\cos 4\pi/3 + i\sin 4\pi/3\big) \\
&= 2^{17}\Big(-\frac{1}{2} - \frac{\sqrt{3}}{2}i\Big) \\
&= 2^{16}(-1 - \sqrt{3}i).
\end{aligned}
$$

De Moivre's Theorem can be a powerful method for doing trigonometry. For example, it can be used to prove trigonometric identities.

Example 8   Prove that, for all real numbers θ,

$$\cos 3\theta = 4 \cos^3 \theta - 3 \cos \theta$$
$$\sin 3\theta = 3 \sin \theta - 4 \sin^3 \theta.$$

Solution: By De Moivre's Theorem, we have

$$(\cos \theta + i \sin \theta)^3 = \cos 3\theta + i \sin 3\theta,$$

and using the Binomial Theorem, we also have

$$(\cos \theta + i \sin \theta)^3 = \cos^3 \theta + 3i \cos^2 \theta \sin \theta - 3 \cos \theta \sin^2 \theta - i \sin^3 \theta.$$

Equating these two expressions for $(\cos \theta + i \sin \theta)^3$ gives

$$\cos 3\theta + i \sin 3\theta = \cos^3 \theta + 3i \cos^2 \theta \sin \theta - 3 \cos \theta \sin^2 \theta - i \sin^3 \theta.$$

Now equate the real and imaginary parts above to obtain

$$\cos 3\theta = \cos^3 \theta - 3 \cos \theta \sin^2 \theta$$
$$\sin 3\theta = 3 \cos^2 \theta \sin \theta - \sin^3 \theta.$$

Finally we use the identity $\sin^2 \theta + \cos^2 \theta = 1$ to obtain

$$\cos 3\theta = \cos^3 \theta - 3 \cos \theta (1 - \cos^2 \theta) = 4 \cos^3 \theta - 3 \cos \theta$$
$$\sin 3\theta = 3(1 - \sin^2 \theta) \sin \theta - \sin^3 \theta = 3 \sin \theta - 4 \sin^3 \theta.$$

Another way in which De Moivre's Theorem can be used is to determine exact expressions for tricky sine and cosine values.

Example 9   Use the fact that $\theta = \frac{2\pi}{5}$ satisfies $\cos 2\theta = \cos 3\theta$ to determine $\cos \frac{2\pi}{5}$.

Solution: In Example 8 we proved that

$$\cos 3\theta = 4 \cos^3 \theta - 3 \cos \theta.$$

Now, from the hint and the special case $\alpha = \beta = \theta$ of (10.5), we have

$$\cos 3\theta = \cos 2\theta = \cos^2 \theta - \sin^2 \theta = \cos^2 \theta - (1 - \cos^2 \theta) = 2 \cos^2 \theta - 1,$$

and equating these two expressions for $\cos 3\theta$, we obtain

$$4 \cos^3 \theta - 3 \cos \theta = 2 \cos^2 \theta - 1.$$

Rearranging this and letting $X = \cos \theta$, we get

$$4X^3 - 2X^2 - 3X + 1 = 0. \qquad\qquad (10.6)$$

Note that $X = 1$ is a solution to this equation, and dividing the cubic polynomial on the left hand side by $X - 1$, we obtain

$$(X - 1)(4X^2 + 2X - 1) = 0.$$

Now, from the quadratic formula, the roots of the quadratic $4X^2 + 2X - 1$ are given by

$$\frac{-2 \pm \sqrt{20}}{8} = \frac{-1 \pm \sqrt{5}}{4}.$$

Hence there are 3 solutions to equation (10.6), namely $X = 1$, $X = \frac{-1+\sqrt{5}}{4}$, $X = \frac{-1-\sqrt{5}}{4}$, and since $\cos \frac{2\pi}{5}$ is a solution to equation (10.6), it must be equal to one of these 3 values. From the fact that $0 < \frac{2\pi}{5} < \frac{\pi}{2}$, we know that $0 < \cos \frac{2\pi}{5} < 1$, so $\cos \frac{2\pi}{5} \neq 1$ and $\cos \frac{2\pi}{5} \neq \frac{-1-\sqrt{5}}{4}$, and we conclude that

$$\cos \frac{2\pi}{5} = \frac{-1 + \sqrt{5}}{4}.$$

REMARK

Note that a polar form contains both sine and cosine of the *same* angle θ, always within the expression $\cos \theta + i \sin \theta$. For convenience, some mathematicians use the notation "cis θ" as a shorthand, instead of writing out the full "$\cos \theta + i \sin \theta$".

Using this notation, the rules for multiplication and taking powers of complex numbers in polar form give us

$$\operatorname{cis} 0 = 1, \qquad \operatorname{cis} \alpha \operatorname{cis} \beta = \operatorname{cis}(\alpha + \beta), \qquad \big(\operatorname{cis} \alpha\big)^n = \operatorname{cis}(n\alpha),$$

for all real numbers α, β, and all integers n. We also know, by the usual rules for powers of any positive real number x, that

$$x^0 = 1, \qquad x^\alpha x^\beta = x^{\alpha+\beta}, \qquad \big(x^\alpha\big)^n = x^{n\alpha},$$

again, for all real numbers α, β, and all integers n.

Motivated by these similarities, is it possible that $\operatorname{cis} \theta = x^\theta$ for some choice of x? The perhaps surprising answer is *yes*, that $\operatorname{cis} \theta = e^{i\theta}$, which can also be written as

$$\cos \theta + i \sin \theta = e^{i\theta}.$$

The special case $\theta = \pi$ of this, after rearrangement, gives the famous formula $e^{i\pi} + 1 = 0$. The precise mathematical meaning of the purely imaginary number $i\theta$ as an exponent and the equalities above are studied in *complex analysis*, an upper-year mathematics course.

10.5   Complex n-th Roots

Definition 10.5.1

complex roots

For a complex number a and positive integer n, the complex numbers z that satisfy the equation

$$z^n = a$$

are called the **complex n-th roots** of a.

De Moivre's Theorem gives us a straightforward way to find complex n-th roots of a. In the statement of the following result, for a non-negative real number r, we use $\sqrt[n]{r}$ to denote the n-th non-negative real root of r (which is a unique non-negative real number).

Theorem 11

(Complex n-th Roots Theorem (CNRT))

For all complex numbers $a = r(\cos\theta + i\sin\theta)$ and natural numbers n, the complex n-th roots of a are given by

$$\sqrt[n]{r}\left(\cos\left(\frac{\theta + 2k\pi}{n}\right) + i\sin\left(\frac{\theta + 2k\pi}{n}\right)\right), \qquad k = 0, 1, 2, \ldots, n-1.$$

Proof: Let $a = r(\cos\theta + i\sin\theta)$ be an arbitrary complex number, and n be an arbitrary natural numnber. Define the two sets

$$A = \left\{ \sqrt[n]{r}\left(\cos\left(\frac{\theta + 2k\pi}{n}\right) + i\sin\left(\frac{\theta + 2k\pi}{n}\right)\right) : k \in \{0, 1, 2, \ldots, n-1\}\right\},$$

$$B = \{z \in \mathbb{C} : z^n = a\}.$$

We will prove that $A = B$ by proving that $A \subseteq B$ and $B \subseteq A$, using universally quantified implications as described in the proof method for $S \subseteq T$ on page 89.

To prove $A \subseteq B$: We prove the implication

For all complex numbers z, if $z \in A$, then $z \in B$.

Hence, let k be an arbitrary integer in $\{0, 1, 2, \ldots, n-1\}$, and assume that

$$z = \sqrt[n]{r}\left(\cos\left(\frac{\theta + 2k\pi}{n}\right) + i\sin\left(\frac{\theta + 2k\pi}{n}\right)\right) \in A.$$

Then, using DeMoivre's Theorem, we obtain

$$z^n = (\sqrt[n]{r})^n\left(\cos\left(n\,\frac{\theta + 2k\pi}{n}\right) + i\sin\left(n\,\frac{\theta + 2k\pi}{n}\right)\right)$$
$$= r\left(\cos(\theta + 2k\pi) + i\sin(\theta + 2k\pi)\right)$$
$$= r\left(\cos\theta + i\sin\theta\right)$$
$$= a.$$

Thus we conclude that $z \in B$.

To prove $B \subseteq A$: We prove the implication

For all complex numbers z, if $z \in B$, then $z \in A$.

Hence, assume that $z^n = a$, so $z \in B$, and suppose that z has polar form $z = s(\cos\phi + i\sin\phi)$. Then, writing equation $z^n = a$ in polar form, recalling that $a = r(\cos\theta + i\sin\theta)$, we have

$$s^n(\cos n\phi + i\sin n\phi) = r(\cos\theta + i\sin\theta),$$

where we have used De Moivre's Theorem on the left hand side. Now two complex numbers in polar form are equal if and only if their moduli are equal and their arguments differ by an integer multiple of 2π. Thus from the equation above we obtain $s^n = r$ and $n\phi - \theta = 2\ell\pi$ for some $\ell \in \mathbb{Z}$, or, equivalently,

$$s = \sqrt[n]{r}, \qquad \phi = \frac{\theta + 2\ell\pi}{n}.$$

Hence, for some $\ell \in \mathbb{Z}$, we have

$$z = \sqrt[n]{r}\left(\cos\left(\frac{\theta + 2\ell\pi}{n}\right) + i\sin\left(\frac{\theta + 2\ell\pi}{n}\right)\right). \tag{10.7}$$

Now, dividing ℓ by n, the Division Algorithm gives

$$\ell = qn + j,$$

for some $q \in \mathbb{Z}$ and $j \in \{0, 1, 2, \ldots, n-1\}$. Substituting this for ℓ in equation (10.7) gives

$$
\begin{aligned}
z &= \sqrt[n]{r}\left(\cos\left(\frac{\theta + 2(qn+j)\pi}{n}\right) + i\sin\left(\frac{\theta + 2(qn+j)\pi}{n}\right)\right) \\
&= \sqrt[n]{r}\left(\cos\left(\frac{\theta + 2j\pi}{n} + 2q\pi\right) + i\sin\left(\frac{\theta + 2j\pi}{n} + 2q\pi\right)\right) \\
&= \sqrt[n]{r}\left(\cos\left(\frac{\theta + 2j\pi}{n}\right) + i\sin\left(\frac{\theta + 2j\pi}{n}\right)\right),
\end{aligned}
$$

where the last equality follows since $q \in \mathbb{Z}$. Thus we conclude that $z \in A$.   □

Note that the Complex n-th Roots Theorem says that every non-zero complex number has exactly n different complex n-th roots.

