

Lecture Notes for Math 74: A Transition to
Upper Division Mathematics

Copyright © Scott N. Armstrong

November 15, 2006

Contents

1	Introduction	3
1.1	Mathematical Definitions	3
1.2	Propositional Logic	5
1.3	Sets and Quantifiers	8
1.4	Exercises and Problems	8
2	Sets and Functions	9
2.1	Sets and Subsets	9
2.2	Operations on Sets	13
2.3	The Power Set of a Set	16
2.4	Ordered pairs and Cartesian products	17
2.5	Functions	18
2.6	Images and Preimages	21
2.7	Injective, Surjective, and Bijective Functions	24
2.8	Sequences, Sums, and Products	26
2.9	Exercises and Problems	27
3	Mathematical Induction	28
3.1	Mathematical Induction	28
3.2	The Well-Ordering Principle	37
3.3	Strong Induction	42
3.4	Exercises and Problems	43
4	Elementary Number Theory	44
4.1	The Division Theorem	44
4.2	The Euclidean Algorithm	49
4.3	Prime Numbers and Euclid's Theorem	60
4.4	The Fundamental Theorem of Arithmetic	62

<i>CONTENTS</i>	2
4.5 Exercises and Problems	64
5 Modular Arithmetic	68
5.1 Congruence Modulo n	68
5.2 Linear Congruences	77
5.3 The Chinese Remainder Theorem	84
5.4 Systems of Linear Congruences	90
5.5 Congruence Classes	92
5.6 Fermat's Little Theorem	101
5.7 An Application: RSA Cryptography	102

Chapter 1

Introduction

1.1 Mathematical Definitions

Definitions play a special role in mathematics. We typically think of a definition as a *description* of the meaning(s) of a word or phrase as it is used by normal people in everyday life. That is, it is the usage of the word that determines its meaning. We can these kinds of definitions *extracted* definitions, and they are the most common kind of definition found in a standard dictionary.

In contrast, a mathematical definition is the birth of a new object or relationship between objects. When we make a mathematical definition, we are not simply making a description, we are making a *decision* about the meaning of a particular word or phrase. The meaning of the word is no longer open to debate. We call these kinds of definitions *stipulative*.

It is likely that you have already encountered the concept of a prime number. If this is the case, then you have developed a *concept image* of the phrase *prime number*. Pause for a moment and try to write down a definition for *prime*.

According to your definition, is 1 a prime number? According to your concept image of a prime number, is 1 prime? Since each of us think in different ways and may have different ideas of what a prime number is, we are likely to come up with different answers to these questions. Moreover, there are no “right” answer to these questions in any objective sense. This is a problem, because we need to eliminate any subjectivity from our mathematics. What we need is to decide once and for all what the word *prime*

means, to eliminate further confusion and so that we do not waste our energy arguing about semantics.

Definition 1.1. A positive integer p is called **prime** if and only if $p > 1$ and the only positive integers that divide p are 1 and p .

In a mathematical context, when we define a word such as *prime* we are making a final agreement as to the meaning of that word. We are not simply writing down a *description* of what a prime number is. Definition 1.1 is a decree establishing what is a prime number and what is not a prime number. The word *prime* means precisely what Definition 1.1 says it means, and nothing more. (One may object to Definition 1.1 on the basis that we have neglected to define the words *positive*, *integer*, and *divide*. This is a fair objection, but one which we will handle later.)

According to Definition 1.1, 1 is not prime, since it is not greater than itself. Those who consider 1 to be a prime number will have to adjust their concept images. Of course, mathematicians do not make their definitions arbitrarily or without careful consideration, and we will see later why it is a good idea to exclude 1 from the set of prime numbers (to skip ahead, see Remark 4.46 and Remark 4.50).

In these notes, when we make a definition, we will always write the term being defined in **boldface**.

We do not choose the names we give to our mathematical concepts arbitrarily. Our choices are often made on the basis of a preexisting concept image associated with a particular term which closely resembles the mathematical concept we wish to codify. But we must be careful to never confuse our concept image with our mathematical definition, and when doing mathematics (e.g., reading a proposition or writing a proof) the mathematical definition that we set is the ultimate authority on the meaning of each concept.

It is often the case that the same word or the same notation can be used in different ways by different authors or in different mathematical contexts (for example, see Remark 2.26). Sometimes authors choose different words for the same concepts. These are simply unfortunate facts of life. Therefore, when we read, write, teach or discuss mathematics we must always be clear which definition we are using in each context; and a student of mathematics must learn to translate between the different choices that different authors make.

Remark 1.2. In Definition 1.1, we used the phrase *if and only if*.

When say “if” in the course of a definition, we usually mean “if and only if.” For instance, in the definition of a prime number above, we mean for the reader to infer more than just “if a number is only divisible by itself and 1, then it is prime.” We also are making an agreement that nothing else will be considered to be prime.

1.2 Propositional Logic

Propositional logic is a system useful for formalizing logical reasoning. It consists of *statements*, which are declarative sentences having a *truth value* (in other words, each statement is either true or false, but not both), and *logical connectives*, which are logical operators that allow us to build new statements from previous ones.

We will usually assign statements the symbols P, Q, R, S, \dots . The logical connectives we will most often use are *not* (\neg), *and* (\wedge), *or* (\vee), *implies* (\Rightarrow), and *if and only if* (\Leftrightarrow).

In the context of propositional logic, it is important to keep in mind that we are primarily concerned with determining the truth or falsity of statements. Therefore, *understanding what a statement means is equivalent to knowing how to determine whether the statement is true or false*. For example, given two statements P and Q , the connective \wedge is an operator that creates a new statement $P \wedge Q$. Obviously, the truth value of $P \wedge Q$ will depend on the truth values of P and Q . Therefore, to define $P \wedge Q$, it is sufficient to write out a *truth table*:

P	Q	$P \wedge Q$
T	T	T
T	F	F
F	T	F
F	F	F

Given statements P and Q , there are four possibilities: either they are both true, both false, P is true and Q is false, or P is false and Q is true. The truth table above has one row for each of these possibilities and then gives the corresponding value of the statement $P \wedge Q$ (we have used the symbols T and F to represent “true” and “false,” respectively). Thus we see that $P \wedge Q$ is true if both P and Q are true, and it is false otherwise. This makes intuitive sense, because we read the statement $P \wedge Q$ aloud as “ P and Q .”

Via the truth table below, we define all the logical connectives that we will need.

P	Q	$\neg P$	$P \wedge Q$	$P \vee Q$	$P \Rightarrow Q$	$P \Leftrightarrow Q$
T	T	F	T	T	T	T
T	F	F	F	T	F	F
F	T	T	F	T	T	F
F	F	T	F	F	T	T

The last two columns of the truth table above often present difficulties for students. First, let's consider the statement $P \Rightarrow Q$. The statement $P \Rightarrow Q$ is called a conditional statement. The statement P is called the *hypothesis* of the conditional statement, and Q is called the *conclusion*. There are many ways to read the statement $P \Rightarrow Q$ in English, including:

“ P implies Q .”

“If P , then Q .”

“ Q if P .”

“ P is a sufficient condition for Q .”

“ Q is a necessary condition for P .”

“ P only if Q .”

“ Q or not P .”

As the truth table illustrates, the statement “If P , then Q ” is true if P and Q are both true, or if P is false. It is false if P is true and Q is false. To see why this makes intuitive sense, consider the following situation. Your acquaintance Bob makes the promise that

If Bob wins the lottery, he will give you half of his winnings.

Under what situation could you accuse Bob of breaking his promise? If he wins the lottery and gives you your promised share, obviously he has kept his word. Likewise, if he doesn't win the lottery, you cannot accuse him of breaking his promise. The only situation in which Bob breaks his promise is when he does win the lottery but does not give you half of his winnings. If we use P to denote the statement “Bob won the lottery” and Q to denote the statement “Bob gives you half his winnings,” then we observe that our intuition matches our truth table.

Remark 1.3. Notice that the statement

“If $2 + 3 = 6$, then $5 + 7 = 21$ ”

is true! If the hypothesis of a conditional statement is false (in our case, $2 + 3$ is not, in fact, equal to 6), the conditional statement is true, regardless of whether or not the conclusion is true.

The statement $P \Leftrightarrow Q$ can be read aloud as any one of the following:

“ P if and only if Q .”
 “ P is a necessary and sufficient condition for Q .”
 “ Q is a necessary and sufficient condition for P .”
 “If P , then Q , and if Q , then P .”
 “ Q if and only if P .”
 “ P implies Q and Q implies P .”
 “ P is logically equivalent to Q .”
 “ Q is logically equivalent to P .”
 “ P and Q are either both true, or else both false.”
 “If P , then Q , and if not P , then not Q .”

It is true if P and Q have the same truth value, and is false otherwise. The symbol \Leftrightarrow is the propositional logic version of an equal sign: the statement $P \Leftrightarrow Q$ asserts that P and Q are logically equivalent statements.

A **tautology** is a statement build out of simpler statements P, Q, R, \dots and logical connectives that is always true, no matter the truth value of P, Q, R, \dots . For example, the statement $P \vee \neg P$ is a tautology:

P	$\neg P$	$P \vee \neg P$
T	F	T
F	T	T

Another example of a tautology is the statement

$$(1.1) \quad [P \wedge Q] \Leftrightarrow [Q \wedge P],$$

as we check below:

P	Q	$P \wedge Q$	$Q \wedge P$	$P \Leftrightarrow Q$
T	T	T	T	T
T	F	F	F	T
F	T	F	F	T
F	F	F	F	T

Exercise 1.1. Using truth tables, show that the following statements are tautologies:

(a) $P \Rightarrow [P \vee Q]$.

(b) $[P \wedge (P \Rightarrow Q)] \Rightarrow Q$.

(c) $[P \Rightarrow Q] \Leftrightarrow [\neg Q \Rightarrow \neg P]$.

(d) $[(P \Rightarrow Q) \wedge (Q \Rightarrow R)] \Rightarrow [P \Rightarrow R]$.

Two important tautologies are called De Morgan's Laws. They are

$$(1.2) \quad \neg [P \wedge Q] \Leftrightarrow [\neg P \vee \neg Q]$$

and

$$(1.3) \quad \neg [P \vee Q] \Leftrightarrow [\neg P \wedge \neg Q].$$

To prove that De Morgan's Laws are tautologies, construct truth tables (exercise).

The opposite of a tautology is a **contradiction**. It is a statement that is always false, regardless of the truth value of the constitutive statements P, Q, R, \dots . For example, the statement $P \wedge \neg P$ is a contradiction.

Exercise 1.2. Come up with some examples of contradictions, each time using a truth table to establish your claim.

1.3 Sets and Quantifiers

1.4 Exercises and Problems

to be added later.

Chapter 2

Sets and Functions

Sets and functions between sets are two of the most foundational notions in mathematics. Every aspiring student of mathematics must become well acquainted with them. In this chapter, we will introduce these concepts and study their basic properties.

2.1 Sets and Subsets

A **set** is an undefined mathematical object. Intuitively, we think of a set as a collection of objects, but strictly speaking it has no definition. A **member** of a set is another undefined mathematical object. We think of a member of a set as being one of the objects that belongs to that set. If x is a member of the set A , then we say that x **belongs** to A and that x is an **element** of A .

We leave the words *set* and *member* undefined because we have not defined the terms *collection* or *object*. Moreover, if we tried to properly define these words, we would face the same difficulty. To make a proper definition, we must already have some (previously defined) words to work with. We have to start somewhere, and so our first “definition” must be left somewhat vague.

Notation 2.1. Suppose that A is a set. If x is an element of A , then we write

$$x \in A.$$

The expression above is read as “ x belongs to A ” or “ x is an member of A .”

Likewise, we write

$$y \notin A$$

if y is not an element of A .

Notation 2.2. We will use two main methods for specifying the sets we will use. The first is simply listing the elements belonging to the set between the curly brackets “{” and “}”. This notation for writing sets works well for many finite sets and some “small” infinite sets. For example, by writing

“let A be the set $\{3, 21, 7\}$ ”

we are assigning the name A to the set whose elements are 3, 7, and 21, and nothing more. Similarly, if we write

$$\text{Let } B = \{1, 2, 3, \dots, 14\}$$

then we are assigning is the name B to the set consisting of the first fourteen positive integers.

Definition 2.3. The set of **natural numbers**, also called the set of **non-negative integers**, is the set

$$\mathbb{N} = \{0, 1, 2, 3, 4, \dots\}.$$

The set of **positive integers** is the set

$$\mathbb{N}^* = \{1, 2, 3, 4, \dots\},$$

and we define the set of **integers** by

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

Warning 2.4. Consider the ambiguous phrase

“Let B denote the set $\{1, 4, \dots\}$.”

Is the reader to infer that

$$B = \{1, 4, 7, 10, 13, 16, \dots\}$$

or that

$$B = \{1, 4, 9, 16, 25, 36, \dots\}?$$

Reckless use of ellipses dots “...” can lead to trouble. If we use them, then we should take care to ensure that our meaning is clear.

Notation 2.5. A notation for specifying sets which is superior to the “element listing” method of Notation 2.2 above is called *set-builder notation*, in which we write an expression of the form

$$(2.1) \quad \{x \mid P(x)\}$$

where $P(x)$ is a predicate in the variable x . The set above consists of those elements x for which $P(x)$ is a true statement. Thus the predicate P provides a criterion for membership in the set. We read equation (2.1) aloud as “the set of elements x such that $P(x)$.” (Notice that the vertical line “ \mid ” stands for the phrase “such that”.)

Example 2.6. Let B denote the set consisting of the first fourteen positive integers. Using set-builder notation, we can write B as

$$(2.2) \quad B = \{n \mid n \in \mathbb{N} \text{ and } 1 \leq n \leq 14\}.$$

The above line is read aloud as “ B is the set of elements n such that n is a natural number and $1 \leq n \leq 14$.” Notice that this notation avoids criticism of Warning 2.4 by not using ellipses dots “...”. Moreover, any set that can be written using the element-listing method can be written in set-builder notation. For example, the set $A = \{3, 7, 21\}$ can be written in set-builder notation as

$$A = \{n \mid n = 3 \text{ or } n = 7 \text{ or } n = 21\}.$$

Set-builder notation provides an easy to check whether an object belongs to a set. For example, to check whether the integer 6 belongs to the set B above, we simply plug $n = 6$ into the predicate

$$n \in \mathbb{N} \text{ and } 1 \leq n \leq 14$$

and see if the statement is true. Since the statement

$$6 \in \mathbb{N} \text{ and } 1 \leq 6 \leq 14$$

is true, we see that $6 \in B$. Similarly, we can check that $-12 \notin B$ since $1 \not\leq -12$ and that $3.7 \notin B$ since $3.7 \notin \mathbb{N}$.

Example 2.7. Some sets that we can write in set-builder notation do not have any elements at all! For example, let

$$A = \{n \mid n \in \mathbb{Z} \text{ and } n \notin \mathbb{Z}\}.$$

Then A is the set consisting of those elements which are both integers and not integers! It is easy to see that A consists of no elements at all. Similarly, the set

$$\{x \mid x \in \mathbb{Z} \text{ and } x^2 < 0\}$$

has no elements whatsoever. A set containing no elements is said to be an **empty set**.

Sometimes we will encounter sets that are parts of larger sets. This notion is made precise in the following definition.

Definition 2.8. If A and B are two sets, we say that A is a **subset** of B if every element of A is also an element of B . We say that A is a **proper subset** of B if A is a subset of B and $A \neq B$. If A is a subset of B , we write

$$A \subseteq B.$$

Sometimes you will see the notation $A \subset B$, which means exactly the same thing as $A \subseteq B$.

Example 2.9. The set \mathbb{N} of natural numbers is a subset of the set \mathbb{Z} of integers.

Lemma 2.10. *Every set A is a subset of itself.*

Proof. The proof is left to the reader (see Exercise 2.6). □

Lemma 2.11. *Suppose that A is an empty set and B is any other set. Then $A \subseteq B$.*

Proof. By way of contradiction, suppose that there is some set B such that $A \not\subseteq B$. Then there is a element $x \in A$ such that $x \notin B$. But A has no elements, so $x \notin A$. Therefore $x \in A$ and $x \notin A$, which is a contradiction. □

Definition 2.12. We say that two sets A and B are **equal** if and only if $A \subseteq B$ and $B \subseteq A$.

Exercise 2.1. Show that every set is equal to itself.

Lemma 2.13. *If A and B are empty sets, then $A = B$.*

Proof. Since A is an empty set, Lemma 2.11 implies that $A \subseteq B$. Likewise, $B \subseteq A$ since B is an empty set. Therefore, $A = B$. □

Definition 2.14. According to Lemma 2.13, there is only one empty set. We call this set **the empty set**. It is denoted by the symbol \emptyset .

Lemma 2.15. *Suppose that A , B , and C are sets. If $A \subseteq B$ and $B \subseteq C$, then $A \subseteq C$.*

Proof. The proof is left to the reader (Exercise 2.3). □

2.2 Operations on Sets

In this section we will define some common operations that we will perform on our sets, and discuss their properties.

Definition 2.16. Let A and B be two sets. The **union** of A and B , denoted by $A \cup B$, is the set consisting of those elements which belong to A or B :

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

Here we use “or” in the inclusive sense; that is, we mean that x is a member of the union of A and B if x belongs to either A or B or both. The **intersection** of A and B is the set $A \cap B$ consisting of those elements which belong to both A and B :

$$A \cap B = \{x \mid x \in A \text{ and } x \in B\}.$$

It is clear from the definitions that $A \cup A = A \cap A = A$, $A \cup B = B \cup A$, $A \cap B = B \cap A$, $A \cap B \subseteq A, B$ and $A, B \subseteq A \cup B$ for all sets A and B .

Lemma 2.17. *Let A and B be sets. Then the following are equivalent:*

1. A is a subset of B ;
2. $A \cup B = B$;
3. $A \cap B = A$.

Proof. We will show that (2) \iff (1) \iff (3).

To show that (1) \implies (2), assume that $A \subseteq B$. We want to deduce from this assumption that $A \cup B = B$. It is clear that $B \subseteq A \cup B$. So we only need to show that $A \cup B \subseteq B$. To that end, let x be an arbitrary element of $A \cup B$. That means that x belongs to either A or B (or both). Either

way, $x \in B$, because if x belongs to A , then it also belongs to B since we are assuming $A \subseteq B$. Therefore $A \cup B \subseteq B$, and thus $A \cup B = B$.

The proof that (2) \implies (1) is easy: assuming that $A \cup B = B$, we easily deduce that $A \subseteq A \cup B = B$.

The proof that (3) \implies (1) is similarly easy: assuming that $A \cap B = A$, we conclude that $A = A \cap B \subseteq B$.

All we have left is to show that (1) \implies (3). So we assume that $A \subseteq B$, and we have to show that $A \cap B = A$. It is obvious that $A \cap B \subseteq A$, so we only need to show that $A \subseteq A \cap B$. So let x be an arbitrary element of A . By our assumption, we see that $x \in B$. Therefore x belongs to both A and B , so $x \in A \cap B$. Thus $A \subseteq A \cap B$, and hence $A = A \cap B$, which completes the proof. \square

We collect the associative and distributive properties of unions and intersections below.

Lemma 2.18 (Associativity Properties of Unions and Intersections). *Let A, B , and C be sets. Then*

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

Proof. The proof is left as an exercise. \square

Lemma 2.19 (Distributive Properties of Unions and Intersections). *Let A, B , and C be sets. Then*

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

Proof. We will show the first equality and leave the second as an exercise. Suppose that $x \in A \cup (B \cap C)$. Then $x \in A$ or $x \in B \cap C$. In the first case, we observe that x is an element of both $A \cup B$ and $A \cup C$, hence $x \in (A \cup B) \cap (A \cup C)$. In the second case, $x \in B$ and $x \in C$, so we see that $x \in A \cup B$ and $x \in A \cup C$. So again, $x \in (A \cup B) \cap (A \cup C)$. Thus

$$A \cup (B \cap C) \subseteq (A \cup B) \cap (A \cup C).$$

For the other direction, let y be an arbitrary element of $(A \cup B) \cap (A \cup C)$. Then y is an element of both of the sets $A \cup B$ and $A \cup C$. We will consider

two cases: $y \in A$ and $y \notin A$. If $y \in A$, then clearly $y \in A \cup (B \cap C)$. On the other hand, if $y \notin A$, then y must belong to both B and C since y belongs to both $A \cup B$ and $A \cup C$. Thus $y \in B \cap C$, which implies that $y \in A \cup (B \cap C)$. So we have shown that

$$(A \cup B) \cap (A \cup C) \subseteq A \cup (B \cap C),$$

which completes the proof that $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$. \square

We will now present a couple of useful facts relating intersections and unions to complements.

Definition 2.20 (De Morgan's Laws). Let A and X be sets. The **complement** $X \setminus A$ of A in X is the set

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

Lemma 2.21. For all sets A, B , and X ,

1. $X \setminus (A \cup B) = (X \setminus A) \cap (X \setminus B)$;
2. $X \setminus (A \cap B) = (X \setminus A) \cup (X \setminus B)$.

Proof. We will prove the first equality and leave the second to the reader. Suppose that x is an element of $X \setminus (A \cup B)$. This means that $x \in X$, but $x \notin A \cup B$. Hence $x \notin A$ and $x \notin B$. From this we deduce that $x \in X \setminus A$ and $x \in X \setminus B$, so that $x \in (X \setminus A) \cap (X \setminus B)$. Thus we have shown that

$$X \setminus (A \cup B) \subseteq (X \setminus A) \cap (X \setminus B).$$

Now consider an arbitrary element $y \in (X \setminus A) \cap (X \setminus B)$. Thus y belongs to both of the sets $(X \setminus A)$ and $(X \setminus B)$. Since y belongs to the first set, we have that $y \in X$ but $y \notin A$. Since y belongs to the second set, we deduce that $y \notin B$. Therefore, $y \in X$ but $y \notin A \cup B$. That is, $y \in X \setminus (A \cup B)$. This demonstrates that

$$(X \setminus A) \cap (X \setminus B) \subseteq X \setminus (A \cup B),$$

completing the proof. \square

2.3 The Power Set of a Set

Given a set X , we can think of each subset of X as being an element of a larger set, the set of all subsets of X . This idea is captured in the following definition.

Definition 2.22. Let X be a set. The **power set** of X , denoted by $\mathcal{P}(X)$, is the set of all subsets of X . That is,

$$\mathcal{P}(X) = \{A \mid A \subseteq X\}.$$

Notice that since X is a subset of itself, it is a member of its own power set: $X \in \mathcal{P}(X)$. Also, $\emptyset \in \mathcal{P}(X)$ for any set X , since the empty set is a subset of every set. Thus

$$\{X, \emptyset\} \subseteq \mathcal{P}(X).$$

We could also say that $\{X, \emptyset\} \in \mathcal{P}(\mathcal{P}(X))$. In English, “the set consisting of the set X and the empty set is a member of the power set of the power set of X .”

Remark 2.23. Notice that there is a difference between an element of a set, say a , and the *set* $\{a\}$ consisting of a and nothing else. They are *not* equal:

$$a \neq \{a\}.$$

The difference between the two is similar to the difference between a stamp and a stamp collection. A stamp collection is not itself a stamp, even if the stamp collection consists of only one stamp! So notice that if we say that $a \in X$ (“ a is a member of the set X ”), then $\{a\} \subseteq X$ and $\{a\} \in \mathcal{P}(X)$, but a is *not* a member of $\mathcal{P}(X)$ and $\{a\}$ is *not* a member of X !

Exercise 2.2. Let $X = \{1, 2\}$, and $Y = \{2, 3, 5\}$.

1. Calculate the power set $\mathcal{P}(X)$ of X .
2. Calculate $\mathcal{P}(Y)$ of Y .
3. What is $\mathcal{P}(X \cup Y)$?

Example 2.24. Let $A = \{a, b, c\}$, where a, b , and c are distinct elements of A (that is, no two of them are equal). Then we calculate that

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

Thus while A has 3 elements, $\mathcal{P}(A)$ has 8. We will see later that it is no coincidence that $2^3 = 8$.

2.4 Ordered pairs and Cartesian products

You have probably already encountered the idea of an *ordered pair* in your previous studies. We think of an ordered pair as something have two *coordinates*, and we use the notation (x, y) to denote an ordered pair. We consider two such ordered pairs (x_1, y_1) and (x_2, y_2) to be equal if they have the same coordinates, in the correct order. In other words, if $x_1 = x_2$ and $y_1 = y_2$. But the order matters. We want the point $(2, 1)$ to be distinct from the point $(1, 2)$!

We need a nice way to make this idea formal. That is, we need a way to precisely define an ordered pair using sets, but in a way that conforms to our vision of above of how ordered pairs should behave. This is more difficult than it appears at first glance. For example, a naive way to define an ordered pair might be to set $(x, y) = \{x, y\}$. But we quickly see that this doesn't work, since $\{1, 2\} = \{2, 1\}$!

Definition 2.25. Let $x \in X$ and $y \in Y$. Then the **ordered pair** (x, y) is the set

$$(x, y) = \{\{x\}, \{x, y\}\}.$$

We say that x and y are the **coordinates** of the ordered pair.

Remark 2.26. It is an unfortunate fact of life that the notation (a, b) is also used to denote the open intervals $\{x \in \mathbb{R} \mid a < x < b\}$. This does not usually pose a problem, since the context will allow us to determine a writer's intended meaning.

