Notes on Discrete Mathematics
Mathematics 256
Calvin College
Fall 2014

Gerard A. Venema

Professor of Mathematics, Calvin College
E-mail address: venema@calvin.edu

Associate Secretary, Mathematical Association of America

Contents

Chapter 1

Logic and Proofs

1.1. Propositional calculus

Definition. A *proposition* is a declarative sentence that is either true or false, but not both.

Note. The word “proposition” is used this way in logic and in computer science. In most other advanced mathematics courses, the word “proposition” is used as another name for “theorem.” This is true in Euclidean geometry, for example. In that case, what we are calling a proposition is instead referred to as a *statement*.

We will use letters like p and q to denote propositions.

Every proposition has a *truth value*: it is either true (T) or false (F).

Definition. Let p be a proposition. The *negation of* p, written $\neg p$, is the statement “It is not the case that p.”

The truth value of $\neg p$ is the opposite of that of p.

Given two propositions p and q, we can form new (compound) propositions from them using the *logical operators* (also called *logical connectives*).

(1) conjunction: $p \wedge q$, p and q

(2) disjunction: $p \vee q$, p or q

(3) exclusive or: $p \oplus q$, p XOR q

(4) conditional statement: $p \to q$, p implies q (or “if p, then q”)

The following *truth table* specifies the meaning of the logical operators.

p	q	$\neg p$	$\neg q$	$p \wedge q$	$p \vee q$	$p \oplus q$	$p \to q$
T	T	F	F	T	T	F	T
T	F	F	T	F	T	T	F
F	T	T	F	F	T	T	T
F	F	T	T	F	F	F	T

In mathematics the word “or”, when used by itself, is always understood in the nonexclusive sense. Hence $p \vee q$ allows the possibility that both p and q are true. The “exclusive or” must be used if we require that exactly one of the two statements is true. This removes the ambiguity that is present in everyday English usage.

In the conditional statement $p \to q$, p is called the *hypothesis* and q is called the *conclusion.*

In this chapter we are studying purely logical relationships. In that context the conditional statement $p \to q$ simply means that q is true provided p is. It should not be interpreted to mean the p somehow causes q to be true. If p happens to be true, then q must also be true (for whatever reason). If p happens not be be true, then nothing is required of q. In other words, if p is false, then $p \to q$ is true regardless of whether q is true or false—see the last two lines of the truth table. In case p is false, we will say that $p \to q$ is *vacuously true.*

There are several ways to make new conditional statements from old ones.

Definition. The *converse* of $p \to q$ is $q \to p$.

Definition. The *contrapositive* of $p \to q$ is $(\neg q) \to (\neg p)$.

The statement $(\neg p) \to (\neg q)$ is called the *inverse* of $p \to q$. It is the contrapositive of the converse. We will not make use of the inverse.

Definition. The proposition $p \leftrightarrow q$ is called a *biconditional.*

$$p \leftrightarrow q \text{ means } (p \to q) \wedge (q \to p).$$

Terminology. The words “necessary” and “sufficient” are also used for conditional statements. If $p \to q$, we say that q *is a necessary condition for* p or that p *is a sufficient condition for* q. In this terminology, $p \leftrightarrow q$ means that p is a necessary and sufficient condition for q. The expression “only if” is often used to mean “implies.” Thus “p only if q” means the same as “if p, then q.” This allows the biconditional $p \leftrightarrow q$ to be written as p if and only if q, which is abbreviated p iff q.

Exercises 1.1

1. Decide whether or not each of the following is a proposition. Give the truth values of those that are propositions.
(a) Do your homework!
(b) How many students are enrolled in Math 256 this semester?
(c) There are 100 students enrolled in Math 256 this semester.
(d) $x + 1 = 5$.
(e) Math is fun.

2. Let p be the proposition "It is raining" and let q be the proposition "I get wet." Express each of the following compound propositions as an English sentence.
(a) $p \vee q$
(b) $p \rightarrow q$
(c) $p \wedge (\neg q)$

3. Let p be the proposition "It is raining" and let q be the proposition "I walk to work." Express each of the following in terms of p, q and logical operators.
(a) It is raining, but I walk to work.
(b) If it is raining, then I do not walk to work.
(c) I walk to work whenever it is not raining.

4. Identify the hypothesis and conclusion of each of the following statements.
(a) If it rains, then I get wet.
(b) If the sun shines, then we go hiking and biking.
(c) If $x > 0$, then there exists a y such that $y^2 = 0$.
(d) If $2x + 1 = 5$, then either $x = 2$ or $x = 3$.

5. State the converse and contrapositive of each of the statements in Exercise 4.

6. Write each of the following statements in "if…, then…" form.
(a) It is necessary to score at least 90% on the test in order to receive an A.
(b) A sufficient condition for passing the test is a score of 50% or higher.
(c) You fail only if you score less than 50%.
(d) You succeed whenever you try hard.

7. State the converse and contrapositive of each of the statements in Exercise 6.

8. Restate each of the following assertions in "if…, then…" form.
(a) Perpendicular lines must intersect.
(b) Any two great circles on a sphere intersect.

(c) Congruent triangles are similar.

9. Identify the hypothesis and conclusion of each of the following statements.

(a) I can take topology if I pass geometry.
(b) I get wet whenever it rains.
(c) A number is divisible by 4 only if it is even.

10. Let p be the proposition "If $2 < 1$, then $3 < 1$." Is p true or false? Explain.

1.2. Equivalence of propositions

Definition. A *compound proposition* is one that is built from simpler propositions using logical operators.

Definition. A *tautology* is a compound proposition that is always true regardless of the truth values of the constituent propositions.

Example. The statement $p \vee (\neg p)$ is a tautology as is $p \to p$.

Definition. Two compound statements p and q are *logically equivalent* if $p \leftrightarrow q$ is a tautology.

Notation. Logical equivalence is denoted by $\equiv$.

Here is an informal, but more useful, statement of the definition of logical equivalence: two compound propositions are logically equivalent if for any values of the constituent propositions they are either both true or both false.

Example. $p \wedge (q \vee (\neg q)) \equiv p$.

In the preceding example, the two compound propositions are made up of different constituent propositions. In most cases of interest, the two will include the same constituent propositions. In that case, logical equivalence can be demonstrated by checking that the truth tables for the two compound propositions are the same. In the next example, observe that columns 3 and 6 of the truth table contain the same values.

Example. The conditional statement $p \to q$ is logically equivalent to its contrapositive $(\neg q) \to (\neg p)$.

p	q	$p \to q$	$\neg q$	$\neg p$	$(\neg q) \to (\neg p)$
T	T	T	F	F	T
T	F	F	T	F	F
F	T	T	F	T	T
F	F	T	T	T	T

Negating conjunction and disjunction. The negation of the assertion that at least one of p and q is true is the assertion that neither of them is true. In other words, the negation of $p \vee q$ is $(\neg p) \wedge (\neg q)$. In the same way, the negation of $p \wedge q$ is $(\neg p) \vee (\neg q)$. We summarize these two observations by saying that "negation interchanges disjunction and conjunction". The rules are formalized in the following laws.

De Morgan's Laws.

(1) $\neg(p \wedge q) \equiv (\neg p) \vee (\neg q)$.

(2) $\neg(p \vee q) \equiv (\neg p) \wedge (\neg q)$.

Here is a truth table that demonstrates De Morgan's first law.

p	q	$p \wedge q$	$\neg(p \wedge q)$	$\neg p$	$\neg q$	$(\neg p) \vee (\neg q)$
T	T	T	F	F	F	F
T	F	F	T	F	T	T
F	T	F	T	T	F	T
F	F	F	T	T	T	T

Negating a conditional statement. The conditional statement $p \to q$ means that q is true whenever p is; thus the negation of $p \to q$ is the assertion that it is possible for p to be true while q is false. *Note that the negation of a conditional statement is not another conditional statement.*

Example. $\neg(p \to q) \equiv p \wedge (\neg q)$.

As usual, this is demonstrated by means of a truth table.

p	q	$p \to q$	$\neg(p \to q)$	p	$\neg q$	$p \wedge (\neg q)$
T	T	T	F	T	F	F
T	F	F	T	T	T	T
F	T	T	F	F	F	F
F	F	T	F	F	T	F

Exercises 1.2

1. Show that $(p \wedge q) \to p$ is a tautology.

2. Is $(p \vee q) \to p$ a tautology? Explain.

3. Construct a truth table that verifies De Morgan's second law.

4. Construct a truth table that verifies the distributive law $p \wedge (q \vee r) \equiv (p \wedge q) \vee (p \wedge r)$.

5. Show that $p \leftrightarrow q \equiv (p \wedge q) \vee (\neg p \wedge \neg q)$.

6. Show that $(p \wedge q) \vee (p \wedge (\neg q)) \equiv p$.

7. Show that $(p \vee q) \to r \equiv (p \to r) \wedge (q \to r)$.

8. Show that $p \to (q \to r) \equiv q \to (\neg p \vee r)$.

9. Write the negation of each of the statements in Exercise 1.1.4.

10. Write the negation of each of the statements in Exercise 1.1.6.

1.3. Predicate logic

Many of the statements we encounter in mathematics involve variables. An assertion such as "$x > 0$" does not, by itself, qualify as a proposition in the technical sense defined in §1.1 because it is neither true nor false until a value has been assigned to the variable x. The statements we will study in this section typically have a *subject*, which is a variable, and a *predicate*, which asserts that the subject has a certain property. For example, in the sentence "*x is greater than 0*", "x" is the subject, and "is greater than 0" is the predicate. For that reason the study of the logical structure of such statements is called *predicate logic* or the *predicate calculus*.

Definition. A *propositional function* is a function that assigns a truth value to a variable; i.e., a propositional function is a function whose range consists of the set $\{\text{True}, \text{False}\}$.

[If you've forgotten what is meant by the range of a function, you can look ahead to the next chapter where the definitions of function, domain, and range that you learned in high school are reviewed.]

Example. The function $P(x) = (x > 0)$ is a propositional function; e.g., $P(5) = \text{True}$, $P(0) = \text{False}$. Notice that $P(x)$ is not a proposition, but $P(0)$ is a proposition.

The domain of a propositional function can usually be determined from the context. For example, the domain of the function P in the preceding example is $\mathbb{R}$, the set of all real numbers. If the domain is not clear from the context, then it should be specified explicitly.

A propositional function can have more than one independent variable; e.g., $Q(x, y)$, $R(x, y, z)$, etc.

Example. $Q(x, y) = (x + y > 0)$ is a propositional function whose domain consists of pairs of real numbers. In this example $Q(-1, 1) = \text{False}$ and $Q(-1, 2) = \text{True}$.

Example. $P(\ell, m) = (\ell \parallel m)$ is a propositional function whose domain consists of ordered pairs of lines (ℓ, m).

Quantifiers. While a propositional function by itself is not a proposition (because $P(x)$ does not have a truth value until x is assigned a value), the statements "$P(x)$ is true for every x" and "there exists an x for which $P(x)$ is true" are propositions. This process of making a proposition out of a propositional function is called *quantification*. There are two quantifiers.

(1) The *universal quantifier*: $\forall x\, P(x)$, which is read "for every x $P(x)$" or "for all x $P(x)$", means that $P(x)$ is true for every value of x. In order for this to make sense there must be a clearly understood *universe of discourse*, which consists of the set of all possible values for x.

(2) The *existential quantifier*: $\exists x\, P(x)$, which is read "there exists an x $P(x)$," means that there is at least one value for x for which the proposition $P(x)$ is true. Again it must be made clear what the possible values of x are.

Observe that $\forall$ generalizes $\wedge$ and $\exists$ generalizes $\vee$. This is reflected in the fact that there are De Morgan's laws for quantifiers that exactly parallel the De Morgan's laws we saw before.

De Morgan's laws for quantifiers.

(1) $\neg(\forall x\, P(x)) \equiv \exists x\, \neg(P(x))$

(2) $\neg(\exists x\, P(x)) \equiv \forall x\, \neg(P(x))$

Propositional functions in conditional statements. Another way in which to make a proposition out of propositional functions is to use them as the hypothesis and conclusion of a conditional statement. For example, *if* $x > 3$, *then* $x^2 > 9$ is a proposition. The statement "if $P(x)$, then $Q(x)$" means that any value of x that makes $P(x)$ true must make $Q(x)$ true as well.

Uniqueness. The existential quantifier is often combined with an assertion about uniqueness. The symbols $\exists! x P(x)$ are read "there exists a unique x such that $P(x)$ is true." This proposition asserts two things: first, that there is an x such that $P(x)$ and, second, that there is only one x such that $P(x)$. The definition is

$$\exists! x\, P(x) = (\exists x\, P(x)) \wedge ((P(x) \wedge P(y)) \rightarrow (x = y))$$

Note that $\exists!$ is *not* a new quantifier, but the conjunction of a quantifier and another statement about uniqueness. The combination has a special notation simply because it occurs so often.

Examples.

(1) Commutative law for addition: $\forall x\, \forall y\ (x + y = y + x)$.

(2) Existence of additive identity: $\exists z\, \forall x\ (z + x = x)$.

(3) Existence of additive inverses: $\forall x\, \exists y\ (x + y = 0)$.

(4) Switch the order of quantifiers in (3): $\exists y\,\forall x\ (x+y=0)$. (???)

(5) Existence of parallels: $\forall \ell\,\forall P\,(P \notin \ell)\,\exists! m\ ((P \in\ m) \wedge (m \parallel \ell))$.

(6) Definition of continuous at x: $\forall \epsilon > 0\ \exists \delta > 0\ ((0 < |x-y| < \delta) \rightarrow (|f(x)-f(y)| < \epsilon))$.

Negation of preceding examples.

(1′) $\exists x\,\exists y\ (x+y \neq y+x)$.

(2′) $\forall z\,\exists x\ (z+x \neq x)$.

(3′) $\exists x\,\forall y\ (x+y \neq 0)$.

(4′) $\forall y\,\exists x\ (x+y \neq 0)$.