Example 10   Find all the cube roots of i.

Solution: Writing i in polar form, we obtain

$$i = 0 + 1\,i = \cos\frac{\pi}{2} + i\sin\frac{\pi}{2}.$$

Then, from the Complex n-th Roots Theorem, the cube roots are given by

$$\cos\left(\frac{\frac{\pi}{2} + 2k\pi}{3}\right) + i\sin\left(\frac{\frac{\pi}{2} + 2k\pi}{3}\right) = \cos\left(\frac{\pi + 4k\pi}{6}\right) + i\sin\left(\frac{\pi + 4k\pi}{6}\right)$$

for $k = 0, 1, 2$. Writing them out separately, we get the following 3 complex cube roots:

$$
\begin{aligned}
k = 0, : \qquad & z_0 = \cos\left(\frac{\pi}{6}\right) + i\sin\left(\frac{\pi}{6}\right) = \frac{\sqrt{3}}{2} + \frac{i}{2}, \\
k = 1 : \qquad & z_1 = \cos\left(\frac{5\pi}{6}\right) + i\sin\left(\frac{5\pi}{6}\right) = -\frac{\sqrt{3}}{2} + \frac{i}{2}, \\
k = 2 : \qquad & z_2 = \cos\left(\frac{3\pi}{2}\right) + i\sin\left(\frac{3\pi}{2}\right) = -i.
\end{aligned}
$$

Example 11   Determine all the complex fourth roots of -16.

Solution: Writing -16 in polar form, we obtain

$$-16 = 16(-1 + 0\,i) = 16(\cos \pi + i \sin \pi).$$

Then, from the Complex n-th Roots Theorem, the fourth roots are given by

$$\sqrt[4]{16}\left(\cos\left(\frac{\pi + 2k\pi}{4}\right) + i \sin\left(\frac{\pi + 2k\pi}{4}\right)\right) = 2\left(\cos\left(\frac{\pi}{4} + \frac{k\pi}{2}\right) + i \sin\left(\frac{\pi}{4} + \frac{k\pi}{2}\right)\right)$$

for $k = 0, 1, 2, 3$. Writing them out separately, we get the following 4 complex fourth roots:

$$k = 0: \qquad z_0 = 2\left(\cos\left(\frac{\pi}{4}\right) + i \sin\left(\frac{\pi}{4}\right)\right) \qquad = 2\left(\frac{1}{\sqrt{2}} + \frac{i}{\sqrt{2}}\right) = \sqrt{2} + i\sqrt{2},$$

$$k = 1: \qquad z_1 = 2\left(\cos\left(\frac{3\pi}{4}\right) + i \sin\left(\frac{3\pi}{4}\right)\right) \qquad = 2\left(\frac{-1}{\sqrt{2}} + \frac{i}{\sqrt{2}}\right) = -\sqrt{2} + i\sqrt{2},$$

$$k = 2: \qquad z_2 = 2\left(\cos\left(\frac{5\pi}{4}\right) + i \sin\left(\frac{5\pi}{4}\right)\right) \qquad = 2\left(\frac{-1}{\sqrt{2}} + \frac{-i}{\sqrt{2}}\right) = -\sqrt{2} - i\sqrt{2},$$

$$k = 3: \qquad z_3 = 2\left(\cos\left(\frac{7\pi}{4}\right) + i \sin\left(\frac{7\pi}{4}\right)\right) \qquad = 2\left(\frac{1}{\sqrt{2}} + \frac{-i}{\sqrt{2}}\right) = \sqrt{2} - i\sqrt{2}.$$

In Figure 10.4 we have graphed the 4 fourth roots of -16 in the complex plane. Note how they all occur on a circle of radius 2, centred at the origin, and that they are equally spaced around this circle.
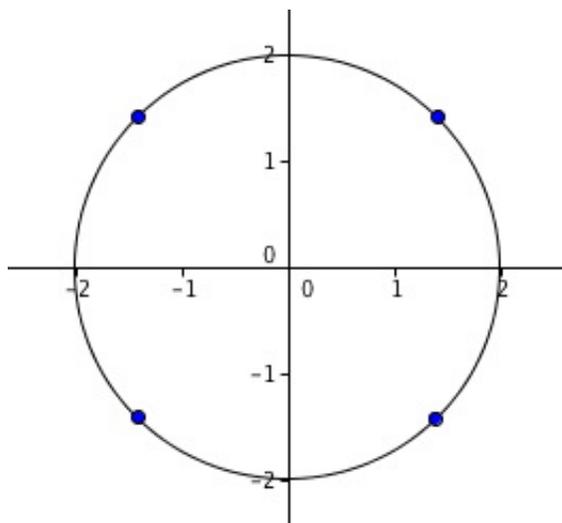

Figure 10.4: The fourth roots of -16 in the complex plane

REMARK

Note that, as in the example above, the n-th roots of a non-zero complex number of modulus r are equally spaced around a circle of radius $\sqrt[n]{r}$ in the complex plane, centred at the origin.

10.6   Square Roots and the Quadratic Formula

When $n = 2$, the Complex n-th Roots Theorem tells us that for a complex number $a \neq 0$, there are exactly two solutions to $z^2 - a = 0$. What about solutions to the equation $az^2 + bz + c = 0$ when a, b, c are complex numbers? It turns out that the solutions are given by the same quadratic formula that we use when a, b, c are real numbers!

Proposition 12   **(Quadratic Formula)**

For all complex numbers a, b and c, with $a \neq 0$, the solutions to $az^2 + bz + c = 0$ are given by
$$z = \frac{-b \pm w}{2a},$$
where w is a solution to $w^2 = b^2 - 4ac$.

Note that instead of writing $\sqrt{b^2 - 4ac}$ in the formula above, we have written "w", where $w^2 = b^2 - 4ac$. This is because the notation $\sqrt{}$ is defined for nonnegative real numbers, but here a, b, c are complex.

EXERCISE

Prove that the quadratic formula holds for any complex quadratic equation (i.e., prove Proposition 12 above).

Example 12   Express the solutions to $2z^2 + 3z + 2 = 0$ in standard form.

Solution: Using the quadratic formula, we get
$$z = \frac{-3 \pm w}{4}, \quad \text{where } w^2 = 9 - 4(4) = -7.$$

One solution to $w^2 = -7$ is $w = \sqrt{7}i$, so the two solutions are
$$\frac{-3}{4} \pm \frac{\sqrt{7}}{4}i.$$

Example 13   Express the solutions to $iz^2 + 3z - 2i = 0$ in standard form.

Solution: Using the quadratic formula, we get
$$z = \frac{-3 \pm w}{2i}, \quad \text{where } w^2 = 9 - 4i(-2i) = 9 - 8 = 1.$$

One solution to $w^2 = 1$ is $w = 1$, so the two solutions are
$$\frac{-3 \pm 1}{2i} = \frac{-3 \pm 1}{2i}\left(\frac{-2i}{-2i}\right) = i \text{ or } 2i.$$

Chapter 11

Polynomials

11.1 Introduction

Our number systems were developed in response to the need to find solutions to equations. In the previous chapter we considered equations of the form

$$ax^2 + bx + c = 0, \quad \text{or} \quad x^n - a = 0,$$

where a, b and c are real or complex numbers, and n is a natural number. In fact, a great deal more is known.

Consider the left hand side of the equations above. They are expressions called polynomials, built using a symbol x and coefficients taken from some set. In this chapter, the coefficients will always come from a special type of set called a **field**, and we will consider only the following four examples of a field:

- the rational numbers $\mathbb{Q}$,

- the real numbers $\mathbb{R}$,

- the complex numbers $\mathbb{C}$,

- the integers modulo a prime $\mathbb{Z}_p$.

A field is a set which has an addition and a multiplication satisfying nine properties. These properties are stated for the complex numbers in Properties of Complex Arithmetic (PCA) on page 160. Though Proposition 2 is stated only for $\mathbb{C}$, the same nine properties hold for $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{Z}_p$. That is, the only change needed for the nine properties in Proposition 2 to apply to $\mathbb{R}$, say, is to replace the $u, v, z \in \mathbb{C}$ at the top by $u, v, z \in \mathbb{R}$, and to specialize $z = a + bi$ in properties 4 and 8 by setting $b = 0$.

In addition, we have already seen some of these nine properties for the field $\mathbb{Z}_p$ as part of our study of modular arithmetic in Chapter 8, specifically in Example 13 on page 135 (restricted to $m = p$, a prime), and Corollary 13.

The integers are not a field because, for example, the only integers with a multiplicative inverse are 1 and -1 (1 is the multiplicative identity for multiplication of integers). Similarly, $\mathbb{Z}_6$ is not a field since [3] does not have a multiplicative inverse.

REMARK

One of the most important facts about a field is the following.

For all fields $\mathbb{F}$, and all $a, b \in \mathbb{F}$, if $ab = 0$, then $a = 0$ or $b = 0$.

Informally, this says that $\mathbb{F}$ has no *zero divisors*. Convince yourself that this statement is true when $\mathbb{F}$ is $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ or $\mathbb{Z}_p$ where p is a prime. Although the integers are not a field, this property is also true for integers. However, it is not true in $\mathbb{Z}_6$, where we have $[2][3] = [0]$, but $[2] \neq [0]$ and $[3] \neq [0]$.

Now we define a polynomial with coefficients in a field.

Definition 11.1.1

polynomial, indeterminate, coefficient, term, $\mathbb{F}[x]$

A **polynomial** in x over the field $\mathbb{F}$ is an expression of the form

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

where $n \geq 0$ is an integer, and

- x is a symbol called an **indeterminate**, and

- $a_0, a_1, \ldots, a_n$ are elements of $\mathbb{F}$.