It does really matter what ordered pairs *are*, so long as they *behave* as we intended. In particular, we designed the definition of ordered pair so that we could prove the following lemma. After its proof, for the most part we will forget our definition of ordered pair and think of ordered pairs as a mathematical object which satisfies Lemma 2.27.

Lemma 2.27. *The ordered pairs (x, y) and (a, b) are equal if and only if $x = a$ and $y = b$.*

Proof. Assume that $x = a$ and $y = b$. Then clearly $\{x\} = \{a\}$ and $\{x, y\} = \{a, b\}$. Thus $(x, y) = \{\{x\}, \{x, y\}\} = \{\{a\}, \{a, b\}\} = (a, b)$.

Conversely, assume that $(x, y) = (a, b)$. That is, $\{\{x\}, \{x, y\}\} = \{\{a\}, \{a, b\}\}$. Then either $\{x\} = \{a\}$ or $\{x\} = \{a, b\}$. In both cases, $x = a$. We have left

to show that $y = b$. We know by our assumption that either $\{x, y\} = \{a\}$ or $\{x, y\} = \{a, b\}$. The first possibility implies that $x = y = a$, from which it follows that

$$\{\{a\}, \{a, b\}\} = \{\{x\}, \{x, y\}\} = \{\{a\}\}.$$

This implies that $b = a$, and so $b = y$ as well. In the second case, $\{x, y\} = \{a, b\}$, we know that either $y = a$ or $y = b$. If $y = a$, then since $x = a$ as well we deduce that $\{a, b\} = \{x, y\} = \{a\}$, from which it follows that $b = a$. Therefore $y = a = b$. We have exhausted all the possibilities, and so the proof is complete. \square

Definition 2.28. Let X and Y be sets. The **Cartesian product** of X and Y is defined to be the set

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

consisting of all ordered pairs whose first coordinate belongs to X and whose second coordinate belongs to Y .

Example 2.29. We denote the Cartesian product of a set X with itself by X^2 . That is, $X^2 = X \times X$. One Cartesian product you are already familiar with is $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$, what we sometimes call the **Euclidean plane**. Now you see why Lemma 2.27 is important: because we do not want the point $(2, 1) \in \mathbb{R}^2$ to be equal to the point $(1, 2) \in \mathbb{R}^2$, or any other point (a, b) of \mathbb{R}^2 besides itself; that is, we want $(2, 1) \neq (a, b)$ unless $a = 2$ and $b = 1$. We designed the definition of *ordered pair* to achieve this.

2.5 Functions

Definition 2.30. A **function** (also called a **map** or a **mapping**) is an ordered triple (X, f, Y) , where X and Y are sets and f is a rule which assigns to each element $x \in X$ a corresponding element $f(x)$ of Y , called the **value** of f at x . The **domain** of the function (X, f, Y) is the first coordinate; that is, the set X . Likewise, we call the set Y the **codomain** of (X, f, Y) .

We think of a function (X, f, Y) as being comprised of a set X of possible inputs (the domain), a set Y of possible outputs (the codomain), and a machine f which takes an input $x \in X$ and delivers the output $f(x) \in Y$ (the “rule”).

Notation 2.31. While a function is technically an ordered triple (X, f, Y) , we will usually refer to the rule f as being the function itself, but all the while keeping in the back of our mind that, technically, the function is the rule f along with the domain X and codomain Y . For example, we will often say something like

f is a function from X to Y

when what we really mean is

(X, f, Y) is a function.

Also, we will make frequent use of the notation

$$f : X \rightarrow Y,$$

which is a compact way of writing “ f is a function from X to Y .”

When we specify a particular function, we need to identify the domain, the codomain, and the rule. We will use the following notation to accomplish this:

$$\begin{aligned} f : X &\rightarrow Y, \\ x &\mapsto f(x). \end{aligned}$$

The top line specifies the domain, the codomain, and gives the name we assign to the rule (here we call it f). The bottom line declares what the rule actually *does*. We will reserve the barred arrow “ \mapsto ” for specifying function rules.

As an example, consider the function

$$\begin{aligned} f : \mathbb{R} &\rightarrow \mathbb{R}, \\ x &\mapsto x^2. \end{aligned}$$

This is the function whose domain and codomain are both \mathbb{R} and whose rule simply takes a real number and squares it. As an alternative, we may simply say “let $f : \mathbb{R} \rightarrow \mathbb{R}$ be the function given by $f(x) = x^2$.” This is a kind of shorthand for the notation above.

Warning 2.32. We must be very careful not to confuse a rule/function f with $f(x)$, its value at x . This has been known to cause a lot of trouble. So while we usually are not confused by calling f a function (when it is actually a rule), we will be careful to avoid calling $f(x)$ a function. Because if $f(x)$ is a function, and $f(x) = x^2$, then x^2 is a function! This makes no sense, because in this context x^2 is just a real number— the square of x — not a function.

Now we give some examples of functions.

Example 2.33. Perhaps the simplest example of a function is the function which does nothing. Let $i_X : X \rightarrow X$ be the function given by the rule $x \mapsto x$. That is, $i_X(x) = x$ for every $x \in X$. We call i_X the **identity map** on X .

Example 2.34. A **constant map** is a function $f : X \rightarrow Y$ whose rule sends every input $x \in X$ to the same output. For example, the function

$$\begin{aligned} f : \mathbb{R} &\rightarrow \mathbb{R}^2, \\ x &\mapsto (4, -7) \end{aligned}$$

is a constant map, since it sends every real number to the point $(4, -7)$ in the plane \mathbb{R}^2 .

Example 2.35. If A is a subset of X , we define the **characteristic function of A in X** to be the function $\chi_A : X \rightarrow \{0, 1\}$ given by

$$\chi_A(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \in X \setminus A. \end{cases}$$

We think of χ_A as being a machine which answers the question “is my input a member of the set A ?” The output is either 1 (yes) or 0 (no).

Example 2.36. Let X and Y be sets, and consider the product $X \times Y$. The function

$$\begin{aligned} p : X \times Y &\rightarrow X, \\ (x, y) &\mapsto x \end{aligned}$$

is called the **projection of $X \times Y$ onto X** . Likewise, the the **projection of $X \times Y$ onto Y** is the map

$$\begin{aligned} q : X \times Y &\rightarrow Y, \\ (x, y) &\mapsto y. \end{aligned}$$

Example 2.37. Suppose that $f : X \rightarrow Y$ is a map, and that A is a subset of X . Then the **restriction of f to A** is the function, denoted by $f|_A$, which

has the same rule as f but can only take inputs from the smaller domain A . That is, we define the function $f|_A$ by

$$\begin{aligned} f|_A : A &\rightarrow Y, \\ x &\mapsto f(x). \end{aligned}$$

It may seem strange to distinguish between f and $f|_A$ since they have the same rules, but keep in mind that functions are not just rules, but ordered triples. When we remember this, it is easy to see why $(X, f, Y) \neq (A, f|_A, Y)$, because their first coordinates are not the same. The **inclusion map of A into X** is defined to be the restriction of the identity map on X to A . That is, the inclusion map of A into X is the function $(i_X)|_A$, given by

$$\begin{aligned} (i_X)|_A : A &\rightarrow X, \\ x &\mapsto x. \end{aligned}$$

Example 2.38. Another example of a function that you have probably encountered before but that you may not recognize as a function is integration. Let

$$\mathcal{C}[0, 1] = \{f \mid f : [0, 1] \rightarrow \mathbb{R} \text{ and } f \text{ is continuous}\}$$

be the set of functions whose domain is the interval $[0, 1] \subseteq \mathbb{R}$, whose codomain is \mathbb{R} , and which are continuous. Now, we have not defined what a continuous function is, but if you have taken calculus courses you have probably been convinced that such things exist and that we can integrate them. Define a map I by

$$\begin{aligned} I : \mathcal{C}[0, 1] &\rightarrow \mathbb{R}, \\ f &\mapsto \int_0^1 f(x) dx. \end{aligned}$$

Notice that the domain of I is a set of functions— that is, I is a function that takes other functions as inputs!

2.6 Images and Preimages

Definition 2.39. Let $f : X \rightarrow Y$ be a function, let $A \subseteq X$ and $B \subseteq Y$. The **image of A under f** is the set, which we denote by $f(A)$, defined by

$$f(A) = \{f(x) \mid x \in A\}.$$

The **range of f** is the set $f(X)$. The **preimage of B under f** is the set

$$f^{-1}[B] = \{x \in X \mid f(x) \in B\}.$$

Do not let the notation we use for images and preimages confuse you. Since $x \in X$ and $A \subseteq X$ are very different objects, likewise $f(x)$ and $f(A)$ are very different objects, even though the notation is similar. While $f(x) \in Y$ is the output of the function f given the input $x \in X$, the image $f(A)$ is a *subset* of Y consisting of those outputs which can be obtained by all the inputs from A . Also, we have not defined an “inverse function” as of yet. While preimages and inverses (when they can be defined) are related, at this point you should pretend that you’ve never seen the symbol f^{-1} before and let the your understanding of the definition of preimage guide your intuition, not the other way around. You can only talk about the preimage of a *subset* of the codomain, you can *not* refer to the “preimage of an element y of the codomain,” an expression which is meaningless.

Example 2.40. Let $h : \mathbb{R} \rightarrow \mathbb{R}$ be the function given by $h(x) = x^2$. The range $h(\mathbb{R})$ of h is the set of positive real numbers:

$$h(\mathbb{R}) = \{w \in \mathbb{R} \mid w \geq 0\}.$$

If $\alpha > 0$ is a large positive real number, then the image of the interval $[\alpha, \infty)$ is the interval $[\alpha^2, \infty)$. The preimages of the sets $\{-1\}$ and $\{2, 9\}$ are

$$f^{-1}[\{-1\}] = \emptyset \quad \text{and} \quad f^{-1}[\{2, 9\}] = \{\sqrt{2}, -\sqrt{2}, 3, -3\}.$$

Proposition 2.41. *Let $f : X \rightarrow Y$ and $A, B \subseteq X$. Then*

$$f(A \cap B) \subseteq A \cap B.$$

Proof. Choose a point $y \in f(A \cap B)$. Then there exists a point $x \in A \cap B$ such that $y = f(x)$. Since $x \in A$ and $y = f(x)$, it is clear that $y \in f(A)$. Likewise, $y \in f(B)$. Hence $y \in f(A) \cap f(B)$. \square

The next two propositions record some important relationships between preimages and images.

Proposition 2.42. *Let $f : X \rightarrow Y$ be a function, and suppose that $A \subseteq X$. Then*

$$A \subseteq f^{-1}[f(A)].$$

Proof. Pick an arbitrary element $x \in A$. Then clearly $f(x) \in \{f(z) \mid z \in A\} = f(A)$. Thus $x \in \{z \in X \mid f(z) \in f(A)\} = f^{-1}[f(A)]$. \square

Proposition 2.43. *Let $f : X \rightarrow Y$ be a function, and suppose that $B \subseteq Y$. Then*

$$f(f^{-1}[B]) \subseteq B.$$

Proof. Let y be an arbitrary element of the set

$$f(f^{-1}[B]) = \{f(x) \in Y \mid x \in f^{-1}[B]\}.$$

Then there exists an $x \in f^{-1}[B]$ such that $y = f(x)$. But since $x \in f^{-1}[B]$, we see that $f(x) \in B$. Thus $y \in B$. \square

Definition 2.44. Suppose that X, Y, Z are functions, and let $f : X \rightarrow Y$ and $g : Y \rightarrow Z$ be maps. We define the **composition of g with f** is the function $g \circ f : X \rightarrow Z$ defined by

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

Example 2.45. Denote the set of nonnegative real numbers by $\mathbb{R}_+ = \{x \in \mathbb{R} \mid x \geq 0\}$. Let $f : \mathbb{R} \rightarrow \mathbb{R}_+$ be given by the rule $f : x \mapsto x^2$ and $g : \mathbb{R}_+ \rightarrow \mathbb{R}$ be given by the rule $x \mapsto \sqrt{x}$. Then the composition of g with f is the absolute value function $x \mapsto |x|$.

Proposition 2.46. *Suppose that X, Y, Z are functions, and let $f : X \rightarrow Y$ and $g : Y \rightarrow Z$ be maps. If A is a subset of X , then*

$$(g \circ f)(A) = g(f(A)).$$

Likewise, if $B \subseteq Z$, then

$$(g \circ f)^{-1}[B] = f^{-1}[g^{-1}[B]].$$

Proof. Suppose that $z \in (g \circ f)(A)$. Then there exists $x \in A$ for which $z = (g \circ f)(x) = g(f(x))$. Clearly $f(x) \in f(A)$, and therefore $z \in g(f(A))$ since $z = g(f(x))$. Thus $(g \circ f)(A) \subseteq g(f(A))$.

Now suppose that $w \in g(f(A))$. Then there exists an element $y \in f(A)$ such that $w = g(y)$. Also, there exists an element $\tilde{x} \in A$ for which $y = f(\tilde{x})$. Therefore, $w = g(y) = g(f(\tilde{x})) = (g \circ f)(\tilde{x})$, and so $w \in (g \circ f)(A)$. This demonstrates that $g(f(A)) \subseteq (g \circ f)(A)$, which completes the proof of the first equality. The second equality is similar, and is left to the reader. \square

2.7 Injective, Surjective, and Bijective Functions

Definition 2.47. A function $f : X \rightarrow Y$ is called **surjective** if $f(X) = Y$. That is, if for every element $y \in Y$ there exists an element $x \in X$ such that $y = f(x)$. Informally, this means that every point of the codomain is “hit” by f . A function $f : X \rightarrow Y$ is called **injective** if $f(x_1) = f(x_2)$ implies that $x_1 = x_2$ for all $x_1, x_2 \in X$. Informally, this means that no two distinct points in the domain are mapped to the same point in the codomain. In some books, surjective functions are called “onto” functions, and injective functions are called “one-to-one” functions. We say that a function is **bijective** if it is both injective and surjective.

Example 2.48. Some functions are neither surjective nor injective. For example, consider the function $f : \mathbb{R} \rightarrow \mathbb{R}$ defined by the rule $x \mapsto x^2$. Then f is not surjective, since there exists a point $y \in \mathbb{R}$ such that $y \neq f(x)$ for every $x \in \mathbb{R}$. For example, for each real number x , $-1 \neq f(x)$. Also, f is not injective, since there exist real numbers x_1 and x_2 such that $x_1 \neq x_2$ but $f(x_1) = f(x_2)$. For example, $f(-2) = 4 = f(2)$.

Example 2.49. An example of the function which is injective but not surjective is the map $g : \mathbb{N} \rightarrow \mathbb{R}$ given by the rule $n \mapsto n$. Clearly g is injective, since if $g(n_1) = g(n_2)$ then $n_1 = g(n_1) = g(n_2) = n_2$. But g is not surjective, since $\pi \notin g(\mathbb{N})$. That is, there is no $n \in \mathbb{N}$ such that $g(n) = \pi$.

Example 2.50. It is easy to construct a function which is surjective but not injective. Let h be the map $h : \mathbb{R} \rightarrow \{1\}$ given $h(x) = 1$. Obviously, h is surjective since $h(\mathbb{R}) = \{1\}$, but h is not injective since $h(0) = h(1) = 1$.

Example 2.51. If X is any set, then the identity map i_X on X is bijective.

We can determine whether functions are injective or surjective by studying their images and preimages, as the next few propositions demonstrate.

Proposition 2.52. *Suppose that $f : X \rightarrow Y$. Then f is injective if and only if*

$$f^{-1}[f(A)] = A$$

for every subset $A \subseteq X$.

Proof. Suppose that f is injective. Choose a subset $A \subseteq X$. We will show that

$$f^{-1}[f(A)] = A.$$

By Proposition 2.42, we have that $A \subseteq f^{-1}[f(A)]$. For the other direction, pick an element $x \in f^{-1}[f(A)]$. This means that $f(x) \in f(A)$. So we can select a point $\tilde{x} \in A$ such that $f(x) = f(\tilde{x})$. Using the fact that f is injective, we deduce that $x = \tilde{x}$. Thus implies that $x \in A$, so we have shown that $f^{-1}[f(A)] \subseteq A$.

Therefore, we have proven that if f is injective then $f^{-1}[f(A)] = A$ for all subsets $A \subseteq X$.

Now assume that $f^{-1}[f(A)] = A$ for all subsets $A \subseteq X$. We will show that f is injective. To that end, select points $x_1, x_2 \in X$ for which $f(x_1) = f(x_2)$. This implies that

$$x_1 \in \{z \in X \mid f(z) = f(x_2)\} = \{z \in X \mid f(z) \in f(\{x_2\})\} = f^{-1}[f(\{x_2\})].$$

Using our assumption, we see that $f^{-1}[f(\{x_2\})] = \{x_2\}$, which implies that $x_1 = x_2$. \square

Proposition 2.53. *Suppose that $f : X \rightarrow Y$. Then f is surjective if and only if*

$$f(f^{-1}[B]) = B$$

for every subset $B \subseteq Y$.

Proof. The proof is left as an exercise. \square

Proposition 2.54. *Suppose that $f : X \rightarrow Y$. Then f is injective if and only if*

$$f(A \cap B) = f(A) \cap f(B).$$

for all subsets $A, B \subseteq X$.

Proof. Assume that f is injective, and that A and B are subsets of X . We will show that $f(A \cap B) = f(A) \cap f(B)$. From Proposition 2.41, we have that $f(A \cap B) \subseteq f(A) \cap f(B)$, so we only need to demonstrate that $f(A) \cap f(B) \subseteq f(A \cap B)$. To that end, choose a point $y \in f(A) \cap f(B)$. This means that y belongs to both $f(A)$ and $f(B)$. Since $y \in f(A)$, we may find a point $x_1 \in A$ such that $y = f(x_1)$. Likewise, since $y \in f(B)$, we may select a point $x_2 \in B$ for which $y = f(x_2)$. Since f is injective, and $f(x_1) = y = f(x_2)$, we

deduce that $x_1 = x_2$. Thus $x_1 \in A \cap B$, and since $y = f(x_1)$, we see that $y \in f(A \cap B)$.

Now suppose that $f(A \cap B) = f(A) \cap f(B)$ for all subsets $A, B \subseteq X$. We must show that f is injective. To that end, select points $x_1, x_2 \in X$ such that $f(x_1) = f(x_2)$. Define $A = \{x_1\}$ and $B = \{x_2\}$. Then using our assumption, we have that

$$f(A \cap B) = f(A) \cap f(B) = \{f(x_1)\} \cap \{f(x_2)\} = \{f(x_1)\}.$$

This implies that $A \cap B$ is nonempty, for otherwise the set $f(A \cap B)$ would be empty. But $\{x_1\} \cap \{x_2\} \neq \emptyset$ implies that $x_1 = x_2$. \square

Proposition 2.55. *Suppose that $f : X \rightarrow Y$. Then f is injective if and only if*

$$f(A \setminus B) = f(A) \setminus f(B).$$

for all subsets $A, B \subseteq X$.

Proof. The proof is left to the reader as an exercise. \square

2.8 Sequences, Sums, and Products

It is often useful to think of objects as being arranged in a sequence: a_1 , then a_2 , then a_3, \dots , where we think of a_k as being the k th object in the sequence. Before proceeding further, we should provide a rigorous definition for the term *sequence*.

Definition 2.56. Let X be a set. A **sequence in X** is a function a whose domain is \mathbb{N}^* or \mathbb{N} and whose codomain is X . It is common in to write a_n instead of $a(n)$ if a is a sequence. Sometimes we also refer to the range $\{a_n \mid n \in \mathbb{N}\}$ of the sequence as the sequence itself. In fact, the latter is sometimes written as $\{a_n\}_{n \in \mathbb{N}}$ or $\{a_n\}_{n=1}^{\infty}$ or even simply as $\{a_n\}$.

Thus even though we use funny notation for sequences, a sequence is just a function: a function whose domain is \mathbb{N} .

Some sequences can be defined in terms of themselves by means of a *recurrence relation*. For instance, thinking again of a sequence as a list of objects, consider the sequence $\{F_n\}$ whose first term is $F_1 = 1$, whose second

term is $F_2 = 1$, and whose every term after that is obtained by adding the previous two. That is,

$$(2.3) \quad \begin{cases} F_1 = F_2 = 1, \\ F_{n+2} = F_{n+1} + F_n \quad (n = 1, 2, 3, \dots). \end{cases}$$

According to our prescription above, we see that $F_3 = F_2 + F_1 = 1 + 1 = 2$, $F_4 = 3$, $F_5 = 5$, and the next few terms after F_5 are 8, 13, 21, 34, 55, 89, 144, 233, 377, 610, 987 and so on. A *recurrence relation* is an expression like equation (2.3), which prescribes the first few terms in a sequence and then gives a rule for generating the next term in the sequence in terms of the previous ones.

When we remember that a sequence is really a function, this recursive way of specifying a sequence may seem troubling. There is a rigorous way of dispensing with this objection, but we will not address it here.

2.9 Exercises and Problems

Exercise 2.3. Prove Lemma 2.15.

Exercise 2.4. Prove Lemma 2.13.

Exercise 2.5. Show that the empty set is a subset of every set.

Exercise 2.6. Prove Lemma 2.10

Chapter 3

Mathematical Induction

An important feature of the set \mathbb{N} of natural numbers is the presence of a natural *order*. Given two natural numbers m and n , then either $m \leq n$ or $n \leq m$ (or both, if $m = n$). Moreover, for every integer m , there is a *successor* to m , namely $m + 1$, which is the *least integer greater than m* . Most of us are so accustomed to this concept, that it may not seem very special. In this chapter we will learn some powerful techniques which exploit this simple concept, and will use them to prove some interesting theorems.

3.1 Mathematical Induction

Axiom 3.1 (Mathematical Induction). Suppose that $\mathcal{I} \subset \mathbb{N}^*$ is a subset of the set \mathbb{N}^* of positive integers with the following two properties:

1. $1 \in \mathcal{I}$; and
2. whenever a positive integer n belongs to \mathcal{I} , its successor $n + 1$ is a member of \mathcal{I} as well.

Then $\mathcal{I} = \mathbb{N}^*$.

Axiom 3.1 is often called *the Principle of Mathematical Induction* or simply *induction*. It is an *axiom*: a statement about the set \mathbb{N}^* that we take to be true without proof.

To get an intuitive idea of what Axiom 3.1 is asserting, consider the following thought experiment. Suppose on a certain street there is row of houses that is infinite in one direction. Suppose we know that the first house

is painted red. Suppose we also know that *if* it should be the case that a certain house is painted red, *then* the next house is also painted red. Then we can deduce the second house must be painted red because the first one is, the third house must be painted red because the second one is, the fourth house must be painted red because the third one is, and on and on. Intuitively, we see that *every* house should be painted red. The Principle of Mathematical Induction effectively makes this idea rigorous.

Example 3.2. The Principle of Mathematical Induction is useful for proving things about the natural numbers. For example, consider the statement

$$(3.1) \quad 2^n > n \text{ for every } n \in \mathbb{N}.$$

It is not difficult to convince oneself that (3.1) is true. Simply check that

$$\begin{aligned} 2^1 &= 2 > 1 & (n = 1) \\ 2^2 &= 4 > 2 & (n = 2) \\ 2^3 &= 8 > 3 & (n = 3) \\ 2^4 &= 16 > 4 & (n = 4) \\ 2^5 &= 32 > 5 & (n = 5) \\ 2^6 &= 64 > 6 & (n = 6) \\ &\vdots \end{aligned}$$

It seems that 2^n will continue to “get bigger faster than” n . However, this intuition is not a rigorous argument. The difficulty is that(?) is really an infinite number of inequalities, and we need a way of proving it in only finitely many words. We can check that $2^n > n$ is true for whatever particular values of n we like, but we cannot check it for *every* value of n no matter how much time we spend trying. Also, if we try to begin our proof by saying “Let n be an arbitrary positive integer...” it will be impossible to argue that $2^n > n$. (Try it!) We simply do not have enough information about n to work with. The Principle of Mathematical Induction provides a solution to this problem.

Proposition 3.3. *For every positive integer n ,*

$$2^n > n.$$

Before we prove Proposition 3.3, we summarize the main ideas of our proof. We saw above that 2^n is “much larger” than n as we increase the value of n . That is, part of the reason that 2^5 is so much larger than 5 is because 2^4 is larger than 4. Put another way, the truth of the statement $2^4 > 4$ is related to the truth of statement $2^5 > 5$. When we encounter a situation like this, we should attempt a proof by induction.

An induction argument requires us to check a single case ($n = 1$) and then show that the truth of the statement for n forces the statement to be true for $n + 1$ as well. We have already done the first part, so we are left with the second part. We need to figure out how exactly the truth of the statement $2^n > n$ “forces” the statement $2^{n+1} > n + 1$ to be true as well. So let us assume that $2^n > n$. Let’s work backwards, manipulating the statement $2^{n+1} > n + 1$ until we get something that we *know* to be true:

$$\begin{aligned} 2^{n+1} &> n + 1 ? \\ 2 \cdot 2^n &> n + 1 ? \end{aligned}$$

Now, we’ve assumed that $2^n > n$ and that n is a positive integer, so

$$2 \cdot 2^n > 2n = n + n \geq n + 1.$$

Now we have an explicit calculation showing that $2^n > n$ implies that $2^{n+1} > n + 1$. We are ready to rewrite our ideas into a proof.