Exercises 1.3

1. Let $P(x)$ be the statement $x = x^2$, where x is a real number. Find the truth value of each of the following.
(a) $P(-1)$
(b) $P(0)$
(c) $P(1)$
(d) $P(2)$
(e) $\forall x\, P(x)$
(f) $\exists x\, P(x)$

2. Let $Q(x,y) = (x+y > 0)$, where x and y are real numbers. Find the truth value of each of the following.
(a) $Q(2,-3)$
(b) $\exists y\, Q(2,y)$
(c) $\forall y\, Q(2,y)$
(d) $\exists x\,\exists y\, Q(x,y)$
(e) $\forall x\,\exists y\, Q(x,y)$
(f) $\forall x\,\forall y\, Q(x,y)$

3. Let $P(x) = (x$ is taking a physics course),
$R(x) = (x$ is taking a religion course), and
$E(x) = (x$ is taking an English course),
where x is a member of the Math 256 class. Express each of the following in terms of those three propositional functions, logical operators, and quantifiers.
(a) A student in the class is taking physics, religion, and English.
(b) Every student in the class is taking religion.
(c) No one in the class is taking both physics and English.
(d) Someone in the class is taking both physics and religion.
(e) Every student in the class is taking at least one of physics, religion, and English.

(f) For each of the three subjects (physics, religion, English), there is at least one student in the course who is taking that subject.

4. Let $C(x, y)$ be the statement "student x is enrolled in course y". Write each of the following logical expressions as a plain English sentence.

(a) $C(\text{Nick}, \text{CPSC 108})$.

(b) $\exists x\ C(x, \text{Spanish 101})$.

(c) $\exists y\ (y \neq \text{ Math 256 }) \wedge C(\text{Kaylee}, y))$.

(d) $\exists x\ (C(x, \text{Music 243}) \wedge C(x, \text{Math 243}))$.

(e) $\exists x\, \exists y\, \forall z\ ((x \neq y) \wedge (C(x, z) \rightarrow C(y, z))$.

(f) $\exists x\, \exists y\, \forall z\ ((x \neq y) \wedge (C(x, z) \leftrightarrow C(y, z))$.

5. Let $I(x)$ be the statement "x has an iPad", let $F(x, y)$ be the statement "x and y are friends", where the domain for both x and y is the set of all students at Calvin College. Translate the statement

$$\forall x(I(x) \vee \exists y(I(y) \wedge F(x, y)))$$

into English.

6. Let $F(x, y)$ be statement "x can fool y", where the domain for both variables is the set of all people in the world. Use quantifiers to express each of the following statements:

(a) Everybody can fool Fred.

(b) There is no one who can fool everybody.

(c) Everyone can be fooled by somebody.

(d) Tim can fool exactly two people.

(e) There is exactly one person whom everybody can fool.

(f) No one can fool himself or herself.

7. Let $R(x, y)$ be the statement "$x + y = x - y$." Find the truth values of the following statements.

(a) $R(2, 0)$

(b) $\forall y\, R(1, y)$

(c) $\forall x\, \exists y\, R(x, y)$

(d) $\forall y\, \exists x\, R(x, y)$

(e) $\exists y\, \forall x\, R(x, y)$

8. Rewrite each of the following statements so that negations appear only within predicates.

(a) $\neg(\forall x\, \forall y\, P(x, y))$

(b) $\neg(\forall y\, \forall x\, (P(x, y) \vee Q(x, y)))$

(c) $\neg(\forall x\, (\exists y\, \forall z\, P(x, y, z) \wedge \exists z\, \forall y\, P(x, y, z)))$

1.4. Proofs

This section briefly lays out basic information about how to write proofs. The next section contains a number of examples of proofs that are written in the style described in this section.

Terminology. An *axiom* (or *postulate*) is a statement that is assumed without proof. In the conditional statement "if p, then q", p is called the *hypothesis* (or antecedent or premise) and q is called the *conclusion* (or consequent). A *proof* for a conditional statement is a logical argument which demonstrates that if the hypothesis is true, then the conclusion is true. A *theorem* is a conditional statement that has a proof. A *lemma* is a minor theorem that is a step in the proof of a major theorem. A *corollary* is a theorem that can be simply proved using a given theorem. Sometimes *proposition* is used as a synonym for theorem. Some authors reserve the word "proposition" for a less important theorem.

Theorem statements. As indicated above, a theorem is a conditional statement (that has a proof). But theorems are often stated informally in a way that does not fit the strict "if. . . , then. . . " form of a conditional statement. In that case it is necessary to restate the theorem in "if. . . , then. . . " form before the proof is written.

Forms of proof. A proof of $h \to c$ can take any one of the following forms.

(1) Direct proof. This is the most straightforward method of proof. Begin by assuming that h is true. Argue in a sequence of logical steps that c must also be true.

(2) Proof by contraposition. A direct proof of $(\neg c) \to (\neg h)$. Since the conditional statement and its contrapositive are logically equivalent, this will suffice.

(3) Proof by contradiction, or *reductio ad absurdum* (RAA). The truth table for $h \to c$ shows that $h \to c$ is true except in the case when h is true and c is false. So an indirect way to prove $h \to c$ is to prove that $h \wedge (\neg c)$ is impossible. We show this by demonstrating that $h \wedge (\neg c)$ leads to a contradiction. In symbols, we show that $(h \wedge (\neg c)) \to (r \wedge (\neg r))$ for some r.

(4) Proof by cases. If there are only a finite number of ways in which the hypothesis can be true, we can prove the theorem by considering each of the cases separately and proving that the conclusion holds in each of them. (See Exercise 1.2.7 for a verification of the logic of this proof form.)

Proofs by contraposition and contradiction are both somewhat indirect and are often confused. Be sure to notice that they are logically different; in a proof by contraposition we assume only $\neg c$ while in a proof by contradiction we assume both h and $\neg c$. When viewed in this way, it is clear why proof by contradiction is often the best way to prove a theorem: in both direct proof and proof by contraposition we assume only one thing (h in a direct proof, $\neg c$ in a proof by contraposition), but in a proof by contraction we assume both h and $\neg c$ and thus have more information to work with in the proof. In order to distinguish the two kinds of assumptions in a proof by contradiction we sometimes refer to $\neg c$ as the *RAA hypothesis*.

Later in the course we will learn a fifth proof form, proof by mathematical induction.

Valid reasons. Each step in the proof must be supported by a reason. Here is the list of the acceptable kinds of reasons.

(1) By hypothesis.
(2) By axiom.
(3) By definition.
(4) By a previously proved theorem.
(5) By a previous step in the current proof.
(6) By one of the rules of logic.

How to write proofs.

(1) Begin by restating the theorem in *if . . . then* form.
(2) Put the label **Proof** at the beginning of the proof and use either QED or □ to mark the end of the proof.
(3) Indicate explicitly which proof form you are using and what your assumptions are.
(4) Write the argument in complete sentences, in paragraph form.
(5) Back up every step in the proof with a reason. The reason can either be stated within the sentence or added in parentheses at the end of the sentence.

1.5. Examples of proofs

The example proofs in this section are meant to serve as models of the proofs you should write in this course. You should use the same style for the proofs you write in all your upper-level mathematics courses at Calvin.

Theorems and proofs must be about something. We will practice writing proofs in the context of elementary number theory. This section spells out exactly what we will assume about number theory and provides some sample proofs. The purpose of the sample proofs is to model what your written proofs should look like; in particular, you should include about the same amount of detail.

Numbers. We will work with the following sets of numbers.

$\mathbb{N} = \{1, 2, 3, 4, \dots\}$, the set of *natural numbers.*

$\mathbb{Z} = \{0, \pm 1, \pm 2, \pm 3, \dots\} = \{\cdots - 2, -1, 0, 1, 2, 3, \dots\}$, the set of *integers.*

$\mathbb{Q} = \{p/q \mid p \in \mathbb{Z}, q \in \mathbb{N}\}$, the set of *rational numbers.*

$\mathbb{R}$, the set of real numbers.

A *real number* is a number that can be written as a decimal (possibly infinite). Recall that a real number is rational (i.e., can be written as the quotient of two integers) if and only if it has a decimal expansion that either terminates or repeats. Real numbers that are not rational are *irrational.*

Assumptions. We will assume the basic facts regarding the operations of addition and multiplication. In particular, both operations are associative ($(a+b)+c = a+(b+c)$ and $(ab)c = a(bc)$) and commutative ($a+b = b+a$ and $ab = ba$) and the distributive law holds ($a(b+c) = ab+ac$). There exists an identity for addition (0) and an identity for multiplicaton (1). Every number has an additive inverse (the inverse of a is $-a$) and every nonzero real number has a multiplicative inverse (the inverse of a is $1/a$).

We will assume three special additional facts about integers.

Positivity Axiom. If m and n are natural numbers, then $m + n$ and $m \cdot n$ are also natural numbers.

Closure Axiom. If m and n are integers, then $m + n$ and $m \cdot n$ are also integers.

Definition. An integer n is *odd* if there exists an integer k such that $n = 2k + 1$. An integer n is *even* if there exists an integer k such that $n = 2k$.

Parity Axiom. Every integer is either even or odd; no integer is both even and odd.

Sample proofs.

Theorem 1.5.1. *The square of an odd number is odd.*

Restatement. *If n is odd, then n^2 is odd.*

Proof. This is a direct proof. Let n be odd (hypothesis). By definition of odd, there is an integer k such that $n = 2k + 1$. Thus

$$n^2 = (2k+1)^2 = 4k^2 + 4k + 1 = 2(2k^2 + 2k) + 1.$$

Since $2k^2 + 2k$ is an integer (by the Closure Axiom), this shows that n^2 is odd (by the definition of odd). □

Corollary 1.5.2. *If n^2 is even, then n is even.*

Proof. Let n be an integer such that n^2 is even (hypothesis). By the Parity Axiom this means that n^2 is not odd. Thus the contrapositive of Theorem 1.5.1 shows that n is not odd. Applying the Parity Axiom again yields the desired conclusion. □

Theorem 1.5.3. *The sum of two rational numbers is rational.*

Restatement. *If a and b are rational numbers, then $a + b$ is a rational number.*

Proof. This is a direct proof. Let a and b be rational numbers (hypothesis). By definition of rational number, this means that there are integers p, q, r, and s such that $a = p/q$ and $b = r/s$. Thus

$$a + b = \frac{p}{q} + \frac{r}{s} = \frac{ps + qr}{qs}.$$

Since $ps + qr$ and qs are integers (by the Closure Axiom), this shows that $a + b$ is rational. □

Theorem 1.5.4. *The sum of a rational number and an irrational number is irrational.*

Restatement. *If a is a rational number and b is an irrational number, then $a + b$ is irrational.*

Restatement. *Assume a is rational. If b is irrational, then $a + b$ is irrational.*

Lemma. *If a is a rational number, then $-a$ is rational.*

Proof. This is a direct proof. Let a be a rational number (hypothesis). By definition of rational number, there exist integers p and q such that $a = p/q$. Since the product of two integers is an integer, $-p$ is also an integer. Hence $-a = (-p)/q$ is rational. □

Proof of Theorem 1.5.4. This is a proof by contraposition. We will assume that a is rational and show that if $a + b$ is rational, then b is rational.

Assume a is rational (hypothesis). By the lemma we know that $-a$ is rational. Thus Theorem 1.5.3 shows that $b = (a + b) + (-a)$ is rational. □

Theorem 1.5.5. *The product of a nonzero rational number and an irrational number is irrational*

Restatement. *If x is rational, $x \neq 0$, and y is irrational, then xy is irrational.*

Proof. This is a proof by contradiction. We will assume x is a nonzero rational number (hypothesis), y is an irrational number (hypothesis), and that xy is rational (RAA hypothesis). From these hypotheses we will derive a contradiction.

Since x and xy are assumed rational, the definition of rational number allows us to write $x = p/q$ and $xy = m/n$ for some $p, q, m, n \in \mathbb{Z}$ with $p \neq 0$. Thus

$$y = \left(\frac{1}{x}\right)(xy) = \left(\frac{q}{p}\right)\left(\frac{m}{n}\right) = \frac{qm}{pn}.$$

But this contradicts the assumption that y is irrational, so we must reject the RAA hypothesis and the proof is complete. □

Theorem 1.5.6. *$\sqrt{2}$ is irrational.*

Restatement. *If $x^2 = 2$, then x is irrational.*

Note. This theorem is known as the Theorem of Theaetetus because it is discussed by Socrates and Theaetetus in Plato's dialog *Theaetetus*—see http://www.cut-the-knot.org/proofs/sq_root.shtml.

Proof. This is a proof by contradiction. We will assume x is a number such that $x^2 = 2$ (hypothesis) and also suppose that x is rational (RAA hypothesis). We will show that these combined hypotheses lead to a contradiction.

Since x is rational, x can be written as p/q, where p and q are both integers. Reduce the fraction p/q to lowest terms. Then p and q have no common factor; in particular, p and q are not both even. Since $p^2 = 2q^2$, p^2 must be even (Corollary 1.5.1) and so p is even. Write $p = 2s$. Then $(2s)^2 = 2q^2$, so $q^2 = 2s^2$ and we see that q is even as well (Corollary 1.5.1 again). We have now arrived at a contradiction: we started with p and q not both even and have concluded that they are both even. This contradiction forces us to reject the hypothesis that x is rational and conclude that x is irrational. □

The next proof is logically correct, but is considered to be bad form. The proof is really a direct proof of the contrapositive with the structure of a proof by contradiction built around it.

Theorem 1.5.7. *If n^2 is odd, then n is odd.*

First proof of Theorem 1.5.7. This is a proof by contradiction. Assume that n^2 is odd and also make the RAA hypothesis that n is even. We will see that this hypothesis leads to a contradiction.

Since n is even, there exists an integer k such that $n = 2k$. Thus $n^2 = (2k)^2 = 2(2k^2)$, which is even. This contradicts the assumption that n^2 is odd. Thus we reject the hypothesis that n is even and conclude that n is odd. □

While the proof above is strictly correct, it is considered to be an example of sloppy thinking. It is better to use a simpler, more direct argument whenever possible. The following is a direct proof of the contrapositive; i.e., it shows that if n is not odd, then n^2 is not odd.