Each individual a_i is called a **coefficient** of the polynomial, and each individual expression of the form $a_i x^i$ is called a **term** of the polynomial.

We use the notation $\mathbb{F}[x]$ to denote the set of all polynomials over the field $\mathbb{F}$.

Example 1

As examples of polynomials, we have

- $2x^3 + (\sqrt{2} - i)x^2 - \frac{7\pi}{2}ix + (5 - 2i) \in \mathbb{C}[x]$. Here $a_3 = 2, a_2 = \sqrt{2} - i, a_1 = -\frac{7\pi}{2}i$ and $a_0 = 5 - 2i$, all of which are complex numbers.

- $-x^2 + \sqrt{7}x - 1 \in \mathbb{R}[x]$. Note that since $\mathbb{R} \subsetneq \mathbb{C}$, all polynomials over $\mathbb{R}$ are also polynomials over $\mathbb{C}$. So $-x^2 + \sqrt{7}x - 1$ is also in $\mathbb{C}[x]$.

- $\frac{1}{2}x^5 - \frac{5}{13}x^4 + x^3 - x^2 + 5x + \frac{3}{2} \in \mathbb{Q}[x]$. This is also a polynomial in $\mathbb{R}[x]$, and hence in $\mathbb{C}[x]$.

- $5x^4 + 0x^3 + 1x^2 + 0x - 2 \in \mathbb{Q}[x]$ (and also in $\mathbb{R}[x]$, and $\mathbb{C}[x]$). We would usually express the term $1x^2$ simply as x^2 and omit the terms $0x^3$ and $0x$, and write the polynomial more simply as $5x^4 + x^2 - 2$.

- $[2]x^3 + [4]x^2 + [0]x + [1] \in \mathbb{Z}_5[x]$. We would usually omit the term $[0]x$, and write the polynomial more simply as $[2]x^3 + [4]x^2 + [1]$.

However, note that $2x^3 + x^2 - \frac{7\pi}{2}ix + \sqrt{5} \notin \mathbb{R}[x]$, since one of the coefficients is not a real number. This means we need to be careful about the field when working with polynomials.

Most often we will work with polynomials over the rational numbers, real numbers or complex numbers. These are called **rational polynomials**, **real polynomials** and **complex polynomials** respectively.

One of the most important properties of a polynomial is its *degree*, which we define next.

Definition 11.1.2
degree of polynomial

Let $a_nx^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0$ be a polynomial in $\mathbb{F}[x]$, where $n \geq 0$ is an integer, and $a_n \neq 0$. Then the polynomial is said to have **degree** n and we write $\deg f(x) = n$. In other words, the degree of a polynomial is the largest power of x that has a non-zero coefficient.

The **zero polynomial** has all of its coefficients equal to zero, and its degree is undefined.

A **constant polynomial** is a polynomial that is either the zero polynomial or a polynomial of degree 0. Polynomials of degree 1, 2 and 3 are called **linear polynomials**, **quadratic polynomials**, and **cubic polynomials**, repectively.

Example 2

As examples of the degree of polynomials, we have

- $2x^3 + (\sqrt{2} - i)x^2 - \frac{7\pi}{2}ix + (5 - 2i)$ is a cubic polynomial.

- $-x^2 + \sqrt{7}x - 1$ is a quadratic polynomial.

- $\frac{1}{2}x^5 - \frac{5}{13}x^4 + x^3 - x^2 + 5x + \frac{3}{2}$ is a polynomial of degree 5.

- $0x^3 + 0x^2 + 3x + 0$ is a linear polynomial, since the largest power of x with a non-zero coefficient is $x = x^1$.

We end the section by defining what it means for two polynomials to be *equal*.

Definition 11.1.3
equal polynomials

The polynomials $a_nx^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0$, and $b_nx^n + b_{n-1}x^{n-1} + \cdots + b_1x + b_0$ in $\mathbb{F}[x]$ are **equal** if and only if $a_k = b_k$ for all $k = 0, 1, \ldots, n$.

11.2   Arithmetic with Polynomials

Arithmetic can be done with polynomials just as you have done in high school. When working with polynomials, we will sometimes use both function notation and summation notation. That is, we use $f(x)$ to denote an element of $\mathbb{F}[x]$, and write

$$f(x) = a_nx^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0 = \sum_{k=0}^{n} a_kx^k.$$

We begin this section by formally defining addition and multiplication of polynomials. The notation "$\max\{m, n\}$" denotes the *maximum* of m and n.

Definition 11.2.1

addition and multiplication of polynomials

Let $f(x) = \sum_{i=0}^{m} a_i x^i$ and $g(x) = \sum_{j=0}^{n} b_j x^j$ be polynomials in $\mathbb{F}[x]$.

- **Addition** of $f(x)$ and $g(x)$ is defined by

$$f(x) + g(x) = \sum_{k=0}^{\max\{m,n\}} (a_k + b_k) x^k,$$

where $a_k = 0$ for $k > m$, and $b_k = 0$ for $k > n$.

- **Multiplication** of $f(x)$ and $g(x)$ is defined by

$$f(x)g(x) = \sum_{i=0}^{m} \sum_{j=0}^{n} a_i b_j x^{i+j} = \sum_{\ell=0}^{m+n} c_\ell x^\ell,$$

where

$$c_\ell = a_0 b_\ell + a_1 b_{\ell-1} + \cdots + a_{\ell-1} b_1 + a_\ell b_0 = \sum_{i=0}^{\ell} a_i b_{\ell-i},$$

for $\ell = 0, 1, \ldots, m + n$ and where again $a_k = 0$ for $k > m$, and $b_k = 0$ for $k > n$.

Example 3

As examples of addition and multiplication of polynomials, we have

- In $\mathbb{R}[x]$, for $f(x) = x^2 + 7x - 1$ and $g(x) = \sqrt{2}x^3 + 4x^2 - 3x$, we obtain

$$f(x) + g(x) = \sqrt{2}x^3 + 5x^2 + 4x - 1,$$
$$f(x)g(x) = \sqrt{2}x^5 + (4 + 7\sqrt{2})x^4 + (25 - \sqrt{2})x^3 - 25x^2 + 3x.$$

- In $\mathbb{C}[x]$, for $f(x) = (1 + i)x^2 - 3ix + 5$ and $g(x) = (4 + 3i)x - 2i$, we obtain

$$f(x) + g(x) = (1 + i)x^2 + 4x + (5 - 2i),$$
$$f(x)g(x) = (1 + 7i)x^3 + (11 - 14i)x^2 + (14 + 15i)x - 10i.$$

- In $\mathbb{Z}_7[x]$, for $f(x) = [3]x^2 + [5]x + [4]$ and $g(x) = [4]x^2 + [2]x + [6]$, we obtain

$$f(x) + g(x) = [0]x^2 + [0]x + [3] = [3],$$
$$f(x)g(x) = [5]x^4 + [5]x^3 + [2]x^2 + [3]x + [3].$$

Next we give a very useful lemma about the degree of a product of two non-zero polynomials over any field $\mathbb{F}$.

Lemma 1

(Degree of a Product (DP))

For all fields $\mathbb{F}$, and all non-zero polynomials $f(x)$ and $g(x)$ in $\mathbb{F}[x]$, we have

$$\deg f(x)g(x) = \deg f(x) + \deg g(x).$$

Proof: Let $\mathbb{F}$ be an arbitrary field, and m, n be arbitrary non-negative integers. Let $f(x) = \sum_{i=0}^{m} a_i x^i$ and $g(x) = \sum_{j=0}^{n} b_j x^j$ be arbitrary polynomials in $\mathbb{F}[x]$ of degree m and n, respectively, so we have $a_m \neq 0$ and $b_n \neq 0$. Now from the definition of multiplication, we have $f(x)g(x) = \sum_{\ell=0}^{m+n} c_\ell x^\ell$, where $c_{m+n} = a_m b_n$. Since a_m and b_n are non-zero elements of the field $\mathbb{F}$, it follows that $c_{m+n} = a_m b_n$ is also non-zero. We therefore conclude that $\deg f(x)g(x) = m + n = \deg f(x) + \deg g(x)$. $\qquad\qquad\square$

Now we consider division of polynomials. It turns out that there is a lot of similarity between division of integers and division of polynomials – when we divide one polynomial into another, we get a quotient polynomial and a remainder polynomial. The result that describes precisely what happens is called the Division Algorithm for Polynomials, and is stated next. This result is not proved in this course. Note that we use the notation $q(x)$ for the *quotient polynomial*, and $r(x)$ for the *remainder polynomial*, to emphasize the similarity with the Division Algorithm for integers that we have already seen, as Proposition 3 on page 94.

Proposition 2   **(Division Algorithm for Polynomials (DAP))**

For all fields $\mathbb{F}$, and all polynomials $f(x)$ and $g(x)$ in $\mathbb{F}[x]$ with $g(x)$ not the zero polynomial, there exist unique polynomials $q(x)$ and $r(x)$ in $\mathbb{F}[x]$ such that

$$f(x) = q(x)g(x) + r(x),$$

where $r(x)$ is the zero polynomial, or $\deg r(x) < \deg g(x)$.

We also define what it means for one polynomial to divide another in a similar way as for the integers.

Definition 11.2.2   For polynomials $f(x)$ and $g(x)$ over $\mathbb{F}$, we say that $g(x)$ **divides** $f(x)$ or $g(x)$ is a **factor** of
divides, factor in $\mathbb{F}[x]$ $f(x)$, and we write $g(x) \mid f(x)$, if there exists a polynomial $q(x)$ such that $f(x) = q(x)g(x)$. That is, $g(x)$ divides $f(x)$ when the remainder $r(x)$ from DAP is the zero polynomial, or when $f(x)$ and $g(x)$ are both the zero polynomial.

Given a polynomial $f(x)$, and a non-zero polynomial $g(x)$, in order to find the quotient and remainder polynomials $q(x)$ and $r(x)$ featured in DAP, we use a process called *long division*. This process starts with the largest powers of x in $f(x)$ and $g(x)$, and is demonstrated in the following pair of examples. The first of these examples of DAP is for polynomials in $\mathbb{R}[x]$, and the second of these examples is for polynomials in $\mathbb{C}[x]$.