Proof of Proposition 3.3. Let \mathcal{I} be the set

$$\mathcal{I} = \{n \in \mathbb{N}^* \mid 2^n > n\}.$$

We need to show that $\mathcal{I} = \mathbb{N}^*$. We will use the principle of mathematical induction to accomplish this. First, notice that $1 \in \mathcal{I}$ since $2^1 = 2 > 1$. Now assume that n is a positive integer that belongs to \mathcal{I} ; that is, $n \in \mathbb{N}^*$ and $2^n > n$. Then using this assumption and the fact that $n \geq 1$, we deduce that

$$2^{n+1} = 2 \cdot 2^n > 2n = n + n \geq n + 1.$$

Thus $n + 1 \in \mathcal{I}$ as well.

We have shown that $1 \in \mathcal{I}$, and that $n + 1 \in \mathcal{I}$ whenever $n \in \mathcal{I}$. Therefore, by Axiom 3.1, we deduce that $\mathcal{I} = \mathbb{N}^*$. \square

As we become for familiar with mathematical induction, we will usually dispense with explicitly mentioning the set \mathcal{I} . For example, we may rewrite the proof of Proposition 3.3 more succinctly as follows.

More succinct proof of Proposition 3.3. We will proceed by induction. It is clear that $2^1 > 1$. If n is a positive integer such that $2^n > n$, then

$$2^{n+1} = 2 \cdot 2^n > 2n = n + n \geq n + 1.$$

Therefore, by induction we conclude that $2^n > n$ for all integers $n \geq 1$. \square

In fact, we may rephrase the principle of mathematical induction in a way that doesn't involve the set \mathcal{I} at all.

Theorem 3.4 (Principle of Mathematical Induction, Predicate Form). *Suppose that $P(n)$ is a predicate in the variable $n \in \mathbb{N}^*$. Suppose also that $P(1)$ is true, and that the statement*

$$P(n) \text{ implies } P(n+1)$$

is true. Then $P(n)$ is true for all positive integers n .

Proof. Let

$$\mathcal{I} = \{n \in \mathbb{N}^* \mid P(n)\}.$$

By assumption, $1 \in \mathcal{I}$. Suppose that $n \in \mathcal{I}$. Then $P(n)$ is true. Thus, by hypothesis, we deduce that $P(n+1)$ is true. Thus $n+1 \in \mathcal{I}$. We have shown that $n+1 \in \mathcal{I}$ whenever $n \in \mathcal{I}$, and that $1 \in \mathcal{I}$. Therefore, by induction, $\mathcal{I} = \mathbb{N}^*$. Hence $P(n)$ is true for every $n \in \mathbb{N}^*$. \square

Induction arguments are ubiquitous in mathematics. Every aspiring student of mathematics must learn to master them. They are found whenever a fact about the integers is true by virtue of the “domino effect,” that is, the truth of the statement for a particular integer somehow implies that the statement is also true for the successive integer.

We will now give several examples of propositions which can be proved using induction arguments.

Proposition 3.5. *For every positive integer n ,*

$$(3.2) \quad \sum_{k=1}^n k = 1 + 2 + \dots + n = \frac{n(n+1)}{2}.$$

Proof. We will argue by induction. For $n = 1$, the formula (3.2) is simply

$$1 = \frac{(1)(2)}{2}.$$

Assume that (3.2) holds for some positive integer n . We need to show that (3.2) also holds for $n + 1$ replacing n ; that is, we want to prove that

$$(3.3) \quad \sum_{k=1}^{n+1} k = 1 + \dots + n + (n + 1) = \frac{(n + 1)(n + 2)}{2}.$$

Using our inductive hypothesis, we calculate

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

This is what we wanted to show, and so the proof is complete. \square

Proposition 3.6. *For every positive integer n ,*

$$(3.4) \quad \sum_{k=1}^n k^2 = 1^2 + 2^2 + \dots + n^2 = \frac{n(n + 1)(2n + 1)}{6}.$$

Proof. To proceed by induction, first notice that for $n = 1$ equation (3.4) simply reads

$$1^2 = \frac{(1)(2)(3)}{6}.$$

Now assume that (3.4) holds for some particular positive integer n . We want to show that (3.4) is true when we replace n by $n + 1$; that is, we want to prove that

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

A calculation shows that

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

Therefore, (3.4) is true if we replace n with $n+1$, and so we conclude by induction that (3.4) is true for every positive integer n . \square

Proposition 3.7. *Suppose that X is a set consisting of $n \geq 0$ distinct elements. Then the power set $\mathcal{P}(X)$ of X consists of 2^n distinct elements.*

Proof. We will proceed by induction on n . The only set with zero elements, namely the empty set \emptyset , has only one subset: itself. Thus $\mathcal{P}(\emptyset) = \{\emptyset\}$ has one distinct element. Similarly, it is clear for all sets with $n = 1$ distinct elements, since the set $\{a\}$ has precisely two subsets, namely the empty set \emptyset and itself $\{a\}$. Therefore $\mathcal{P}(\{a\}) = \{\emptyset, \{a\}\}$ has precisely $2 = 2^1$ distinct elements. Now suppose that for every set with n distinct elements, its power set has 2^n distinct elements. Let X be an arbitrary set with $n+1$ distinct members. We claim that $\mathcal{P}(X)$ has 2^{n+1} distinct elements.

We may write $X = \{a_1, \dots, a_n, a_{n+1}\}$. Let Y be the set $Y = \{a_1, \dots, a_n\}$. Then Y has n distinct elements, so $\mathcal{P}(Y)$ has 2^n distinct elements. Now every subset of X either contains a_{n+1} or does not contain a_{n+1} . We will count these groups of subsets separately. There are clearly 2^n subsets of X which do not contain a_{n+1} , since these are precisely the subsets of Y and we assumed that $\mathcal{P}(Y)$ has 2^n distinct elements. There are also 2^n subsets of X which do contain a_{n+1} , since each of these can be written in a unique way as $A \cup \{a_{n+1}\}$ where A is a subset of Y . That is, for every subset of Y we get one subset of X that contains a_{n+1} simply by adding the element a_{n+1} ,

and for every subset of X that contains a_{n+1} we get a subset of Y simply by subtracting a_{n+1} . Therefore, in total there are

$$2^n + 2^n = 2 \cdot 2^n = 2^{n+1}$$

distinct subsets of X . This is what we wanted to prove, and the result now follows by induction. \square

Definition 3.8. Let m be an integer, and d be a positive integer. We say that d **divides** m if there exists an integer k such that $m = kd$. We also say that d is a **divisor** of m and that m is a **multiple** of d if d divides m . We use the notation

$$d \mid m$$

if d divides m .

Proposition 3.9. *For every positive integer n , the integer $n^3 + 2n$ is a multiple of 3.*

Proof. Clearly the statement is true for $n = 1$, since

$$(1)^3 + 2(1) = 3.$$

Now suppose that $3 \mid (k^3 + 2k)$ for some positive integer k . Then there exists an integer s such that $k^3 + 2k = 3s$. We must show that 3 divides $(k+1)^3 + 2(k+1)$ as well. To that end, we calculate

$$\begin{aligned} (k+1)^3 + 2(k+1) &= (k^3 + 3k^2 + 3k + 1) + (2k + 2) \\ &= (k^3 + 2k) + (3k^2 + 3k + 3) \\ &= 3s + 3(k^2 + k + 1) \\ &= 3(s + k^2 + k + 1). \end{aligned}$$

Thus 3 divides $(k+1)^3 + 2(k+1)$, and so an appeal to induction completes the proof. \square

We may begin our induction arguments at whichever integer we

Theorem 3.10 (Principle of Mathematical Induction, Version 3). *Suppose that m is an integer, and that \mathcal{I} is a subset of the set $\mathbb{Z}_{\geq m} = \{k \in \mathbb{Z} \mid k \geq m\}$. Suppose also that $m \in \mathcal{I}$, and that whenever an integer n belongs to \mathcal{I} , the integer $n + 1$ also belongs to \mathcal{I} . Then $\mathcal{I} = \mathbb{Z}_{\geq m}$.*

Proof. The argument is a straightforward adjustment of Axiom 3.1. We define a new set \mathcal{J} by

$$\mathcal{J} = \{k \in \mathbb{N}^* \mid k - m + 1 \in \mathcal{I}\}.$$

It is clear that $1 \in \mathcal{J}$ since $m \in \mathcal{I}$. Also, if $n \in \mathcal{J}$, then $n - m + 1 \in \mathcal{I}$. This implies that $(n + 1) - m + 1 = (n - m + 1) + 1 \in \mathcal{I}$, so $n + 1 \in \mathcal{J}$. Now we apply Axiom 3.1 to get that $\mathcal{J} = \mathbb{N}^*$. It is easy to see that this implies that $\mathcal{I} = \mathbb{Z}_{\geq m}$. \square

Example 3.11. Consider the set

$$A = \{n \in \mathbb{N} \mid 2^n > n^2\}.$$

We certainly expect A to contain *most* positive integers, since our intuition tells us that when n gets large, the number 2^n gets larger faster than n^2 . We can check that $0, 1 \in A$, and $2, 3, 4 \notin A$, and then $5, 6, 7, 8 \in A$. For integers $n \geq 8$, it seems that 2^n will always be much larger than n^2 . For instance, $2^8 = 256 > 64 = 8^2$. So we expect that $5, 6, 7, 8, \dots \in A$. Our proof will be similar to that of Proposition 3.3, but it has a slight wrinkle as we will see below.

Our goal is to show that if $k \in A$ and $k \geq 5$ then $k + 1 \in A$. Then we can apply Theorem 3.4, and our proof will be complete. Notice that we *must* use our assumption that $k \geq 5$ somewhere in our proof, since without this assumption the statement is not true! For instance, $1 \in A$ but $2 \notin A$. We begin by sketching out some ideas, which we will rewrite afterwards into a nice proof. First of all, our assumptions are that $2^k > k$ and that $k \geq 5$. Our goal is to show that $2^{k+1} > (k + 1)$. It's probably a good idea to first write

$$2^{k+1} = 2 \cdot 2^k,$$

because we need to say something about 2^{k+1} and we have information about 2^k , and the equation above is the link between the two numbers. The only thing we really know about 2^k is that it is greater than k^2 , so we should probably write

$$2^{k+1} = 2 \cdot 2^k > 2k^2.$$

Now, we are trying to show that 2^{k+1} is greater than $(k + 1)^2$. So if we could somehow show that $2k^2 \geq (k + 1)^2$ we would be done! So let's try to prove

this inequality. If we expand $(k+1)^2$ and rearrange some terms, we see that $2k^2 \geq (k+1)^2$ is equivalent to the inequality

$$k^2 \geq 2k + 1.$$

Well, is this true? If it is, we'd be done. It certainly is false if $k = 2$, for instance. But remember that we assumed that $k \geq 5$, and we haven't used this assumption yet. Actually, an easy induction exercise shows that $k^2 \geq 2k + 1$ for all $k \geq 3$. But we will show it without induction. The inequality is equivalent to

$$k^2 - 2k - 1 \geq 0$$

which is true if and only if

$$k^2 - 2k + 1 \geq 2$$

which is the same as

$$(k-1)^2 \geq 2$$

which is clearly true whenever $k \geq 3$. Now we have all the ideas we need for the proof.

Lemma 3.12. *For every integer $n \geq 3$,*

$$n^2 > 2n + 1$$

Proof. Let $k = n - 1$. Then $k \geq 2$, and $k^2 = k \cdot k \geq 2k \geq 4$. That is, $(n-1)^2 \geq 4$. Therefore $n^2 - 2n + 1 \geq 4$, and so

$$n^2 \geq 2n + 3 > 2n + 1.$$

□

Armed with this fact, we now prove the result we set out to prove.

Proposition 3.13. *The set*

$$\{n \in \mathbb{N} \mid 2^n > n^2\} = \{n \in \mathbb{N} \mid n \geq 5\} \cup \{0, 1\}.$$

Proof. Let A be the set of natural numbers n for which $2^n > n^2$. It is easy to check that $0, 1 \in A$ while $2, 3, 4 \notin A$. We need to show that $n \in A$ for every $n \geq 5$. Our proof will be based on induction in the form of Theorem 3.10. Clearly $5 \in A$ since $2^5 = 32 > 25 = 5^2$. Now assume that $n = k$ is an integer that satisfies $2^k > k^2$ and $k \geq 5$. Then using our assumption and Lemma 3.12 we see that

$$\begin{aligned} 2^{k+1} &= 2 \cdot 2^k \\ &> 2k^2 \\ &= k^2 + k^2 \\ &\geq k^2 + (2k + 1) \\ &= (k + 1)^2. \end{aligned}$$

An appeal to induction now completes the proof. \square

Notice that the proof above hides much of the thinking that went into its construction. This is an unfortunate fact of life: it is easy to read a proof and miss the really key insights that made it possible. When reading a proof, it is important to detect the key ingredients and ideas that make the proof *work*.

3.2 The Well-Ordering Principle

In this section we will discover a counterpart to mathematical induction, called the *Well-Ordering Principle*. It is a way of rephrasing the idea of mathematical induction in a way that is convenient for some problems.

Definition 3.14. Let A be a subset of \mathbb{Z} . Then we say that $l \in A$ is a **least element** of A if $l \leq n$ for every element $n \in A$. In other words, l is a least element of A if l belongs to A and every other element of A is greater than or equal to l .

In our definition above, we defined *a* least element of a set A , not *the* least element of a set A . Our reason for doing so is that, until we prove otherwise, we cannot assume it is impossible for a set to have two distinct least elements. We will now prove otherwise— that is, we will prove that if a set has a least element, that least element is *unique*. Our proof will resemble other proofs of “uniqueness” claims— we will assume there are two least elements and use them against each other.

Lemma 3.15. *Let $A \subset \mathbb{Z}$. Then A has at most one least element.*

Proof. Suppose that l and k are both least elements of A . We will show that, in fact, $l = k$. Since k is a least element of A , it must be that $k \in A$. Thus, since l is a least element of A , we deduce that

$$l \leq k.$$

Similarly, since $l \in A$ and k is a least element of A , we see that

$$k \leq l.$$

Therefore $l = k$. □

Owing to Lemma 3.15, if a set $A \subset \mathbb{Z}$ has a least element l , we may call l *the* least element of A .

Notation 3.16. Suppose A is a nonempty subset of \mathbb{N} . If A has a least element, then we denote the least element of A by

$$\min A.$$

We will also refer to $\min A$ as the **minimum** of A .

We are now ready to state the Well-Ordering Principle.

Theorem 3.17 (The Well-Ordering Principle). *Let A be a nonempty subset of the set of natural numbers. Then A has a least element.*

Remark 3.18. Notice that the word *nonempty* appears in the hypothesis of the Well-Ordering Principle. This is for the obvious reason that the empty set \emptyset , while being a subset of \mathbb{N} , does not have a least element.

The proof of the Well-Ordering Principle is based on Axiom 3.1. In fact, the Well-Ordering Principle is *equivalent* to the Principle of Mathematical Induction, in the sense that each can be proven from the other. In some books, the Principle of Mathematical Induction is proven using the Well-Ordering Principle, which is taken as an axiom. This means that, in principle, anything that can be proven with the Well-Ordering Principle can be proven instead with mathematical induction, and vice-versa. The choice of which one to use is largely a matter of taste. Some theorems have a simpler proof if we use one over the other.

Proof. INSERT PROOF HERE. \square

We will now reprove Proposition 3.3 using the Well-Ordering Principle.

Proposition 3.19. *For every natural number n ,*

$$2^n > n.$$

Proof. Let

$$B = \{n \in \mathbb{N} \mid 2^n \leq n\}.$$

Then B is the set of natural numbers for which our assertion is *false*. We want to show that $B = \emptyset$. We will assume that B is *not* the empty set, and then derive a contradiction.

Suppose $B \neq \emptyset$. Then by the well-ordering principle B has a least element. Let $l = \min B$. Since $2^0 = 1 > 0$ and $2^1 = 2 > 1$, we see that 0 and 1 are not elements of B . Therefore, since $l \in B$, it must be true that $l \geq 2$. We will show that $l - 1$ belongs to B , which will contradict the fact that $l = \min B$.

Since $l \in B$, we see that

$$\begin{aligned} 2^{l-1} &= \frac{1}{2} \cdot 2^l \\ &\leq \frac{1}{2} \cdot l \quad (\text{since } l \in B) \\ &\leq \frac{1}{2}(l + l - 2) \quad (\text{since } l \geq 2) \\ &= l - 1. \end{aligned}$$

Therefore $l - 1 \in B$, which is the desired contradiction. \square

Remark 3.20. The proof of Proposition 3.19 is an example of a typical way the Well-Ordering Principle is used in practice. This common type of argument can be summarized as follows:

1. Suppose that we want to show that the predicate $\mathcal{P}(n)$ is true for all positive integers $n \in \mathbb{N}$. Then consider the “bad set” B of natural numbers for which $\mathcal{P}(n)$ is false: $B = \{n \in \mathbb{N} \mid \text{not } \mathcal{P}(n)\}$. We want to show that B is empty.
2. Argue by contradiction: assume B is not empty.

3. Apply the well-ordering principle and select the least element l of B .
4. Using the fact that $l \in B$, argue somehow that there is a positive integer less than l that must also be in B . This contradicts the fact that l is the least element of B , which completes the proof.

Using a similar argument, we will now give a more surprising consequence of the well-ordering principle.

Proposition 3.21. *The real number $\sqrt{2}$ is irrational.*

Proof. We will proceed according to the steps in Remark 3.20. A real number $x \in \mathbb{R}$ is rational if it can be expressed as the ratio of two integers; that is, $x \in \mathbb{Q}$ if and only if there are integers $m, n \in \mathbb{Z}$ with $n \geq 1$ such that

$$x = \frac{m}{n}.$$

Multiplying by n , we see that this is equivalent to the requirement that nx is an integer for some integer $n \geq 1$. Therefore, we desire to show that the set

$$B = \left\{ n \in \mathbb{N}^* \mid n\sqrt{2} \in \mathbb{Z} \right\}$$

is the empty set. That is, there is no positive integer n for which $n\sqrt{2}$ is an integer.

Suppose on the contrary that B is *not* the empty set. Then by the well-ordering principle, there exists a least element n_0 of B . Let $k = n_0(\sqrt{2} - 1)$. We will show that k is a positive integer belonging to B that is less than n_0 , which give us our desired contradiction. First, notice that k must be a positive real number since $\sqrt{2} - 1 > 0$ and $n_0 \geq 1$. Also, since

$$k = n_0(\sqrt{2} - 1) = n_0\sqrt{2} - n_0,$$

k is the difference of two integers and must therefore be an integer itself. Thus k is a positive integer. To see that k belongs to B , we must check that $k\sqrt{2}$ is an integer. We calculate

$$k\sqrt{2} = n_0(\sqrt{2} - 1)\sqrt{2} = n_0 \left((\sqrt{2})^2 - \sqrt{2} \right) = 2n_0 - n_0\sqrt{2},$$

which is indeed an integer, since it is the difference of two integers. Therefore $k \in B$. We now need to show only that k is less than n_0 . To see this, note that $\sqrt{2} < 2$ and calculate

$$k = n_0(\sqrt{2} - 1) < n_0(2 - 1) = n_0.$$

Therefore, we have found an element of B that is less than the least element of B , which is a contradiction. \square

Definition 3.22. Let $A \subseteq \mathbb{Z}$. We say that $k \in \mathbb{Z}$ is an **upper bound** for A if every element of A is less than k ; that is, if $m \leq k$ for every $m \in A$. We say that A is **bounded above** if there exists an upper bound for A . We call an element $g \in \mathbb{Z}$ a **greatest element** of A if $g \in A$ and g is an upper bound for A .

Example 3.23. Consider the set

$$A = \{-41, -13, 4, 47, 81, 82\}.$$

Then ten million, 103, 86 and 82 are all upper bounds for A . However, the only greatest element of A is 82, since it is the only upper bound for A that also belongs to A .

Lemma 3.24. *Let $A \subseteq \mathbb{Z}$. Then A has at most one greatest element.*

Proof. We proof is similar to that of Lemma 3.15, so we leave it as an exercise. \square

Notation 3.25. Suppose $A \subseteq \mathbb{Z}$. If A has a greatest element, then by Lemma 3.24, that greatest element is unique. Therefore, we may refer to *the* greatest element of A , which we will denote by

$$\max A.$$

Not every subset of \mathbb{Z} has a greatest element! For instance, \mathbb{Z} itself has no greatest element, and neither does the empty set \emptyset . The next proposition characterizes those subsets of \mathbb{Z} that have greatest elements.

Proposition 3.26. *Suppose that $A \subseteq \mathbb{Z}$. Then A has a greatest element if and only if A is nonempty and bounded above.*

Proof. Clearly if A is empty, then A has no greatest element. Similarly, if A is not bounded above, then there is no upper bound for A , hence no greatest element of A .

Conversely, suppose that A is nonempty and bounded above. Let

$$U = \{k \in \mathbb{Z} \mid k \geq n \text{ for every } n \in A\}$$

be the set of upper bounds for A . Since A is bounded above, U is nonempty. Since every element of A is a lower bound for U , and A is nonempty, the set U is bounded below. [proof not complete] \square

3.3 Strong Induction

The Principle of Strong Induction is a generalization of the Principle of Mathematical Induction, which is sometimes called “weak induction.” The Principle of Strong Induction is “strong” because it allows us to make a stronger assumption in the induction step of our proofs, which will allow us to prove theorems that would be difficult to prove otherwise.

Theorem 3.27 (The Principle of Strong Induction). *Suppose that $\mathcal{I} \subset \mathbb{N}^*$ is a subset of the set of positive integers with the following two properties:*

1. $1 \in \mathcal{I}$; and
2. Whenever the set $\{1, 2, 3, \dots, n\}$ is a subset of \mathcal{I} , then $n + 1$ is an element of \mathcal{I} .

Then $\mathcal{I} = \mathbb{N}^*$.

Comparing the principle of strong induction above with the principle of weak induction (Axiom 3.1), we see that our hypotheses on the set \mathcal{I} are less restrictive than they were for weak induction. The conclusions are the same: $\mathcal{I} = \mathbb{N}^*$. Therefore, the principle of strong induction is more powerful than the principle of weak induction (although one could argue that this is not true on the grounds that the proof of the principle of strong induction is based on an application of weak induction, we will see that in practice it is justified).

We will postpone the proof of Theorem 3.27 and focus on learning how to apply it. We begin with an interesting example. Recall from Definition 1.1 that a positive integer $p \geq 2$ is called **prime** if the only divisors of p are 1 and p .

Theorem 3.28. *Every positive integer $n \geq 2$ is either a prime itself or a product of primes.*

Before we prove Theorem 3.28, let’s illustrate the difficulty we would encounter if we tried to prove it using weak induction. What does the fact that $22 = 2 \cdot 11$ have to do with the fact that $21 = 3 \cdot 7$? There simply is no link between the two facts that enables us to get a viable induction off the ground. In other words, it is not clear how knowing that n is a product of primes would help us to prove that $n + 1$ is a product of primes. This is

where *strong* induction helps out. We do not have to prove that $n + 1$ is a product of primes knowing only that n is prime— we may make the stronger assumption that all of the integers $2, 3, 4, \dots, n$ are products of primes and use this stronger assumption to prove that $n + 1$ is prime.

Proof of Theorem 3.28. We proceed using strong induction. Clearly 2 is prime. Suppose that for some fixed positive integer $k \geq 2$, each of the integers $2, \dots, k$ is either prime or a product of primes. We will show that $k + 1$ is as well.

If $k + 1$ is prime, then we have nothing left to prove, so suppose that $k + 1$ is not prime. Then there exist positive integers p and q such that $k + 1 = pq$ and p is neither equal to 1 nor equal to $k + 1$. This implies that $q \neq 1$ and $q \neq k + 1$. Therefore,

$$2 \leq p, q \leq k.$$

By our induction hypothesis, we know that both p and q are either prime or products of primes. Either way, $k + 1 = pq$ is a product of primes. The proof is now complete by an appeal to the principle of strong induction. \square

3.4 Exercises and Problems

Problem 3.1. Recall that the **Fibonacci sequence** is defined recursively by

$$\begin{cases} F_1 = F_2 = 1, \\ F_{n+2} = F_{n+1} + F_n. \end{cases}$$

(a) Show that for every integer $n \geq 5$,

$$\frac{8}{5} \leq \frac{F_{n+1}}{F_n} \leq \frac{13}{8}.$$

(b) Show that for every integer $n \geq 1$,

$$F_{n+1}^2 - F_n F_{n+2} = (-1)^n$$

(c) Show that for every $n \geq 1$,

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

Chapter 4

Elementary Number Theory

4.1 The Division Theorem

Definition 4.1. Let m and d be integers. Then we say that d **divides** m and that d is a **divisor** of m and that m is a **multiple** of d if $d \geq 1$ and there exists an integer $k \in \mathbb{Z}$ such that $m = kd$.

Notation 4.2. We write

$$d \mid m$$

if d is a divisor of m .

To get acquainted with the notion of divisibility, we will now prove some simple lemmas.

Lemma 4.3. *Suppose that m and d are integers such that d divides m . If $m \neq 0$, then $d \leq |m|$.*

Proof. Since $d \mid m$, there exists an integer k such that $m = kd$. Since $m \neq 0$ we deduce that $k \neq 0$. Therefore, $|k| \geq 1$. Thus

$$d = |d| = \frac{|m|}{|k|} \leq |m|.$$

□

Why did we have to assume that $m \neq 0$ in the above lemma? Because every positive integer divides 0, as the following lemma asserts.