Second proof of Theorem 1.5.7. This is a proof by contraposition. Suppose n is not odd. Then n is even by the Parity Axiom. Thus there exists an integer k such that $n = 2k$ (definition of even). Thus $n^2 = (2k)^2 = 2(2k^2)$, which is even. By the Parity Axiom again, this means that n^2 is not odd and the proof is complete. □

Exercises 1.5

1. Prove: The sum of two odd numbers is even.

2. Prove: The sum of an even and an odd number is odd.

3. Prove: The product of two odd numbers is odd.

4. Prove: The product of an even and an odd number is even.

5. Prove: If the product of two numbers is even, then at least one of the two numbers is even.

6. Prove: n is odd if and only if n^2 is odd.

7. Prove or disprove: The product of two rational numbers is a rational number.

8. Prove or disprove: The product of two irrational numbers is an irrational number.

9. Prove or disprove: The reciprocal of an irrational number is an irrational number.

10. Prove: n is even if and only if $7n + 3$ is odd.

11. Prove: Every odd number is the difference between two perfect squares.

12. Prove: There is no positive perfect cube less than 500 that is the sum of two positive perfect cubes.

Chapter 2

Sets, functions, and cardinality

2.1. Sets

A *set* is an unordered collection of objects. The objects in the set are called the *elements* of the set. Write $a \in A$ to indicate that a is an element of A. Two sets are *equal*, written $A = B$, if they contain exactly the same elements. Set A is a *subset* of B, written $A \subseteq B$, if every element of A is an element of B. We say that A is a *proper subset* of B, written $A \subset B$, if $A \subseteq B$ but $A \neq B$.

The subset relationship can also be described in symbols.

$$A \subseteq B \text{ means } (x \in A) \rightarrow (x \in B).$$
$$A = B \text{ means } A \subseteq B \text{ and } B \subseteq A.$$

One important set is the set with no elements in it. This set is called the *empty set* or the *null set* and is written $\emptyset$. Notice that $\emptyset \subseteq S$ for every set S.

Be sure you distinguish between $\in$ and $\subseteq$. (For example, $1 \in \mathbb{Z}$ and $\{1\} \subseteq \mathbb{Z}$, but $\{1\} \notin \mathbb{Z}$.) The distinction between $\subset$ and $\subseteq$ is less important.

Two ways to describe a set.

(1) Use a roster; e.g., $A = \{1, 2, -1, 0, 2, -1, 1, 0\}$.

(2) Use set builder notation; e.g., $A = \{n \in \mathbb{Z} \mid -1 \leq n \leq 2\}$.

Several ways to make new sets from old.

(1) Power set: $\mathscr{P}(A) = \{S \mid S \subseteq A\}$.

(2) Cartesian Product: $A \times B = \{(a, b) \mid a \in A \text{ and } b \in B\}$.

(3) Union: $A \cup B = \{x \mid (x \in A) \vee (x \in B)\}$.

(4) Intersection: $A \cap B = \{x \mid (x \in A) \wedge (x \in B)\}$.

(5) Set difference: $A - B = \{x \mid (x \in A) \wedge (x \notin B)\}$.

(6) Complement: $\overline{A} = U - A = \{x \mid x \notin A\}$. ($U$ is the universal set.)

Two sets A and B are said to be *disjoint* if $A \cap B = \emptyset$.

De Morgan's Laws, v.3.0.

(1) $\overline{A \cup B} = \overline{A} \cap \overline{B}$.

(2) $\overline{A \cap B} = \overline{A} \cup \overline{B}$.

Exercises 2.1

1. Make a roster of the elements of these sets.
(a) $\{x \mid x \text{ is a real number and } x^2 = 9\}$
(b) $\{r \mid r \text{ is a rational number and } r^2 = 2\}$
(c) $\{n \mid n \text{ is a natural number less than } 9\}$

2. Use set builder notation to describe the following sets.
(a) $\{3, 6, 9, 12, 15\}$
(b) $\{-15, -12, -9, -6, -3, 0, 3, 6, 9, 12, 15\}$
(c) $\{f, g, h, i, j, k\}$

3. True or false?
(a) $\emptyset \in \{0\}$
(b) $\emptyset \subseteq \{0\}$
(c) $\emptyset \in \{\emptyset\}$
(d) $\emptyset \in \{\{\emptyset\}\}$
(e) $\emptyset \subseteq \{\{\emptyset\}\}$
(f) $\{\emptyset\} \in \{\emptyset\}$
(g) $\{\emptyset\} \in \{\{\emptyset\}\}$
(h) $\{\emptyset\} \in \{\emptyset, \{\emptyset\}\}$
(i) $\{\emptyset\} \subseteq \{\{\emptyset\}\}$

4. Let $A = \{1, 2, 3, 4, 5\}$ and let $B = \{2, 4, 6\}$. Make a roster of each of the following sets.
(a) $A \cup B$
(b) $A \cap B$
(c) $A - B$
(d) $B - A$
(e) $A \times B$
(f) $\mathscr{P}(B)$

2.2. Functions

Definition. A *function* f consists of two sets A and B, called the *domain* and *range* of f, and a rule that assigns to each element $a \in A$ a unique element $f(a) \in B$.

Write $f : A \to B$ to indicate that f is a function with domain A and range B. We usually say "f is a function from A to B."

In calculus the "rule" in the function is almost always described by an equation or formula of some kind. That need not be the case in other parts of mathematics (such as this course).

Example 2.2.1. Let $A = \{0, 2, 4\}$ and let $B = \{3, 7, 11\}$. We will give several different descriptions of a rule that assigns elements of B to elements of A. All of these ways of describing the rule result in the same assignment of an element of B to a given element of A, so they all define the same function.

(1) One way to describe the rule is to simply list all the function values. Define $f : A \to B$ by $f(0) = 3, f(2) = 3$, and $f(4) = 11$.

(2) A second way is to use a diagram. Define $f : A \to B$ by the following diagram.

$$\begin{tikzcd}
0 \arrow[d, "f"'] & 2 \arrow[dl, "f"] & 4 \arrow[d, "f"'] \\
3 & 7 & 11
\end{tikzcd}$$

(3) A third way is to use a formula. Define $f : A \to B$ by $f(n) = n^2 - 2n + 3$.

(4) The formula $f(k) = 2k^2 - 2k + 3$ works just as well.

Example 2.2.2. If A is any set, the *identity* function $\iota_A : A \to A$ is defined by $\iota_A(a) = a$ for every $a \in A$.

Definition. A function $f : A \to B$ is *one-to-one* if

$$(a_1 \neq a_2) \to (f(a_1) \neq f(a_2)).$$

One-to-one is often shortened to 1-1. It is logically equivalent (contrapositive) to say $f : A \to B$ is one-to-one if $(f(a_1) = f(a_2)) \to (a_1 = a_2)$. This second version of the definition is often easier to work with.

Example 2.2.3. The function $f : \mathbb{N} \to \mathbb{N}$ defined by $f(n) = |n|$ is one-to-one. The function $g : \mathbb{Z} \to \mathbb{Z}$ defined by $g(n) = |n|$ is not one-to-one.

Definition. A function $f : A \to B$ is *onto* if

$$\forall\, b \in B \;\exists\, a \in A \;(f(a) = b).$$

Example 2.2.4. The function $f : \mathbb{N} \to \mathbb{N}$ defined by $f(n) = |n|$ is onto. The function $g : \mathbb{Z} \to \mathbb{Z}$ defined by $g(n) = |n|$ is not onto.

Example 2.2.5. The function of Example 2.2.1 is neither one-to-one nor onto.

Example 2.2.6. If $A \subseteq B$, the function $I : A \to B$ defined by $I(a) = a$ is called the *inclusion* function. An inclusion function is not onto unless $A = B$. By contrast, every identity function is onto.

Definition. A *one-to-one correspondence* is a function that is both one-to-one and onto.

Example 2.2.7. Prove that $f : \mathbb{Z} \to \mathbb{Z}$ defined by $f(n) = 9 - n$ is a one-to-one correspondence.

Proof. First we prove that f is one-to-one. We will use the second version of the definition of one-to-one. Assume n_1 and n_2 are two integers such that $f(n_1) = f(n_2)$ (hypothesis). Then $9 - n_1 = 9 - n_2$, so $-n_1 = -n_2$ and $n_1 = n_2$ (algebra). Thus f is one-to-one.

Next we prove that f is onto. We must show that $\forall m \in \mathbb{Z}\, \exists n \in \mathbb{Z}$ such that $f(n) = m$ (definition). Let $m \in \mathbb{Z}$ (hypothesis). Define $n = 9 - m$. Then $f(n) = 9 - n = 9 - (9 - m) = m$. Therefore f is onto and we have proved that f is a one-to-one correspondence. □

Definition. If $f : A \to B$ and $g : B \to C$, the *composite* function $g \circ f : A \to C$ is defined by $g \circ f(a) = g(f(a))$. We call $g \circ f$ the *composition* of f and g.

Definition. An *inverse* for $f : A \to B$ is a function $g : B \to A$ such that $f \circ g = \iota_B$ and $g \circ f = \iota_A$. A function that has an inverse is called *invertible.*

Theorem 2.2.8. *A function is invertible if and only if it is a one-to-one correspondence. The inverse of an invertible function is unique.*

Exercises 2.2

1. Determine whether $f : \mathbb{Z} \to \mathbb{Z}$ is onto if f is defined by

(a) $f(n) = n - 5$.
(b) $f(n) = 5n - 1$.
(c) $f(n) = n^3$.
(d) $f(n) = n^2 + 5$.

2. Determine whether each of the functions in the preceding exercise is one-to-one.

3. Determine whether $f : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$ defined by $f(m, n) = n + m$ is onto. Is f one-to-one? Explain.

4. Find a function $f : \{1, 3, 5\} \to \{2, 4, 6, 8\}$ that is one-to-one. Is there a function $f : \{1, 3, 5\} \to \{2, 4, 6, 8\}$ that is onto? Explain.

5. Find a function $f : \{1, 3, 5, 7\} \to \{2, 4, 6\}$ that is onto. Is there a function $f : \{1, 3, 5, 7\} \to \{2, 4, 6\}$ that is one-to-one? Explain.

6. Is there a function $f : \{1, 3, 5, 7\} \to \{2, 4, 6, 8\}$ that is onto but not one-to-one? Explain. Is there a function $f : \{1, 3, 5, 7\} \to \{2, 4, 6, 8\}$ that is one-to-one but not onto? Explain.

7. Find functions $f : \mathbb{N} \to \mathbb{N}$ that have the following properties.
(a) f is onto, but not one-to-one.
(b) f is one-to-one, but not onto.
(c) f is neither one-to-one nor onto.
(d) f is a one-to-one correspondence.

8. Prove that $f : \mathbb{Q} \to \mathbb{Q}$ defined by $f(r) = 3r - 7$ is a one-to-one correspondence.

9. Is $f : \mathbb{Z} \to \mathbb{Z}$ defined by $f(n) = 3n - 7$ (same formula as in the preceding exercise) one-to-one? Is it onto? Explain.

10. Prove that $f : (0, 1) \to (1, \infty)$ defined by $f(x) = 1/x$ is a one-to-one correspondence.

2.3. The cardinality of a set

The cardinality of a set is the number of elements in the set. We know how to count the elements in a finite sets, but it is not so obvious how to do that for an infinite set. In this section we will learn to distinguish different sizes of infinite sets. All the results in the section are due to the German mathematician Georg Cantor (1845–1918).

We begin by giving a rigorous definition of cardinality for finite sets.

Definition. The empty set has cardinality 0. A set S is said to have *cardinality* n, where n is a natural number, if there is a one-to-one correspondence $f : S \to \{1, 2, \dots, n\}$.

Definition. A *finite set* is a set that is either empty or has cardinality n for some natural number n. A set that is not finite is said to be *infinite*.

The cardinality of S is denoted either $|S|$ or $\operatorname{card} S$.

Example 2.3.1. $|\emptyset| = 0$. $|\mathscr{P}(\emptyset)| = 1$. $|\{1, 2, 0, 1, 0, 3, 2, 1\}| = 4$.

Example 2.3.2. $\mathbb{N} = \{1, 2, 3, 4, 5, \dots\}$, the set of positive integers, is infinite.

Example 2.3.3. If A and B are finite sets, then $|\mathscr{P}(A)| = 2^{|A|}$ and $|A \times B| = |A| \cdot |B|$.

It was relatively easy to define the cardinality of a finite set, but we can't do the same thing for infinite sets because we do not have words for different infinities. Just saying "$\mathbb{N}$ is infinite" or "$\mathbb{R}$ is infinite" is too vague—we want to distinguish different infinite cardinalities. Instead of trying to give a definition of the cardinality of an arbitrary set, we take an indirect approach and define what it means for two sets to have the same cardinality.

Definition. Two sets A and B *have the same cardinality*, written $|A| = |B|$, if there exists a one-to-one correspondence $f : A \to B$.

Definition. A set is *countably infinite* if it has the same cardinality as $\mathbb{N}$. A set is *countable* if it is either finite or countably infinite. A set is *uncountable* if it is not countable.

Notation. The symbol $\aleph_0$ is used to denote the cardinality of $\mathbb{N}$. Thus the assertion $|S| = \aleph_0$ means that S is countably infinite which, in turn, means that there is a one-to-one correspondence $S \to \mathbb{N}$.

There is a one-to-one correspondence $f : S \to \mathbb{N}$ if and only if the elements of S can be listed in a sequence $f(1)f(2)f(3)f(4)f(5)\ldots$. Hence we can take as an informal definition that a set is countably infinite if the elements of the set can be listed in an infinite sequence. The elements of a sequence can be counted off, one at a time. This is the origin of the term "countable."