Example 4   **(Long Division of Polynomials over $\mathbb{R}$)**

What are the quotient and remainder polynomials when $f(x) = 3x^4 + x^3 - 4x^2 - x + 5$ is divided by $g(x) = x^2 + 1$ in $\mathbb{R}[x]$?

Before we begin, we would expect from the Division Algorithm for Polynomials that the remainder polynomial is either the zero polynomial, or has degree at most one. Now we carry out the long division:

$$
\begin{array}{r}
3x^2 \; + \;\; x \; - \;\; 7 \\
x^2 \; + \; 1 \; \overline{\big)\; 3x^4 \; + \; x^3 \; - \; 4x^2 \; - \;\; x \; + \;\; 5} \\
3x^4 \qquad\quad + \; 3x^2 \\
\hline
x^3 \; - \; 7x^2 \; - \;\; x \; + \;\; 5 \\
x^3 \qquad\qquad + \;\; x \\
\hline
- \; 7x^2 \; - \; 2x \; + \;\; 5 \\
- \; 7x^2 \qquad\qquad - \;\; 7 \\
\hline
- \; 2x \; + \; 12
\end{array}
$$

Thus, the quotient polynomial is $q(x) = 3x^2 + x - 7$ and the remainder polynomial is $r(x) = -2x + 12$, of degree 1, and we can check that indeed $f(x) = q(x)g(x) + r(x)$.

Example 5   **(Long Division of Polynomials over $\mathbb{C}$)**

What are the quotient and remainder polynomials when
$f(z) = iz^3 + (2 + 4i)z^2 + (3 - i)z + (40 - 4i)$ is divided by $g(z) = iz + (2 - 2i)$ in $\mathbb{C}[z]$?

From DAP, we would expect that the remainder polynomial is a constant polynomial (either the zero polynomial, or it has degree 0).

$$
\begin{array}{r}
z^2 \; + \qquad\qquad 6z \; + \; (11 + 9i) \\
iz \; + \; (2 - 2i) \; \overline{\big)\; iz^3 \; + \; (2 + 4i)z^2 \; + \quad (3 - i)z \; + \; (40 - 4i)} \\
iz^3 \; + \; (2 - 2i)z^2 \\
\hline
6iz^2 \; + \quad (3 - i)z \; + \; (40 - 4i) \\
6iz^2 \; + \; (12 - 12i)z \\
\hline
(-9 + 11i)z \; + \; (40 - 4i) \\
(-9 + 11i)z \; + \; (40 - 4i) \\
\hline
0
\end{array}
$$

Thus, the quotient polynomial is $q(z) = z^2 + 6z + (11 + 9i)$ and the remainder is the zero polynomial. Therefore, $g(z)$ divides $f(z)$, and we can check that indeed $f(z) = q(z)g(z)$.

Note that when dividing $f(x)$ by $g(x)$, you can always verify the correctness of your quotient polynomial $q(x)$ and remainder polynomial $r(x)$ by checking that $f(x) = q(x)g(x) + r(x)$.

11.3   Roots of Complex Polynomials and the Fundamental Theorem of Algebra

Definition 11.3.1

polynomial
equation, root

A **polynomial equation** is an equation of the form

$$
a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 = 0
$$

which will often be written as $f(x) = 0$, where $f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 \in \mathbb{F}[x]$. An element $c \in \mathbb{F}$ is called a **root** of the polynomial $f(x)$ if $f(c) = 0$ (equivalently, if c is a solution of the polynomial equation $f(x) = 0$).

We now apply the Division Algorithm for Polynomials to prove a very useful result for polynomials over an arbitrary field $\mathbb{F}$.

Proposition 3   **(Remainder Theorem (RT))**

For all fields $\mathbb{F}$, all polynomials $f(x) \in \mathbb{F}[x]$, and all $c \in \mathbb{F}$, the remainder polynomial when $f(x)$ is divided by $x - c$ is the constant polynomial $f(c)$.

Proof: Let $\mathbb{F}$ be an arbitrary field, $f(x)$ be an arbitrary polynomial in $\mathbb{F}[x]$, and c be an arbitrary element of $\mathbb{F}$. Applying the Division Algorithm for Polynomials with $g(x) = x - c$, there exist unique polynomials $q(x)$ and $r(x)$ such that

$$f(x) = q(x)(x - c) + r(x),$$

where $\deg r(x) < \deg(x - c) = 1$ or $r(x)$ is the zero polynomial. Therefore, the remainder $r(x)$ is a constant polynomial (which could be zero), and we will denote it by r_0. Hence we have

$$f(x) = q(x)(x - c) + r_0,$$

and substituting $x = c$ into this equation gives $f(c) = r_0$. $\square$

Example 6   Find the remainder when $f(z) = 3z^{12} - 8iz^5 + (4 + i)z^2 + z + 2 - 3i$ is divided by $z + i$.

Solution: Instead of doing long division, we use the Remainder Theorem, by calculating

$$
\begin{aligned}
f(-i) &= 3(-i)^{12} - 8i(-i)^5 + (4 + i)(-i)^2 + (-i) + 2 - 3i \\
&= 3 - 8i(-i) + (4 + i)(-1) - i + 2 - 3i \\
&= 3 - 8 - 4 - i - i + 2 - 3i \\
&= -7 - 5i.
\end{aligned}
$$

Hence the remainder is $-7 - 5i$.

The Remainder Theorem immediately implies the following corollary about linear factors of a polynomial.

Corollary 4   **(Factor Theorem (FT))**

For all fields $\mathbb{F}$, all polynomials $f(x) \in \mathbb{F}[x]$, and all $c \in \mathbb{F}$, the linear polynomial $x - c$ is a factor of the polynomial $f(x)$ if and only if $f(c) = 0$ (equivalently, c is a root of the polynomial $f(x)$).

The Factor Theorem gives us a linear factor $x - c$ of a polynomial $f(x)$ over the field $\mathbb{F}$ whenever c is a root of $f(x)$. The following remarkable result says that when $\mathbb{F} = \mathbb{C}$, every polynomial of positive degree has at least one root. This was proved in 1799 by the brilliant mathematician Carl Friedrich Gauss.

Theorem 5

(Fundamental Theorem of Algebra (FTA))

For all complex polynomials $f(z)$ with $\deg f(z) \geq 1$, there exists a $z_0 \in \mathbb{C}$ such that $f(z_0) = 0$.

Though we can prove that a root exists for any polynomial of positive degree in $\mathbb{C}[x]$, we can't construct a root in general when the degree is greater than four. The proofs of this fact and the Fundamental Theorem of Algebra are both demanding and are left for later courses.

Next we prove that, from the Fundamental Theorem of Algebra, we can deduce that every complex polynomial of positive degree n has n linear factors.

Proposition 6

(Complex Polynomials of Degree n Have n Roots (CPN))

For all integers $n \geq 1$, and all complex polynomials $f(z)$ of degree n, there exist complex numbers $c \neq 0$ and $c_1, c_2, \ldots, c_n$ such that

$$f(z) = c(z - c_1)(z - c_2) \cdots (z - c_n).$$

Moreover, the roots of $f(z)$ are $c_1, c_2, \ldots, c_n$.

Proof: The proof is by induction on n, where $P(n)$ is the statement

For all complex polynomials $f(z)$ of degree n, there exist complex numbers $c_1, c_2, \ldots, c_n$ and $c \neq 0$, such that

$$f(z) = c(z - c_1)(z - c_2) \cdots (z - c_n).$$

Moreover, the roots of $f(z)$ are $c_1, c_2, \ldots, c_n$.

Base Case For $n = 1$, we have $f(z) = az + b$ for complex numbers a and b with $a \neq 0$. In this case, we can write $f(z) = c(z - c_1)$ where $c = a$ and $c_1 = -\frac{b}{a}$. Now, $f(w) = 0$ for a complex number w if and only if $c(w - c_1) = 0$ and thus since $c \neq 0$, w is a root of $f(z)$ if and only if $w = c_1$. That is, c_1 is the one and only root of $f(z)$. This proves $P(1)$.

Inductive Step Let k be an arbitrary natural number.

Assume the inductive hypothesis, $P(k)$. That is ,we assume

For all complex polynomials $f(z)$ of degree k, there exist complex numbers $c_1, c_2, \ldots, c_k$ and $c \neq 0$, such that

$$f(z) = c(z - c_1)(z - c_2) \cdots (z - c_k).$$

Moreover, the roots of $f(z)$ are $c_1, c_2, \ldots, c_k$.

We wish to prove $P(k+1)$, which is the statement

> For all complex polynomials $f(z)$ of degree $k+1$, there exist complex numbers $c_1, c_2, \ldots, c_{k+1}$ and $c \neq 0$ such that
> $$f(z) = c(z - c_1)(z - c_2) \cdots (z - c_{k+1}).$$
> Moreover, the roots of $f(z)$ are $c_1, c_2, \ldots, c_{k+1}$.

Consider a complex polynomial $f(z)$ of degree $k + 1$. By the Fundamental Theorem of Algebra, $f(z)$ has a complex root. Call this root c_{k+1}. By the Factor Theorem, we know that $f(z) = q(z)(z - c_{k+1})$ for some complex polynomial $q(z)$. This quotient $q(z)$ must have degree k by the lemma, Degree of a Product, and therefore, by the inductive hypothesis, there exist complex numbers $c_1, c_2, \ldots, c_k$ and $c \neq 0$ such that

$$q(z) = c(z - c_1)(z - c_2) \cdots (z - c_k).$$

Moreover, the roots of $q(z)$ are $c_1, c_2, \ldots, c_k$. Substituting, we obtain

$$f(z) = q(z)(z - c_{k+1}) = c(z - c_1)(z - c_2) \cdots (z - c_k)(z - c_{k+1}).$$

Since $f(z) = q(z)(z - c_{k+1})$, a complex number w is a root of $f(z)$ if and only if $q(w)(w - c_{k+1}) = 0$, if and only if $w = c_{k+1}$ or $q(w) = 0$. That is, the roots of $f(z)$ are precisely c_{k+1}, and $c_1, c_2, \ldots, c_k$, the roots of $q(z)$.