Lemma 4.4. *Let d be a positive integer. Then $d \mid 0$.*

Proof. Notice that $0 = 0 \cdot d$. □

Lemma 4.5. *Let d be a positive integer. Then $d \mid d$.*

Proof. Notice that $d = 1 \cdot d$. □

Lemma 4.6. *Let d and m be positive integers, and let $n \in \mathbb{Z}$. If $d \mid m$ and $m \mid n$, then $d \mid n$.*

Proof. There exist integers k_1 and k_2 such that

$$m = k_1d \quad \text{and} \quad n = k_2m.$$

Then we have $n = k_2(k_1d) = (k_1k_2)d$. □

Lemma 4.7. *If d is a positive integer that divides the integers m and n , then*

$$d \mid (am + bn)$$

for any integers $a, b \in \mathbb{Z}$.

Proof. We may write $m = k_1d$ and $n = k_2d$ for integers $k_1, k_2 \in \mathbb{Z}$. Then

$$am + bn = a(k_1d) + b(k_2d) = (ak_1 + bk_2)d.$$

□

Definition 4.8. Let $x \in \mathbb{R}$ be a real number. Then the **floor** of x is the greatest integer less than or equal to x . We denote the floor of x by $\lfloor x \rfloor$. In a similar way, we define the **ceiling** $\lceil x \rceil$ of x as the smallest integer greater than or equal to x .

Lemma 4.9. *Let $x \in \mathbb{R}$. Then there is a unique integer $n \in \mathbb{Z}$ such that*

$$x - 1 < n \leq x.$$

Moreover, $n = \lfloor x \rfloor$.

Proof. Exercise. □

Theorem 4.10 (The Division Theorem). *Let n be a positive integer, and let $m \in \mathbb{Z}$. Then there exists unique integers $q, r \in \mathbb{Z}$ such that*

$$(4.1) \quad m = qn + r \text{ and } 0 \leq r < n.$$

Moreover,

$$q = \left\lfloor \frac{m}{n} \right\rfloor.$$

Proof. Define

$$q = \left\lfloor \frac{m}{n} \right\rfloor$$

and set $r = m - qn$. We will show that q and r satisfy (4.1). It is clear from the definition of r that $m = qn + r$. It remains to show that $0 \leq r < n$. By the definition of q and Lemma 4.9, we have the inequality

$$\frac{m}{n} - 1 < q \leq \frac{m}{n}.$$

Multiplying both sides by n (and remembering that $n \geq 1$) and rearranging, we obtain

$$0 \leq m - qn < n.$$

Thus $0 \leq r < n$.

To show that q and r are unique, suppose that \tilde{q} and \tilde{r} is another pair of integers satisfying

$$(4.2) \quad m = \tilde{q}n + \tilde{r} \text{ and } 0 \leq \tilde{r} < n.$$

We will show that $q = \tilde{q}$ and $r = \tilde{r}$. Using (4.1) and (4.2), we have that $qn + r = \tilde{q}n + \tilde{r}$, and by rearranging we get

$$(4.3) \quad n(q - \tilde{q}) = \tilde{r} - r.$$

Since $0 \leq r, \tilde{r} < n$, it follows that $-n < r - \tilde{r} < n$. Therefore, using (4.3), we see that

$$-1 < q - \tilde{q} < 1.$$

Therefore, $q - \tilde{q} = 0$. Using (4.3) again, we see that $r - \tilde{r} = 0$. Thus we have shown that $q = \tilde{q}$ and $r = \tilde{r}$, and so the proof is complete. \square

The unique remainder r that satisfies the conclusion of the Division Theorem will prove to be quite useful and we will need to refer to it again and again. This motivates the following definition.

Definition 4.11. Let $m \in \mathbb{Z}$ and $n \in \mathbb{N}^*$. We define the integer $m \bmod n$ to be the unique r satisfying the conclusion of the Division Theorem.

Remark 4.12. From the proof of the Division Theorem, we can easily see that

$$m \bmod n = m - n \left\lfloor \frac{m}{n} \right\rfloor.$$

Definition 4.13. We say that an integer k is **even** if k is a multiple of 2. We say that an integer k is **odd** if k is not even.

Lemma 4.14. *An integer m is even if and only if $m \bmod 2 = 0$.*

Proof. If m is even, then there is an integer q such that $m = 2q$. Thus if we set $r = 0$ then $m = 2q + r$ and $0 \leq r < 2$. Thus $m \bmod 2 = 0$.

Conversely, suppose that $m \bmod 2 = 0$. Then is an integer k such that

$$m = 2k + (m \bmod 2) = 2k,$$

so m is even. □

Corollary 4.15. *An integer m is odd if and only if there exists an integer k such that $m = 2k + 1$.*

Proof. The proof is left to the reader (Exercise 4.1). □

Lemma 4.16. *Let m be an integer. Then m is even if and only if m^2 is even.*

Proof. Suppose m is even. Then $m = 2k$ for some integer k . Thus

$$m^2 = (2k)^2 = 2(2k^2),$$

so m^2 is even.

Conversely, suppose that m is odd. Then by Corollary 4.15, there is an integer k such that $m = 2k + 1$. Thus

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

so by Corollary 4.15 again, m^2 is odd. □

Lemma 4.17. *Let b be an integer such that $b \geq 2$. Then for any positive integer n , there is a unique integer $m \in \mathbb{N}^*$ for which*

$$(4.4) \quad b^{m-1} \leq n < b^m.$$

Proof. Let $n \in \mathbb{N}^*$ be fixed. Consider the set

$$K = \{k \in \mathbb{N} \mid b^k > n\}.$$

We claim that $n \in K$. Indeed, for since $b \geq 2$, Proposition 3.3 implies that $b^n \geq 2^n > n$. Thus K is nonempty. Let m be the least element of K . We will show that m satisfies (4.4). Clearly $n < b^m$ since $m \in K$. Moreover, since $m - 1 \notin K$ (because m is the least element of K), we have that $b^{m-1} \leq n$. Uniqueness is left as an exercise. \square

Theorem 4.18 (Base Representation Theorem). *Let b be an integer such that $b \geq 2$. Then for every positive integer a , there exists unique integers n and $d_0, d_1, \dots, d_n \in \{0, 1, \dots, b - 1\}$ such that*

$$(4.5) \quad a = \sum_{j=0}^n d_j b^j \quad \text{and} \quad d_n > 0.$$

Proof. We will show existence, and leave uniqueness as an exercise. We proceed by strong induction on a . Clearly the theorem is true for $a = 1$, since we may take $n = 0$ and $d_0 = 1$. Now suppose that for some positive integer k , the result is true for every $a = 1, \dots, k - 1$. We will prove that the result holds for k as well.

According to Lemma 4.17 there is a unique integer $n \in \mathbb{N}$ for which $b^n \leq k < b^{n+1}$. According to the division theorem, there are integers q and r such that

$$k = qb^n + r \quad \text{and} \quad 0 \leq r < b^n.$$

We claim that $1 \leq q \leq b - 1$. If $q \leq 0$, then we would have $k \leq r < b^n$, a contradiction to our choice of n . On the other hand, if $q \geq b$, then we would have $k \geq b^{n+1} + r \geq b^{n+1}$, again contradicting our choice of n . Hence we have demonstrated our claim that $1 \leq q \leq b - 1$.

Suppose for the moment that $r = 0$. Then we would have nothing left to show. Indeed, for by choosing $d_n = q$ and $d_j = 0$ for every $0 \leq j \leq n - 1$, we would have that $0 \leq d_j \leq b - 1$ and $d_n > 0$, and that

$$k = qb^n = \sum_{j=0}^n d_j b^j.$$

On the other hand, if $r > 0$, then by our induction hypothesis we may write

$$r = \sum_{j=0}^m d_j b^j \quad \text{and} \quad d_m > 0$$

for integers $d_0, \dots, d_m \in \{0, \dots, b-1\}$. Since $r < b^n$, we have that $m < n$. Set $d_n = q$ and $d_j = 0$ for every j such that $m < j < n$. Therefore, we have that

$$k = qb^n + r = qb^n + \sum_{j=0}^m d_j b^j = \sum_{j=0}^n d_j b^j.$$

This completes the proof of existence. □

Notation 4.19. Here we note a shorthand for writing the base representation of an integer in base b . If a , b , n and $d_0 \dots, d_n$ are as in Theorem 4.18, then we denote the base b representation of a by writing

$$a = (d_n \cdots d_1 d_0)_b.$$

For example, since

$$1023 = 1 \times 10^3 + 0 \times 10^2 + 2 \times 10^1 + 3 \times 10^0,$$

we may write

$$1023 = (1023)_{10}.$$

Also, since

$$1023 = \sum_{j=0}^9 2^j = 4 \times 6^3 + 4 \times 6^2 + 2 \times 6 + 3,$$

we write $1023 = (1111111111)_2 = (4423)_6$.

4.2 The Euclidean Algorithm

Example 4.20. The set of divisors of 132 is $\{1, 2, 3, 4, 6, 11, 12, 22, 33, 44, 66, 132\}$, and the set of divisors of 198 is $\{1, 2, 3, 6, 9, 11, 18, 22, 33, 66, 99, 198\}$. The intersection of these sets is the set

$$D = \{1, 2, 3, 6, 11, 22, 33, 66\},$$

which consists of the of *common divisors* of 132 and 198— those positive integers that divide both 132 and 198. The *greatest common divisor* of 132 and 198 is apparently 66, since it is the greatest element of D . Notice that every other element of D is a divisor of 66, and that

$$66 = 198a + 132b$$

for $a = 1$ and $b = -1$ (we could also have chosen $a = -1$ and $b = 2$ or $a = 3$ and $b = -4$). In this section, we will generalize these simple observations.

Lemma 4.21. *Let m and n be integers which are not both zero. Then the set*

$$A = \{k \in \mathbb{N}^* \mid k \mid m \text{ and } k \mid n\}$$

is nonempty and bounded above.

Proof. Since 1 divides any integer, it is clear that $1 \in A$, and thus A is nonempty. Since either m or n is not zero, we may assume without loss of generality that $n \neq 0$. Let $k \in A$. Then $k \mid n$, and so $k \leq |n|$ by Lemma 4.3. Thus we have shown that A is bounded above by $|n|$. \square

Definition 4.22. Let m and n be integers, not both of which are zero, and let

$$A = \{k \in \mathbb{N}^* \mid k \mid m \text{ and } k \mid n\}.$$

We say that d is a **common divisor** of m and n if $d \in A$; that is, d is a common divisor of m and n if d divides both m and n . By the previous lemma, the set A of common divisors of m and n is nonempty and bounded above. Thus A has a greatest element by Proposition 3.26. We define the **greatest common divisor** of m and n , which we denote by $\gcd(m, n)$, to be the greatest element of A . That is, the greatest common divisor of m and n is the integer

$$\gcd(m, n) = \max \{k \in \mathbb{N}^* \mid k \mid m \text{ and } k \mid n\}.$$

Remark 4.23. It makes no sense to write $\gcd(0, 0)$, because every integer is a divisor of 0, and hence there is no greatest common divisor of 0 and 0. Therefore, let us agree that whenever we refer to the greatest common divisor of m and n or write expression $\gcd(m, n)$ that *we are implicitly making the assumption that m and n are not both zero*. Henceforth, we will often not make explicit mention of this assumption.

Definition 4.24. We say that integers m and n are **relatively prime** if $\gcd(m, n) = 1$. In other words, m and n are relatively prime if 1 is their only common divisor.

Lemma 4.25. *Let m and n be integers (not both zero). Then the integers*

$$\frac{m}{\gcd(m, n)} \quad \text{and} \quad \frac{n}{\gcd(m, n)}$$

are relatively prime.

Proof. Let $a = m/\gcd(m, n)$ and $b = n/\gcd(m, n)$. First, we should check that a and b are actually integers. But this is clear enough, since $\gcd(m, n)$ is a divisor of both m and n . We must show that a and b are relatively prime. To accomplish this, we will prove that 1 is the only common divisor of a and b .

Let d be a common divisor of a and b . That is, $d \in \mathbb{N}^*$ such that $d \mid a$ and $d \mid b$. Then there are integers r and s such that

$$a = rd \quad \text{and} \quad b = sd.$$

Using the definitions of a and b , we may rewrite these equalities as

$$m = rd \cdot \gcd(m, n) \quad \text{and} \quad n = sd \cdot \gcd(m, n).$$

Therefore, $d \cdot \gcd(m, n)$ is a common divisor m and n . Therefore,

$$d \cdot \gcd(m, n) \leq \gcd(m, n)$$

by the definition of the greatest common divisor of m and n . However, since $d \in \mathbb{N}^*$, we also have $d \cdot \gcd(m, n) \geq \gcd(m, n)$. Hence

$$d \cdot \gcd(m, n) \geq \gcd(m, n),$$

which implies that $d = 1$. □

Lemma 4.25 provides us with another proof of the irrationality of $\sqrt{2}$.

Theorem 4.26. *The real number $\sqrt{2}$ is irrational.*

Proof. Suppose that $\sqrt{2}$ is in fact rational. Then there exists $m \in \mathbb{Z}$ and $n \in \mathbb{N}^*$ such that

$$\sqrt{2} = \frac{m}{n}.$$

Notice that $n \neq 0$ (so not both m and n are zero). If we set $\tilde{m} = m/\gcd(m, n)$ and $\tilde{n} = n/\gcd(m, n)$, then by Lemma 4.25 \tilde{m} and \tilde{n} are relatively prime. Also, it is clear that

$$\frac{\tilde{m}}{\tilde{n}} = \frac{\frac{m}{\gcd(m, n)}}{\frac{n}{\gcd(m, n)}} = \frac{m}{n} = \sqrt{2}.$$

Squaring both sides and rearranging we get $\tilde{m}^2 = 2\tilde{n}^2$. Therefore, \tilde{m}^2 is even. By Lemma 4.16, \tilde{m} is even. So we may write $\tilde{m} = 2k$ for some integer k . Thus

$$2\tilde{n}^2 = \tilde{m}^2 = (2k)^2 = 4k^2,$$

and so $\tilde{n}^2 = 2k^2$. Therefore, \tilde{n}^2 is even and so \tilde{n} is even (by Lemma 4.16). Thus \tilde{m} and \tilde{n} are both even, a contradiction to the fact that they are relatively prime. \square

The primary concern of this section is to prove the following theorem.

Theorem 4.27 (The Euclidean Algorithm). *Let m and n be two integers, not both of which are zero. Then there exists integers a and b such that*

$$\gcd(m, n) = am + bn.$$

We will prove Theorem 4.27 in two different ways. The proof we will give is more elegant and theoretically pleasing, but the second proof will “constructive” (we will explain what we mean later). Our first proof of the Euclidean algorithm is based on the concept of an *ideal*, which we will now define.

Definition 4.28. An **ideal** in \mathbb{Z} is a nonempty subset $\mathcal{I} \subseteq \mathbb{Z}$ that has the following properties:

1. Whenever two integers m and n belong to \mathcal{I} , their sum $m + n$ belongs to \mathcal{I} .
2. Whenever m belongs to \mathcal{I} , the integer $-m$ also belongs to \mathcal{I} .

Example 4.29. The set $2\mathbb{Z}$ of even integers is an ideal in \mathbb{Z} . Indeed, for the set of even integers is nonempty, the sum of any two even integers is an even integer, and any multiple of an even integer is an even integer.

Example 4.30. Let $n \in \mathbb{N}$. We will show that the set $n\mathbb{Z}$ of multiples of n is an ideal in \mathbb{Z} . Clearly $n\mathbb{Z} \neq \emptyset$. By Lemma 4.7, the sum of two elements of $n\mathbb{Z}$ belongs to $n\mathbb{Z}$, and by Lemma 4.6, any multiple of an element of $n\mathbb{Z}$ is another element of $n\mathbb{Z}$. Thus $n\mathbb{Z}$ is an ideal in \mathbb{Z} .

It may seem as though Example 4.30 only contains relatively simple ideals. However, it is a surprising fact that every (nontrivial) ideal in \mathbb{Z} is one of the ideals in Example 4.30!

Lemma 4.31. *If $\mathcal{I} \subseteq \mathbb{Z}$ is an ideal and $a \in \mathcal{I}$, then every multiple of a also belongs to \mathcal{I} .*

Proof. The proof is left as an exercise. \square

Proposition 4.32. *Let \mathcal{I} be an ideal in \mathbb{Z} . Then either $\mathcal{I} = \{0\}$ or there exists a natural number a such that*

$$\mathcal{I} = \{ka \mid k \in \mathbb{Z}\} = a\mathbb{Z}.$$

Proof. Since \mathcal{I} is nonempty, we may choose $m \in \mathcal{I}$. Notice that

$$0 = 0 \cdot m \in \mathcal{I}.$$

Then, assuming that $\mathcal{I} \neq \{0\}$, we deduce that \mathcal{I} has some nonzero element n . Since $-n = (-1)n \in \mathcal{I}$, we may just as well assume that n is positive. Thus the set $\mathcal{I} \cap \mathbb{N}^*$ of positive elements of \mathcal{I} is nonempty. By the well ordering principle, $\mathcal{I} \cap \mathbb{N}^*$ has a least element a . We will show that

$$\mathcal{I} = \{ka \mid k \in \mathbb{Z}\}.$$

By Lemma ??, every multiple of a belongs to \mathcal{I} since $a \in \mathcal{I}$. Hence $\{ka \mid k \in \mathbb{Z}\} \subset \mathcal{I}$. It remains to show that

$$(4.6) \quad \mathcal{I} \subset \{ka \mid k \in \mathbb{Z}\}.$$

To prove equation (4.6), pick an arbitrary element $t \in \mathcal{I}$. By the division theorem, there are integers q and r such that

$$(4.7) \quad t = qa + r \quad \text{and} \quad 0 \leq r < a.$$

Since $(-q)a \in \mathcal{I}$ and $t \in \mathcal{I}$, we know that $r = t - qa \in \mathcal{I}$. Thus, since a is the least positive element of \mathcal{I} and $r < a$, we deduce that r is not positive. Since $0 \leq r$, the only remaining possibility is that $r = 0$. Going back to equation (4.7) and plugging in $r = 0$, we see that t is a multiple of a . Thus $t \in \{ka \mid k \in \mathbb{Z}\}$, and we have demonstrated (4.6), completing the proof. \square

Armed with Proposition 4.32, we will now give our first proof of Theorem 4.27.

First proof of The Euclidean Algorithm. Let m and n be integers, not both of which are zero. We must show that there are integers a and b such that

$$(4.8) \quad \gcd(m, n) = am + bn.$$

To that end, we introduce the set

$$J = \{sm + tn \mid s, t \in \mathbb{Z}\}.$$

We will first show that J is an ideal. Clearly J is nonempty since

$$0 = 0 \cdot m + 0 \cdot n \in J.$$

Let k_1 and k_2 be elements of J . Then we may write $k_1 = s_1m + t_1n$ and $k_2 = s_2m + t_2n$ for integers $s_1, s_2, t_1, t_2 \in \mathbb{Z}$. Thus

$$k_1 + k_2 = (s_1 + s_2)m + (t_1 + t_2)n \in J.$$

Likewise, if $r \in \mathbb{Z}$ then we have

$$rk_1 = (rs_1)m + (rt_1)n \in J.$$

Thus J is an ideal.

By Proposition 4.32, there exists a positive integer $g \in J$ such that

$$(4.9) \quad J = \{kg \mid k \in \mathbb{Z}\}.$$

Since $g = 1 \cdot g \in J$, there exist integers a and b for which

$$g = am + bn.$$

We will show that $g = \gcd(m, n)$. Since $m = 1 \cdot m + 0 \cdot n \in J$ and $n = 0 \cdot m + 1 \cdot n \in J$, we deduce via equation (4.9) that both m and n are multiples of g . That is, g is a common divisor of m and n . To see that g is the greatest common divisor of m and n , let d be another common divisor of m and n . Then there are integers p and q such that $m = pd$ and $n = qd$. Then

$$g = am + bn = a(pd) + b(qd) = d(ap + bq).$$

Thus g is a multiple of d . By Lemma 4.3, we see that $d \leq g$. Thus $g = \gcd(m, n)$. \square

Remark 4.33. In our first proof of Theorem 4.27 above we actually proved considerably more than the statement of Theorem 4.27. A close inspection of the proof will reveal that we have shown that

$$(4.10) \quad \gcd(m, n) = \min\{sm + tn \mid s, t \in \mathbb{Z} \text{ and } sm + tn > 0\}.$$

Also, we may deduce from the argument above that every common divisor of m and n divides their greatest common divisor. The latter fact is important, so we record it (and reprove it) in the following corollary.

Corollary 4.34. *Let m and n be integers, not both of which are zero. Then every common divisor of m and n is also a divisor of $\gcd(m, n)$.*

Proof. By the Theorem 4.27, there are integers a and b such that

$$\gcd(m, n) = am + bn.$$

If d divides m and n , then by Lemma 4.7, d also divides $am + bn = \gcd(m, n)$. \square

Corollary 4.35. *Let $m, n \in \mathbb{Z}$. Then m and n are relatively prime if and only if there exists $s, t \in \mathbb{Z}$ such that*

$$sm + tn = 1.$$

Proof. If m and n are relatively prime, then the desired conclusion is a direct consequence of Theorem 4.27. Conversely, suppose that there exists $s, t \in \mathbb{Z}$ for which

$$sm + tn = 1.$$

Let d be a common divisor of m and n . Then by Lemma 4.7, d also divides 1. Thus $d = 1$. \square

Corollary 4.36. *Suppose that $m, n \in \mathbb{Z}$ and $d \in \mathbb{N}^*$ such that d and m are relatively prime, and*

$$d \mid mn.$$

Then d divides n .

Proof. Since d divides mn there is an integer q such that

$$(4.11) \quad mn = qd.$$

Since d and m are relatively prime, by Corollary 4.35, there exist integers $s, t \in \mathbb{Z}$ such that

$$1 = sd + tm.$$

Multiplying both sides by n and using equation (4.11), we get

$$n = n(sd + tm) = sdn + tmn = sdn + tqd = d(sn + tq).$$

Thus d divides n . \square

Corollary 4.37. *Suppose that a and b are relatively prime integers such that both a and b divide n . Then the product ab also divides n .*

Proof. Since a divides n , there exists an integer k for which

$$(4.12) \quad n = ak.$$

Since b divides $n = ak$, by Corollary 4.36, b divides k . Thus there is an integer q for which $k = bq$. Substituting this into (4.12), we get

$$n = a(bq) = (ab)q.$$

Thus ab divides n . □

Remark 4.38. The hypothesis that a and b are relatively prime in Corollary 4.37 is necessary. For example, 6 and 3 divide 12, but $18 = 6 \cdot 3$ does not divide 12.

Our second proof of the Euclidean algorithm will be based on the following simple lemmas.

Lemma 4.39. *Suppose that m is a positive integer. Then*

$$\gcd(m, 0) = m.$$

Proof. Clearly $m \leq \gcd(m, 0)$ since m is a common divisor of m and 0. If d is any common divisor of m and 0, then clearly $d \leq m$ by Lemma 4.3. Thus $\gcd(m, 0) \leq m$. □

Lemma 4.40. *Suppose that m is a positive integer and n is an integer such that $0 \leq n < m$. Then*

$$\gcd(m, n) = \gcd(n, m \bmod n).$$

Proof. Let $r = m \bmod n$. Then $0 \leq r < n$ and there exists an integer q such that

$$m = qn + r.$$

We will show that the common divisors of m and n are precisely the common divisors of n and r . To that end, let d be a common divisor of m and n . We may select k and l such that $m = kd$ and $n = ld$. Then

$$r = m - qn = (kd) - q(ld) = (k - ql)d,$$

and so d is a divisor of r . Thus d is a common divisor of n and r .

Now suppose that \tilde{d} is a common divisor of n and r . We may select \tilde{k} and \tilde{l} such that $n = \tilde{k}d$ and $r = \tilde{l}d$. Then

$$m = qn + r = q(\tilde{k}d) + \tilde{l}d = (q\tilde{k} + \tilde{l})d,$$

and so \tilde{d} divides m . Therefore, \tilde{d} is a common divisor of m and n .

We have shown that the set of common divisors of m and n is equal to the set of common divisors of n and r . Therefore, $\gcd(m, n) = \gcd(n, r)$. \square

Before using Lemma 4.39 and Lemma 4.40 to give another proof of Theorem 4.27, we will work out an example that will highlight the main ideas of the proof.