Example 2.3.4. The set of even numbers $E = \{2, 4, 6, 8, \ldots\}$ is countably infinite; i.e., $|E| = \aleph_0$. In order to prove this assertion we must find a one-to-one correspondence from $\mathbb{N}$ to E. Define the function $f : \mathbb{N} \to E$ by $f(n) = 2n$. That f is a one-to-one correspondence can be proved rigorously using a proof like that in Example 2.2.7 or can be demonstrated informally by inspection of the following diagram. Each number in the top row is matched with exactly one number in the second row and each number in the second row is matched with exactly one number in the top row.

$$\begin{tikzcd}
1 \arrow[d, <->, "f"'] & 2 \arrow[d, <->] & 3 \arrow[d, <->] & 4 \arrow[d, <->] & 5 \arrow[d, <->] & 6 \arrow[d, <->] & 7 \arrow[d, <->] & 8 \arrow[d, <->] & \cdots \\
2 & 4 & 6 & 8 & 10 & 12 & 14 & 16 & \cdots
\end{tikzcd}$$

Example 2.3.5. The set $B = \{100, 101, 102, 103, \ldots\}$ is countably infinite. Again, in order to demonstrate this we must produce a one-to-one correspondence $g : \mathbb{N} \to B$. Either the formula $g(n) = n + 99$ or the diagram below can be used.

$$\begin{tikzcd}
1 \arrow[d, leftrightarrow, "g"'] & 2 \arrow[d, leftrightarrow] & 3 \arrow[d, leftrightarrow] & 4 \arrow[d, leftrightarrow] & 5 \arrow[d, leftrightarrow] & 6 \arrow[d, leftrightarrow] & 7 \arrow[d, leftrightarrow] & 8 \arrow[d, leftrightarrow] & \cdots \arrow[d, phantom, "\cdots"] \\
100 & 101 & 102 & 103 & 104 & 105 & 106 & 107 & \cdots
\end{tikzcd}$$

Example 2.3.6. The set of all integers, $\mathbb{Z}$, is countably infinite. Once again this is demonstrated by showing the existence of a one-to-one correspondence $h : \mathbb{N} \to \mathbb{Z}$. It is a little tricker than the previous examples because $\mathbb{Z}$ is infinite in two directions. But alternating the positive and negative integers allows us to put the integers in a sequence. The following diagram illustrated the construction of a one-to-one correspondence.

$$\begin{tikzcd}
1 \arrow[d, leftrightarrow, "h"'] & 2 \arrow[d, leftrightarrow] & 3 \arrow[d, leftrightarrow] & 4 \arrow[d, leftrightarrow] & 5 \arrow[d, leftrightarrow] & 6 \arrow[d, leftrightarrow] & 7 \arrow[d, leftrightarrow] & 8 \arrow[d, leftrightarrow] & \cdots \arrow[d, phantom, "\cdots"] \\
0 & 1 & -1 & 2 & -2 & 3 & -3 & 4 & \cdots
\end{tikzcd}$$

We can also define h by means of a formula.

$$h(x) = \begin{cases} \frac{n}{2} & \text{if } n \text{ is even} \\ -\left(\frac{n-1}{2}\right) & \text{if } n \text{ is odd} \end{cases}$$

Theorem 2.3.7 (Cantor)**.** *The set of rational numbers is countable.*

Sketch of the proof. We first show that the set of positive rational numbers, which we denote $\mathbb{Q}_+$, is countable. Arrange the positive rational numbers in a two-dimensional array as indicated in Figure 2.1. The first row contains all positive fractions that have numerator 1, the second row contains all the positive fractions that have numerator 2, and so on. Every positive rational number is included in the table; in fact, every positive rational number is included many times since we have listed all the unreduced fractions.

We want to list the elements of $\mathbb{Q}_+$ in a sequence. If we tried to list them one row at a time or one column at a time, we would never get to the numbers in the next row or column. But Figure 2.1 shows how to follow a zig-zag pattern along the diagonals to list all of the positive rationals in a single sequence. In order to construct a one-to-one correspondence, we list the positive rationals in this order, but we omit any rational that is equal to one that was already listed. The result is that only the fractions that are circled in Figure 2.1 are listed.

A similar proof shows that the set of negative rationals is countable. In order to list the entire set of rationals in a sequence, we first list 0, then

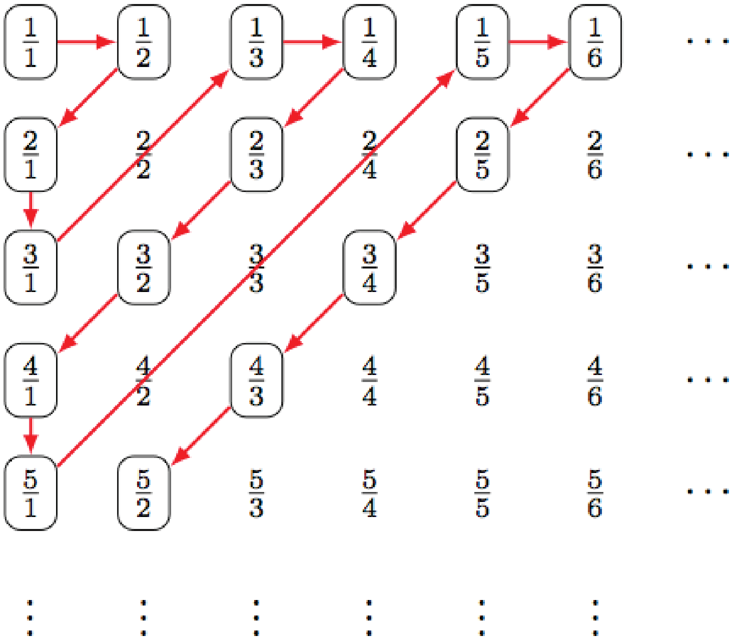

$\frac{1}{1}$ $\frac{1}{2}$ $\frac{1}{3}$ $\frac{1}{4}$ $\frac{1}{5}$ $\frac{1}{6}$ ⋯
$\frac{2}{1}$ $\frac{2}{2}$ $\frac{2}{3}$ $\frac{2}{4}$ $\frac{2}{5}$ $\frac{2}{6}$ ⋯
$\frac{3}{1}$ $\frac{3}{2}$ $\frac{3}{3}$ $\frac{3}{4}$ $\frac{3}{5}$ $\frac{3}{6}$ ⋯
$\frac{4}{1}$ $\frac{4}{2}$ $\frac{4}{3}$ $\frac{4}{4}$ $\frac{4}{5}$ $\frac{4}{6}$ ⋯
$\frac{5}{1}$ $\frac{5}{2}$ $\frac{5}{3}$ $\frac{5}{4}$ $\frac{5}{5}$ $\frac{5}{6}$ ⋯
⋮ ⋮ ⋮ ⋮ ⋮ ⋮

Figure 2.1. A one-to-one correspondence between $\mathbb{N}$ and $\mathbb{Q}_+$

the first positive rational, then the first negative rational, then the second positive rational, then the second negative rational, then the third positive rational, and so on. □

The examples presented so far might give the impression that all infinite sets are countable. But that is not the case. We now prove that there are uncountable sets. In particular, we will show that the set of real numbers is uncountable. We first present the basic argument in the context of sequences of 0's and 1's, where it is cleanest and clearest.

We will use the symbol $\mathscr{S}$ to denote the set of infinite sequences of 0's and 1's. An element of $\mathscr{S}$ is a sequence like $001100110011\cdots$.

Theorem 2.3.8 (Cantor)**.** *The set $\mathscr{S}$ is uncountable.*

Proof. The conclusion is negative; it says that $\mathscr{S}$ is *not* countable. The set is clearly infinite, so this means we must show that there is no one-to-one correspondence $\mathbb{N} \to \mathscr{S}$. Specifically, what we will show is that there is no onto function $\mathbb{N} \to \mathscr{S}$. Here is a restatement of what we intend to prove: *If $f : \mathbb{N} \to \mathscr{S}$ is any function whatsoever, then there exists a sequence $s \in \mathscr{S}$ such that $s \neq f(n)$ for any n.*

Assume $f : \mathbb{N} \to \mathscr{S}$ is a function (hypothesis). For each $n \in \mathbb{N}$, $f(n)$ is a sequence of 0's and 1's. We will denote the kth term in $f(n)$ by b_{nk}. Thus

$$\begin{aligned} f(1) &= b_{11}b_{12}b_{13}b_{14}b_{15}b_{16}b_{17}b_{18}b_{19}\ldots \\ f(2) &= b_{21}b_{22}b_{23}b_{24}b_{25}b_{26}b_{27}b_{28}b_{29}\ldots \end{aligned}$$

etc. Define s to be the sequence $s = c_1c_2c_3c_4c_5\ldots$, where

$$c_n = \begin{cases} 0 & \text{if } b_{nn} = 1, \text{ and} \\ 1 & \text{if } b_{nn} = 0. \end{cases}$$

Then $s \neq f(n)$ for any n because the nth term of s is different from the nth term of $f(n)$. This completes the proof. □

The proof above is known as "Cantor's diagonal argument." The reason is that in order to define s we look at the diagonal entries in the list

$$\begin{aligned} f(1) &= \boxed{b_{11}}b_{12}b_{13}b_{14}b_{15}b_{16}b_{17}b_{18}b_{19}\ldots \\ f(2) &= b_{21}\boxed{b_{22}}b_{23}b_{24}b_{25}b_{26}b_{27}b_{28}b_{29}\ldots \\ f(3) &= b_{31}b_{32}\boxed{b_{33}}b_{34}b_{35}b_{36}b_{37}b_{38}b_{39}\ldots \\ f(4) &= b_{41}b_{42}b_{43}\boxed{b_{44}}b_{45}b_{46}b_{47}b_{48}b_{49}\ldots \\ f(5) &= b_{51}b_{52}b_{53}b_{54}\boxed{b_{55}}b_{56}b_{57}b_{58}b_{59}\ldots \\ f(6) &= b_{61}b_{62}b_{63}b_{64}b_{65}\boxed{b_{66}}b_{67}b_{68}b_{69}\ldots \\ f(7) &= b_{71}b_{72}b_{73}b_{74}b_{75}b_{76}\boxed{b_{77}}b_{78}b_{79}\ldots \\ f(8) &= b_{81}b_{82}b_{83}b_{84}b_{85}b_{86}b_{87}\boxed{b_{88}}b_{89}\ldots \\ f(9) &= b_{91}b_{92}b_{93}b_{94}b_{95}b_{96}b_{97}b_{98}\boxed{b_{99}}\ldots \\ &\vdots \end{aligned}$$

The nth term in s is determined by taking the nth term on the diagonal and changing it from a 1 to a 0 or from a 0 to a 1. This seemingly simple idea has proved to have a large number of deep and poweful applications.

Corollary 2.3.9 (Cantor's Power Set Theorem)**.** $\mathscr{P}(\mathbb{N})$ *is uncountable.*

Proof. We will contruct a one-to-one correspondence $f : \mathscr{P}(\mathbb{N}) \to \mathscr{S}$. Let $A \in \mathscr{P}(\mathbb{N})$. Define a sequence $s = b_1b_2b_3b_4b_5b_6\cdots \in \mathscr{S}$ by

$$b_n = \begin{cases} 1 & \text{if } n \in A, \text{ and} \\ 0 & \text{if } n \notin A \end{cases}$$

and define $f(A) = s$. The verification that this defines a one-to-one correspondence is left as an exercise. □

Corollary 2.3.10 (Cantor)**.** $\mathbb{R}$ *is uncountable.*

Sketch of proof. Any real number can be represented in base two as a sequence of 0's and 1's. Since there are uncountably many sequences, there are also uncountably many real numbers. □

The last proof is rather sketchy. It can be made precise and rigorous, but the details are a bit complicated. One complication results from the fact that base-two representations of real numbers are not unique. The same is true for decimal representations. For example

$$1.000000000\cdots = 0.999999999\cdots .$$

One question Cantor was not able to answer was whether there are subsets of the real numbers that are uncountable but whose cardinality is different from that of $\mathbb{R}$.

Definition. Say that $|A| < |B|$ if there exists a one-to-one function $f : A \to B$ but there is no one-to-one and onto function from A to B.

In this terminology, Corollary 2.3.9 asserts that $|\mathbb{N}| < |\mathscr{P}(\mathbb{N})|$.

Continuum Hypothesis. There is no set S such that

$$|\mathbb{N}| < |S| < |\mathscr{P}(\mathbb{N})|.$$

Cantor was never able to prove his Continuum Hypothesis (CH). In 1940, Kurt Gödel proved that it is consistent with the axioms of set theory for CH to be true. In 1963 Paul Cohen proved that it is consistent with the axioms of set theory for CH to be false. Thus it is not possible to decide whether CH is true of false using the axioms of set theory; CH is independent of the axioms. Kurt Gödel had proved in 1931 that such undecidable statements must always exist in any axiomatic system, but it was still surprising that such a seemingly simple statement would turn out to be undecidable.

The smallest infinite cardinal is named $\aleph_0$ and $\aleph_1$ is the name for the next largest cardinal number. The cardinal number of the real numbers is c (for continuum). Another way to state the Continuum Hypothesis is to assert that $c = \aleph_1$.

The following generalization of the Power Set Theorem shows that there is no limit to how large a cardinal number can be. There are infinitely many different sizes of infinity and there is no largest infinity!

Theorem 2.3.11 (Generalized Power Set Theorem)**.** *If S is any set, then $|S| < |\mathscr{P}(S)|$.*

Proof. The function $g : S \to \mathscr{P}(S)$ defined by $g(a) = \{a\}$ is one-to-one. Let $f : S \to \mathscr{P}(S)$ be any function. We will use a proof by contradiction to show that f is not onto. Assume f is onto (RAA hypothesis). Define $A = \{x \in S \mid x \notin f(x)\}$. Since f is onto, there exists $a \in S$ such that $f(a) = A$. If $a \in A$, then $a \notin A$ (definition of A). If $a \notin A$, then $a \notin f(a)$ (because $A = f(a)$), so $a \in A$ (definition of A). Since either $a \in A$ or $a \notin A$, we have a contradiction. □

Generalized Continuum Hypothesis. If A is any set, then there is no set S such that $|A| < |S| < |\mathscr{P}(A)|$.

Exercises 2.3

1. Let $S = \{0, 1, 2\}$. What is the cardinality of $\mathscr{P}(\mathscr{P}(S))$?

2. What is the cardinality of the set of strings of 0's and 1's that have length exactly n? What is the cardinality of the set of strings of 0's and 1's that have length less than or equal to n?

3. Determine whether each of the following sets is countable or uncountable. For those that are countable, describe a specific one-to-one correspondence between the set and $\mathbb{N}$.

(a) The negative integers.
(b) The integers that are multiples of 10.
(c) The integers larger than 1,000,000.
(d) The real numbers between 0 and 1.
(e) The rational numbers between 0 and 1.
(f) $\mathbb{N} \times \mathbb{N}$.
(g) The set of all finite strings of 0's and 1's.

4. The Hilbert Hotel has $\aleph_0$ rooms. On a certain night all of the rooms are occupied, but David Hilbert, the manager, does not turn on the "No Vacancy" sign.

(a) Suppose a new guest arrives. How can Mr. Hilbert move his current guests around to make room for the new arrival?
(b) Suppose $\aleph_0$ new guests arrive. How can Mr. Hilbert move his current guests around to make room for all the new arrivals?
(c) Suppose $\aleph_0$ buses arrive, each carrying $\aleph_0$ people who want rooms. Can Mr. Hilbert make room for all of them? Explain.