The result is true for $n = k + 1$, and hence holds for all $n \geq 1$ by the Principle of Mathematical Induction. $\square$

Of course, the n roots of a complex polynomial of degree n may not be distinct. Put another way, a number c can be a root of a polynomial "more than once", in which case the corresponding linear polynomial $x - c$ will appear more than once as a factor of the polynomial.

Definition 11.3.2
multiplicity of a root

The **multiplicity** of a root c of a polynomial $f(x)$ is the largest positive integer k such that $(x - c)^k$ is a factor of $f(x)$.

Example 7

The complex number $1 - i$ is a root of multiplicity 2 of the complex polynomial

$$z^3 - (2 - 3i)z^2 - (2 + 4i)z + 2 = (z - (1 - i))^2(z + i).$$

The above results Fundamental Theorem of Algebra and Complex Polynomials of Degree n Have n Roots (CPN) deal with polynomials over the complex numbers only. We now return to polynomials over an arbitrary field $\mathbb{F}$ with the following result, which describes roots a polynomial over an arbitrary field can have. We omit the proof, which is similar to the proof of CPN, and can be obtained by applying induction on n, together with the lemma, Degree of a Product, and the Factor Theorem.

Proposition 7

For all fields $\mathbb{F}$, all integers $n \geq 1$, and all $f(x) \in \mathbb{F}[x]$ of degree n, the polynomial $f(x)$ has at most n roots.

We end the section by introducing some new terminology about whether a polynomial can be written as a product of factors of positive degree.

Definition 11.3.3

reducible, irreducible polynomial

A polynomial in $\mathbb{F}[x]$ of positive degree is a **reducible polynomial** in $\mathbb{F}[x]$ when it can be written as the product of two polynomials in $\mathbb{F}[x]$ of positive degree. Otherwise, we say that the polynomial is an **irreducible polynomial** in $\mathbb{F}[x]$.

The lemma, Degree of a Product, tells us that all linear polynomials are irreducible over any field. Also, from CPN it follows that linear polynomials are the *only* irreducible polynomials over the complex numbers $\mathbb{C}$.

What are the possible degrees of irreducible polynomials over other fields? Linear polynomials are not the only irreducible polynomials over an arbitrary field. Are there irreducible polynomials whose degree is greater than 1 over $\mathbb{R}$? What about $\mathbb{Q}$, or $\mathbb{Z}_p$ for prime numbers p?

In the following example, we give a cubic polynomial over $\mathbb{Z}_5$ that is irreducible. It is known that for all primes p, there are irreducible polynomials of any given positive degree, so there is no upper bond on the degree of irreducible polynomials in $\mathbb{Z}_p[x]$. Proving this result is difficult, and is left for a later course.

Example 8

Consider the polynomial $f(x) = [1]x^3 + [1]x + [1]$ over $\mathbb{Z}_5$. Observe that

$$f([0]) = [1], \quad f([1]) = [3], \quad f([2]) = [1], \quad f([3]) = [1], \quad f([4]) = [4],$$

so there are no solutions to the polynomial equation $f(x) = [0]$ in $\mathbb{Z}_5$, and hence $f(x)$ has no roots in $\mathbb{Z}_5$. Thus, by the Factor Theorem, the polynomial $f(x)$ has no factor of degree 1. Assume, for the sake of contradiction, that $f(x)$ is reducible over $\mathbb{Z}_5$. Then $f(x)$ can be written as the product of two polynomials of positive degree. From the lemma, Degree of a Product, these two polynomials must have degrees 1 and 2, since the only way of writing 3 as the sum of two positive integers is $3 = 1 + 2$ (or, equivalently, $3 = 2 + 1$). This means that $f(x)$ has a factor of degree 1, which is a contradiction, and we conclude that $f(x)$ is irreducible.

In the next two sections, we will consider polynomials over $\mathbb{R}$ and $\mathbb{Q}$. Among the results, we will prove in both cases that there are irreducible polynomials of degree greater than 1.

REMARK

The Unique Factorization Theorem in Chapter 6 tells us that every natural number can be written as a product of prime numbers uniquely, apart from the order of the factors.

A similar result holds for polynomials in $\mathbb{F}[x]$ over any field $\mathbb{F}$, though we won't prove it in this course. *Monic polynomials* are polynomials of positive degree in which the coefficient of the largest power of x is 1 (that is, the multiplicative identity in the field $\mathbb{F}$). The unique factorization result for polynomials is that every monic polynomial in $\mathbb{F}[x]$ can be written as a product of monic irreducible polynomials uniquely, apart from the order of the factors.

Of course, if $f(x)$ is a polynomial of positive degree in which the coefficient of the largest power of x is $c \neq 1$, then we can write $f(x) = c\,g(x)$ where $g(x)$ is a monic polynomial,

and then use the unique factorization result to write $g(x)$ as a product of monic irreducible polynomials.

11.4   Real Polynomials and the Conjugate Roots Theorem

In this section we consider polynomials over the real numbers. Of course, polynomials over the real numbers are also polynomials over the complex numbers, and we begin by considering the complex roots of these polynomials. The following result says that the complex roots that are not purely real come in pairs.

Theorem 8   **(Conjugate Roots Theorem (CJRT))**

For all polynomials $f(x)$ with real coefficients, if $c \in \mathbb{C}$ is a root of $f(x)$, then $\bar{c} \in \mathbb{C}$ is a root of $f(x)$.

Proof: Let $f(x)$ be an arbitrary polynomial with real coefficients, and assume that c is a root of $f(x)$. Then $f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$, where $n \geq 0$ is an integer, and $a_n, a_{n-1}, \ldots, a_1, a_0$ are all real numbers, and we have $f(c) = 0$, which gives

$$a_n c^n + a_{n-1} c^{n-1} + \cdots + a_1 c + a_0 = 0.$$

Taking the complex conjugate of both sides gives

$$\overline{a_n c^n + a_{n-1} c^{n-1} + \cdots + a_1 c + a_0} = \bar{0},$$

and using Properties of Conjugates, we obtain

$$\overline{a_n}\, \bar{c}^n + \overline{a_{n-1}}\, \bar{c}^{n-1} + \cdots + \overline{a_1}\, \bar{c} + \overline{a_0} = \bar{0}.$$

Since $\bar{a} = a$ whenever a is purely real, we now have

$$a_n\, \bar{c}^n + a_{n-1}\, \bar{c}^{n-1} + \cdots + a_1\, \bar{c} + a_0 = 0.$$

That is, $f(\bar{c}) = 0$, and hence $\bar{c}$ is a root of $f(x)$. $\square$

Example 9   Let $f(x) = x^4 - x^3 - 5x^2 - x - 6$. Given that i is a root of $f(x)$, write $f(x)$ as a product of linear factors in $\mathbb{C}[x]$, if possible.

Solution: Since $f(x)$ is a polynomial with real coefficients, we can use the Conjugate Roots Theorem. Thus, i and $-i$ are both roots and, by the Factor Theorem, $(x - i)$ and $(x + i)$ are factors of $f(x)$. The product of these two factors is $x^2 + 1$. Dividing $f(x)$ by $x^2 + 1$ yields a quotient of $x^2 - x - 6$, which can be written as the product $(x - 3)(x + 2)$. Thus

$$f(x) = (x - i)(x + i)(x - 3)(x + 2).$$

The Conjugate Roots Theorem has the following very useful corollary.

Corollary 9   **(Real Quadratic Factors (RQF))**

For all polynomials $f(x)$ with real coefficients, if $c \in \mathbb{C}$ is a root of $f(x)$, and $\text{Im}(c) \neq 0$, then there exists a real quadratic polynomial $g(x)$ and a real polynomial $q(x)$ such that $f(x) = g(x)q(x)$. Moreover, the quadratic factor $g(x)$ is irreducible in $\mathbb{R}[x]$.

Proof: Assume $c \in \mathbb{C}$ is a root of $f(x)$ and $\text{Im}(c) \neq 0$. Then by the Factor Theorem

$$f(x) = (x - c)q_1(x), \quad \text{for some } q_1(x) \in \mathbb{C}[x].$$

Now, by the Conjugate Roots Theorem, $\bar{c}$ is also a root of $f(x)$. Hence we have

$$f(\bar{c}) = (\bar{c} - c)q_1(\bar{c}) = 0.$$

Since $\text{Im}(c) \neq 0$, then $\bar{c} \neq c$, or $\bar{c} - c \neq 0$ which in turn means $q_1(\bar{c}) = 0$. That is, $\bar{c}$ is a root of $q_1(x)$ and so by using the Factor Theorem again, we get that

$$q_1(x) = (x - \bar{c})q_2(x), \quad \text{where } q_2(x) \in \mathbb{C}[x].$$

Substituting in our first equation for $f(x)$, we get

$$f(x) = (x - c)(x - \bar{c})q_2(x) = g(x)q_2(x),$$

where $g(x) = (x - c)(x - \bar{c})$. By Properties of Conjugates and Properties of Modulus,

$$g(x) = x^2 - (c + \bar{c})x + c\bar{c} = x^2 - 2\text{Re}(c)x + |c|^2.$$

Since $-2\text{Re}(c) \in \mathbb{R}$ and $|c|^2 \in \mathbb{R}$, $g(x)$ is a real quadratic polynomial. All that remains is to show that $q_2(x)$ is in $\mathbb{R}[x]$. From above, in $\mathbb{C}[x]$, we have that

$$f(x) = g(x)q_2(x) + r_2(x),$$

where $r_2(x)$ is the zero polynomial. Using the Division Algorithm for Polynomials (DAP) in $\mathbb{R}[x]$, we get

$$f(x) = g(x)q(x) + r(x),$$

where $q(x)$ is in $\mathbb{R}[x]$ and the remainder $r(x)$ is the zero polynomial or $\deg r(x) < \deg g(x)$. Now, every real polynomial is a complex polynomial, so we can also view this as a statement in $\mathbb{C}[x]$. As for any field, DAP over $\mathbb{C}$ tells us that the quotient and remainder are unique. Therefore $r(x) = r_2(x)$ is the zero polynomial and $q(x) = q_2(x)$ has real coefficients.