Example 4.41. Let $m = 25650$ and $n = 11172$. We can use Lemma 4.40 repeatedly to calculate $\gcd(m, n)$ as follows: first, notice that

$$(4.13) \quad 25650 = 2 \cdot 11172 + 3306$$

and use Lemma 4.40 to get that

$$\gcd(m, n) = \gcd(25650, 11172) = \gcd(11172, 3306).$$

Dividing 11172 by 3306, we get

$$11172 = 3 \cdot 3306 + 1254,$$

and applying Lemma 4.40 again, we see that

$$\gcd(11172, 3306) = \gcd(3306, 1254).$$

Continuing in this way, we see that

$$\begin{aligned} \gcd(3306, 1254) &= \gcd(1254, 798) \quad (\text{since } 3306 = 2 \cdot 1254 + 798) \\ &= \gcd(798, 456) \quad (\text{since } 1254 = 1 \cdot 798 + 456) \\ &= \gcd(456, 342) \quad (\text{since } 798 = 1 \cdot 456 + 342) \\ &= \gcd(342, 114) \quad (\text{since } 456 = 1 \cdot 342 + 114) \\ &= \gcd(114, 0) \quad (\text{since } 342 = 3 \cdot 114 + 0) \\ &= 114 \quad (\text{by Lemma 4.39}). \end{aligned}$$

So Lemmas 4.39 and 4.40 have allowed us to calculate $\gcd(25650, 11172) = 114$ in an extremely efficient way. Moreover, if we follow our work backwards, we can find integers a and b such that

$$\gcd(25650, 11172) = 114 = a \cdot 25650 + b \cdot 11172$$

as guaranteed by the Euclidean Algorithm. Using our quotients and remainders we calculated at each step above, we have that

$$\begin{aligned} 114 &= 456 - 1 \cdot 342 \\ &= 456 - 1 \cdot (798 - 1 \cdot 456) = -1 \cdot 798 + 2 \cdot 456 \\ &= -1 \cdot 798 + 2 \cdot (1254 - 1 \cdot 798) = 2 \cdot 1254 - 3 \cdot 798 \\ &= 2 \cdot 1254 - 3 \cdot (3306 - 2 \cdot 1254) = -3 \cdot 3306 + 8 \cdot 1254 \\ &= -3 \cdot 3306 + 8 \cdot (11172 - 3 \cdot 3306) = 8 \cdot 11172 - 27 \cdot 3306 \\ &= 8 \cdot 11172 - 27 \cdot (25650 - 2 \cdot 11172) = -27 \cdot 25650 + 62 \cdot 11172. \end{aligned}$$

Thus $a = -27$ and $b = 62$ satisfy $\gcd(m, n) = am + bn$, which reads

$$114 = -27 \cdot 25650 + 62 \cdot 11172.$$

Thus we have found a way to efficiently calculate $\gcd(m, n)$ as well as to find integers a and b satisfying the conclusion of the Euclidean Algorithm.

Using the ideas contained in Example 4.41, we now give our second proof of the Euclidean algorithm. This proof is less elegant than our original proof, but is very important because it gives a method for calculating the integers a and b whose existence are asserted.

Second proof of The Euclidean Algorithm. First, we will show that it suffices to prove Theorem 4.27 under the additional assumption that $0 \leq n < m$. Indeed, for if j is any integer, then j and $-j$ have the same divisors. Therefore, we may assume without loss of generality that m and n in the hypothesis of the Theorem 4.27 are both nonnegative integers. Moreover, if $m = n$, then we have that

$$\gcd(m, n) = \gcd(m, m) = m = 1 \cdot m + 0 \cdot n$$

and there is nothing left to show. Thus we may suppose with loss of generality that $0 \leq n < m$.

The proof will proceed by strong induction on m . If $m = 1$, then $n = 0$, and we have

$$\gcd(m, n) = \gcd(1, 0) = 1 = 1 \cdot 1 + 0 \cdot 0,$$

and so the conclusion of Theorem 4.27 holds for $a = 1$ and $b = 0$. For our induction hypothesis, assume that for some positive integer m , it is true that for every positive integer $k < m$ and every integer n such that $0 \leq n < k$ there exists $a, b \in \mathbb{Z}$ such that $\gcd(k, n) = ak + bn$. We must show that for every $0 \leq \tilde{n} < m$ there exists $\tilde{a}, \tilde{b} \in \mathbb{Z}$ such that $\gcd(m, \tilde{n}) = \tilde{a}m + \tilde{b}\tilde{n}$.

To prove this, suppose that \tilde{n} is some integer that satisfies $0 \leq \tilde{n} < m$. Let $r = m \bmod \tilde{n}$, and let $q \in \mathbb{Z}$ such that

$$(4.14) \quad m = q\tilde{n} + r.$$

By Lemma 4.40, we have that

$$(4.15) \quad \gcd(m, \tilde{n}) = \gcd(\tilde{n}, r).$$

Since $0 \leq r < \tilde{n} < m$, by our induction hypothesis, we may select $a, b \in \mathbb{Z}$ such that

$$\gcd(\tilde{n}, r) = a\tilde{n} + br.$$

Solve (4.14) for r and substitute into the above equation to get

$$\gcd(\tilde{n}, r) = a\tilde{n} + b(m - q\tilde{n}) = bm + (a - bq)\tilde{n}.$$

Set $\tilde{a} = b$ and $\tilde{b} = a - bq$, and use (4.15) to get

$$\gcd(m, \tilde{n}) = \gcd(\tilde{n}, r) = bm + (a - bq)\tilde{n} = \tilde{a}m + \tilde{b}\tilde{n}.$$

This is what we wanted to show, and so the proof is complete. \square

Remark 4.42. While our first proof of the Euclidean Algorithm may seem cleaner and more theoretically pleasing, our second proof outlines a method for calculating a pair a and b that satisfy $\gcd(m, n) = am + bn$. Mathematicians would therefore say that our second proof is a “constructive” one. It is one thing to simply assert that a pair $a, b \in \mathbb{Z}$ satisfying $\gcd(m, n) = am + bn$ exists (like our first proof does), but it is quite another to give a method for finding such an a and b , a difference that is important for many applications.

4.3 Prime Numbers and Euclid's Theorem

Definition 4.43. An integer p is called **prime** if $p \geq 2$ and if $d|p$ implies that $d = 1$ or $d = p$. An integer n is called **composite** if $n \geq 2$ and n is not prime.

The following theorem is one of the oldest and most famous theorems in mathematics.

Theorem 4.44 (Euclid's Theorem). *There are infinitely many primes.*

We will give two proofs of Euclid's theorem. Our first proof is Euclid's original proof, given in [insert book, year]. First, we will point remind ourselves of what we have previously shown, namely that every integer greater than 1 has a prime factor.

Lemma 4.45. *Every positive integer greater than 1 has a prime divisor.*

Proof. Let $n > 1$ be an integer. Then by Theorem 3.28, either n is a prime itself or a product of two or more primes. Either way, there is a prime that divides n . \square

Remark 4.46.

First proof of Euclid's Theorem. The proof is by way of contradiction. Suppose that p_1, p_2, \dots, p_n are the only primes. We will show that there is a prime that is not one of the p_i 's. Set

$$M = 1 + \prod_{i=1}^n p_i = 1 + p_1 p_2 \cdots p_n.$$

By Lemma 4.45, there is a prime p that divides M . Suppose that $p = p_j$ for some $j \in \{1, \dots, n\}$. Then p also divides the product $p_1 p_2 \cdots p_n$. By Lemma 4.7, we have that

$$p | (M - p_1 p_2 \cdots p_n) = 1.$$

Since 1 is the only divisor of itself, this implies that $p = 1$, which is a contradiction because 1 is not prime. \square

Second proof of Euclid's Theorem. We break the proof into several steps.

1. Define a sequence a_n recursively by

$$(4.16) \quad \begin{cases} a_1 = 2, \\ a_{n+1} = a_n^2 - a_n + 1 \quad (n \geq 1). \end{cases}$$

We will first show that for all $m \geq 2$ and all $1 \leq n < m$,

$$(4.17) \quad a_m = 1 + (a_n - 1) \prod_{j=n}^{m-1} a_j = 1 + a_{m-1} a_{m-2} \cdots a_n (a_n - 1).$$

We will prove of formula (4.17) by induction on m . For $m \geq 2$, the only n satisfying $1 \leq n < m$ is $n = 1$. By our definition of a_n , we see that

$$a_2 = a_1^2 - a_1 + 1 = 1 + a_1(a_1 - 1),$$

and so (4.17) holds for $m = 2$. Now suppose that (4.17) holds for some particular $m = k$ and all $1 \leq n < k$. We will show that (4.17) holds for $m = k + 1$ and all $0 \leq n < k + 1$. There are two cases: $n = k$ and $1 \leq n < k$. For $n = k$, we simply apply (4.16) to immediately deduce (4.17). If $1 \leq n < k$, then we calculate

$$\begin{aligned} a_{k+1} &= 1 + a_k(a_k - 1) \quad (\text{by (4.16)}) \\ &= 1 + a_k \left(\left(1 + (a_n - 1) \prod_{j=n}^{k-1} a_j \right) - 1 \right) \quad (\text{by induction hypothesis}) \\ &= 1 + a_k(a_n - 1) \prod_{j=n}^{k-1} a_j = 1 + (a_n - 1) \prod_{j=n}^k a_j. \end{aligned}$$

Thus, by induction, formula (4.17) holds for all $m, n \in \mathbb{N}^*$ for which $1 \leq n < m$.

2. We will now show that

$$(4.18) \quad \gcd(a_m, a_n) = 1 \quad \text{whenever } m \neq n.$$

If $m \neq n$ then we may assume without loss of generality that $1 \leq n < m$. Rewriting (4.17), we get

$$a_m + b \cdot a_n = 1$$

where we have set $b = (a_n - 1) \prod_{j=n+1}^{m-1} a_j$. By Corollary 4.35, we deduce (4.18).

3. A simple argument will show that

$$(4.19) \quad a_m \geq 2 \quad \text{for all } m \geq 1.$$

For $m = 1$, we have $a_1 = 2$. If (4.19) holds for some $m = k$, then

$$a_{k+1} = a_k(a_k - 1) + 1 \geq 2 \cdot 1 + 1 \geq 2.$$

By induction, we deduce (4.19).

4. By Lemma 4.45 there is a prime p_m that is a divisor of a_m for each positive integer m . Since $\gcd(a_m, a_n) = 1$, we see that

$$p_m \neq p_n \quad \text{whenever } m \neq n$$

(otherwise some p_m would be a common divisor of a_m and a_n for some $m \neq n$, which would imply that $p_m = 1$). Therefore, the sequence

$$p_1, p_2, p_3, \dots$$

is an infinite sequence of distinct primes. □

4.4 The Fundamental Theorem of Arithmetic

Proposition 4.47. *Let $m, n \in \mathbb{Z}$ and p be a prime. If p divides the product mn then p divides m or p divides n .*

Proof. Suppose that p does not divide m . We will show that p must divide n . If we can show that

$$(4.20) \quad \gcd(m, p) = 1,$$

then we may apply Corollary 4.36 to conclude that since p divides mn , p divides n . To show (4.20), let d be a common divisor of m and p . Then, since $d \mid p$, $d = 1$ or $d = p$. However, $d \neq p$ since $d \mid m$ but p does not divide m . Thus the only possibility is that $d = 1$. Therefore, $\gcd(m, p) = 1$. Now we may apply Corollary 4.36 to conclude that p divides n . □

Lemma 4.48. *Let p and p_1, p_2, \dots, p_k be primes. If p divides the product $p_1 p_2 \cdots p_k$, then $p = p_j$ for some index $j \in \{1, \dots, k\}$.*

Proof. The proof is left to the reader (see Problem 4.7). □

The following theorem states that any positive integer greater than 1 can be written as a product of primes in a unique way (up to the order the primes appear in the product).

Theorem 4.49 (The Fundamental Theorem of Arithmetic). *Suppose that n is an integer greater than 1. Then there is a unique positive integer m and unique primes p_1, p_2, \dots, p_m such that*

$$n = \prod_{j=1}^m p_j \quad \text{and} \quad p_1 \leq p_2 \leq \cdots \leq p_m.$$

Proof. 1. We have already demonstrated the existence of m and p_1, p_2, \dots, p_m in Theorem 3.28. Our proof of uniqueness will proceed by strong induction on n .

2. We will first show uniqueness for the base case $n = 2$. Suppose that \tilde{m} is a positive integer and $\tilde{p}_1, \dots, \tilde{p}_{\tilde{m}}$ are primes such that

$$2 = \prod_{j=1}^{\tilde{m}} \tilde{p}_j \quad \text{and} \quad \tilde{p}_1 \leq \tilde{p}_2 \leq \cdots \leq \tilde{p}_{\tilde{m}}.$$

If $\tilde{m} > 1$, then

$$\begin{aligned} 2 &= \prod_{j=1}^{\tilde{m}} \tilde{p}_j \\ &\geq \tilde{p}_1 \cdot \tilde{p}_2 \\ &\geq 4, \end{aligned}$$

a contradiction. Thus we must have $\tilde{m} = 1$. Therefore $2 = \prod_{j=1}^1 \tilde{p}_j = \tilde{p}_1$, and so there is only one way to write 2 as a product of primes.

3. Now assume that k is a positive integer greater than 2 such that the conclusion of the theorem is true for all n such that $2 \leq n < k$. Suppose that k can be written as a product of primes in two different ways. That is,

suppose that there are positive integers m and \tilde{m} and primes p_1, \dots, p_m and $\tilde{p}_1, \dots, \tilde{p}_{\tilde{m}}$ such that

$$(4.21) \quad k = \prod_{j=1}^m p_j \quad \text{and} \quad p_1 \leq \dots \leq p_m$$

and

$$(4.22) \quad k = \prod_{j=1}^{\tilde{m}} \tilde{p}_j \quad \text{and} \quad \tilde{p}_1 \leq \dots \leq \tilde{p}_{\tilde{m}}.$$

We must show that $m = \tilde{m}$ and $p_j = \tilde{p}_j$ for all indices j .

4. We will first show that $p_1 = \tilde{p}_1$. Since p_1 divides $k = \tilde{p}_1 \tilde{p}_2 \cdots \tilde{p}_{\tilde{m}}$, by Lemma 4.48 there is an index i such that $p_1 = \tilde{p}_i$. Therefore,

$$\tilde{p}_1 \leq \tilde{p}_i = p_1.$$

By reversing the roles of p_1 and \tilde{p}_1 and repeating the argument we just made, we can show as well that $p_1 \leq \tilde{p}_1$. Therefore, we have shown that $p_1 = \tilde{p}_1$.

5. Let $s = \frac{k}{p_1}$. Since $p_1 \mid k$, we see that s is a positive integer. Since p_1 is a prime, it is greater than 1 and thus $s < k$. Dividing Equations (4.21) and (4.22) by $p_1 = \tilde{p}_1$, we get that

$$s = \prod_{j=2}^m p_j \quad \text{and} \quad p_2 \leq \dots \leq p_m$$

and

$$s = \prod_{j=2}^{\tilde{m}} \tilde{p}_j \quad \text{and} \quad \tilde{p}_2 \leq \dots \leq \tilde{p}_{\tilde{m}}.$$

Since $1 \leq s < k$, our induction hypothesis implies that $m = \tilde{m}$ and $p_j = \tilde{p}_j$ for all indices $j \in \{2, \dots, m\}$. Since we have already shown that $p_1 = \tilde{p}_1$, the proof is complete. \square

Remark 4.50.

4.5 Exercises and Problems

Exercise 4.1. Prove Corollary 4.15.

Problem 4.2.

- (a) Show that $10^n \bmod 3 = 1$ for every $n \in \mathbb{N}$.
- (b) Show that an integer is a multiple of 3 if and only if the sum of the digits in its base 10 representation is a multiple of 3; in other words, show that if

$$a = \sum_{j=0}^n d_j 10^j$$

then a is a multiple of 3 if and only if

$$b = \sum_{j=0}^n d_j$$

is a multiple of 3.

Problem 4.3. Let a be a positive integer and let

$$a = \sum_{j=0}^n d_j 10^j$$

be the base 10 representation of a . Show that a is a multiple of 11 if and only if

$$\sum_{j=0}^n (-1)^j d_j$$

is a multiple of 11.

Problem 4.4. Let b be a positive integer such that $b \geq 2$.

- (a) Suppose that $c_0, c_1, \dots, c_k \in \{0, 1, \dots, b-1\}$. Show that

$$\sum_{j=0}^k c_j b^j < b^{k+1}.$$

- (b) Suppose that $n, \tilde{n} \in \mathbb{N}$ and $d_0, d_1, \dots, d_n, \tilde{d}_0, \tilde{d}_1, \dots, \tilde{d}_{\tilde{n}} \in \{0, 1, \dots, b-1\}$ such that

$$\sum_{j=0}^n d_j b^j = \sum_{j=0}^{\tilde{n}} \tilde{d}_j b^j,$$

as well as $d_n > 0$ and $d_{\tilde{n}} > 0$. Show that $n = \tilde{n}$ and $d_j = \tilde{d}_j$ for all $j \in \{0, \dots, n\}$ (Hint: First use part (a) to show that $n = \tilde{n}$. Then let k be the largest integer for which $d_k \neq \tilde{d}_k$, and use (a) again to get a contradiction.)

Problem 4.5. Let $m, n \in \mathbb{N}^*$.

- (a) Show that $\gcd(m, n) = \gcd(m + n, n)$.
- (b) Under what conditions is $\gcd(m, n) = \gcd(mn, n)$?
- (c) If k is a positive integer, show that $\gcd(km, kn) = k \cdot \gcd(m, n)$.

Problem 4.6. Let $m, n, k \in \mathbb{N}^*$.

- (a) Suppose that m and n are relatively prime, and k and n are relatively prime. Show that km and n are relatively prime.
- (b) Show that m and n are relatively prime if and only if m^2 and n^2 are relatively prime.
- (c) Show by induction that m and n are relatively prime if and only if m^q and n^q are relatively prime for every integer $q \geq 1$.

Problem 4.7. Prove Lemma 4.48.

Problem 4.8. Mimic the proof of Example 2.20 in the book to prove that if $n \in \mathbb{N}$ then $\sqrt{n} \in \mathbb{Q}$ if and only if n is a perfect square (Hint: Use some facts you proved in Problem 4.6).

Problem 4.9. Show that if $s, t \in \mathbb{Z}$ then $s\sqrt{2} + t\sqrt{3}$ is irrational unless $s = t = 0$.

Problem 4.10. Let $m, n \in \mathbb{N}$ and $s, t \in \mathbb{Z}$. When is the real number

$$s\sqrt{m} + t\sqrt{n}$$

rational and when is it irrational? Invent a proposition that answers this question *completely*, and then prove it.

Exercise 4.11. Let m and n be positive integers.

- (a) Come up with a good definition of a *common multiple* of m and n . Also define the *least common multiple* of m and n , and prove a lemma that shows that your definition makes sense. Denote the least common multiple of m and n by $\text{lcm}(m, n)$.
- (b) Show that $\text{lcm}(m, n)$ is the unique common multiple of m and n that is a divisor of every common multiple of m and n . (Hint: Think in an “ideal” way.)
- (c) Show that $mn = \text{gcd}(m, n) \cdot \text{lcm}(m, n)$. (Hint: First show that $mn/\text{gcd}(m, n)$ is an integer and a common multiple of m and n .)

Chapter 5

Modular Arithmetic

In this chapter, we will discuss a new kind of arithmetic, which will enable us to solve many interesting number theory problems.

5.1 Congruence Modulo n

Definition 5.1. Let $a, b \in \mathbb{Z}$ and $n \in \mathbb{N}^*$. Then we say that a is **congruent to b modulo n** if

$$a \bmod n = b \bmod n.$$

The positive integer n is called the **modulus** of such a congruence. If a is congruent to b modulo n , then we write

$$a \equiv b \pmod{n}.$$

Example 5.2. Since $57 \bmod 7 = 1 = 36 \bmod 7$, we see that

$$57 \equiv 36 \pmod{7}.$$

Similarly,

$$\begin{aligned} 29 &\equiv 16 \pmod{13} \\ 37 &\equiv 13 \pmod{24} \\ 118 &\equiv -12 \pmod{10} \\ 9112 &\equiv 12 \pmod{100} \\ -34 &\equiv 2 \pmod{18} \\ 104 &\equiv 0 \pmod{4}. \end{aligned}$$

Warning 5.3. We should not be too reckless with our notation. For instance, notice very carefully that while

$$57 \equiv 36 \pmod{7},$$

if we carelessly remove the parentheses around the modulus and turn our congruence sign into an equal sign the statement is *false*:

$$57 \neq 36 \pmod{7}.$$

This is for the simple reason that $36 \pmod{7} = 1$, and obviously $57 \neq 1$. The moral of the story is that

$$a \equiv b \pmod{n} \text{ does \textbf{not} mean the same thing as } a = b \pmod{n}.$$

However, as we will discover below (see Lemma 5.7), it *is* true that

$$a \equiv (a \pmod{n}) \pmod{n}.$$

We think of congruence as being the “arithmetic mod n ” equivalent of equality. It is for this reason that the congruence sign \equiv looks similar to an equal sign. The next lemma will show that in fact congruence modulo n has some of the same important properties as equality in usual arithmetic.

Lemma 5.4. *For any positive integer n , congruence modulo n is a reflexive, symmetric, and transitive relation on the integers. In other words, for every $a, b, c \in \mathbb{Z}$,*

$$(i) \ a \equiv a \pmod{n};$$

$$(ii) \ \text{if } a \equiv b \pmod{n} \text{ then } b \equiv a \pmod{n}; \text{ and}$$

$$(iii) \ \text{if } a \equiv b \pmod{n} \text{ and } b \equiv c \pmod{n} \text{ then } a \equiv c \pmod{n}.$$

Proof. The proof is very easy, and so is left to the reader. For example, to prove (iii), notice that if $a \pmod{n} = b \pmod{n}$ and $b \pmod{n} = c \pmod{n}$, then clearly $a \pmod{n} = c \pmod{n}$. \square

In each of the congruences in Example 5.2, the modulus divided the difference of the two integers in the congruence. This is true in general, and in fact provides a useful characterization of congruence modulo n , a fact which we will use again and again. This is the content of the following lemma.

Lemma 5.5. *Let $a, b \in \mathbb{Z}$ and $n \in \mathbb{N}^*$. Then $a \equiv b \pmod{n}$ if and only if n divides $a - b$.*

Proof. There exist integers $q_1, q_2 \in \mathbb{Z}$ such that

$$\begin{aligned} a &= q_1n + (a \bmod n), \quad \text{and} \\ b &= q_2n + (b \bmod n). \end{aligned}$$

Subtracting the second line from the first, we get

$$(5.1) \quad a - b = (q_1 - q_2)n + (a \bmod n - b \bmod n).$$

Now suppose that n divides $a - b$. Then there exists an integer k such that

$$a - b = nk.$$

We must show that $a \bmod n = b \bmod n$. Without loss of generality, suppose that $a \bmod n \geq b \bmod n$. Set $r = a \bmod n - b \bmod n$. It is clear that $0 \leq r < n$. We need to show that $r = 0$. Equation (5.1) reads

$$a - b = (q_1 - q_2)n + r.$$

Since we have also written $a - b = kn + 0$ above, employing the uniqueness part of the Division Theorem, we get that $k = (q_1 - q_2)$ and $r = 0$. Since $r = 0$, we have proven that $a \equiv b \pmod{n}$.

Conversely, suppose that $a \equiv b \pmod{n}$. This means that $a \bmod n = b \bmod n$. Then according to (5.1), n divides $a - b$. \square

Remark 5.6. Congruence modulo 1 is a dull concept. This is due to the fact that every integer is congruent to every other integer modulo 1 (by Lemma 5.5 and the fact that 1 is a divisor of every integer). For this reason, we will always work with moduli greater than 1.

The following Lemma is a form of the division algorithm stated in a way that we will find useful in this chapter.

Lemma 5.7. *Let $m \in \mathbb{Z}$ and $n \in \mathbb{N}^*$. Then there is a unique integer r that satisfies*

$$m \equiv r \pmod{n} \quad \text{and} \quad 0 \leq r < n.$$

Moreover, $r = m \bmod n$.

Exercise 5.1. Prove Lemma 5.7.

Example 5.8. To illustrate the significance of Lemma 5.7, consider the integer $m = 17^6 + 14^{123}$ and a modulus $n = 4$. In this case, what Lemma 5.7 says is that m , while being a very large integer, is congruent to one and only one of the integers 0, 1, 2, or 3, modulo 4. Roughly speaking, *there are only 4 varieties of integers modulo 4*. Moreover, to figure out which of the integers 0, 1, 2, 3 that m is congruent to modulo 4, we may simply divide m by 4 and take the remainder. This will prove to be a very useful observation when used in combination with Theorem 5.9 below. For example, it will allow us to reduce some questions about infinitely many integers to a finite number of cases (see Example 5.11).

What makes congruence modulo n an interesting concept is the following theorem. Without it, this chapter would not exist. It allows for the development of a separate *arithmetic* in modulo n , which is a powerful tool for solving problems and proving theorems about our usual arithmetic.

Theorem 5.9 (The Fundamental Theorem of Modular Arithmetic).

Suppose that $n \in \mathbb{N}^$ and $a, b, c, d \in \mathbb{Z}$ such that*

$$a \equiv b \pmod{n} \quad \text{and} \quad c \equiv d \pmod{n}.$$

Then

$$a + c \equiv b + d \pmod{n} \quad \text{and} \quad ac \equiv bd \pmod{n}.$$

Proof. By Lemma 5.5, n is a common divisor of $(a - b)$ and $(c - d)$. Thus n divides the integer

$$(a - b) + (c - d) = (a + c) - (b + d).$$

Using Lemma 5.5 again, we deduce that $a + c \equiv b + d \pmod{n}$. Similarly, n also divides the integer

$$a(c - d) + d(a - b) = (ac - ad) + (ad - bd) = ac - bd.$$

By Lemma 5.5 again, we see that $ac \equiv bd \pmod{n}$. □

Remark 5.10. From now on, we will routinely make use of Lemma 5.5 without explicit mention.