5. Let $B = \{8, 16, 32, 64, 128, 256, 512, \dots\}$. Find a formula for a specific one-to-one correspondence $\mathbb{N} \to B$. Give a rigorous proof (like that in Example 2.2.7) that the function you have found is a one-to-one correspondence.

6. Find a systematic way to list all the rational numbers between 0 and 1 in a sequence.

7. Prove that the function $f : \mathscr{P}(\mathbb{N}) \to \mathscr{S}$. defined in the proof of Corollary 2.3.9 is a one-to-one correspondence.

8. **[Optional Extra Credit Challenge]** Find a one-to-one correspondence from the half open interval $[0, 1)$ to the closed interval $[0, 1]$. The two intervals are intervals of real numbers; for example,

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

Chapter 3

Topics in Elementary Number Theory

In mathematics "number theory" refers to the arithmetic of integers.

3.1. Division of integers

We begin by revisiting a topic you studied in middle school: division of whole numbers.

The Division Algorithm. If $a \in \mathbb{Z}$ and $d \in \mathbb{N}$, then there exist unique integers q and r such that $a = dq + r$ and $0 \leq r < d$.

Terminology. In the Division Algorithm, d is the *divisor*, q is the *quotient*, and r is the *remainder*.

Notation. $q = a \textbf{ div } d$ and $r = a \textbf{ mod } d$.

Example. $105 \textbf{ mod } 9 = 6$. $-105 \textbf{ mod } 9 = 3$.

Definition. Let $a, b \in \mathbb{Z}$ with $a \neq 0$. We say that a *divides* b, written $a \mid b$, provided there exists $c \in \mathbb{Z}$ such that $b = ac$. If $a \mid b$, then a is a *factor* of b, or a is a *divisor* of b, or b is a *multiple* of a.

Observation. $b \mid a$ if and only if the remainder when a is divided by b is 0. If you want to determine whether $b \mid a$, you should divide a by b and check whether the remainder is 0.

Examples. $7 \mid 42$. $42 \nmid 7$. For every $a \in \mathbb{Z}$, $1 \mid a$ and $a \mid 0$.

Theorem 3.1.1 (Properties of "divides."). *Let a, b, and c be integers.*

(*i*) *If* $a \mid b$ *and* $b \mid c$*, then* $a \mid c$*.*

(*ii*) *If* $a \mid b$ *and* $a \mid c$*, then* $a \mid mb + nc$ *for every* $m, n \in \mathbb{Z}$*.*

Proof. Assume, first, that $a \mid b$ and $b \mid c$ (hypothesis). Then there exist $d, e \in \mathbb{Z}$ such that $b = ad$ and $c = be$ (definition). Therefore $c = be = (ad)e = a(de)$, so $a \mid c$. This proves (*i*).

Now assume $a \mid b$ and $a \mid c$ (hypothesis). Then there exist $d, e \in \mathbb{Z}$ such that $b = ad$ and $c = ae$ (definition). Therefore $mb+nc = mad+nae = (md+ne)a$, so $a \mid mb + nc$. This proves (*ii*). □

Definition. Let $a, b \in \mathbb{Z}$. A number d is a *common divisor* of a and b if $d \mid a$ and $d \mid b$.

Definition. Assume $a, b \in \mathbb{Z}$ are not both zero. The *greatest common divisor* of a and b, denoted $\gcd(a, b)$, is the largest positive integer d such that $d \mid a$ and $d \mid b$.

The middle school algorithm. To find $\gcd(a, b)$.

Step 1: Find all the prime factors of a and all the prime factors of b.

Step 2: For each prime factor that appears in both decompositions, select the smaller exponent.

Step 3: $\gcd(a, b)$ is the product of the primes to the powers selected in Step 2.

Example. Find $\gcd(360, 378)$. First write $360 = 2^3 \cdot 3^2 \cdot 5^1$ and $378 = 2^1 \cdot 3^3 \cdot 7^1$. Then $\gcd(360, 378) = 2^1 \cdot 3^2 = 18$.

The middle school algorithm is not practical for use with large numbers, but the following algorithm is.

The Euclidean algorithm. To find the greatest common divisor of two nonnegative integers a and b, not both 0.

Step 0: Relabel, if necessary, so that $a \geq b$.

Step 1: If $b = 0$, then $\gcd(a, b) = a$.

Step 2: If $b \neq 0$, divide b into a to write $a = bq + r$ where $0 \leq r < b$ (the Division Algorithm).

Step 3: By the lemma below, $\gcd(a, b) = \gcd(b, r)$. Since $r < b \leq a$, we have reduced the problem to that of finding the greatest common divisor of strictly smaller numbers.

Step 4: Continue until $r = 0$.

Step 5: $\gcd(a, b)$ is the last nonzero remainder.

This algorithm is fast and efficient, even for very large numbers. In Chapter 4 we will prove that the number of operations required is $O(\log b)$, where b is the smaller of the two numbers. Specifically, we will prove that the number of operations required is 5 times the number of decimal digits in b.

Example. Find $\gcd(378, 360)$. Divide 360 into 378 to get $378 = 360 \cdot 1 + 18$. Divide 18 into 360 to get $360 = 18 \cdot 20 + 0$. Then $\gcd(378, 360) = \gcd(360, 18) = \gcd(18, 0) = 18$.

The Euclidean Algorithm depends on the following simple lemma.

Lemma. *If $a = bq + r$, then* $\gcd(a, b) = \gcd(b, r)$*.*

Proof. Suppose d is a common divisor of b and r. Then d divides a by Theorem 3.1.1(ii). Thus d divides both a and b. Now suppose d' is a common divisor of a and b. We can write $r = a - bq$, so d' divides r by Theorem 3.1.1(ii) again. Thus d' divides both b and r. This means that (a, b) and (b, r) share the same set of common divisors. As a result they have the same greatest common divisor. □

In future applications of the greatest common divisor we will need to know that $\gcd(a, b)$ can be written as a combination of a and b as described in the following theorem.

Theorem 3.1.2. *If $a, b \in \mathbb{N}$, then there exist $s, t \in \mathbb{Z}$ such that* $\gcd(a, b) = sa + tb$*. Furthermore,* $\gcd(a, b)$ *is the smallest positive integer that can be written as such a combination.*

Proof. The fact that $\gcd(a, b)$ can be written this way follows from the fact that there is an algorithm (The Extended Euclidean Algorithm, below) that can be used to determine s and t. Let c be the smallest positive integer that such that $c = sa + tb$. Then any common divisor of a and b also divides c by Theorem 3.1.1(ii). Thus any c that can be written in the form $sa + tb$ is a multiple of $\gcd(a, b)$, which makes $\gcd(a, b)$ the smallest possible. □

Extended Euclidean Algorithm. To find s and t such that $\gcd(a, b) = sa + tb$.

Step 1: Apply the Euclidean algorithm to a and b, writing down the resulting equations in order.

Step 2: "Solve" each of the equations for the remainder.

Step 3: The gcd is the last nonzero remainder; start with the equation containing it and work back towards the first equation, replacing each remainder with its equivalent in the preceding equation.

Example. Express $\gcd(456, 111)$ as a combination of 456 and 111.

$$\begin{array}{lcl} 456 = 4 \cdot 111 + 12 & \text{so} & 12 = 1 \cdot 456 - 4 \cdot 111 \\ 111 = 9 \cdot 12 + 3 & \text{so} & 3 = 1 \cdot 111 - 9 \cdot 12 \\ \;12 = 4 \cdot 3 + 0. & & \end{array}$$

Since 3 is the last nonzero remainder, it is the gcd. Use the last equation with a nonzero remainder to express 3 as a combination of 11 and 12, and then work up through the previous equations to express 3 as a combination of progressively large remainders.

$$\begin{aligned} 3 &= 111 - 9 \cdot 12 \\ &= 111 - 9 \cdot (456 - 4 \cdot 111) \\ &= 1 \cdot 111 - 9 \cdot 456 + 36 \cdot 111 \\ &= 37 \cdot 111 - 9 \cdot 456. \end{aligned}$$

We will see that the Extended Euclidean Algorithm is a useful tool in many calculations. But it also has important theoretical consequences. The following corollary to Theorem 3.1.2 is an example of one of those consequences.

Corollary 3.1.3. *If* $\gcd(a, b) = 1$ *and* $a \mid bc$*, then* $a \mid c$*.*

Proof. This is a direct proof. Assume $\gcd(a, b) = 1$ and $a \mid bc$ (hypotheses). The Extended Euclidean Algorithm allows us to write $1 = sa + tb$ and the definition of "divides" implies that there exists $d \in \mathbb{Z}$ such that $bc = ad$. Thus

$$c = sac + tbc = sac + tad = a(sc + td),$$

which shows that $a \mid c$. □

Exercises 3.1

1. True or False?

(a) $17 \mid 68$
(b) $68 \mid 17$
(c) $17 \mid 1$
(d) $1 \mid 17$
(e) $17 \mid 0$
(f) $0 \mid 17$

2. Evaluate.

(a) $158 \textbf{ mod } 7$.
(b) $-97 \textbf{ mod } 11$.
(c) $155 \textbf{ mod } 19$.
(d) $-221 \textbf{ mod } 23$.

3. Which of the following statements is true?
- If $3 \mid mn$, then either $3 \mid m$ or $3 \mid n$.
- If $6 \mid mn$, then either $6 \mid m$ or $6 \mid n$.

What explains the difference?

4. Prove: If $a \in \mathbb{Z}$, then $a \mid 0$.

5. For each of the following pairs (a, b), use the Euclidean Algorithm to find $\gcd(a, b)$ and then use the Extended Euclidean Algorithm to find $s, t \in \mathbb{Z}$ such that $\gcd(a, b) = sa + tb$.

(a) $(252, 356)$.
(b) $(144, 89)$.
(c) $(314, 159)$.
(d) $(1001, 100001)$.
(e) $(4144, 7696)$.

3.2. Modular arithmetic

Modular arithmetic is the arithmetic of remainders. In this context, the number we are dividing by is called the *modulus* and is usually denoted by m. There are two ways in which to define equivalence modulo m: we can either say that two numbers are equivalent modulo m if their difference is divisible by m or we can say that they are equivalent modulo m if they give the same remainder when divided by m. We make the first definition the official definition and then prove that the second is equivalent.

Definition. Say that *a is congruent to b modulo m*, written $a \equiv b \pmod{m}$, if $m \mid b - a$.

Theorem 3.2.1. *$a \equiv b \pmod{m}$ if and only if a **mod** $m = b$ **mod** m.*

Proof. Assume, first, that $a \equiv b \pmod{m}$ (hypothesis). Then $m \mid b - a$ (definition). Use the Division Algorithm to write $a = mq_1 + r_1$ and $b = mq_2 + r_2$, where $0 \le r_1 < m$ and $0 \le r_2 < m$. Then $b - a = m(q_2 - q_1) + (r_2 - r_1)$, so $m \mid r_2 - r_1$ (Theorem 3.1.1, Part (*ii*)). But the fact that $0 \le r_1 < m$ and $0 \le r_2 < m$ means that $|r_2 - r_1| < m$ (algebra), so we can conclude $r_2 - r_1 = 0$. Since $r_1 = a$ **mod** m and $r_2 = b$ **mod** m, this completes the first half of the proof.

Now assume a **mod** $m = b$ **mod** m. Then $a = mq_1 + r$ and $b = mq_2 + r$ (same r), so $b - a = m(q_2 - q_1)$ and we see that $m \mid b - a$. □

Theorem 3.2.2. *If $a \equiv b \pmod{m}$ and $c \equiv d \pmod{m}$, then*

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

Theorem 3.2.2 justifies doing arithmetic modulo m.

Proof. Assume $m \mid b-a$ and $m \mid d-c$ (hypothesis). We must show that $m \mid (b+d)-(a+c)$ and $m \mid bd-ac$. Since $(b+d)-(a+c)=(b-a)+(d-c)$, the fact that $m \mid (b+d)-(a+c)$ follows directly from Theorem 3.1.1 Part (*ii*). Now $bd-ac = bd-ad+ad-ac = d(b-a)+a(d-c)$, so $m \mid bd-ac$ by the same theorem. □

Definition. A *linear congruence* is a congruence of the form $ax \equiv b \pmod{m}$. The constants a, b, and m are given; the problem is to find all integers x for which the equivalence holds.

Theorem 3.2.3. *The linear congruence $ax \equiv b \pmod{m}$ has a solution if and only if $\gcd(a,m)$ divides b. The number of solutions that are different modulo m is exactly $\gcd(a,m)$.*

Corollary 3.2.4. *The congruence $ax \equiv b \pmod{m}$ has a unique solution for every b if and only if $\gcd(a,m)=1$.*

Note. In the statement above, "unique" means "unique modulo m." In other words, there is exactly one solution x that satisfies $0 \le x < m$.

Algorithm. To solve the linear congruence $ax \equiv b \pmod{m}$.

Step 0: Find $d=\gcd(a,m)$. If $d \mid b$, then define $a'=a/d, b'=b/d$, and $m'=m/d$. It will then be the case that $\gcd(a',m')=1$.

Step 1: Use the Extended Euclidean Algorithm to find $s,t \in \mathbb{Z}$ such that $1 = sa'+tm'$. Observe that s is a solution to the congruence $a's \equiv 1 \pmod{m}$.

Step 2: Multiply by b' to obtain a preliminary solution $z=b's$ to the congruence $a'x \equiv b' \pmod{m'}$.

Step 3: Reduce modulo m' to find the smallest nonnegative solution $x = z \ \mathbf{mod}\ m'$.

Step 4: Add multiples of m' to x to find the other solutions.

Example 3.2.5. Consider $23x \equiv 5 \pmod{120}$. Since $\gcd(23,120)=1$, this problem has a unique solution. Use the Extended Euclidean Algorithm to write $1 = 47\cdot 23 - 9\cdot 120$. From that equation we see that $23\cdot 47 \equiv 1 \pmod{120}$. Multiply both sides by 47 to obtain $23\cdot 47\cdot 5 \equiv 1\cdot 5 \pmod{120}$. Thus $x = 47\cdot 5 = 235$ is a solution. Reduce modulo 120 to find $x=115$.

Definition. An integer c such that $ac \equiv 1 \pmod{m}$ is called an *inverse* of a modulo m.