We finish by proving that $g(x)$ is irreducible in $\mathbb{R}[x]$. Note that $g(x) \in \mathbb{R}[x]$, so we also have $g(x) \in \mathbb{C}[x]$. Now since c is not purely real, we have $c \neq \bar{c}$, so c and $\bar{c}$ are both roots of $g(x)$. But from CPN, we know that $g(x)$ cannot have any additional roots in $\mathbb{C}$, so $g(x)$ has no purely real root. Hence, by the Factor Theorem, $g(x)$ has no linear factor in $\mathbb{R}[x]$. Assume for the sake of contradiction that $g(x)$ is reducible in $\mathbb{R}[x]$. Then $g(x)$ can be written as the product of two polynomials in $\mathbb{R}[x]$ of positive degree, and by the lemma, Degree of a Product, both polynomials must have degree 1. This means that $g(x)$ has a linear factor in $\mathbb{R}[x]$, which is a contradiction, and we conclude that $g(x)$ is irreducible in $\mathbb{R}[x]$. $\square$

Note how the proof of Real Quadratic Factors used information about non-real roots to prove a result about real polynomials. The result is now used to characterize the factorization of real polynomials, as given in the following theorem. We omit the proof, which can be obtained by induction on the degree n of the real polynomial, together with the above result Real Quadratic Factors.

Theorem 10   **(Real Factors of Real Polynomials (RFRP))**

For all real polynomials $f(x)$ of positive degree, $f(x)$ can be written as a product of real linear and real quadratic factors.

Note that the above results tell us that there are irreducible polynomials over the real numbers of degrees 1 and 2, but none of degree greater than 2.

11.5   Integer Polynomials and the Rational Roots Theorem

In this section we consider polynomials over the rational numbers. When the polynomial has *integer* coefficients, the following famous result gives a good starting point for possible *rational* roots of the polynomial. Recall from Example 15 in Chapter 6 that every rational number can be written in the form $\frac{p}{q}$, where p and q are coprime.

Theorem 11   **(Rational Roots Theorem (RRT))**

For all polynomials $f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_2 x^2 + a_1 x + a_0$ with integer coefficients and $n \geq 1$, if $\frac{p}{q}$ is a rational root of $f(x)$ with $\gcd(p, q) = 1$, then $p \mid a_0$ and $q \mid a_n$.

Proof: Let $f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_2 x^2 + a_1 x + a_0$, $n \geq 1$, be an arbitrary polynomial with integer coefficients, and assume $\frac{p}{q}$ is a root of $f(x)$. Then $f(\frac{p}{q}) = 0$, so we have

$$a_n \left(\frac{p}{q}\right)^n + a_{n-1} \left(\frac{p}{q}\right)^{n-1} + \cdots + a_2 \left(\frac{p}{q}\right)^2 + a_1 \left(\frac{p}{q}\right) + a_0 = 0.$$

Multiplying on both sides by q^n gives

$$a_n p^n + a_{n-1} p^{n-1} q + \cdots + a_2 p^2 q^{n-2} + a_1 p q^{n-1} + a_0 q^n = 0,$$

and rearranging, we obtain

$$a_n p^n = -q \left(a_{n-1} p^{n-1} + \cdots + a_2 p^2 q^{n-3} + a_1 p q^{n-2} + a_0 q^{n-1}\right).$$

All of the symbols in this equation are integers, so from the definition of divisibility we have

$$q \mid a_n p^n.$$

Now $\gcd(q, p) = 1$, so from Example 14 on page 106, we obtain $\gcd(q, p^n) = 1$. Hence by Coprimeness and Divisibility we get $q \mid a_n$.

The proof that $p \mid a_0$ is similar, and is omitted.   □

REMARK

Note that the above result is stated for the special case of rational polynomials in which all coefficients are integers. However, this result can be applied to any rational polynomial, for the following reason: Let $g(x)$ be an arbitrary rational polynomial, and suppose that m is a common denominator for all of the rational coefficients in $g(x)$. Then $f(x) = m\,g(x)$ is a polynomial with integer coefficients. Moreover, $f(x)$ and $g(x)$ have exactly the same roots, since $f(c) = 0$ if and only if $g(c) = 0$.

The Rational Roots Theorem says that in order to find the rational roots of a polynomial $f(x)$ with integer coefficients, we only need to examine a *finite* set of rational numbers, those whose numerator divides the constant term and whose denominator divides the leading coefficient. Note that the Rational Root Theorem is not an "if and only if" statement, so it only provides a list of rational numbers that are *candidates* to be roots. It does not guarantee that any of these rational numbers are roots.

Example 10   Find all rational roots of $f(x) = 2x^4 + x^3 + 6x + 3$, and express $f(x)$ as a product of irreducible polynomials in $\mathbb{Q}[x]$.

Solution: We will use the Rational Roots Theorem. The divisors of 2 are ± 1 and ± 2. The divisors of 3 are ± 1 and ± 3. Hence, the candidates for rational roots are

$$\pm 1, \pm \tfrac{1}{2}, \pm 3, \pm \tfrac{3}{2}.$$

Now test each of these candidates.

x	1	-1	$\frac{1}{2}$	$-\frac{1}{2}$	3	-3	$\frac{3}{2}$	$-\frac{3}{2}$
$f(x)$	12	-2	$\frac{25}{4}$	0	210	120	$\frac{51}{2}$	$\frac{3}{4}$

Thus, the only rational root is $-\frac{1}{2}$, so $2x + 1 = \frac{1}{2}\left(x - \left(-\frac{1}{2}\right)\right)$ is a factor, and by long division we obtain

$$f(x) = (2x + 1)(x^3 + 3). \tag{11.1}$$

The quotient $q(x) = x^3 + 3$ has no rational roots, since any rational root of $q(x)$ must be a rational root of $f(x)$, and since $q(-\frac{1}{2}) = \frac{23}{8} \neq 0$, so $-\frac{1}{2}$ is not a root of $q(x)$. Hence by the Factor Theorem, $q(x)$ has no linear factors.

Assume, for the sake of contradiction, that $q(x)$ is reducible in $\mathbb{Q}[x]$. Then from the lemma, Degree of a Product, $q(x)$ (which has degree 3) can be written as a product of two factors in $\mathbb{Q}[x]$, one of degree 1 and the other of degree 2. Hence $q(x)$ has a linear factor, which is a contradiction.

We conclude that $q(x) = x^3 + 3$ is irreducible in $\mathbb{Q}[x]$, so (11.1) gives a way of expressing $f(x)$ as a product of irreducible polynomials in $\mathbb{Q}[x]$.

Note that in the above example we have given a cubic polynomial that is irreducible in $\mathbb{Q}[x]$. It is known that there are irreducible polynomials of every positive degree in $\mathbb{Q}[x]$, so there is no upper bound on the degree of irreducible polynomials over the rational numbers. Proving this result is difficult, and is left for a later course.

In our next example, we show how the Rational Roots Theorem gives a compact proof that $\sqrt{2}$ is irrational. You may wish to compare this with our earlier proof on page 57 in Chapter 3, together with our remark about that proof following Example 15 in Chapter 6.

Example 11   Prove that $\sqrt{2}$ is irrational, using the Rational Roots Theorem.

Solution: Let $f(x) = x^2 - 2$, so we have $f(\sqrt{2}) = 0$. Hence $\sqrt{2}$ is a root of $f(x)$.

Now we apply the Rational Roots Theorem to $f(x)$. Since the divisors of 1 are ± 1, and the divisors of 2 are ± 1 and ± 2, the candidates for rational roots of $f(x)$ are

$$\pm 1, \pm 2.$$

But $f(1) = f(-1) = -1 \neq 0$, and $f(2) = f(-2) = 2 \neq 0$, which means that $f(x)$ has no rational roots. Since $\sqrt{2}$ is a root of $f(x)$ but $f(x)$ has no rational roots, we conclude that $\sqrt{2}$ is irrational.

We can also use the Rational Roots Theorem to prove that more complicated looking real numbers are irrational.

Example 12   Prove that $\frac{\sqrt{3}-2\sqrt{2}}{\sqrt{5}}$ is irrational, using the Rational Roots Theorem.

Solution: Let $a = \frac{\sqrt{3}-2\sqrt{2}}{\sqrt{5}}$. Multiply on both sides by $\sqrt{5}$ and square, to obtain

$$5a^2 = 11 - 4\sqrt{6}.$$

Rearranging this equation, we get $5a^2 - 11 = -4\sqrt{6}$, and squaring again gives

$$25a^4 - 110a^2 + 121 = 96.$$

Rearranging, we obtain $25a^4 - 110a^2 + 25 = 0$, and dividing on both sides by 5, we get $5a^4 - 22a^2 + 5 = 0$. Now, if we let

$$f(x) = 5x^4 - 22x^2 + 5,$$

then what we have shown above is that $f(a) = 0$, so $a = \frac{\sqrt{3}-2\sqrt{2}}{\sqrt{5}}$ is a root of $f(x)$.

Now we apply the Rational Roots Theorem to $f(x)$. Since the divisors of 5 are ± 1 and ± 5, the candidates for rational roots of $f(x)$ are

$$\pm 1, \pm \tfrac{1}{5}, \pm 5.$$

But $f(1) = f(-1) = -12 \neq 0$, $f(\frac{1}{5}) = f(-\frac{1}{5}) = \frac{516}{125} \neq 0$, and $f(5) = f(-5) = 2580 \neq 0$, which means that $f(x)$ has no rational roots. Since a is a root of $f(x)$ but $f(x)$ has no rational roots, we conclude that $a = \frac{\sqrt{3}-2\sqrt{2}}{\sqrt{5}}$ is irrational.

EXERCISE

Prove that if p is a prime, then $\sqrt[n]{p}$ is irrational for any integer $n > 1$.

11.6   More Examples for Roots and Factoring

Example 13   For each of the following, you are given an $\mathbb{F}[x]$ and several roots. Find a polynomial in the given $\mathbb{F}[x]$ of lowest possible degree that has the given roots.

- $\mathbb{R}[x]$, $3 + \sqrt{2}i$, 5.