To give the reader some understanding of the power of the theory we have developed so far, we will work out some examples.

Example 5.11. In Proposition 3.9, we proved using induction that 3 divides $n^3 + 2n$ for every positive integer n . We will now employ the methods of modular arithmetic (Lemma 5.7 and Theorem 5.9) to give a simple proof that $n^3 + 2n$ is divisible by 3 for *every* integer n . Interestingly, our proof will not rely on induction!

Our goal will be to show that for every $n \in \mathbb{Z}$,

$$(5.2) \quad n^3 + 2n \equiv 0 \pmod{3}.$$

By Lemma 5.5, this is equivalent to the statement that $n^3 + 2n$ is divisible by 3 for every integer $n \in \mathbb{Z}$.

First, we use Lemma 5.7 to see that every integer n is congruent to either 0, 1, or 2 modulo 3. This neatly breaks the problem up into only three cases. In each case, we will use Theorem 5.9 repeatedly.

Case 1: $n \equiv 0 \pmod{3}$. By Theorem 5.9, we see that

$$n^2 \equiv n \cdot n \equiv 0 \cdot 0 \equiv 0 \pmod{3}.$$

Using Theorem 5.9 again, we get that

$$n^3 \equiv n \cdot n^2 \equiv 0 \cdot 0 \equiv 0 \pmod{3}.$$

Likewise, we may use Theorem 5.9 to get that

$$2n \equiv 2 \cdot 0 \equiv 0 \pmod{3}.$$

Therefore, using Theorem 5.9 *again* we deduce that

$$n^3 + 2n \equiv 0 + 0 \equiv 0 \pmod{3}.$$

Thus we have demonstrated that (5.2) holds in the case that $n \equiv 0 \pmod{3}$.

Case 2: $n \equiv 1 \pmod{3}$. Mimicking our calculation above, we use Theorem 5.9 repeatedly to get that

$$\begin{aligned} n^2 &\equiv n \cdot n \equiv 1 \cdot 1 \equiv 1 \pmod{3}, \\ n^3 &\equiv n \cdot n^2 \equiv 1 \cdot 1 \equiv 1 \pmod{3}, \\ 2n &\equiv 2 \cdot 1 \equiv 2 \pmod{3}, \end{aligned}$$

and thus

$$n^3 + 2n \equiv 1 + 2 \equiv 3 \equiv 0 \pmod{3}.$$

Notice that we used Theorem 5.9 four times in the previous sentence.

Case 3: $n \equiv 2 \pmod{3}$. Mimicking our calculations above, we use Theorem 5.9 over and over again to get that

$$\begin{aligned} n^2 &\equiv n \cdot n \equiv 2 \cdot 2 \equiv 4 \equiv 1 \pmod{3}, \\ n^3 &\equiv n \cdot n^2 \equiv 2 \cdot 1 \equiv 2 \pmod{3}, \\ 2n &\equiv 2 \cdot 2 \equiv 4 \equiv 1 \pmod{3}, \end{aligned}$$

and therefore

$$n^3 + 2n \equiv 2 + 1 \equiv 0 \pmod{3}.$$

This completes the proof, since we have shown that $n^3 + 2n \equiv 0 \pmod{3}$ for every integer n .

Remark 5.12. In the last example we exploited one of the key ideas of modular arithmetic: that when we are doing arithmetic (multiplying, adding and subtracting integers) with respect to congruence modulo n , by applying Theorem 5.9 repeatedly we can replace any integer with any other integer as long as the two integers are congruent modulo n . The following corollaries, whose proofs rely on repeated use of Theorem 5.9, precisely record this idea. They will simplify the arguments and calculations of Example 5.11 and allow us to avoid repeated use of Theorem 5.9 in the future.

Corollary 5.13. *Suppose $a, b \in \mathbb{Z}$ and $n \in \mathbb{N}$ such that $a \equiv b \pmod{n}$. Then*

$$a^k \equiv b^k \pmod{n}$$

for every $k \in \mathbb{N}$.

Proof. We will proceed by induction on k . By assumption, $a^1 \equiv b^1 \pmod{n}$. Now suppose that

$$a^k \equiv b^k \pmod{n}.$$

Using Theorem 5.9, we see that

$$a^{k+1} \equiv a \cdot a^k \equiv b \cdot b^k \equiv b^{k+1} \pmod{n}.$$

This completes the proof. □

Corollary 5.14. *Let*

$$p(x) = \sum_{k=0}^m c_k x^k$$

be a polynomial with integer coefficients $c_0, c_1, \dots, c_m \in \mathbb{Z}$. Suppose that $a, b \in \mathbb{Z}$ and $n \in \mathbb{N}^$ such that $a \equiv b \pmod{n}$. Then*

$$p(a) \equiv p(b) \pmod{n}.$$

Exercise 5.2. Prove Corollary 5.14.

Exercise 5.3. Show that $n^4 + n^2 + 2n$ is divisible by 4 for every integer n . Use Corollary 5.14 to improve the method employed in Example 5.11.

Example 5.15. By Lemma 5.7, we know that every integer is congruent to a “small” integer modulo n (by “small” we mean a nonnegative integer less than the modulus n). This allows us to break problems into a finite number of cases (as in Example 5.11) as well as handle problems involving *very large* integers. For instance, consider whether the integer

$$k = 1987635232343243134131234892045642176314$$

has the property that $k^{17} + 17k$ is a multiple of 3. By Lemma 5.5, we only need to figure out whether $k^{17} + 17k \equiv 0 \pmod{3}$ or not. As we will see, we can figure this out without actually calculating the very large integer $k^{17} + 17k$ (lucky for us, because it would take a while to calculate). It takes only a couple moments to calculate $k \pmod{3} = 2$. Thus, by Lemma 5.7, we see that $k \equiv 2 \pmod{3}$. By Corollary 5.14, we see that

$$k^{17} + 17k \equiv 2^{17} + 17 \cdot 2 \equiv (-1)^{17} + 34 \equiv -1 + 34 \equiv 33 \equiv 0 \pmod{3}.$$

Whew, that was easy!

Example 5.16. In this example, we will calculate the last two decimal digits of the very large integer 3^{693} . We obviously want to avoid calculating 3^{693} explicitly (it has over 300 decimal digits), and so we will have to make clever use of Theorem 5.9. First of all, notice that our task is equivalent to calculating

$$r = 3^{693} \pmod{100}.$$

(If you don't see why this is so, think about the division theorem.) We will begin by calculating smaller powers of 3 (modulo 100) and working our way up to larger and larger powers using Theorem 5.9:

$$\begin{aligned}
 3^1 &\equiv 3 \pmod{100}, \\
 3^2 &\equiv 9 \pmod{100}, \\
 3^4 &\equiv 81 \pmod{100}, \\
 3^8 &\equiv (3^4)^2 \equiv (81)^2 \equiv 6561 \equiv 61 \pmod{100}, \\
 3^{16} &\equiv (3^8)^2 \equiv (61)^2 \equiv 3721 \equiv 21 \pmod{100}, \\
 3^{32} &\equiv (3^{16})^2 \equiv (21)^2 \equiv 441 \equiv 41 \pmod{100}, \\
 3^{64} &\equiv (3^{32})^2 \equiv (41)^2 \equiv 1681 \equiv 81 \pmod{100}, \quad (81 \text{ again!}) \\
 3^{128} &\equiv (3^{64})^2 \equiv (81)^2 \equiv 61 \pmod{100}, \\
 3^{256} &\equiv (3^{128})^2 \equiv (61)^2 \equiv 21 \pmod{100}, \\
 3^{512} &\equiv (3^{256})^2 \equiv (21)^2 \equiv 41 \pmod{100}.
 \end{aligned}$$

Using our calculations above and Theorem 5.9, we calculate

$$\begin{aligned}
 3^{693} &\equiv 3^{512} \cdot 3^{181} \pmod{100} \\
 &\equiv 3^{512} \cdot 3^{128} \cdot 3^{53} \pmod{100} \\
 &\equiv 3^{512} \cdot 3^{128} \cdot 3^{32} \cdot 3^{16} \cdot 3^4 \cdot 3^1 \pmod{100} \\
 &\equiv (41)(61)(41)(21)(81)(3) \pmod{100} \\
 &\equiv (41)^2(81)(61)(21)(3) \pmod{100} \\
 &\equiv (81)^2(61)(21)(3) \pmod{100} \\
 &\equiv (61)^2(21)(3) \pmod{100} \\
 &\equiv (21)^2(3) \pmod{100} \\
 &\equiv (41)(3) \pmod{100} \\
 &\equiv 123 \equiv 23 \pmod{100}.
 \end{aligned}$$

Thus the last two decimal digits of 3^{693} are 2 and 3.

Exercise 5.4. (a) Calculate $17^{771} \pmod{25}$.

(b) Find the last two digits in the base-6 representation of 5^{65656} .

Example 5.17. As another application of modular arithmetic, we will show that the polynomial equation

$$41x^5 + 7x^4 - 37x^3 + 3x^2 - 21x + 13 = 2006$$

has no integer solutions. To that end, let p be the polynomial

$$(5.3) \quad p(x) = 41x^5 + 7x^4 - 37x^3 + 3x^2 - 21x + 13.$$

We will show that $p(n) \not\equiv 2 \pmod{3}$ for every integer n . Having accomplished this, we will have completed the proof, since $2006 \equiv 2 \pmod{3}$. First of all, notice that

$$41 \equiv 2, \quad 7 \equiv 1, \quad -37 \equiv 2, \quad 3 \equiv 0, \quad -21 \equiv 0, \quad \text{and} \quad 13 \equiv 1 \pmod{3}.$$

Thus, by Theorem 5.9,

$$(5.4) \quad p(n) \equiv 2n^5 + n^4 + 2n^3 + 1 \pmod{3}$$

for any integer n . If n is an integer congruent to 0 modulo 3, then by Corollary 5.14 and Equation (5.4)

$$p(n) \equiv 2 \cdot 0^5 + 0^4 + 2 \cdot 0^3 + 1 \equiv 1 \pmod{3}.$$

If $n \equiv 1 \pmod{3}$, then using Corollary 5.14 again we see that

$$p(n) \equiv 2 \cdot 1^5 + 1^4 + 2 \cdot 1^3 + 1 \equiv 6 \equiv 0 \pmod{3}.$$

Lastly, if $n \equiv 2 \pmod{3}$, then

$$p(n) \equiv 2 \cdot 2^5 + 2^4 + 2 \cdot 2^3 + 1 \equiv 97 \equiv 1 \pmod{3}.$$

Using Lemma 5.7, we deduce that $p(n) \not\equiv 2$ for every integer n . Therefore $p(n) \neq 2006$ for every integer n , and so the polynomial equation (5.3) has no integer solutions.

Remark 5.18. The key to the success of our argument in Example 5.17 was choosing the right modulus. If we had chosen 2 for the modulus instead of 3 that our argument would have failed to work.

Exercise 5.5. Show that the polynomial $q(x) = 31x^4 - 4x^3 - 17x + 60$ has no integer roots. (Hint: Work in mod 7.)

5.2 Linear Congruences

In this section we will study *linear congruences*, which are expressions of the form

$$(5.5) \quad ax \equiv b \pmod{n}.$$

In this context, the integers a, b and the modulus $n > 0$ are given and the variable x is the unknown. We say that an integer x is a **solution** of the congruence (5.5) if it satisfies (5.5). In our study of equation (5.5), we will seek to answer the following basic questions:

1. Does equation (5.5) have any solutions at all?
2. If equation (5.5) does have a solution, is it unique in any sense?
3. If equation (5.5) does have solution(s), how do we calculate (all of) them?

The next two examples will demonstrate that the answer to the first question depends on the given integers a, b and the modulus n .

Example 5.19. Consider the linear congruence

$$(5.6) \quad 3x \equiv 7 \pmod{13}.$$

Clearly $x = 11$ is a solution, since $3 \cdot 11 = 33 \equiv 7 \pmod{13}$. Due to Theorem 5.9, every integer congruent to 11 modulo 13 will also be a solution. For if $y \equiv 11 \pmod{13}$, then

$$3y \equiv 3 \cdot 11 \equiv 33 \equiv 7 \pmod{13}.$$

Thus there are in fact infinitely many solutions to (5.6).

Example 5.20. Now consider the linear congruence

$$(5.7) \quad 3x \equiv 7 \pmod{12}.$$

We will argue that (5.7) has no solutions. Let $k \in \mathbb{Z}$ be arbitrary, and suppose that b is an integer satisfying

$$3k \equiv b \pmod{12}.$$

Then 12 divides the integer $3k - b$, and so there exists an integer $t \in \mathbb{Z}$ such that

$$3x - b = 12t.$$

Hence

$$b = 3x - 12t = 3(x - 4),$$

so b is a multiple of 3.

Therefore, if the linear congruence

$$3x \equiv b \pmod{12}.$$

has a solution, it is necessary that b is a multiple of 3. Since 7 is not a multiple of 3, it follows that (5.7) has no solutions.

Examples 5.19 and 5.20 have shown us that in some cases the linear congruence (5.5) will have infinitely many solutions, and in other cases it will have no solutions. To completely answer the first question posed above, we need to *characterize* exactly when (5.5) has solutions and when it does not. That is, we would like to discover a simple test that will determine, given a, b and n , whether or not equation (5.5) has solutions.

We will try writing down the circumstance we wish to characterize, and see if we can show that it is equivalent to simpler situations that we may be familiar with:

Equation (5.5) has a solution

$$\begin{aligned} &\iff \exists x \in \mathbb{Z} \text{ such that } ax \equiv b \pmod{n} \\ &\iff \exists x \in \mathbb{Z} \text{ such that } n \mid (ax - b) \\ &\iff \exists x \in \mathbb{Z} \text{ such that } \exists k \in \mathbb{Z} \text{ such that } kn = (ax - b) \\ &\iff \exists x, k \in \mathbb{Z} \text{ such that } b = (ax - kn) \\ &\iff b \in \{ax + ny \mid x, y \in \mathbb{Z}\} \\ &\iff b \in \{k \gcd(a, n) \mid k \in \mathbb{Z}\} \\ &\iff \gcd(a, n) \mid b. \end{aligned}$$

Now we record our observations in the form of a theorem, and carefully write out all the details of the proof, which is based on the preceding deduction.

Theorem 5.21 (Linear Congruence Theorem). *Suppose that $a, b, n \in \mathbb{Z}$ such that $n > 1$. Then there exists a solution $x \in \mathbb{Z}$ of the linear congruence*

$$(5.8) \quad ax \equiv b \pmod{n}$$

if and only if $\gcd(a, n)$ divides b .

Proof. Suppose that $x \in \mathbb{Z}$ such that $ax \equiv b \pmod{n}$. Then n divides $ax - b$. This implies that $\gcd(a, n)$ divides $ax - b$ as well. Since $\gcd(a, n)$ divides a , it also divides ax . Thus $\gcd(a, n)$ divides $ax - (ax - b) = b$ as well.

Conversely, suppose that $\gcd(a, n)$ divides b . Then we may find an integer k such that $k \gcd(a, n) = b$. By Theorem 4.27, there exists integers $s, t \in \mathbb{Z}$ such that

$$sa + tn = \gcd(a, n).$$

Hence

$$a(sk) + n(tk) = k \cdot (sa + tn) = k \gcd(a, n) = b.$$

Define $x = sk$. Then

$$ax - b = (-tk)n,$$

and so n divides $ax - b$. Therefore, $ax \equiv b \pmod{n}$. □

Example 5.22. The proof of Theorem 5.21 provides us with a method for finding the solutions of (5.8) in practice. Examining the proof, we see that one solution of the congruence

$$ax \equiv b \pmod{n}$$

is

$$x = s \cdot \frac{b}{\gcd(a, n)}$$

where s is an integer that satisfies $sa + tn = \gcd(a, n)$ for some integer t . To illustrate the technique, consider the linear congruence

$$(5.9) \quad 18x \equiv 12 \pmod{42}.$$

We first carry out the Euclidean Algorithm on the integers 42 and 18:

$$\begin{array}{rcl} 1 \cdot 42 & + 0 \cdot 18 & = 42 \\ 0 \cdot 42 & + 1 \cdot 18 & = 18 \\ 1 \cdot 42 & + (-2) \cdot 18 & = 6 \quad (\text{Line 1} - 2 \cdot \text{Line 2}). \end{array}$$

Since 6 divides 18, we see that

$$\gcd(42, 18) = 6 = 1 \cdot 42 + (-2) \cdot 18.$$

In our notation above, we see that $s = -2$. Thus, we expect that one solution of (5.9) should be

$$x = -2 \cdot \frac{12}{6} = -4.$$

It is easy to check that

$$18x = 18 \cdot (-4) = -72 \equiv 12 \pmod{42},$$

so $x = -4$ is indeed a solution.

Example 5.23. Consider the linear congruence

$$(5.10) \quad 18x \equiv 100 \pmod{39}.$$

By performing the Euclidean Algorithm, we see that

$$\gcd(39, 18) = 3.$$

Since 3 does not divide 100, according to Theorem 5.21 there is no solution of (5.10).

Exercise 5.6. Find a solution (if any exist) of the linear congruence

$$45x \equiv 999 \pmod{108}.$$

We pause here to point out that in the case that a and n are relatively prime, equation (5.8) has a solution x for *every* given integer b .

Corollary 5.24. *Suppose that a and n are relatively prime integers with $n > 1$. Then for every integer b , there exists a solution $x \in \mathbb{Z}$ of the linear congruence*

$$ax \equiv b \pmod{n}.$$

Proof. If a and n are relatively prime, then $1 = \gcd(a, n)$ and so $\gcd(a, n)$ divides every integer b . \square

Exercise 5.7. Suppose that p is a prime. Show that for every $b \in \mathbb{Z}$ and every $a \in \mathbb{Z}$ such that $a \not\equiv 0 \pmod{p}$, the linear congruence

$$ax \equiv b \pmod{p}$$

has a solution $x \in \mathbb{Z}$.

Exercise 5.8. Show that the linear congruence

$$ax \equiv 1 \pmod{n}$$

has a solution if and only if a and n are relatively prime.

Continuing our study of the equation

$$(5.11) \quad ax \equiv b \pmod{n},$$

we will next consider the conditions under which solutions of (5.11) are unique. We will also need to carefully consider what we mean by the word *unique*, since, as was pointed out in Example 5.19, solutions of (5.5) are never unique in the strictest sense. To see why, suppose that $x \in \mathbb{Z}$ is a solution of (5.11), where $a, b, n \in \mathbb{Z}$ and $n > 1$. Pick another integer y that is congruent to x modulo n :

$$y \equiv x \pmod{n}.$$

Due to Theorem 5.9, we see that

$$ay \equiv ax \equiv b \pmod{n},$$

so y is another solution of equation (5.11). Therefore, there are necessarily *infinitely* many solutions of equation (5.11) if there are any (because there are infinitely many integers that congruent to any one particular integer modulo n).

It might seem as though we have thoroughly answered the second question posed at the beginning of this section, but we will get a more interesting answer if we were to ask a “better” question. We know that solutions of (5.11) cannot be unique, but this fact begs the question, what do solutions of (5.11) have in common? We have seen that any integer that is congruent modulo n to a solution of (5.11) is also a solution. Does the converse hold? That is, are any two solutions congruent modulo n ?

To study this question, let's consider the situation that

$$(5.12) \quad ax \equiv ay \pmod{n}$$

and try to discover what we can say about x and y . It would be nice if we could “cancel out” the integer a to deduce that $x \equiv y \pmod{n}$. However, as the next example shows, this is not true in general.

Example 5.25. Take $a = 6$, $x = 4$, $y = 19$ and $n = 10$, and observe that

$$ax \equiv 6 \cdot 4 \equiv 24 \equiv 4 \equiv 114 \equiv 6 \cdot 19 \equiv ay \pmod{n}.$$

However, $4 \not\equiv 19 \pmod{10}$, and so unfortunately

$$x \not\equiv y \pmod{n}.$$

Therefore, the statement $ax \equiv ay \pmod{n}$ does *not* in general imply that $x \equiv y \pmod{n}$.

Example 5.25 notwithstanding, there is one circumstance in which we may cancel out the integer a from both sides of (5.12). We take this up in our next lemma.

Lemma 5.26. *Suppose that a and n are relatively prime integers with $n > 1$. Then for any integers x and y ,*

$$ax \equiv ay \pmod{n}$$

if and only if

$$x \equiv y \pmod{n}.$$

Proof. Suppose that $ax \equiv ay \pmod{n}$. Then n divides $ax - ay = a(x - y)$. By Corollary 4.36 and the fact that a and n are relatively prime we deduce that n divides $x - y$. Thus $x \equiv y \pmod{n}$.

Conversely, suppose that $x \equiv y \pmod{n}$. Then Theorem 5.9 implies that $ax \equiv ay \pmod{n}$. \square

Looking back at Example 5.25 with the insight of Lemma 5.26, we see that our counterexample was made possible by our choice of $a = 6$ and $n = 10$, which are not relatively prime. We now know that concluding that $x \equiv y \pmod{n}$ from the hypothesis that $ax \equiv ay \pmod{n}$ is too much to hope for in the case that $\gcd(a, n) \neq 1$. However, using Lemma 4.25 we are still able to generalize Lemma 5.26 to the case that $\gcd(a, n) \neq 1$.

Theorem 5.27 (Congruence Cancellation Theorem). *Suppose that $a, n \in \mathbb{Z}$ such that $n > 1$. Set $m = n/\gcd(a, n)$. Then for all integers x and y ,*

$$ax \equiv ay \pmod{n}$$

if and only if

$$x \equiv y \pmod{m}.$$

Proof. Let $c = a/\gcd(a, n)$, and notice that Lemma 4.25 implies that c and m are relatively prime integers. We will first show that for all integers x and y , the statement

$$ax \equiv ay \pmod{n}$$

is equivalent to

$$cx \equiv cy \pmod{m}.$$

Indeed, for

$$\begin{aligned} ax \equiv ay \pmod{n} &\iff n \mid a(x - y) \\ &\iff \exists k \in \mathbb{Z} \text{ such that } nk = a(x - y) \\ &\iff \exists k \in \mathbb{Z} \text{ such that } mk = c(x - y) \\ &\iff m \mid c(x - y) \\ &\iff cx \equiv cy \pmod{m}. \end{aligned}$$

Since c and m are relatively prime, we deduce from Lemma 5.26 that $cx \equiv cy \pmod{m}$ if and only if $x \equiv y \pmod{m}$. \square

Corollary 5.28. *Suppose that $a, b, n \in \mathbb{Z}$ such that $n > 1$. Set $m = n/\gcd(a, n)$. Then solutions of the linear congruence*

$$(5.13) \quad ax \equiv b \pmod{n}$$

are unique modulo m . Moreover, if $x \in \mathbb{Z}$ is one solution of equation (5.13), then an integer y satisfies $ay \equiv b \pmod{n}$ if and only if $y \equiv x \pmod{m}$.

Proof. Suppose that $x \in \mathbb{Z}$ satisfies

$$ax \equiv b \pmod{n}.$$

By Theorem 5.27, if $y \in \mathbb{Z}$ then

$$ay \equiv b \equiv ax \pmod{n}$$

if and only if

$$y \equiv x \pmod{m}.$$

\square

Remark 5.29. The phrase “solutions are unique modulo m ” means that if x and y are two solutions, then

$$x \equiv y \pmod{m}.$$

Example 5.30. Returning to Example 5.22, we will find *all* solutions of the linear congruence

$$(5.14) \quad 18x \equiv 12 \pmod{42}.$$

We have already found that $x = -4$ is one solution of (5.14) and that $\gcd(18, 42) = 6$. Noting that $42/6 = 7$, we see that the set of solutions of (5.14) is

$$\{y \in \mathbb{Z} \mid y \equiv -4 \pmod{7}\} = \{-4 + 7t \mid t \in \mathbb{Z}\}.$$

That is, an integer is a solution of (5.14) if and only if it is congruent to -4 modulo 7.

5.3 The Chinese Remainder Theorem

In this section, we will be interested in solving linear congruence systems of the form

$$(5.15) \quad \begin{cases} x \equiv b_1 \pmod{n_1} \\ x \equiv b_2 \pmod{n_2}. \end{cases}$$

Here, the integers b_1 and b_2 and the moduli $n_1, n_2 > 1$ are given and the integer x is the unknown. We seek to find an integer x that satisfies *both* congruences in (5.15). Equation (5.15) is called a *system* because a solution is required to solve two or more separate congruences simultaneously.

Example 5.31. Like the linear congruences we studied in Section 5.2, the system (5.15) may have no solutions. For example, consider the system

$$\begin{cases} x \equiv 3 \pmod{7} \\ x \equiv 4 \pmod{7}. \end{cases}$$

There is clearly no integer that is congruent to both 3 and 4 modulo 7, so this system has no solution.

We will primarily concern ourselves with the case that the moduli n_1 and n_2 are relatively prime. As we will see, this will ensure that there exist solutions of (5.15). Our main result is Theorem 5.33 below, called the Chinese Remainder Theorem.

The following theorem, which is interesting in its own right, will provide the basis for the uniqueness part of Theorem 5.33.