Example 3.2.6. Another way to arrive at the solution in Example 3.2.5 is to observe that the equation $1 = 47\cdot 23 - 9\cdot 120$ shows that $47\cdot 23 \equiv 1 \pmod{120}$, so 47 is the inverse of 23 modulo 120. Thus we can multiply both sides of the original congruence by 47 to get $x \equiv 235 \pmod{120}$.

Example 3.2.7. Consider $36x \equiv 5 \pmod{120}$. Since $\gcd(36, 120) = 12$ and 12 does not divide 5, this congruence has no solutions.

Example 3.2.8. Consider $36x \equiv 24 \pmod{120}$. Since 24 is divisible by $12 = \gcd(36, 120)$, this congruence has solutions. (In fact it has 12 different solutions modulo 120.) Divide through by 12 to get the related congruence $3x \equiv 2 \pmod{10}$. Since $1 = 10 - 3 \cdot 3$, $x = -3$ is an inverse for 3 modulo 10. Thus $x = -6$ is a solution to $3x \equiv 2 \pmod{10}$. Add 10 to give $x = 4$. This is one solution to the original problem. The others are obtained by adding multiples of 10. So the complete solution set (modulo 120) is $\{4, 14, 24, 34, 44, 54, 64, 74, 84, 94, 104, 114\}$.

The kind of problem illustrated in the next example leads to a system of linear congruences

Example 3.2.9. In the first century A.D., the Chinese mathematician Sun-Tsu posed the following problem: There are certain things whose number is unknown. When divided by 3, the remainder is 2; when divided by 5, the remainder is 3; when divided by 7, the remainder is 2. What is the number of things?

Chinese Remainder Theorem. If $m_1, m_2, \ldots, m_n$ are pairwise relatively prime and $a_1, a_2, \ldots, a_n$ are arbitrary, then the system

$$\begin{array}{rcl} x & \equiv & a_1 \pmod{m_1} \\ x & \equiv & a_2 \pmod{m_2} \\ & \vdots & \\ x & \equiv & a_n \pmod{m_n} \end{array}$$

has a unique solution modulo $m = m_1 m_2 \ldots m_n$.

Algorithm. To solve a system of linear congruences.

- **Step 1:** Define $M_k = m/m_k$ for each k. Observe that the hypotheses of the Chinese Remainder Theorem imply that $\gcd(M_k, m_k) = 1$.
- **Step 2:** For each k, find a solution y_k to $M_k y_k \equiv 1 \pmod{m_k}$.
- **Step 3:** A solution to the system is $x = a_1 M_1 y_1 + a_2 M_2 y_2 + \cdots + a_n M_n y_n$.
- **Step 4:** Reduce modulo m to find the “best” solution.

Example 3.2.10. The problem in Example 3.2.9 is to solve the following system of congruences:

$$\begin{array}{rcl} x & \equiv & 2 \pmod{3} \\ x & \equiv & 3 \pmod{5} \\ x & \equiv & 2 \pmod{7} \end{array}$$

Thus $m = 105, M_1 = 35, M_2 = 2$, and $M_3 = 15$. Solve each $M_k y_k \equiv 1 \pmod{m_k}$ to obtain $y_1 = 2$, $y_2 = 1$, and $y_3 = 1$. Thus the solution given by Step 3 of the algorithm is $x = 233$. This answer should be reduced modulo 105, which gives $x = 23$.

The final theorem in our study of modular arithmetic is an important ingredient in the public key cryptography that will be studied later in the chapter. It uses the definition of prime number, so we state that first.

Definition. A *prime number* is an integer $p > 1$ whose only positive divisors are p and 1. A *composite number* is an integer n such that $n > 1$ and n is not prime. Two integers a and b are *relatively prime* if $\gcd(a, b) = 1$.

Fermat's Little Theorem. If p is prime and a is a positive integer, then

$$a^p \equiv a \pmod{p}.$$

Corollary. *If p is prime and p does not divide a, then*

$$a^{p-1} \equiv 1 \pmod{p}.$$

We will prove Fermat's Little Theorem in the next chapter when we cover mathematical induction. If $\gcd(a, p) = 1$, then a has an inverse modulo p; multiply both sides of the congruence in Fermat's Little Theorem to obtain the corollary.

Exercises 3.2

1. Find all solutions to the following linear congruences.
(a) $4x \equiv 5 \pmod{8}$.
(b) $2x \equiv 7 \pmod{17}$.
(c) $8x \equiv 11 \pmod{35}$.
(d) $6x \equiv 9 \pmod{75}$.
(e) $7x \equiv 13 \pmod{100}$.
(f) $49x \equiv 4000 \pmod{999}$.
(g) $50x \equiv 65 \pmod{105}$.

2. Verify that 937 is the inverse of 13 modulo 2436.

3. Find the inverse of 7 modulo 26.

4. Find the inverse of 4 modulo 9.

5. Which of the numbers $1, 2, 3, \ldots, 10$ has an inverse modulo 11?

6. Which of the numbers $1, 2, 3, \ldots, 11$ has an inverse modulo 12? Find the inverse in case it exists.

7. Prove: If n is an odd positive integer, then $n^2 \equiv 1 \pmod{8}$.

8. Use Fermat's Little Theorem to compute 3^{302} **mod** 5, 3^{302} **mod** 7, and 5^{203} **mod** 7.

3.3. Prime numbers

The following fundamentally important theorem will be proved in the next chapter. We will assume it without proof for now.

The Fundamental Theorem of Arithmetic. Every natural number greater than 1 either is prime or can be written as a product of prime numbers. The prime factorization is unique, except for the order of the factors.

Corollary. *Every natural number greater than 1 is divisible by at least one prime number.*

The corollary is used to prove a classic theorem of Euclid.[1]

Euclid's Theorem. There are infinitely many prime numbers.

This is an example of a theorem that is stated in such a way that it does not appear to have any hypotheses. We will follow Euclid and prove the theorem by showing that the set of prime numbers is *potentially infinite* in the sense that no finite list of numbers can exhaust the set of primes. When viewed this way, the theorem can be restated in if-then form.

Restatement. *If* $\{p_1, p_2, \ldots, p_n\}$ *is any finite set of prime numbers, then there is a prime number* q *such that* $q \neq p_i$ *for any* i*.*

Proof. Let $\{p_1, p_2, \ldots, p_n\}$ be a finite set of prime numbers. Define

$$M = (p_1 p_2 \cdots p_n) + 1.$$

By the corollary to the Fundamental Theorem of Arithmetic, there is a prime number q such that $q \mid M$. None of the primes $p_1, p_2, \ldots, p_n$ divides M (because dividing M by any p_i leaves a remainder of 1), so $q \neq p_i$ for any i. □

Remark. The number M in the preceding proof may be a prime number, but it is not necessarily prime. For example,

$$M = (2 \cdot 3 \cdot 5 \cdot 7 \cdot 11) + 1 = 2311$$

is prime while

$$K = (2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdot 13) + 1 = 30031 = 59 \cdot 509$$

is composite. □

[1] Here is what Euclid actually wrote: *Prime numbers are more than any assigned multitude of prime numbers.* (Proposition 20 in Book IX of Euclid's *Elements*.)

Euclid's theorem simply tells us that there is no largest prime number. We can say quite a bit more about how, on the average, the primes are distributed among the integers. The following theorem was first proved by the French mathematician Jacques Hadamard (1865 – 1963).

Definition. For each $x \in \mathbb{R}$, define $\pi(x)$ to be the number of primes less than or equal to x. The function π is called the *prime-counting function.*

The Prime Number Theorem. The ratio of $\pi(x)$ to $x/\ln x$ approaches 1 as x grows large; i.e.,

$$\lim_{x\to\infty} \frac{\pi(x)}{x/\ln x} = 1.$$

Even though we know a lot about how many primes there are, we do not have any formula that explicitly generates a list of infinitely many primes. Two attempts in that direction are due to two other French mathematicians, Pierre de Fermat (1601 – 1665) and Marin Mersenne (1588 – 1648).

Fermat primes. A *Fermat prime* is a prime number of the form $F_n = 2^{(2^n)} + 1$ for some $n \geq 0$. The only known Fermat primes are the ones Fermat himself found: F_0, F_1, F_2, F_3, and F_4. □

Mersenne primes. The nth *Mersenne number* is $M_n = 2^n - 1$. A *Mersenne prime* is a Mersenne number that is prime. For example, $M_2 = 2^2 - 1 = 3$ and $M_5 = 2^5 - 1 = 31$ are Mersenne primes while $M_6 = 2^6 - 1 = 63$ is a Mersenne number that is not prime. In order for $2^n - 1$ to be prime, it is necessary (but not sufficient) that n itself be prime—see Theorem 3.3.1 below.

At present, exactly 48 Mersenne primes are known. The last one was discovered on January 25, 2013, by Curtis Cooper at the University of Central Missouri. Cooper's prime is $2^{57,885,161} - 1$, a number whose decimal representation is 17,425,170 digits long. See http://www.mersenne.org/ for the latest on the Great Internet Mersenne Prime Search (GIMPS). □

The following theorem explains why the Mersenne numbers are prime candidates.

Theorem 3.3.1. *If a and n are both greater than or equal to 2 and $a^n - 1$ is prime, then $a = 2$ and n is prime.*

Proof. Let $a \geq 2$ and $n \geq 2$ be two integers such that $a^n - 1$ is prime (hypothesis).

We first give a direct proof that $a = 2$. Observe that $a^n - 1 = (a-1)(a^{n-1} + a^{n-2} + \cdots + a + 1)$. If $a^n - 1$ is prime, then each factor is either 1 or $a^n - 1$. Thus either $a - 1 = 1$ or $a - 1 = a^n - 1$. But the fact that $a \geq 2$ and $n \geq 2$

implies $a^n > a$ and so $a - 1 \neq a^n - 1$. Hence we must have $a - 1 = 1$, which implies $a = 2$.

Next we use a proof by contraposition to show that n is prime. Suppose n is not prime. Then $n = rs$ for some $r, s \in \mathbb{N}$. Hence $2^n - 1 = 2^{rs} - 1 = (2^r - 1)(2^{r(s-1)} + 2^{r(s-2)} + \ldots 2^r + 1)$. This proves the contrapositive of the second assertion. □

Exercises 3.3

1. Find the prime factorization of each of the following.
(a) 88
(b) 126
(c) 204
(d) 1001
(e) 1111
(f) 909,090

2. A template for the *Sieve of Eratothenes*[2] is included with these notes. Cross out all multiples of 2 except 2 itself. Then cross out all multiples of 3 except 3 itself. (Half of them are already crossed out.) The next number that is not crossed out is 5. Cross out all multiples of 5 except 5 itself. Continue this process.
(a) Explain why this process will terminate with multiples of 7. (Assuming you start with numbers up to 120.)
(b) Explain why all the numbers that were not crossed out must be prime.
(c) List the first 28 prime numbers.
(d) Prime numbers that differ by 2 are called *twin primes.* Find eight pairs of twin primes.[3]
(e) Find seven consecutive integers that are all composite. (A *prime gap* of length seven.)
(f) *Goldbach's Conjecture* asserts that every even positive integer n can be written as the sum of two primes.[4] Verify Goldbach's conjecture for $n = 120$ and $n = 130$.

3. A positive integer is said to be *perfect* if it is equal to the sum of all its positive divisors other than itself.
(a) Verify that 6 and 28 are perfect.

[2] Named for Eratosthenes of Cyrene, 276 BC – 194 BC.

[3] It is not known whether there are infinitely many twin primes. The assertion that there are infinitely many twin primes is known as the *Twin Prime Conjecture.*

[4] No proof has been found for Goldbach's Conjecture.

(b) Show that $2^{p-1}(2^p - 1)$ is a perfect number whenever $2^p - 1$ is a prime.[5]
[Hint: Use the formula for the sum of a finite geometric sequence: $1 + 2 + 4 + \ldots + 2^n = 2^{n+1} - 1$.]

4. The *Euler ϕ-function* is defined this way: If $n \in \mathbb{N}$, $\phi(n)$ is the number of positive integers less than n that are relatively prime to n. For example, $\phi(4) = 2$ because 1 and 3 are relatively prime to 4.
(a) Calculate $\phi(12)$, $\phi(13)$, $\phi(100)$, and $\phi(101)$.
(b) Explain why n is prime if and only if $\phi(n) = n - 1$.

3.4. RSA public key cryptography

RSA is an algorithm for public-key encryption, which means that the instructions for encoding a message may be made public and the decoding algorithm is still believed to be secure. The algorithm is named for its inventors, Ron Rivest, Adi Shamir, and Len Adleman, who first published it in 1977 while working at MIT. The algorithm was patented by MIT in 1983, but the patent expired in 2000. This algorithm was one of the first public key systems and it is still widely used in electronic commerce.

Key generation.

- Choose two large (e.g., 1024–2048 bit) primes p and q.
- Compute $n = pq$ and $m = (p - 1)(q - 1)$.
- Choose an exponent e such that $1 < e < m$ and $\gcd(e, m) = 1$.
- Compute d such that $1 < d < m$ and $ed \equiv 1 \pmod{m}$.

The primes p and q should be roughly the same size but not so close that they can be found by trying integers near $\sqrt{n}$. Usually $q < p < 2q$. The digits of p and q can be generated randomly and then the numbers checked for primeness using a probabilistic algorithm based on Fermat's Little Theorem. There is no definite rule regarding how the encoding number e must be chosen (except that it must be relatively prime to m), but there are certain standard values that are commonly used because they speed up the calculations. The decoding number d is computed using the extended Euclidean algorithm.

[5] It is not known whether there are infinitely many perfect numbers. Euler proved that every even perfect number must have the form $2^{p-1}(2^p - 1)$. Thus there are infinitely many Mersenne primes if and only if there are infinitely many even perfect numbers. It is not known whether there are any odd perfect numbers.

Public key encryption. The public key is the pair (n, e); these numbers may be published and shared with anyone. A plaintext message consists of a number P, $0 < P < n$. Encode P by

$$C = P^e \ \mathbf{mod}\ n.$$

This can be done efficiently using the fast modular exponentiation algorithm (the square-and-multiply algorithm).