  Solution: Since we are looking for a polynomial in $\mathbb{R}[x]$, we can use the Conjugate Roots Theorem for complex roots. Hence, $3+\sqrt{2}i \notin \mathbb{R}$ will be paired with its conjugate $3-\sqrt{2}i$. The product of the corresponding factors will produce a real quadratic. Hence one such polynomial is

  $$\begin{aligned} f(x) &= (x - (3 + \sqrt{2}i))(x - (3 - \sqrt{2}i))(x - 5) \\ &= (x^2 - 6x + 11)(x - 5) \\ &= x^3 - 11x^2 + 41x - 55. \end{aligned}$$

- $\mathbb{C}[x]$, $3 + \sqrt{2}i$, 5.

  Solution: Since both $3 + \sqrt{2}i$ and 5 are complex numbers, the corresponding linear factors are in $\mathbb{C}[x]$, so one such polynomial is

  $$f(x) = (x - (3 + \sqrt{2}i))(x - 5) = x^2 - (8 + \sqrt{2}i)x + (15 + 5\sqrt{2}i).$$

- $\mathbb{R}[x]$, $1 - \sqrt{5}$, $2i$, 0.

  Solution: Since we are looking for a polynomial in $\mathbb{R}[x]$, we can use the Conjugate Roots Theorem for complex roots. The only root not in $\mathbb{R}$ is $2i$ so we need to pair this root with its conjugate $-2i$. The product of the corresponding factors will produce a real quadratic. Hence one such polynomial is

  $$\begin{aligned} f(x) &= (x - (1 - \sqrt{5}))(x - 2i)(x + 2i)(x - 0) \\ &= (x - (1 - \sqrt{5}))(x^2 + 4)x \\ &= (x - (1 - \sqrt{5}))(x^3 + 4x) \\ &= x^4 - (1 - \sqrt{5})x^3 + 4x^2 - 4(1 - \sqrt{5})x. \end{aligned}$$

- $\mathbb{Z}_7[x]$, $[2]$, $[1]$.

  Solution: Both $[2]$, $[1]$ correspond to linear factors so one such polynomial is

  $$f(x) = ([1]x - [2])([1]x - [1]) = [1]x^2 - [3]x + [2] = [1]x^2 + [4]x + [2].$$

Example 14   Write each of the following polynomials $f(x)$ as a product of irreducible polynomials in $\mathbb{F}[x]$. Cite appropriate propositions to justify your reasoning.

- $f(x) = x^2 - x - 6$ in $\mathbb{Q}[x]$.

  Solution: The quadratic formula gives the roots 3 and -2. These are values in $\mathbb{Q}$ so $f(x)$ has linear factors $x - 3$ and $x + 2$ by the Factor Theorem. Hence,

  $$f(x) = (x - 3)(x + 2)$$

is a product of irreducible polynomials in $\mathbb{Q}[x]$.

- $f(x) = x^2 - x + 6$ in $\mathbb{Q}[x]$.

  Solution: The quadratic formula gives only complex roots in this instance. Since complex numbers do not belong to $\mathbb{Q}$ there are no linear factors in $\mathbb{Q}[x]$. Therefore by the lemma, Degree of a Product, $f(x) = x^2 - x + 6$ is irreducible $\mathbb{Q}[x]$.

- $f(x) = 2x^2 - 6ix - 4$ in $\mathbb{C}[x]$.

  Solution: Applying the quadratic formula gives two roots, i and $2i$, and hence

  $$f(x) = c(x^2 - 3ix - 2) = c(x - i)(x - 2i),$$

  for some complex number c is a product of irreducible polynomials in $\mathbb{C}[x]$. It is easy to see from the leading coefficient that $c = 2$.

- $f(x) = 2x^3 - 3x^2 + 2x + 2$ in $\mathbb{R}[x]$.

  Solution: Since all of the coefficients are integers, we can use the Rational Roots Theorem. The divisors of a_0 are ± 1 and ± 2, and the divisors of a_n are ± 1 and ± 2, so the only candidates for rational roots are

  $$\pm 1, \pm 2, \pm \tfrac{1}{2}.$$

  Now test each of these candidates.

x	1	-1	2	-2	$\frac{1}{2}$	$-\frac{1}{2}$
$f(x)$	3	-5	10	-30	$\frac{5}{2}$	0

  Hence $-\frac{1}{2}$ is a rational root, so $2x + 1 = 2\left(x - \left(-\frac{1}{2}\right)\right)$ is a factor, and from long division we obtain

  $$f(x) = (2x + 1)(x^2 - 2x + 2).$$

  The quadratic formula gives two non-real roots for $x^2 - 2x + 2$, so it has no real linear factors. Therefore, by the lemma, Degree of a Product, we have a product of irreducible polynomials in $\mathbb{R}[x]$.

- $f(z) = z^4 + 27z$ in $\mathbb{C}[z]$.

  Solution: Since $f(z)$ is a complex polynomial of degree four, it will have four linear factors. Now $f(z) = z(z^3 + 27)$. Factoring $z^3 + 27$ can be done with the aid of the Complex n-th Roots Theorem applied to $z^3 = -27$. First, we write -27 in polar form as

  $$-27 = 27(\cos \pi + i \sin \pi).$$

  Using the Complex n-th Roots Theorem, the solutions are

  $$\sqrt[3]{27}\left(\cos\left(\tfrac{\pi + 2k\pi}{3}\right) + i \sin\left(\tfrac{\pi + 2k\pi}{3}\right)\right) = 3\left(\cos\left(\tfrac{\pi}{3} + \tfrac{2k\pi}{3}\right) + i \sin\left(\tfrac{\pi}{3} + \tfrac{2k\pi}{3}\right)\right),$$

  for $k = 0, 1, 2$. Writing them out separately, the three distinct roots are

  $$k = 0, \qquad z_0 = 3\left(\cos\left(\tfrac{\pi}{3}\right) + i \sin\left(\tfrac{\pi}{3}\right)\right) = 3\left(\tfrac{1}{2} + \tfrac{\sqrt{3}}{2}i\right) = \tfrac{3}{2} + \tfrac{3\sqrt{3}}{2}i,$$

  $$k = 1, \qquad z_1 = 3\left(\cos \pi + i \sin \pi\right) = -3,$$

  $$k = 2, \qquad z_2 = 3\left(\cos\left(\tfrac{5\pi}{3}\right) + i \sin\left(\tfrac{5\pi}{3}\right)\right) = 3\left(\tfrac{1}{2} + -\tfrac{\sqrt{3}}{2}i\right) = \tfrac{3}{2} - \tfrac{3\sqrt{3}}{2}i.$$

Thus, we have the following product of irreducible polynomials in $\mathbb{C}[z]$:

$$f(z) = z(z+3)\left(z - \left(\tfrac{3}{2} + \tfrac{3\sqrt{3}}{2}i\right)\right)\left(z - \left(\tfrac{3}{2} - \tfrac{3\sqrt{3}}{2}i\right)\right).$$

Example 15   Factor $f(x) = 3x^4 - 5x^3 + x^2 - 5x - 2$ over $\mathbb{R}[x]$ and $\mathbb{C}[x]$ into a product of irreducible polynomials.

Solution: Since all of the coefficients are integers, we can use the Rational Roots Theorem. The divisors of a_0 are ± 1 and ± 2, and the divisors of a_n are ± 1 and ± 3, so the only candidates for rational roots are

$$\pm 1, \pm 2, \pm\tfrac{1}{3}, \pm\tfrac{2}{3}.$$

Now test each of these candidates.

x	1	-1	2	-2	$\frac{1}{3}$	$-\frac{1}{3}$	$\frac{2}{3}$	$-\frac{2}{3}$
$f(x)$	-8	12	0	100	$-\frac{100}{27}$	0	$-\frac{52}{9}$	$\frac{104}{27}$

Since 2 and $-\frac{1}{3}$ are roots, $x - 2$ and $x + \frac{1}{3}$ (or $3x + 1$) are factors. We can perform long division with $f(x)$ and the divisor $(x - 2)(3x + 1) = 3x^2 - 5x - 2$ to get

$$f(x) = (x - 2)(3x + 1)(x^2 + 1).$$

As we saw previously, $x^2 + 1$ is irreducible in $\mathbb{R}[x]$, so we get

$$f(x) = (x - 2)(3x + 1)(x^2 + 1) \in \mathbb{R}[x],$$

and

$$f(x) = (x - 2)(3x + 1)(x - i)(x + i) \in \mathbb{C}[x].$$

Example 16   Let $f(z) = z^6 + 4z^4 + z^2 + 4$. Given that $f(2i) = 0$, factor $f(z)$ into a product of irreducible polynomials over $\mathbb{C}[z]$.

Solution: Over $\mathbb{C}$, a polynomial of degree six will have six linear factors. Since all of the coefficients of $f(z)$ are real, the Conjugate Roots Theorem applies. Since $2i$ is a root, $-2i$ is also a root. Thus

$$(z - 2i)(z + 2i) = z^2 + 4$$

is a factor of $f(z)$. Long division produces

$$f(z) = z^6 + 4z^4 + z^2 + 4 = (z^2 + 4)(z^4 + 1).$$

We now factor $z^4 + 1$ using the Complex n-th Roots Theorem applied to $z^4 = -1$. First, we write -1 in polar form as

$$-1 = 1(\cos \pi + i \sin \pi).$$

So the four distinct roots are

$$\sqrt[4]{1}\left(\cos\left(\frac{\pi + 2k\pi}{4}\right) + i\sin\left(\frac{\pi + 2k\pi}{4}\right)\right) = \cos\left(\frac{\pi}{4} + \frac{k\pi}{2}\right) + i\sin\left(\frac{\pi}{4} + \frac{k\pi}{2}\right),$$

for $k = 0, 1, 2, 3$. Writing them out separately, the roots are

$$k = 0, \qquad z_0 = \cos\left(\frac{\pi}{4}\right) + i\sin\left(\frac{\pi}{4}\right) = \frac{1}{\sqrt{2}} + \frac{i}{\sqrt{2}},$$

$$k = 1, \qquad z_1 = \cos\left(\frac{3\pi}{4}\right) + i\sin\left(\frac{3\pi}{4}\right) = -\frac{1}{\sqrt{2}} + \frac{i}{\sqrt{2}},$$

$$k = 2, \qquad z_2 = \cos\left(\frac{5\pi}{4}\right) + i\sin\left(\frac{5\pi}{4}\right) = -\frac{1}{\sqrt{2}} + -\frac{i}{\sqrt{2}},$$

$$k = 3, \qquad z_3 = \cos\left(\frac{7\pi}{4}\right) + i\sin\left(\frac{7\pi}{4}\right) = \frac{1}{\sqrt{2}} + -\frac{i}{\sqrt{2}}.$$

Hence we get the following product of irreducible polynomials in $\mathbb{C}[z]$:

$$f(z) = (z - 2i)(z + 2i)(z - z_0)(z - z_1)(z - z_2)(z - z_3),$$

where the z_i are defined above.