Theorem 5.32 (Relatively Prime Moduli Law). *Assume that $n_1 > 1$ and $n_2 > 1$ are relatively prime integers. Then for all integers a and b ,*

$$a \equiv b \pmod{n_1 n_2}$$

if and only if

$$a \equiv b \pmod{n_1} \quad \text{and} \quad a \equiv b \pmod{n_2}.$$

Proof. Suppose that $a \equiv b \pmod{n_1 n_2}$. Then $n_1 n_2$ divides $a - b$. Therefore, both n_1 and n_2 divide $a - b$. Thus $a \equiv b \pmod{n_1}$ and $a \equiv b \pmod{n_2}$.

Conversely, suppose that $a \equiv b \pmod{n_1}$ and $a \equiv b \pmod{n_2}$. Then both n_1 and n_2 divide $a - b$. Since n_1 and n_2 are relatively prime, Corollary 4.37 implies that the product $n_1 n_2$ divides $a - b$ as well. Thus $a \equiv b \pmod{n_1 n_2}$. \square

Theorem 5.33 (Chinese Remainder Theorem). *Suppose that $n_1 > 1$ and $n_2 > 1$ are relatively prime integers. Let $b_1, b_2 \in \mathbb{Z}$. Then there exists an integer $x \in \mathbb{Z}$ which solves the system*

$$\begin{cases} x \equiv b_1 \pmod{n_1} \\ x \equiv b_2 \pmod{n_2}. \end{cases}$$

Moreover, x is unique modulo $n_1 n_2$.

Proof. By Theorem 5.21 there is an integer t such that

$$n_1 t \equiv b_2 - b_1 \pmod{n_2}.$$

Define

$$x = b_1 + n_1 t.$$

Then clearly $x \equiv b_1 \pmod{n_1}$ since $n_1 t = x - b_1$. Also, from our choice of t we see that

$$x = b_1 + n_1 t \equiv b_2 \pmod{n_2}.$$

To prove uniqueness, suppose that y is another integer satisfying both $y \equiv b_1 \pmod{n_1}$ and $y \equiv b_2 \pmod{n_2}$. Then $x \equiv y \pmod{n_1}$ and $x \equiv y \pmod{n_2}$. Applying Theorem 5.32, we see that

$$x \equiv y \pmod{n_1 n_2}.$$

\square

Example 5.34. From the proof of Theorem 5.33, we will extract a method for finding solutions of (5.15). As an illustration, consider the system

$$(5.16) \quad \begin{cases} x \equiv 17 \pmod{28} \\ x \equiv 104 \pmod{75}. \end{cases}$$

Performing the Euclidean Algorithm, we calculate

$$\begin{aligned} 1 \cdot 75 + 0 \cdot 28 &= 75 \\ 0 \cdot 75 + 1 \cdot 28 &= 28 \\ 1 \cdot 75 - 2 \cdot 28 &= 19 \\ -1 \cdot 75 + 3 \cdot 28 &= 9 \\ 3 \cdot 75 - 8 \cdot 28 &= 1. \end{aligned}$$

Since 75 and 28 are relatively prime, Theorem 5.33 guarantees that there exists integer solutions of (5.16). Mimicking the proof of Theorem 5.33, we look for a solution of (5.16) of the form

$$x = 17 + 28t.$$

We know from the proof of Theorem 5.33 that x will be a solution provided that t solves the linear congruence $28t \equiv (104 - 17) \pmod{75}$, rewritten as

$$28t \equiv 87 \pmod{75}.$$

Using our calculation above, we see that

$$87 \cdot (3 \cdot 75 - 8 \cdot 28) = 87,$$

and therefore $t = -8 \cdot 87 = -696$ satisfies

$$28t = 87 - (3 \cdot 87) \cdot 75,$$

and so $28t \equiv 87 \pmod{75}$. Therefore, $x = 17 + 28t = 17 + 28 \cdot 54 = -19471$ is a solution of the system (5.16). We can get another, more pleasant solution by noticing that $54 \equiv -696 \pmod{75}$. Setting $s = 54$, we therefore deduce that s satisfies $28s \equiv 87 \pmod{75}$. Hence $\tilde{x} = 17 + 28s = 1529$ is another solution of the system (5.16). Moreover, since $28 \cdot 75 = 2100$, the uniqueness part of Theorem 5.33 tells us that the set of *all* solutions of (5.16) is

$$\{y \in \mathbb{Z} \mid y \equiv 1529 \pmod{2100}\} = \{1529 + 2100t \mid t \in \mathbb{Z}\}.$$

Exercise 5.9. Find all solutions of the linear system

$$\begin{cases} x \equiv 36 & (\text{mod } 91) \\ x \equiv 36 & (\text{mod } 333) \end{cases}.$$

Example 5.35. In Example 5.16, we learned how to calculate the remainder of a really big powers of some integer when divided by some other integer, say $a^k \text{ mod } N$. The method involves first calculating $a \text{ mod } N$, then $a^2 \text{ mod } N$, then $a^4 \text{ mod } N$, and $a^8 \text{ mod } N$, etc. Then the Fundamental Theorem of Modular Arithmetic is used to “stitch” these together, giving us $a^k \text{ mod } N$. Now we will make use of Theorem 5.33 to show how we can improve on this method.

Consider the problem of calculating

$$R = 37^{85} \text{ mod } 187.$$

By Lemma 5.7, we know that R is the unique integer satisfying

$$(5.17) \quad \begin{cases} R \equiv 37^{85} & (\text{mod } 187) \\ 0 \leq R < 187. \end{cases}$$

Using the factorization $187 = 11 \cdot 17$, it is equivalent to find the unique integer R satisfying

$$(5.18) \quad \begin{cases} R \equiv 37^{85} & (\text{mod } 11) \\ R \equiv 37^{85} & (\text{mod } 17) \\ 0 \leq R < 187. \end{cases}$$

First, we will reduce 37^{85} with respect to the moduli 11 and 17. We calculate

$$\begin{aligned} 37^1 &\equiv 37 \equiv 4 & (\text{mod } 11) \\ 37^2 &\equiv 4^2 \equiv 16 \equiv 5 & (\text{mod } 11) \\ 37^4 &\equiv 5^2 \equiv 25 \equiv 3 & (\text{mod } 11) \\ 37^8 &\equiv 3^2 \equiv 9 & (\text{mod } 11) \\ 37^{16} &\equiv 9^2 \equiv 81 \equiv 4 & (\text{mod } 11) \\ 37^{32} &\equiv 4^2 \equiv 16 \equiv 5 & (\text{mod } 11) \\ 37^{64} &\equiv 5^2 \equiv 3 & (\text{mod } 11) \end{aligned}$$

and

$$\begin{aligned}
 37^1 &\equiv 37 \equiv 3 \pmod{17} \\
 37^2 &\equiv 3^2 \equiv 9 \pmod{17} \\
 37^4 &\equiv 9^2 \equiv 81 \equiv -4 \pmod{17} \\
 37^8 &\equiv (-4)^2 \equiv 16 \equiv -1 \pmod{17} \\
 37^{16} &\equiv (-1)^2 \equiv 1 \pmod{17} \\
 37^{32} &\equiv 1^2 \equiv 1 \pmod{17} \\
 37^{64} &\equiv 1^2 \equiv 1 \pmod{17}.
 \end{aligned}$$

Therefore we have

$$37^{85} \equiv 37^{64} \cdot 37^{16} \cdot 37^4 \cdot 37 \equiv 3 \cdot 4 \cdot 3 \cdot 4 \equiv 12 \cdot 12 \equiv 1 \cdot 1 \equiv 1 \pmod{11}$$

and

$$37^{85} \equiv 37^{64} \cdot 37^{16} \cdot 37^4 \cdot 37 \equiv 1 \cdot 1 \cdot -4 \cdot 3 \equiv -12 \equiv 5 \pmod{17}.$$

Looking back at (5.18), observe that we have reduced our problem to the system

$$(5.19) \quad \begin{cases} R \equiv 1 \pmod{11} \\ R \equiv 5 \pmod{17} \\ 0 \leq R < 187. \end{cases}$$

Following the method of Example 5.34, we solve the system by first solving the linear congruence $11t \equiv 5 - 1 \equiv 4 \pmod{17}$. Notice that $t = 5$ is a solution. Thus we expect $R = 1 + 11 \cdot 5 = 56$ to be a solution of our system (5.19) (it is easy to check that it is). Note as well that $0 \leq 56 < 187$. Thus $R = 56$ is the unique integer satisfying (5.19), and therefore is also the unique integer satisfying (5.17). Therefore, $56 = 37^{85} \pmod{187}$.

Let's reflect on the advantages of this method as compared with that of Example 5.16, in which we calculate $37^{85} \pmod{187}$ by brute force—calculating $37^2 \pmod{187}$, and then $37^4 \pmod{187}$ and $37^8 \pmod{187}$, etc. This is difficult because 187 is a relatively large modulus, which means the integers we have to square at each step will be relatively large, and afterwards we have to divide the result by 187. We would need a calculator or a lot of time. While

our new method using the Chinese Remainder Theorem may seem to require tedious calculations, at no time did we need a calculator, and our task took no more than a few minutes to complete.

Exercise 5.10. Calculate

$$(705)^{655} \pmod{275}.$$

The Relatively Prime Moduli Law and the Chinese Remainder Theorem can be generalized to the case that n_1 and n_2 are not necessarily relatively prime. According to Exercise 4.11, if n_1 and n_2 are relatively prime then

$$\text{lcm}(n_1, n_2) = n_1 n_2.$$

It turns out that if we may alter Theorems 5.32 and 5.33 by dropping the assumption that n_1 and n_2 are relatively prime and replacing $n_1 n_2$ by $\text{lcm}(n_1, n_2)$.

Theorem 5.36 (Relatively Prime Moduli Law, Version 2). *Let $n_1, n_2 \in \mathbb{N}^*$. Then for all integers a and b ,*

$$a \equiv b \pmod{\text{lcm}(n_1, n_2)}$$

if and only if

$$a \equiv b \pmod{n_1} \quad \text{and} \quad a \equiv b \pmod{n_2}.$$

Exercise 5.11. Prove Theorem 5.36. (Hint: One direction is easy. For the other direction, use Exercise 4.11 to “reduce” this result to that of Theorem 5.32. See also Lemma 4.25.)

Theorem 5.37 (Chinese Remainder Theorem, Version 2). *Suppose that $n_1 > 1$ and $n_2 > 1$ and $b_1, b_2 \in \mathbb{Z}$. Then there exists an integer $x \in \mathbb{Z}$ which solves the system*

$$\begin{cases} x \equiv b_1 & \pmod{n_1} \\ x \equiv b_2 & \pmod{n_2} \end{cases}$$

if and only if $\text{gcd}(n_1, n_2)$ divides $b_1 - b_2$. Moreover, x is unique modulo $\text{lcm}(n_1, n_2)$.

Exercise 5.12. Prove Theorem 5.37. (Hint: Reduce the system above to a single linear congruence, as we did in the proof of Theorem 5.33. Under what conditions is there a solution of the linear congruence?)

5.4 Systems of Linear Congruences

In this section, we will be interested in solving systems of linear congruence of the form

$$(5.20) \quad \begin{cases} a_1x \equiv b_1 & (\text{mod } n_1) \\ a_2x \equiv b_2 & (\text{mod } n_2) \end{cases}$$

Here, the integers a_1, a_2, b_1 and b_2 and the moduli $n_1, n_2 > 1$ are given and the integer x is the unknown. The presence of the integers a_1 and a_2 causes this system to be a bit more complicated than the system we studied in the previous section (notice that if we let $a_1 = a_2 = 1$, then the system (5.20) reduces to precisely the one we studied in Section 5.3).

Like the equations we studied in Section 5.2 and Section 5.3, the system (5.20) may have no solution. However, in our present case there are, roughly speaking, *two* ways in which (5.20) may fail to have solutions. First of all, it is possible that one the congruences in (5.20) to individually fail to have any solution.

Example 5.38. The system

$$(5.21) \quad \begin{cases} 18x \equiv 100 & (\text{mod } 39) \\ 12x \equiv 27 & (\text{mod } 35) \end{cases}$$

has no solution even though the moduli 39 and 35 are relatively prime. This is because the linear congruence $18x \equiv 100 \pmod{39}$ has no solution (see Example 5.23). Clearly there can be no x satisfying the system (5.21) if there is no x satisfying $18x \equiv 100 \pmod{39}$.

Secondly, as we have already seen (Example 5.31), the system (5.20) can fail to have solutions if the moduli n_1 and n_2 fail to be relatively prime. We expect that any theorem asserting the existence of solutions of (5.20) must take both of these problems into account.

Theorem 5.39 (Linear Congruence Systems Theorem). *Let a_1, a_2, b_1, b_2 and $n_1, n_2 > 1$ be given integers. Suppose that $\gcd(a_1, n_1)$ divides b_1 , and that $\gcd(a_2, n_2)$ divides b_2 . Suppose also that the moduli n_1 and n_2 are relatively prime. Then there exists a solution $x \in \mathbb{Z}$ of the linear congruence system*

$$(5.22) \quad \begin{cases} a_1x \equiv b_1 & (\text{mod } n_1) \\ a_2x \equiv b_2 & (\text{mod } n_2) \end{cases}.$$

Proof. Our argument will be based on the following idea: by first solving each of the congruences individually, we can then reduce the system (5.22) to a system of the form (5.15). We can then apply the Chinese Remainder Theorem to finish off the proof.

By Theorem 5.21, there exist integers c_1 and c_2 such that $a_1c_1 \equiv b_1 \pmod{n_1}$ and $a_2c_2 \equiv b_2 \pmod{n_2}$. By the Chinese Remainder Theorem, there exists an integer x that is a solution to the system

$$\begin{cases} x \equiv c_1 & \pmod{n_1} \\ x \equiv c_2 & \pmod{n_2}. \end{cases}$$

We now claim that x solves the system (5.22). Indeed, for clearly

$$a_1x \equiv a_1c_1 \equiv b_1 \pmod{n_1}$$

and

$$a_2x \equiv a_2c_2 \equiv b_2 \pmod{n_2}.$$

□

Example 5.40. Using the proof of Theorem 5.39 as a guide, we will find a solution $x \in \mathbb{Z}$ of the system

$$(5.23) \quad \begin{cases} 14x \equiv 22 & \pmod{32} \\ 5x \equiv 8 & \pmod{39}. \end{cases}$$

First of all, by performing the Euclidean Algorithm twice (calculations omitted) we see that

$$2 = \gcd(32, 14) = -3 \cdot 32 + 7 \cdot 14$$

and

$$1 = \gcd(39, 5) = -1 \cdot 39 + 8 \cdot 5.$$

Multiplying the first equation by 4 and the second equation by 22 yields integers $c_1 = 4 \cdot 7 = 28$ and $c_2 = 22 \cdot 8 = 176$ which satisfy $14c_1 \equiv 22 \pmod{32}$ and $5c_2 \equiv 8 \pmod{39}$. We next must solve the system

$$(5.24) \quad \begin{cases} x \equiv 28 & \pmod{32} \\ x \equiv 176 & \pmod{39}. \end{cases}$$

Following the procedures outlined in Section 5.3 (see Example 5.34), we look for a solution of (5.24) of the form

$$x = 28 + 32t$$

where t solves the linear congruence $32t \equiv 176 - 28 \pmod{39}$. This is equivalent to the linear congruence

$$(5.25) \quad 32t \equiv 31 \pmod{39}.$$

Performing the Euclidean Algorithm again (calculation omitted), we get

$$1 = \gcd(39, 32) = -9 \cdot 39 + 11 \cdot 32.$$

Multiplying by 31, we see that the integer $11 \cdot 31 = 341$ is a solution of equation (5.25). Thus $t = 29$ is also a solution, since $29 \equiv 341 \pmod{39}$. Therefore, the integer $x = 28 + 32 \cdot 29 = 956$ is a solution of the system (5.24). Therefore, it is also a solution of the system (5.23).

5.5 Congruence Classes

Definition 5.41. Suppose that $a, n \in \mathbb{Z}$ with $n > 1$. Then the **congruence class of a modulo n** is the set

$$(5.26) \quad [a]_n = \{k \in \mathbb{Z} \mid k \equiv a \pmod{n}\}.$$

A set C is a **congruence class modulo n** if $C = [b]_n$ for some integer b . If C is a congruence class modulo n , then we say that an integer b is a **representative** of C if $C = [b]_n$.

Example 5.42. The congruence class of 17 modulo 7 is the set

$$\begin{aligned} [17]_7 &= \{k \in \mathbb{Z} \mid k \equiv 17 \pmod{7}\} \\ &= \{\dots, -11, -4, 3, 10, 17, 24, 31, \dots\}. \end{aligned}$$

Lemma 5.43. For every integer a and modulus n ,

$$a \in [a]_n.$$

Proof. Since $a \equiv a \pmod{n}$, clearly $a \in [a]_n$. □

Lemma 5.44. *Let $a, b, n \in \mathbb{Z}$ with $n > 1$. Then the following statements are equivalent:*

- (i) *The congruence classes $[a]_n$ and $[b]_n$ are equal;*
- (ii) *The integer a is an element of the congruence class $[b]_n$;*
- (iii) *The integers a and b are congruent modulo n .*

Proof. By Lemma 5.43, it is clear that $[a]_n = [b]_n$ implies that $a \in [b]_n$. Thus (i) \implies (ii).

If $a \in [b]_n$, then $a \equiv b \pmod{n}$ by Definition 5.41 (see equation (5.26)). Thus (ii) \implies (iii).

Suppose that $a \equiv b \pmod{n}$. Let $x \in [a]_n$. Then

$$x \equiv a \equiv b \pmod{n},$$

and so $x \in [b]_n$. Hence $[a]_n \subset [b]_n$. Arguing the same way but with the roles of a and b reversed, we see that $[b]_n \subset [a]_n$. Therefore, $[a]_n = [b]_n$. Therefore, (iii) \implies (i). \square

Corollary 5.45. *Suppose that C_1 and C_2 are congruence classes modulo n . Then either $C_1 = C_2$ or $C_1 \cap C_2 = \emptyset$.*

Proof. Suppose that $C_1 \cap C_2$ is not empty. Then C_1 and C_2 have a common element, say a . By Lemma 5.44, $C_1 = [a]_n = C_2$. \square

Corollary 5.46. *An integer a is a representative of a congruence class C modulo n if and only if $a \in C$.*

Exercise 5.13. (a) Show by example that the conclusion of Corollary 5.45 may fail if C_1 and C_2 are congruence classes with respect to different moduli.

(b) Show that if $n \neq m$ then

$$[a]_n \neq [b]_m$$

for all integers a and b .

(c) Suppose that the modulus n divides the modulus m . Show that

$$[a]_m \subset [a]_n$$

for every integer a .

(d) Show that

$$[a]_m \cap [a]_n = [a]_{\text{lcm}(m,n)}$$

for every integer a .

Exercise 5.14. Show that $a \bmod n$ is the least nonnegative representative of $[a]_n$.

Definition 5.47. If $n > 1$ is an integer, we define the set \mathbb{Z}_n , called the **set of integers modulo n** , by

$$\mathbb{Z}_n = \{[a]_n \mid a \in \mathbb{Z}\}.$$

Lemma 5.48. *The set \mathbb{Z}_n has exactly n distinct elements, which are*

$$[0]_n, [1]_n, \dots, [n-1]_n.$$

Proof. We will first show that every element of \mathbb{Z}_n is one of the congruence classes in the list

$$(5.27) \quad [0]_n, [1]_n, \dots, [n-1]_n.$$

Let $A \in \mathbb{Z}_n$. Then we may write $A = [a]_n$ for some integer a . By Lemma 5.7, there is an integer r such that $a \equiv r \pmod{n}$ and $0 \leq r \leq n-1$. Notice that $[r]_n$ is in the list (5.27). Moreover, Lemma 5.44 implies that $A = [a]_n = [r]_n$.

Now we will show that no two members of the list (5.27) are equal. For suppose that r and s are integers such that $0 \leq r, s \leq n-1$ and $[r]_n = [s]_n$. Then by the uniqueness part of Lemma 5.7, we see that $r = s$. \square

Remark 5.49. While Theorem 4.10, Lemma 5.7 and Lemma 5.48 may appear to be very different statements, all three are really the “same” theorem written in three different ways. Often in mathematics we encounter the same idea in different contexts, and it is important for students to develop an ability to recognize seemingly different occurrences of the same idea. The reader should spend the time necessary to become convinced that these three propositions are indeed the “same” theorem.

We would like to give \mathbb{Z}_n an addition and multiplication structure that reflects how modular arithmetic works. Indeed, the purpose of creating congruence classes and \mathbb{Z}_n is to formalize precisely some properties of modular

arithmetic that guide our intuition but are formally incorrect. For example, we think of the integers 13 and 29 as being “the same” modulo 8. We really want to write something that reflects this, like “ $13 = 29$.” However, since 13 and 29 are *not* the same integer, we had to settle for the statement $13 \equiv 29 \pmod{8}$. However, now we can write:

$$[13]_8 = [29]_8.$$

This time, the equality sign is the literal truth. Due to Lemma 5.44, we can write $[a]_n = [b]_n$ instead of the equivalent statement $a \equiv b \pmod{n}$.

If this is not convincing, consider the bit of intuition spelled out in Example 5.8. We wanted to say something like “there are only n integers modulo n ”. Well, now we have proven this bit of philosophy rigorously in Lemma 5.48. Before we defined congruence classes, the world of modular arithmetic still required us to deal with the infinite set of integers. Now we can simply work with the finite set \mathbb{Z}_n .

However, in order to encode expressions like

$$(5.28) \quad 15 + 19 \equiv 6 \pmod{7}$$

and

$$(5.29) \quad 5 \cdot 6 = 2 \pmod{7}$$

into our new framework, we still must define an *addition* and *multiplication* on the set \mathbb{Z}_n . That is, we need to invent a way of adding and multiplying congruence classes. Glancing back to at equations (5.28) and (5.29), we would *like* to be able to write

$$(5.30) \quad [15]_7 + [19]_7 \equiv [6]_7$$

and

$$(5.31) \quad [5]_7 \cdot [6]_7 \equiv [2]_7,$$

so whatever scheme we come up with for adding and multiplying congruence classes should make these expressions true.

It seems like a good idea to simply define the sum and the product of the congruence classes $[a]_7$ and $[b]_7$ to be the congruence classes $[a + b]_7$ and $[a \cdot b]_7$, respectively. This would make both of the expressions (5.30) and

(5.31) true statements. However, there is a slight problem we need to worry about. A single congruence class has many representatives. For example, $[15]_7 = [36]_7$ and $[19]_7 = [-23]_7$. So whatever we define $[15]_7 + [19]_7$ to be, it better agree with our definition of $[36]_7 + [-23]_7$, otherwise our definition will not make sense! The following lemma, based on the Fundamental Theorem of Modular Arithmetic is just what we need to fix this problem.

Lemma 5.50. *Suppose A and B are two congruence classes modulo n . Suppose as well that a_1, a_2 are representatives of A , and b_1, b_2 are representatives of B . Then*

$$[a_1 + b_1]_n = [a_2 + b_2]_n$$

and

$$[a_1 \cdot b_1]_n = [a_2 \cdot b_2]_n.$$

Proof. The proof is simply a matter of translating Theorem 5.9 into our current framework. Since a_1 and a_2 are representatives of A , according to Lemma 5.44 we have

$$A = [a_1]_n = [a_2]_n.$$

Likewise, we may write

$$B = [b_1]_n = [b_2]_n.$$

By Lemma 5.44 again, we see that $a_1 \equiv a_2 \pmod{n}$ and $b_1 \equiv b_2 \pmod{n}$. Apply Theorem 5.9 to get

$$a_1 + b_1 \equiv a_2 + b_2 \pmod{n}$$

and

$$a_1 \cdot b_1 \equiv a_2 \cdot b_2 \pmod{n}.$$

Translating back into congruence class terminology with the use of Lemma 5.44, we deduce that $[a_1 + b_1]_n = [a_2 + b_2]_n$ and $[a_1 \cdot b_1]_n = [a_2 \cdot b_2]_n$. \square

Remark 5.51. Roughly speaking, Lemma 5.50 says that if $A = [a]_n$ and $B = [b]_n$, then the congruence classes $[a + b]_n$ and $[a \cdot b]_n$ do not depend on which representatives a and b of A and B were chosen. Lemma 5.50 should be considered a version of Theorem 5.9, merely rewritten into our new language of congruence classes. This justifies our earlier intuition that Theorem 5.9 is what makes modular arithmetic possible—in this case, it allows us to define addition and multiplication of congruence classes.

Definition 5.52. If $A = [a]_n$ and $B = [b]_n$ are congruence classes modulo n , then we define the **sum** of A and B to be the congruence class

$$A + B = [a + b]_n.$$

Likewise, the **product** of A and B is the congruence class

$$A \cdot B = [a \cdot b]_n.$$

Exercise 5.15. Suppose that $n > 1$ is an integer and $A, B, C \in \mathbb{Z}_n$.

- (a) Show that $(A \cdot B) \cdot C = A \cdot (B \cdot C)$.
- (b) Show that $A \cdot B = B \cdot A$.
- (c) Show that $A \cdot (B + C) = A \cdot B + A \cdot C$.
- (d) Show that \mathbb{Z}_n has a unique additive identity.
- (e) Show that \mathbb{Z}_n has a unique multiplicative identity.