Private key decryption. Your private key is the number d, and it must be kept secret. The primes p and q must also be kept secret since anyone who knows p, q, and e can compute d. Although it is less obvious, it is possible to determine p and q from m and n, so m must also be kept secret. Decode ciphertext C by

$$P = C^d \ \mathbf{mod}\ n.$$

A simple example. Take $p = 67$, $q = 79$, and $e = 137$. Then $n = 5293$ and $m = 5148$. It is not difficult to compute $d = 977$. To encrypt the plaintext $P = 256$, use the formula

$$C = (256)^{137} \ \mathbf{mod}\ 5293 = 4361.$$

The person receiving our message would decode it by the formula

$$D = (4361)^{977} \ \mathbf{mod}\ 5293 = 256.$$

Why RSA works. To understand why RSA works we must understand why

$$P = (P^e)^d \ \mathbf{mod}\ n.$$

Recall that d was chosen so that $ed \equiv 1 \pmod{m}$. Thus there exists an integer k such that $de = 1 + km = 1 + k(p-1)(q-1)$. Therefore

$$(P^e)^d = P^{de} = P^{1+k(p-1)(q-1)} = P \cdot P^{k(p-1)(q-1)}$$

By Fermat's Little Theorem, $P \cdot P^{p-1} \equiv P \pmod{p}$. Applying this result $k(q-1)$ times gives $P \cdot P^{k(p-1)(q-1)} \equiv P \pmod{p}$. In a similar way we see that $P \cdot P^{k(p-1)(q-1)} \equiv P \pmod{q}$. Thus $x = P$ and $x = P^{de}\ \mathbf{mod}\ n$ are both solutions to the system of congruences

$$x \equiv P \pmod{p}$$
$$x \equiv P \pmod{q}$$

By the Chinese Remainder Theorem the solution is unique modulo $pq = n$. Furthermore, both P and $(P^e)^d\ \mathbf{mod}\ n$ are smaller than n, so we can conclude that $P = (P^e)^d\ \mathbf{mod}\ n$. (Note that this last statement is where we used the fact that the plaintext message is smaller than n.)

How secure is it? Our belief that RSA is secure is based on two assumptions. First, that the only way to determine d from n and e is to factor n, and, second, that there is no efficient algorithm that will factor n. As far as we know, both assumptions are correct, but neither of them has been proved. In 1993, Richard Shor published an algorithm which will factor an integer in polynomial time on a quantum computer. So the system appears to be secure for now, but advances in computing might make it insecure in the future.

The system must be implemented carefully in order to avoid attacks on its security. For example, simply converting each character in a message to a number and encoding those numbers would result in a code that is easy to break. Since the number of characters is small, an attacker could simply use the public key to encrypt all possible code words and create a look-up table. This problem is avoided by either converting a relatively large piece of the message into one number to be encoded or by using some sort of padding scheme to convert the small numbers into large ones.

Signing a message. One useful feature of RSA is that it allows the person who knows the private key to prove that without revealing the private key itself. In order to do so, she can take her signature and encode it using her decoding algorithm. In other words, she can take her signature S and encode it as $S^d \textbf{ mod } n$. Then anyone receiving it can decode it using the public key since $S^{de} = S \textbf{ mod } n$.

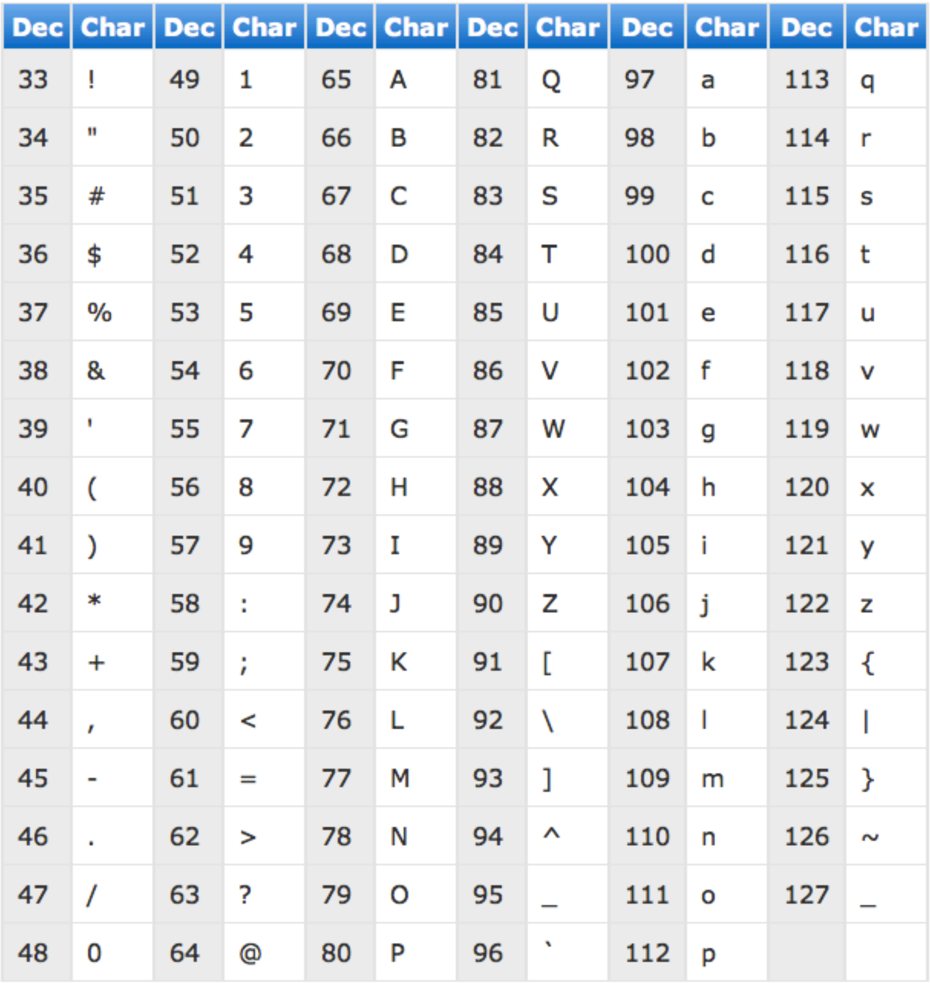
Another example. We can use Mathematica to work with much larger numbers. This gives us a better sense of how the algorithm works, but it is still not realistic in the sense that the numbers that are used in real applications are much larger still.

Let us take $p = 461$, $q = 541$, and $e = 137$. Then $n = 249401$ and $m = 248400$. To find d we must solve the congruence $137d \equiv 1 \pmod{248400}$. The Mathematica command `ExtendedGCD[137,248400]` returns the result `{1, {65273, -36}}`, which means that 1 is the greatest common divisor and that

$$1 = 65273 \cdot 137 + (-36) \cdot 248400.$$

We can conclude that $d = 65273$.

Each character in the plaintext message can be converted to a three-digit decimal number by using the ASCII table in Figure 3.1. (A blank space has ASCII value 32.) We will group two characters together and concatenate their ASCII values to form a six-digit number (which will be smaller than 249401). These six-digit numbers can therefore be encoded using the RSA algorithm with n and e as in the previous paragraph.



Dec	Char	Dec	Char	Dec	Char	Dec	Char	Dec	Char	Dec	Char
33	!	49	1	65	A	81	Q	97	a	113	q
34	"	50	2	66	B	82	R	98	b	114	r
35	#	51	3	67	C	83	S	99	c	115	s
36	$	52	4	68	D	84	T	100	d	116	t
37	%	53	5	69	E	85	U	101	e	117	u
38	&	54	6	70	F	86	V	102	f	118	v
39	'	55	7	71	G	87	W	103	g	119	w
40	(	56	8	72	H	88	X	104	h	120	x
41	)	57	9	73	I	89	Y	105	i	121	y
42	*	58	:	74	J	90	Z	106	j	122	z
43	+	59	;	75	K	91	[	107	k	123	{
44	,	60	<	76	L	92	\	108	l	124	|
45	-	61	=	77	M	93	]	109	m	125	}
46	.	62	>	78	N	94	^	110	n	126	~
47	/	63	?	79	O	95	_	111	o	127	_
48	0	64	@	80	P	96	`	112	p		

Figure 3.1. The ASCII table

Let us say, for example, that we wish to encode the following message: `This is secret.` Since there are an odd number of characters, we will add a blank at the beginning to make the number of characters even and then group as follows:

`␣T|hi|s␣|is|␣s|ec|re|t.`

The message is initially encoded as the following string of six-digit numbers:

32084 104105 115032 105115 32115 101099 114101 116046

To encode these numbers, we apply the formula $C = P^{137}$ ($\mathbf{mod}$ 249401) to each of these numbers. This is implemented in Mathematica as, for example,

```
PowerMod[32084,137,249401].
```

The coded message that results is

105051 211080 62364 194266 209903 193975 30136 965.

Exercises 3.4

1. Find the decoding number d for each of the following choices of p, q, and e.
(a) $p = 11, q = 13$, and $e = 47$.
(b) $p = 11, q = 13$, and $e = 31$. (Is this a good choice for e?)
(c) $p = 29, q = 37$, and $e = 25$.
(d) $p = 61$, $q = 53$, and $e = 17$.

2. Use the standard identification of the letters $\{A, B, \ldots, Z\}$ with the numbers $\{0, 1, \ldots, 25\}$ and the values of p, q, and e in 1(c) to encode the following words. (Encode a letter at a time.)
(a) HELLO
(b) MATH

3. Use the standard identification of the letters $\{A, B, \ldots, Z\}$ with the numbers $\{0, 1, \ldots, 25\}$ and the values of p, q, and e in 1(c) to decode the following messages.
(a) 154 717
(b) 1 48 1051 585

4. Use the standard identification of the letters $\{A, B, \ldots, Z\}$ with the numbers $\{0, 1, \ldots, 25\}$ and group the letters in pairs. Take $p = 43$, $q = 59$, and $e = 13$.
(a) Encode INVADE.
(b) Decode 0292 1947 0204.

5. Group the letters in pairs and use their ASCII values, as in the last example. Take $p = 523$, $q = 653$, and $e = 1223$.
(a) Encode: `Mathematics is easy.`
(b) Calculate the decoding number d.
(c) Decode: 335097 159164 338325 121324 64392 65037 316351 204876 247037 67441 77287 248806.

List of useful Mathematica commands

- `FactorInteger[`n`]` : the prime factors of n and their exponents
- `Mod[`k, n`]` : k $\mathbf{mod}$ n (the remainder when n is divided by k)

- `GCD[`m, n`]` : the greatest common divisor of m and n
- `ExtendedGCD[`m, n`]` : the gcd and the coefficients in gcd $= sm + tn$
- `PowerMod[`a, b, m`]` : a^b **mod** m

To obtain a copy of Mathematica, go to

`http://www.calvin.edu/it/core/desktop_services/personal_purchase.html#math`

and click on Mathematica or simply search for Mathematica on the Calvin College website.

Chapter 4

Induction and Recursion

4.1. Mathematical induction

Mathematical induction is a form of proof that is used to prove propositions about the positive integers. Specifically, it can be employed to prove statements of the form $\forall n\ P(n)$, where $P(n)$ is some propositional function whose domain is the set of positive integers $\mathbb{N}$.

A proof by mathematical induction has two parts.

Part 1: Base Case. Prove $P(1)$.

Part 2: Inductive Step. Prove that $P(k) \to P(k+1)$ for every $k \in \mathbb{N}$.

In the inductive step, $P(k)$ is assumed as a hypothesis. This assumption is called the *inductive hypothesis*. Here is an example of how a proof by induction is written.

Example. The sum of the first n odd positive integers is n^2.

Observe that the first odd positive integer is $1 = 2 \cdot 1 - 1$, the second is $3 = 2 \cdot 2 - 1$, the third is $5 = 2 \cdot 3 - 1$, $\ldots$, and the nth odd positive integer is $2n - 1$.

Restatement. *If* $n \in \mathbb{N}$, *then* $1 + 3 + \cdots + (2n-1) = n^2$.

Proof. This is a proof by mathematical induction.

Base Case. Let $n = 1$. There is only one term in the sum and the equation reduces to $1 = 1$ in this case.

Inductive Step. Assume $1 + 3 + \cdots + (2k-1) = k^2$ for some $k \in \mathbb{N}$ (the inductive hypothesis). We must show that

$$1 + 3 + \cdots + (2(k+1) - 1) = (k+1)^2.$$

But

$$\begin{aligned} 1 + 3 + \cdots + (2(k+1) - 1) &= 1 + 3 + \cdots + (2k-1) + (2k+1) \\ &= k^2 + (2k+1) \quad \text{(by the inductive hypothesis)} \\ &= (k+1)^2 \quad \text{(by algebra)}, \end{aligned}$$

so the proof is complete. □

Example. If $n \in \mathbb{N}$, then $n^3 - n$ is divisible by 3.

Proof. This is a proof by mathematical induction.

Base Case. Let $n = 1$. Then $n^3 - n = 0$, which is divisible by 3. (It might be reassuring to check another, less trivial, base case: If $n = 2$, then $n^3 - n = 8 - 2 = 6$, which is divisible by 3.)

Inductive Step. Assume $k^3 - k$ is divisible by 3 (the inductive hypothesis). We must prove that $(k+1)^3 - (k+1)$ is divisible by 3.

Now $(k+1)^3 - (k+1) = k^3 + 3k^2 + 3k + 1 - k - 1 = (k^3 - k) + 3(k^2 + k)$ (algebra). By the inductive hypothesis, $k^3 - k$ is divisible by 3 and $3(k^2 + k)$ is divisible by 3 by definition of divisible. Thus $(k+1)^3 - (k+1)$ is divisible by 3 by Theorem 3.1.1, Part (*ii*). □

Many of the theorems from number theory can be proved using mathematical induction. We now illustrate this by giving proofs of the uniqueness part of the Fundamental Theorem of Arithmetic and Fermat's Little Theorem; later in the section we will use a modified version of mathematical induction, called strong induction, to prove the existence part of the Fundamental Theorem of Arithmetic.

Lemma 4.1.1. *If p is prime and $p \mid a_1 a_2 \cdots a_n$, where each a_i is an integer, then $p \mid a_i$ for some i.*

Proof. This is a proof by mathematical induction.

Base Case. If $n = 1$, then the hypothesis is that $p \mid a_1$, so we can conclude $p \mid a_1$.