Example 17   Let $f(x) = [1]x^2 + [1] \in \mathbb{Z}_3[x]$. Factor $f(x)$ into a product of irreducible polynomials.

After substituting all three elements of $\mathbb{Z}_3$, we see that $f(x)$ does not have any roots. Therefore by the Factor Theorem it doesn't have any linear factors, and by the lemma, Degree of a Product, there is no work to do. (Recall that the Fundamental Theorem of Algebra only applies to polynomials over $\mathbb{C}$.)

Example 18   Let $f(x) = [1]x^4 + [2]x^2 + [1] \in \mathbb{Z}_3[x]$. Factor $f(x)$ into a product of irreducible polynomials.

As in the previous example, we can deduce that $f(x)$ does not have any roots. This tells us again that $f(x)$ doesn't have any linear factors. So by the lemma, Degree of a Product, we only need to check if $f(x)$ can be written as a product of quadratic polynomials. In fact, $f(x) = ([1]x^2 + [1])^2$, so it can be factored into two quadratics. Can you see how we might have found this factorization? In general, this is a difficult problem – mathematicians do not have efficient methods for factoring in $\mathbb{Z}_p$, where p is a prime.

Chapter 12

Additional Material

12.1   Prime Numbers and the Riemann Hypothesis

The Riemann hypothesis, proposed by Bernhard Riemann in 1859, is a famous conjecture that has far-reaching consequences in many areas of mathematics. A proof of the Riemann hypothesis is one of the seven *Millenium Prize Problems* in mathematics. The first person to provide a proof will be awarded a prize of U.S. $1 million by the Clay Mathematics Institute.

This section explains the significance of the Riemann hypothesis to the problem of counting primes. All logarithms in this section are with respect to the base e.

Definition 12.1.1

prime counting function

For an integer $x \geq 2$, the number of primes in the interval $[2, x]$ is denoted by the **prime counting function $\pi(x)$**.

The following table gives the values of $\pi(x)$ for $2 \leq x \leq 20$.

x	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
$\pi(x)$	1	2	2	3	3	4	4	4	4	5	5	6	6	6	6	6	7	8	8

Mathematicians are interested in developing formulas for $\pi(x)$, i.e., formulas for the number of primes in the first x positive integers. To help us prove results for $\pi(x)$, we will denote the n-th prime number by p_n, for integers $n \geq 1$. Hence we have $p_1 = 2$, $p_2 = 3$, $p_3 = 5$, $p_4 = 7$, $p_5 = 11$, and so on.

We begin by proving Lemma 1, which gives an *upper bound* for p_n. The proof uses the summation formula in the following Exercise.

EXERCISE

Prove that $2^0 + 2^1 + 2^2 + \cdots + 2^{n-1} = 2^n - 1$ for all integers $n \geq 1$.

Lemma 1   For all positive integers n, we have $p_n \le 2^{2^{n-1}}$.

Proof: We prove this result by strong induction on n, where $P(n)$ is the statement

$$p_n \le 2^{2^{n-1}}.$$

Base case The statement $P(1)$ is $p_1 \le 2^{2^{1-1}} = 2$. Since $p_1 = 2$, we see that $P(1)$ is true.

Inductive Step Let k be an arbitrary natural number.

Assume the inductive hypothesis, $P(1) \wedge P(2) \wedge \ldots \wedge P(k)$. That is, we assume that $p_i \le 2^{2^{i-1}}$, for all integers $i = 1, 2, \ldots, k$.

We wish to prove $P(k+1)$, which is the statement

$$p_{k+1} \le 2^{2^k}.$$

Now, by the proof of Euclid's Theorem, we know that the integer $N = p_1 p_2 p_3 \cdots p_k + 1$ is divisible by a prime q that is greater than p_k. Hence

$$
\begin{aligned}
p_{k+1} \ &\le \ q \\
&\le \ N \\
&= \ p_1 p_2 p_3 \cdots p_k + 1 \\
&\le \ 2^{2^0} 2^{2^1} 2^{2^2} \cdots 2^{2^{k-1}} + 1, \quad \text{by the inductive hyptothesis} \\
&= \ 2^{1+2+4+\cdots+2^{k-1}} + 1 \\
&= \ 2^{2^k - 1} + 1, \quad \text{by the Exercise above} \\
&\le \ 2^{2^k - 1} + 2^{2^k - 1} \\
&= \ 2^{2^k}.
\end{aligned}
$$

The result is true for $n = k + 1$, and so holds for all $n \ge 1$ by the Principle of Strong Induction.   □

Now we apply Lemma 1 to obtain the following result, which gives a *lower bound* for the prime counting function $\pi(x)$.

Theorem 2   For all integers $x \ge 2$, we have $\pi(x) > \log(\log x)$.

Proof: Let x be an arbitrary integer such that $x \ge 2$, and let k be the integer satisfying

$$2^{2^{k-1}} \le x < 2^{2^k}.$$

We note that $k \ge 1$. By Lemma 1, we have $p_k \le 2^{2^{k-1}}$ and hence $\pi(x) \ge k$. Now, taking logarithms of both sides of the inequality $x < 2^{2^k}$ yields $\log x < 2^k \log 2$. Since $\log 2 < 1$, we have $\log x < 2^k$, and taking logarithms again yields $\log(\log x) < k \log 2$ and therefore $\log(\log x) < k$. It follows that $\pi(x) > \log(\log x)$.   □

The next result we prove gives a better lower bound for $\pi(x)$ than the lower bound of Theorem 2.

Theorem 3   For all integers $x \geq 2$, we have $\pi(x) \geq \log x/(2 \log 2)$.

Proof: Let x be an arbitrary integer such that $x \geq 2$. Define $k = \pi(x)$, so $p_1, p_2, \ldots, p_k$ gives all the prime numbers that are less than or equal to x. Now, for each integer $n \in [1, x]$, we can uniquely write

$$n = a^2 b,$$

where a and b are positive integers, and b is squarefree (i.e., not divisible by the square of a prime). Thus we have

$$a^2 \leq n \leq x,$$

and so $a \leq \sqrt{x}$. Since $b \leq n \leq x$, every prime factor of b must be at most x, and therefore at most p_k. Thus, since b is squarefree, we can write

$$b = p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k}, \quad \text{where each } e_i \in \{0, 1\}. \tag{12.1}$$

Now, there are at most $\sqrt{x}$ positive integers a satisfying $a^2 \leq x$. Also, there are at most 2^k positive integers b satisfying (12.1). Hence there are at most $\sqrt{x} 2^k$ integers $n \in [1, x]$ of the form $n = a^2 b$ where a, b are positive integers and b is squarefree. Since there are exactly x integers in $[1, x]$, we must have

$$\sqrt{x}\, 2^k \geq x,$$

and therefore $2^k \geq \sqrt{x}$. Taking logarithms of both sides gives $k \log 2 \geq \frac{1}{2} \log x$, and hence we obtain $\pi(x) \geq \log x/(2 \log 2)$. $\qquad\square$

The Prime Number Theorem, proven independently by Jacques Hadamard and Charles Jean de la Vallée Poussin in 1896, describes the asymptotic distribution of the prime numbers. In other words, it gives a formula for $\pi(x)$ that increases in accuracy as x gets larger. A proof of the Prime Number Theorem is demanding and left for later courses.

Theorem 4   **(Prime Number Theorem)**

$$\lim_{x \to \infty} \frac{\pi(x)}{x/\log x} = 1.$$

Finally, we get to the Riemann hypothesis, which has several equivalent formulations. The formulation we give below is a conjectured estimate for $\pi(x)$. It states that $\pi(x)$ is closely estimated by the logarithmic integral $\mathrm{Li}(x)$, i.e., the area under the curve $y = 1/\log t$ between $t = 2$ and $t = x$, the error in the approximation being at most $(\sqrt{x} \log x)/8\pi$.

Conjecture 5   **(Riemann Hypothesis)**

Let $\mathrm{Li}(x) = \int_2^x \frac{dt}{\log t}$ for $x \geq 2$. For all integers $x \geq 3000$, we have

$$|\pi(x) - \mathrm{Li}(x)| < (\sqrt{x} \log x)/8\pi.$$

Let us compare the estimates for $\pi(x)$ that are suggested by Theorems 2, 3, 4 and the Riemann hypothesis. We will consider the number of primes that are $\leq 10^{27}$. The exact value for $\pi(10^{27})$, obtained by a computer search, is given by

$$\pi(10^{27}) = 16\,352\,460\,426\,841\,680\,446\,427\,399. \tag{12.2}$$

Theorem 2 gives the extremely weak bound

$$\pi(10^{27}) \geq 5.$$

Theorem 3 gives the slightly better bound

$$\pi(10^{27}) \geq 45,$$

but this is still very far away from the actual value given in (12.2) above. The Prime Number Theorem suggests the estimate

$$\pi(10^{27}) \approx \frac{10^{27}}{\log 10^{27}} \approx 16\,084\,980\,811\,231\,549\,172\,264\,034,$$

which is much closer to (12.2), agreeing to the first 2 decimal digits. Finally, the Riemann hypothesis suggests the estimate

$$\pi(10^{27}) \approx \int_{2}^{10^{27}} \frac{dt}{\log t} \approx 16\,352\,460\,426\,842\,189\,113\,085\,404.$$

One can see that this value is a very close estimate to (12.2), agreeing to the first 14 decimal digits. This demonstrates the important role that the Riemann hypothesis plays in understanding the distribution of the prime numbers.