Definition 5.53. An congruence class $A \in \mathbb{Z}_n$ is called a **unit in \mathbb{Z}_n** if there exists another element $B \in \mathbb{Z}_n$ such that

$$(5.32) \quad A \cdot B = [1]_n.$$

If equation (5.32) holds, then we say that B is a **multiplicative inverse of A in \mathbb{Z}_n** .

Example 5.54. We will find all the units in \mathbb{Z}_4 . Clearly $[0]_4$ cannot be a unit, since $[0]_4 \cdot B = [0]_4$ for every $B \in \mathbb{Z}_4$. Notice that $[1]_4$ is a unit, and is its own multiplicative inverse, since

$$[1]_4 \cdot [1]_4 = [1]_4.$$

Also, $[3]_4$ is a unit (and its own inverse) since

$$[3]_4 \cdot [3]_4 = [9]_4 = [1]_4.$$

The congruence class $[2]_4$ is not a unit since

$$[2]_4 \cdot [b]_4 = \begin{cases} [0]_4 & \text{if } b = 0, 2 \\ [2]_4 & \text{if } b = 1, 3. \end{cases}$$

Therefore the units of \mathbb{Z}_4 are $[1]_4$ and $[3]_4$.

Exercise 5.16. (a) Show that a unit in \mathbb{Z}_n has a unique multiplicative inverse.

(b) Show that the multiplicative inverse of a unit in \mathbb{Z}_n is a unit in \mathbb{Z}_n .

Notation 5.55. Due to Exercise 5.16, if $A \in \mathbb{Z}_n$ is a unit then we may refer to *the* multiplicative inverse of A , which we will hereafter denote by A^{-1} . For example,

$$[7]_{31}^{-1} = [9]_{31}$$

since $[7]_{31} \cdot [9]_{31} = [63]_{31} = [1]_{31}$.

Exercise 5.17. Show that if A is a unit, then $(A^{-1})^{-1} = A$.

The following theorem may be thought of as version Theorem 5.27, rephrased in terms of congruence classes.

Theorem 5.56. *Let $A \in \mathbb{Z}_n$ be a congruence class modulo n , and let $a \in A$. Then A is a unit in \mathbb{Z}_n if and only if a and n are relatively prime.*

Proof. By Lemma 5.44, $A = [a]_n$. Thus we have

$$\begin{aligned} A \text{ is a unit} &\iff \exists B \in \mathbb{Z}_m \text{ such that } A \cdot B = [1]_n \\ &\iff \exists b \in \mathbb{Z} \text{ such that } [a]_n \cdot [b]_n = [1]_n \\ &\iff \exists b \in \mathbb{Z} \text{ such that } [ab]_n = [1]_n \\ &\iff \exists b \in \mathbb{Z} \text{ such that } ab \equiv 1 \pmod{n}. \end{aligned}$$

So we see that $A = [a]_n$ is a unit in \mathbb{Z}_n if and only if the linear congruence

$$(5.33) \quad ax \equiv 1 \pmod{n}$$

has a solution $x \in \mathbb{Z}$. By Exercise 5.8, equation (5.33) has a solution if and only if a and n are relatively prime. \square

Corollary 5.57. *Suppose that p is prime. Then every element of \mathbb{Z}_p is a unit except $[0]_p$. In particular, \mathbb{Z}_p has $p - 1$ units.*

Exercise 5.18. Show that $A \in \mathbb{Z}_n$ is a unit if and only if the only element $B \in \mathbb{Z}_n$ for which $A \cdot B = [0]_n$ is $B = [0]_n$.

We will often find it necessary to refer to the subset of \mathbb{Z}_n consisting of the units in \mathbb{Z}_n . So we hereafter denote by U_n the set of units in \mathbb{Z}_n . In other words,

$$U_n = \{A \in \mathbb{Z}_n \mid A \text{ is a unit}\}.$$

For example, according to Example 5.54,

$$U_4 = \{[1]_4, [3]_4\}.$$

More generally, according to Theorem 5.56,

$$U_n = \{[a]_n \mid \gcd(a, n) = 1\}.$$

In particular, if p is prime then by Corollary 5.57

$$U_p = \mathbb{Z}_p \setminus \{[0]_p\} = \{[1]_p, [2]_p, \dots, [p-1]_p\}.$$

Exercise 5.19. Show that for any $n > 1$, the set U_n of units in \mathbb{Z}_n is a nonempty, proper subset of \mathbb{Z}_n .

Example 5.58. According to Theorem 5.56, \mathbb{Z}_{24} has exactly 8 units:

$$(5.34) \quad U_{24} = \{[1]_{24}, [5]_{24}, [7]_{24}, [11]_{24}, [13]_{24}, [17]_{24}, [19]_{24}, [23]_{24}\}.$$

Something interesting happens when we multiply each unit in \mathbb{Z}_{24} by a particular unit. For example, let's multiply every unit in \mathbb{Z}_{24} by the unit $[7]_{24}$:

$$\begin{aligned} [7]_{24} \cdot [1]_{24} &= [7 \cdot 1]_{24} = [7]_{24}, \\ [7]_{24} \cdot [5]_{24} &= [7 \cdot 5]_{24} = [35]_{24} = [11]_{24}, \\ [7]_{24} \cdot [7]_{24} &= [7 \cdot 7]_{24} = [49]_{24} = [1]_{24}, \\ [7]_{24} \cdot [11]_{24} &= [7 \cdot 11]_{24} = [77]_{24} = [5]_{24}, \\ [7]_{24} \cdot [13]_{24} &= [7 \cdot 13]_{24} = [91]_{24} = [19]_{24}, \\ [7]_{24} \cdot [17]_{24} &= [7 \cdot 17]_{24} = [119]_{24} = [23]_{24}, \\ [7]_{24} \cdot [19]_{24} &= [7 \cdot 19]_{24} = [133]_{24} = [13]_{24}, \\ [7]_{24} \cdot [23]_{24} &= [7 \cdot 23]_{24} = [161]_{24} = [17]_{24}. \end{aligned}$$

Notice that when we multiply each unit by $[7]_{24}$, we get another unit. Also, notice that multiplying the units in \mathbb{Z}_{24} by $[7]_{24}$ just rearranges the list! Every unit in \mathbb{Z}_{24} appears exactly once on the right hand side in the string of calculations above. In the next two propositions, we record these observations.

Lemma 5.59. *The set of units in \mathbb{Z}_n is closed under multiplication. That is, if $A, B \in U_n$, then $A \cdot B \in U_n$.*

Proof. Suppose that A and B are units in \mathbb{Z}_n . Using Exercise 5.15, we see that

$$(A \cdot B) \cdot (A^{-1} \cdot B^{-1}) = (A \cdot A^{-1}) \cdot (B \cdot B^{-1}) = [1]_n \cdot [1]_n = [1]_n.$$

Thus $A \cdot B$ is a unit with multiplicative inverse $A^{-1} \cdot B^{-1}$. \square

Theorem 5.60. *Fix a unit $A \in U_n$. Then for every unit $B \in U_n$, there exists a unique unit $C \in U_n$ such that*

$$(5.35) \quad B = A \cdot C.$$

In fact, $C = A^{-1} \cdot B$.

Proof. It is clear that $C = A^{-1} \cdot B$ satisfies equation (5.35). If $\tilde{C} \in U_n$ is another unit that satisfies $B = A \cdot \tilde{C}$, then we multiply both sides by the congruence class A^{-1} to get

$$A^{-1} \cdot B = A^{-1} \cdot (A \cdot \tilde{C}) = (A^{-1} \cdot A) \cdot \tilde{C} = [1]_n \cdot \tilde{C} = \tilde{C}.$$

Thus $\tilde{C} = C$. \square

Remark 5.61. It is important to understand the significance of Lemma 5.59 and Theorem 5.60 and their relationship to Example 5.58. Suppose that

$$(5.36) \quad U_1, U_2, \dots, U_k$$

is a list of all the units in \mathbb{Z}_n . Select one particular unit from the list, say U_j . What Lemma 5.59 says is that every congruence class in the list

$$(5.37) \quad (U_j \cdot U_1), (U_j \cdot U_2), \dots, (U_j \cdot U_k)$$

is also a unit. What Theorem 5.60 asserts is that every unit in \mathbb{Z}_n appears precisely once in the list (5.37). That is, (5.37) is merely a rearrangement of the list (5.36). This is what we witnessed in Example 5.58. To review, the units in \mathbb{Z}_{24} are

$$(5.38) \quad [1]_{24}, [5]_{24}, [7]_{24}, [11]_{24}, [13]_{24}, [17]_{24}, [19]_{24}, [23]_{24}.$$

As we calculated in Example 5.58, if we multiply every unit above by the fixed unit $[7]_{24}$, we get

$$[7]_{24}, [11]_{24}, [1]_{24}, [5]_{24}, [19]_{24}, [23]_{24}, [13]_{24}, [17]_{24},$$

which is a rearrangement of (5.38).

5.6 Fermat's Little Theorem

Theorem 5.62. *Suppose that p is a prime, and a is an integer such that $a \not\equiv 0 \pmod{p}$. Then*

$$a^{p-1} \equiv 1 \pmod{p}.$$

Proof. By Corollary 5.57, the units of \mathbb{Z}_p are

$$(5.39) \quad [1]_p, [2]_p, \dots, [p-1]_p.$$

Since $a \not\equiv 0 \pmod{p}$, the congruence class $[a]_p \in U_p$ by Corollary 5.57. Multiplying every unit in (5.39) by $[a]_p$, we get

$$(5.40) \quad [1 \cdot a]_p, [2 \cdot a]_p, \dots, [(p-1) \cdot a]_p.$$

By Theorem 5.60, the list (5.40) is a rearrangement of (5.39). Therefore,

$$\begin{aligned} [1 \cdot 2 \cdots (p-1)]_p &= [1]_p \cdot [2]_p \cdots [p-1]_p \\ &= [a]_p \cdot [2a]_p \cdots [(p-1)a]_p \\ &= [a^{p-1} \cdot 1 \cdot 2 \cdots (p-1)]_p. \end{aligned}$$

Thus we have

$$(5.41) \quad 1 \cdot 2 \cdots (p-1) \equiv a^{p-1} \cdot 1 \cdot 2 \cdots (p-1) \pmod{p}.$$

Since $1 \cdot 2 \cdots (p-1)$ is relatively prime to p , we use Theorem 5.27 to cancel $1 \cdot 2 \cdots (p-1)$ from both sides of (5.41) to deduce that

$$1 \equiv a^{p-1} \pmod{p}.$$

□

The hypothesis of Theorem 5.62 requires that a is not a multiple of p . In the following corollary, we rephrase Fermat's Little Theorem in a way that removes this restriction.

Corollary 5.63. *Suppose that p is a prime. Then for every integer a ,*

$$a^p \equiv a \pmod{p}.$$

Proof. If $a \equiv 0 \pmod{p}$, then obviously $a^p \equiv 0 \equiv a \pmod{p}$. Otherwise, due to Theorem 5.62, we have that $a^{p-1} \equiv 1 \pmod{p}$. We now multiply both sides of the congruence by a to get $a^p \equiv a \pmod{p}$. \square

Exercise 5.20. (a) Suppose that p is prime, and a is an integer that is not a multiple of p . Suppose also that $s, t \in \mathbb{N}^*$. Show that

$$a^s \equiv a^t \pmod{p}$$

if $s \equiv t \pmod{p-1}$.

(b) Is the converse true? In other words, does $a^s \equiv a^t \pmod{p}$ imply that $s \equiv t \pmod{p-1}$?

Exercise 5.21. (a) Using the factorization $341 = 11 \cdot 31$, show that

$$2^{341} \equiv 2 \pmod{341}$$

but that

$$3^{341} \not\equiv 3 \pmod{341}.$$

(b) In this problem, we will show that the converse of Fermat's Little Theorem is false. Using the factorization $561 = 3 \cdot 11 \cdot 17$, show that

$$a^{561} \equiv a \pmod{561}$$

for every integer a .

(c) Suppose that p_1, \dots, p_k are distinct primes, and denote their product by $N = p_1 \cdots p_k$. Show that

$$a^N \equiv a \pmod{N}$$

for every integer a if $p_j - 1$ divides $N - 1$ for every index j .

5.7 An Application: RSA Cryptography

In this section, we describe a surprising practical application of the theory we have developed thus far in this chapter. *Cryptography* is a field of applied mathematics and computer science concerned with the communication of

then the message (5.42) would be encrypted as

Csmklr irq siislj gk Axmooqco si dsuk!

While this cipher is more complicated than (5.43), there are still many methods for easily attacking it. For messages longer than a few sentences, there is a devastating “statistical” approach. The most common letter used in English is **e**, followed by **t**, **a** and **i**. If we know that Bob and Alice are communicating in English, we may look at the encrypted message for the most common four letters, and assume that they represent **e**, **t**, **a** and **i**, respectively. This should decode enough of the message that we could guess entire words from the encrypted text, which would reveal more and more of the cipher until the entire message was decrypted. We also expect that any message written in English will contain common words like **the**, **and**, **at** and **that**. Making a few guess and working out the consequences can reveal much of the cipher.

In World War II, the Germans used a (much more sophisticated) encryption method based on a substitution cipher. Their technique was so complicated that it could not be done efficiently by hand. (In war, messages need to be encrypted and decrypted quickly!) They used a mechanical device, called the *Enigma machine*, to encrypt and decrypt their messages. The Germans changed their cipher every day and took precautions to preclude the possibility of using the statistical approach described above (roughly, they used a different substitution cipher for the n th letter of the message, for every n). Even so, their technique still had too many weaknesses, and British mathematicians had remarkable success in decrypting intercepted transmissions from the German navy and air force, which gave the overwhelmed British a much needed advantage in the early years of the war. In particular, codebreakers are credited with providing the British navy with the location of the *Bismark*, a powerful German battleship, which was promptly overwhelmed and sunk by the British navy in May of 1941.

Modern computers have rendered the encryption methods described above obsolete. However, they have also made possible more powerful encryption algorithms which have so far proven all but impossible to crack. We will describe one method, named the *RSA algorithm* after Ron Rivest, Adi Shamir, and Leonard Adleman, who invented it in 1977.

RSA is called a *public-key* encryption algorithm because it allows Alice and Bob to make part of their key public. While the “substitution cipher”

encryption techniques we described above rely on Bob and Alice exchanging secret knowledge of the cipher, RSA does not require Bob and Alice to exchange any information in secret! All of their communications can be published on the internet, and yet Bob can securely transmit secret information to Alice. They simply do not care if anyone intercepts and reads their messages, *even the unencrypted messages which describe their encryption method.*

Suppose that Bob wants to transmit the message (5.42) to Alice using RSA. First, Bob needs to transform his message into an integer. One way (among many) of accomplishing this is to use the American Standard Code for Information Interchange (ASCII), which assigns each character (including punctuation symbols and spaces) a three-digit number. For example, the character L is represented by 076, the character a is assigned 097, u is assigned 117, n is assigned 110, etc. Since Bob's message (5.42) has 38 characters, it would be encoded by a $38 \cdot 3 = 114$ digit integer m beginning with the digits 076097117110...

Here's the method Alice and Bob arrange for Bob to communicate his message m secretly to Alice:

Step 1. Alice randomly picks two very large primes p and q such that $p \neq q$. The larger the primes, the more secure the communication between Bob and Alice will be. Alice multiplies p and q together and calls the result N :

$$N = pq.$$

The integer N is called a **pseudoprime** since it is the product of two distinct primes. In practice, N is very large—usually over 100 digits.

Step 2. Alice randomly picks an integer E such that

$$0 < E < (p-1)(q-1) \quad \text{and} \quad \gcd(E, (p-1)(q-1)) = 1.$$

The integer E that Alice chooses is called the *public exponent*.

Step 3. Alice solves the linear congruence

$$E \cdot k \equiv 1 \pmod{(p-1)(q-1)}.$$

for the unknown integer k . This congruence has a solution due to Corollary 5.24 and the fact that Alice chose E relatively prime to $(p-$

$1)(q-1)$. Alice selects the least nonnegative solution k , which is called the *private key*. In practice, the calculation of k is based on the methods we described in Section 5.2 (which were based in turn on the Euclidean Algorithm), and can be accomplished relatively efficiently on modern computers. The private key k is what will allow Alice to decrypt Bob's encrypted message.

Step 4. Alice reveals the integers N and E in public, possibly publishing them on her personal webpage. She keeps the integers p, q and k secret. She also publishes instructions for how Bob should use N and E to encrypt his message.

Step 5. Following Alice's instructions, Bob takes his message m , and verifies that $0 < m < N$ (if $m \geq N$, then Bob cuts his message m up into smaller messages m_1, m_2, \dots , and sends them one at a time). Bob then calculates the integer

$$R = m^E \pmod{N}.$$

This computation is relatively efficient on modern computers (and is based on the technique we outlined in Example 5.16). The integer R is the *encrypted message*. Bob then publicly reveals R .

Step 6. Alice uses her private key k to decrypt the encrypted message R . According to Theorem 5.64 below (the proof of which relies on Fermat's Little Theorem),

$$m = R^k \pmod{N}.$$

Thus by calculating $R^k \pmod{N}$ (using the method of Example 5.16), Alice has recovered Bob's message m .

Theorem 5.64. *Suppose that p and q are distinct primes, and E is a positive integer relatively prime to $(p-1)(q-1)$ and satisfying $0 < E < (p-1)(q-1)$. Suppose that k is a positive integer satisfying*

$$E \cdot k \equiv 1 \pmod{(p-1)(q-1)}.$$

Set $N = pq$. Then for any integer m satisfying $0 < m < N$,

$$m = (m^E \pmod{N})^k \pmod{N}.$$

Proof. Since $0 < m < N$, by Lemma 5.7 we simply need to verify that

$$m \equiv (m^E \bmod N)^k \pmod{N}.$$

According to Lemma 5.7, $(m^E \bmod N) \equiv m^E \pmod{N}$. Thus the Fundamental Theorem of Modular Arithmetic implies that

$$(m^E \bmod N)^k \equiv (m^E)^k \equiv m^{Ek} \pmod{N}.$$

Thus it suffices to show that

$$(5.45) \quad m \equiv m^{Ek} \pmod{N}.$$

Due to the fact that $Ek \equiv 1 \pmod{(p-1)(q-1)}$, there is an integer w such that

$$Ek = 1 + w(p-1)(q-1).$$

Moreover, $w > 0$ since $E, k > 0$. Therefore, using Fermat's Little Theorem we see that if m is not divisible by p then

$$m^{Ek} \equiv m \cdot m^{w(p-1)(q-1)} \equiv m \cdot (m^{p-1})^{w(q-1)} \equiv m \pmod{p}.$$

If m is divisible by p , then clearly $m \equiv 0 \equiv m^{Ek} \pmod{p}$. Either way, we deduce that

$$m^{Ek} \equiv m \pmod{p}.$$

Arguing in the same way, we conclude that

$$m^{Ek} \equiv m \pmod{q}.$$

Since p and q are distinct primes, they are relatively prime. Applying the Relatively Prime Moduli Law, we see that

$$m^{Ek} \equiv m \pmod{pq}.$$

Recalling that $N = pq$, we have shown (5.45), as desired. \square

Remark 5.65. The key to the success of the RSA algorithm lies in the difficulty in factoring large integers. Alice is able to decrypt Bob's message because she possesses the private key k . She found k by solving the linear congruence

$$E \cdot k \equiv 1 \pmod{(p-1)(q-1)}.$$

However, only Alice can solve this linear congruence because only Alice knows the correct modulus $(p - 1)(q - 1)$ to use. This is because only Alice knows the identity of the primes p and q . Recall that while Alice published the product $N = pq$, she kept the primes p and q secret.

What is to stop an enemy from calculating p and q by factoring N ? In theory, this is actually a valid attack on RSA. If an enemy can factor N , then they will have no trouble decrypting Bob's message. *The effectiveness of RSA lies in the fact that no one, so far, has figured out how to factor large integers very efficiently.* While it is easy to write a simple computer program that factors integers, if N is very large it might take a modern computer *years* to factor N .

Example 5.66. While, in practice, RSA is implemented using very large primes, it is useful to see an example worked out with small primes. For instance, suppose Alice picks the primes

$$p = 13 \text{ and } q = 17.$$

She next calculate the modulus $N = 13 \cdot 17 = 221$. We will pretend, for the sake of easy calculations, that 221 is a large, unfactorable integer. Alice next picks the public exponent $E = 25$, which is relatively prime to $(p - 1)(q - 1) = 12 \cdot 16 = 192$. To calculate the private key k , Alice must solve the linear congruence

$$(5.46) \quad 25k \equiv 1 \pmod{192}.$$

To accomplish this, she carries out the Euclidean Algorithm:

$$\begin{aligned} 1 \cdot 192 + 0 \cdot 25 &= 192 \\ 0 \cdot 192 + 1 \cdot 25 &= 25 \\ 1 \cdot 192 - 7 \cdot 25 &= 17 \\ -1 \cdot 192 + 8 \cdot 25 &= 8 \\ 3 \cdot 192 - 23 \cdot 25 &= 1. \end{aligned}$$

Thus -23 satisfies the congruence (5.46). However, Alice needs a positive solution, so she picks $k = -23 + 192 = 169$ as her private key.

Finally, Alice makes the integers N and E public by publishing the following instructions on the internet:

To send me a message in the form of an integer m , verify that $0 \leq m \leq 220$ and send me the encrypted message $m^{25} \pmod{221}$.

Bob wants to send Alice the integer $m = 173$, so he calculates $R = m^{25} \pmod{221}$ according to the method outlined in Example 5.16:

$$\begin{aligned} 173^1 &\equiv 173 \pmod{221} \\ 173^2 &\equiv 29929 \equiv 94 \pmod{221} \\ 173^4 &\equiv 94^2 \equiv 8836 \equiv -4 \pmod{221} \\ 173^8 &\equiv (-4)^2 \equiv 16 \pmod{221} \\ 173^{16} &\equiv 16^2 \equiv 35 \pmod{221} \end{aligned}$$

and therefore,

$$173^{25} \equiv 173^{16} \cdot 173^8 \cdot 173^1 \equiv 35 \cdot 16 \cdot 173 \equiv 96880 \equiv 82 \pmod{221}.$$

Thus Bob sends Alice the encrypted message

$$R = 173^{25} \pmod{221} = 82.$$

To decrypt Bob's encrypted message R , Alice must compute $R^k \pmod{221}$. This could easily be done the way Bob performed his calculation, by the method of Example 5.16. However, since Alice knows the factorization of N , she has the advantage of being able to use Fermat's Little Theorem and the Chinese Remainder Theorem to ease her calculations (which is nice, since 169 is a rather large exponent). Since $k = 169 \equiv 1 \pmod{12}$ and $R = 82 \equiv 4 \pmod{13}$, it follows that

$$82^{169} \equiv 4^{169} \equiv 4^1 \equiv 4 \pmod{13}.$$

Likewise, since $k = 169 \equiv 9 \pmod{16}$ and $R = 82 \equiv 14 \pmod{17}$, it follows that

$$82^{169} \equiv 14^{169} \equiv 14^9 \pmod{17}.$$

It will take a little bit of work to calculate $14^9 \pmod{17}$:

$$\begin{aligned} 14^1 &\equiv -3 \pmod{17} \\ 14^2 &\equiv (-3)^2 \equiv 9 \pmod{17} \\ 14^4 &\equiv 9^2 \equiv 81 \equiv -4 \pmod{17} \\ 14^8 &\equiv (-4)^2 \equiv 16 \equiv -1 \pmod{17}. \end{aligned}$$

Therefore,

$$14^9 \equiv 14^8 \cdot 14^1 \equiv (-3) \cdot (-1) \equiv 3 \pmod{17}.$$

We (looking over Alice's shoulder) have seen that $x = R^k = 82^{169}$ is a solution of the system

$$(5.47) \quad \begin{cases} x \equiv 4 \pmod{13} \\ x \equiv 3 \pmod{17}. \end{cases}$$

Since solutions of this system are unique modulo $13 \cdot 17 = 221$, Alice will find $R^k \pmod{221}$ by finding another solution of the system. She looks for a solution of the form $x = 3 + 17t$ where t is an integer satisfying the linear congruence

$$17t = 1 \pmod{13}.$$

Using the Euclidean Algorithm (calculation omitted), Alice finds one such solution $t = 10$. Thus $x = 3 + 17t = 173$ solves the system (5.47). Therefore

$$173 \equiv R^k \pmod{221}.$$

Since $0 \leq 173 < 221$, it follows that $173 = R^k \pmod{221}$. Therefore, Alice has decrypted Bob's message: $m = 173$.

Exercise 5.22. Let $p = 23$ and $q = 31$. Carry out the RSA Algorithm by following the steps below.

- (a) Find the smallest possible public exponent E .
- (b) Find the private key k .
- (c) Pretending you are Bob, encrypt the message $m = 347$.
- (d) Pretending you are Alice, decrypt the encrypted message $R = 551$.

Exercise 5.23. Margo wants her friends to be able to transmit information securely to her. So she posts the following instructions on the internet:

Anyone wishing to securely transmit to me their favorite number, say m , may follow the following instructions (as long as m is three-digits or less): publish $m^{109} \pmod{1073}$ on the internet. Only I will be able to figure out what m is.

Shyam, following Margo's instructions, publishes the number 5 on his personal webpage. Unfortunately for Margo and Shyam, their communication is not as secure as Margo had advertised. By breaking Margo's code, find Shyam's favorite three-digit number. Make sure you write in complete sentences and explain your reasoning (although you may suppress nasty arithmetic calculations).