Inductive Step. Assume that the following inductive hypothesis holds: If p is prime and $p \mid a_1 a_2 \cdots a_k$, where each a_i is an integer, then $p \mid a_i$ for some i. Now suppose p is prime and $p \mid a_1 a_2 \cdots a_{k+1} = (a_1 a_2 \cdots a_k) a_{k+1}$. By Corollary 3.1.3, either $p \mid a_1 a_2 \cdots a_k$ or $p \mid a_{k+1}$. In the first case, $p \mid a_i$ for some i by the inductive hypothesis, and in the second case we also have that p divides one of the a_i (namely a_{k+1}). □

The existence part of the Fundamental Theorem of Arithmetic asserts that every integer greater than 1 is either prime or can be written as a product of prime factors. The uniqueness part asserts that this prime factorization is unique, except for the order in which the factors are listed.

Proof of the Uniqueness of Prime Factorization.

Suppose $a = p_1 p_2 \cdots p_n$ and $a = q_1 q_2 \cdots q_m$, where p_i and q_i are primes. We will first prove that exactly the same factors occur in each factorization and then we will prove that each factor has the same multiplicity in the two factorizations.

Start with p_1. Since $p_1 \mid a$, we must have that $p_1 \mid q_1 q_2 \cdots q_m$. By Lemma 4.1.1, this means that there is an i such that $p_1 \mid q_i$. But q_i is prime, so $p_1 = q_i$. In the same way we can prove that every one of the p's is equal to one of the q's and that each of the q's is equal to one of the p's.

It remains to prove that each prime factor occurs with the same multiplicity in each factorization. We give a proof by contradiction. Suppose there are more factors of p_1 in $p_1 p_2 \cdots p_n$ than in $q_1 q_2 \cdots q_m$. We can start with the equation $p_1 p_2 \cdots p_n = q_1 q_2 \cdots q_m$ and cancel as many factors of p_1 from both sides as possible. Since there were more factors of p_1 in the first product than the second, this will result in an equation that has at least one factor of p_1 on the left and no factors of p_1 on the right. But this is impossible, because if p_1 divides the left hand side it must also divide the right hand side. □

Our proof of Fermat's Little Theorem is based on the Binomial Theorem from high school algebra.

Binomial Theorem (from high school algebra)**.**

$$(a+b)^n = a^n + \binom{n}{1} a^{n-1} b + \binom{n}{2} a^{n-2} b^2 + \cdots + b^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k,$$

where

$$\binom{p}{k} = \frac{p!}{k!(p-k)!}.$$

Remember that the binomial coefficients $\binom{p}{k}$ in the theorem are the entries in Pascal's Triangle. One important observation about the numbers in the triangle is that every entry in a prime numbered row is divisible by that prime.

Lemma. *If p is prime and $0 < t < p$, then p divides $\binom{p}{t}$.*

$$\begin{array}{c}
1 \\
1 \quad 1 \\
1 \quad 2 \quad 1 \\
1 \quad 3 \quad 3 \quad 1 \\
1 \quad 4 \quad 6 \quad 4 \quad 1 \\
1 \quad 5 \quad 10 \quad 10 \quad 5 \quad 1 \\
1 \quad 6 \quad 15 \quad 20 \quad 15 \quad 6 \quad 1 \\
1 \quad 7 \quad 21 \quad 35 \quad 35 \quad 21 \quad 7 \quad 1 \\
\iddots \qquad\qquad\qquad \vdots \qquad\qquad\qquad \ddots
\end{array}$$

Proof. Note that p divides $p! = \binom{p}{t} t!(p-t)!$. But p does not divide either $t!$ or $(p-t)!$ because every prime factor of $t!$ and $(p-t)!$ is smaller than p. Hence p divides $\binom{p}{t}$ by Lemma 4.1.1. □

Fermat's Little Theorem. If p is prime and a is a positive integer, then

$$a^p \equiv a \pmod{p}.$$

Proof. The proof is by induction on a.

Base case. If $a = 1$, then $a^p \equiv 1^p \equiv 1 \pmod{p}$.

Inductive step. Assume $k^p \equiv k \pmod{p}$ (inductive hypothesis). The Binomial Theorem allows us to express $(k+1)^p$ as

$$(k+1)^p = k^p + \binom{p}{1} k^{p-1} + \cdots + \binom{p}{p-1} k + 1.$$

By the lemma above, p divides each of the terms in this sum except for the first and the last. Thus $(k+1)^p \equiv k^p + 1 \pmod{p}$ and the inductive hypothesis gives $(k+1)^p \equiv k + 1 \pmod{p}$. □

Corollary 4.1.2. *If p is prime and p does not divide a, then*

$$a^{p-1} \equiv 1 \pmod{p}.$$

Proof. By Theorem 3.2.3, a has an inverse modulo p. Multiplying both sides of the congruence in Fermat's Theorem by this inverse yields the new congruence. □

Strong Induction. The term *strong induction* refers to a variation on mathematical induction in which we assume a stronger inductive hypothesis. Rather than showing $P(k) \to P(k+1)$ in the inductive step, we prove that

$$[P(1) \wedge P(2) \wedge \cdots \wedge P(k)] \to P(k+1).^1$$

Like an ordinary proof by mathematical induction, a proof by strong induction has two parts.

Part 1: Base Case. Prove $P(1)$.

[1] "Strong" induction is actually a weaker form of proof than ordinary induction in the sense that the inductive step requires a stronger hypothesis in order to reach the same conclusion.

Part 2: Inductive Step. Prove that $[P(1) \wedge P(2) \wedge \cdots \wedge P(k)] \to P(k+1)$ for every $k \in \mathbb{N}$.

As an example we will prove the existence part of Fundamental Theorem of Arithmetic.

Proof of the Existence of Prime Factorization. We must prove that every integer greater than 1 is either prime or can be written as a product of primes. This is a proof by strong induction.

Base Case. The smallest integer greater than 1 is 2, so that is our base case. The number 2 is prime, so the conclusion holds in this case.

Inductive Step. Assume k is an integer greater than 1 and that every integer in the range $2, \ldots, k$ is either prime or can be written as a product of primes. Consider $k+1$. If $k+1$ is a prime, then the conclusion holds. If $k+1$ is not prime, then it can be written as a product $k+1 = a \cdot b$, where $a > 1$ and $b > 1$. Since a and b are both less than or equal to k, each of them is either prime or a product of primes by the inductive hypothesis. It follows that $a \cdot b$ is a product of primes. □

Other variations on induction are possible. See, for example, the proof of Theorem 4.2.1 in the next section.

Exercises 4.1

1. Use mathematical induction to prove the following formulas are true for each $n \in \mathbb{N}$.

(a) $1 + 2 + \cdots + n = \dfrac{n(n+1)}{2}$.

(b) $1^2 + 2^2 + \cdots + n^2 = \dfrac{n(n+1)(2n+1)}{6}$.

(c) $1^3 + 2^3 + \cdots + n^3 = \left(\dfrac{n(n+1)}{2}\right)^2$.

(d) $1 + a + a^2 + \cdots + a^n = \dfrac{a^{n+1} - 1}{a - 1}$ for any a with $0 < a < 1$ or $a > 1$. [This proves the formula used in Exercise 3.3.3(b).]

(e) $1 \cdot 1! + 2 \cdot 2! + 3 \cdot 3! + \cdots + n \cdot n! = (n+1)! + 1$.

2. Use mathematical induction to prove that if S is a finite set with $\operatorname{card}(S) = n$, then $\operatorname{card}(\mathscr{P}(S)) = 2^n$.

3. Prove that $2n + 1 < n^2$ for every $n \geq 3$.

4. Prove that $n^2 < 2^n$ for every $n \geq 5$. [Hint: Use the result of the previous exercise.]

5. Prove that $2^n < n!$ for every $n \geq 4$.

6. Prove that $n! < n^n$ for every $n \geq 2$.

7. Prove that $n^3 - n$ is divisible by 6 for every $n \geq 1$.

8. Prove that $x^n - 1$ is divisible by $x - 1$ for every $n \geq 1$.

4.2. Recursion

A recursive definition of a function has a base case and a rule that determines the value of the function at a nonnegative integer from its values at smaller integers.

Example 1. Define a geometric sequence recursively by $a_0 = 3$ and $a_n = 2a_{n-1}$ for $n \geq 1$. This recursive definition generates the sequence $3, 6, 12, 24, 48, \ldots$. It is quite easy to "solve" the recursion and get the explicit formula $a_n = 3 \cdot 2^n$. More generally, if $\{a_n\}$ is a sequence satisfying the initial condition $a_0 = a$ and the recursive first order linear equation $a_n = ra_{n-1}$, then the solution is $a_n = ar^n$.

Example 2. The Fibonacci numbers are defined recursively by $f_0 = 0$, $f_1 = 1$, and $f_n = f_{n-1} + f_{n-2}$ for $n \geq 2$. In this case it is not so easy to guess an explicit formula for f_n. We will do that later in the semester, but the methods involved will require the use of vectors and matrices. For now it will be enough to find a lower bound for f_n.

Definition. The *golden ratio* is the irrational number $\alpha = (1+\sqrt{5})/2$. It is the larger of the two roots of the quadratic equation $x^2 = x+1$. The golden ratio is irrational; a decimal approximation is

$$\alpha \approx 1.6180339887498948482045868343656381177203091798057628621354 5.$$

Theorem 4.2.1. *If $n \geq 3$, then $f_n > \alpha^{n-2}$.*

Proof. This is a proof by strong induction.

Base Case. Since the recursive formula has order 2, we must check two base cases: $n = 3$ and $n = 4$. In case $n = 3$ we have $f_n = f_3 = 2$ and

$$\alpha^{n-2} = \alpha = \frac{1+\sqrt{5}}{2} < \frac{1+3}{2} = 2$$

since $\sqrt{5} < 3$. For $n = 4$ we have $f + n = f_4 = 3$ and

$$\alpha^{n-2} = \alpha^2 = \alpha + 1 = \frac{1+\sqrt{5}}{2} + 1 = \frac{3+\sqrt{5}}{2} < \frac{3+3}{2} = 3.$$

Inductive Step. Assume the inductive hypothesis $f_i > \alpha^{i-2}$ for every $i \le k$. We must show that $f_{k+1} > \alpha^{(k+1)-2} = \alpha^{k-1}$.

$$\begin{aligned} f_{k+1} &= f_k + f_{k-1} \qquad \text{(by the definition of Fibonnaci number)} \\ &> \alpha^{k-2} + \alpha^{k-3} \qquad \text{(by the inductive hypothesis)} \\ &= \alpha^{k-3}(\alpha + 1) \qquad \text{(by factoring)} \\ &= \alpha^{k-3}\alpha^2 \qquad \text{(by the definition of } \alpha\text{)} \\ &= \alpha^{k-1}. \end{aligned}$$ □

Example 3. The Euclidean Algorithm can be viewed as a recursive definition of the greatest common divisor function. Given a positive integer a and a nonnegative integer $b \le a$, define

$$\gcd(a, b) = \begin{cases} a & \text{if } b = 0 \\ \gcd(b, a \textbf{ mod } b) & \text{otherwise.} \end{cases}$$

The Fibonacci numbers represent the worst case scenario for the Euclidean Algorithm in the sense that if we attempt to use the Euclidean Algorithm to calculate the greatest common divisor of two consecutive Fibonacci numbers, all the quotients are 1 so that remainders decrease at the slowest possible rate. This observation, together with Theorem 4.2.1, was exploited by Gabriel Lamé (1795–1870) to find a bound on the number of operations required to find a greatest common divisor by means of the Euclidean Algorithm.

Theorem 4.2.2 (Lamé)**.** *If $a > b > 0$, then the number of divisions required to find* $\gcd(a, b)$ *using the Euclidean Algorithm is at most five times the number of decimal digits in b.*

Proof. Let us use n to denote the number of divisions required to find $\gcd(a, b)$. We will show below that if we start by dividing by b and n divisions are possible before we reach a zero remainder, then b must be at least as large as the $(n + 1)$st Fibonacci number (i.e., $b \ge f_{n+1}$). Let us assume that result for now and use it to complete the proof of Lamé's Theorem. Combining with Theorem 4.2.1 gives $b \ge f_{n+1} > \alpha^{n-1}$. Take logarithms of both sides of that inequality to obtain $(n - 1)\log_{10} \alpha < \log_{10} b$, or

$$n < \frac{1}{\log_{10} \alpha} \log_{10} b + 1.$$

Now $1/\log_{10} \alpha \approx 4.78497$, so we can conclude $n < 5 \log_{10} b + 1$.

Suppose b has k digits in its decimal expansion. Then $b < 10^k$ and therefore $\log_{10} b < k$. The inequality in the previous paragraph gives $n < 5k + 1$. But both n and $5k$ are integers so we must have $n \leq 5k$, which is the conclusion of Lamé's Theorem.

To complete the proof of the theorem, we must demonstrate that $b \geq f_{n+1}$. Recall that n is the number of divisions required to find $\gcd(a, b)$. If we use r_0 to denote a and r_1 to denote b, we can write the equations associated with the divisions as

$$\begin{aligned} r_0 &= r_1 q_1 + r_2 \\ r_1 &= r_2 q_2 + r_3 \\ &\vdots \\ r_{n-2} &= r_{n-1} q_{n-1} + r_n \\ r_{n-1} &= r_n q_n + 0. \end{aligned}$$

Since the remainders are getting smaller and r_n is the last nonzero remainder, we have $0 < r_n < r_{n-1} < \cdots < r_2 < r_1 < r_0$. Each q_i satisfies $q_i \geq 1$.

Claim. For each $i \geq 0$, $r_{n-i} \geq f_{i+2}$.

Once this claim is established we can take $i = n - 1$ and conclude that $b = r_1 = r_{n-(n-1)} \geq f_{(n-1)+2} = f_{n+1}$, which is what we wish to prove.

The claim is proved by strong induction. Since $0 < r_n < r_{n-1}$, we have $r_n \geq 1$ and $r_{n-1} \geq 2$. Thus $r_n \geq f_2 = 1$ and $r_{n-1} \geq f_3 = 2$. This proves the two base cases, $i = 0$ and $i = 1$, of the claim.

Now assume as an inductive hypothesis that the claim is correct for $i = j - 2$ and $i = j - 1$. The equation that r_{n-j} satisfies is

$$r_{n-j} = r_{n-(j-1)} q_{n-(j-1)} + r_{n-(j-2)}.$$

Since $q_{n-(j-1)} \geq 1$, we have $r_{n-j} \geq r_{n-(j-1)} + r_{n-(j-2)}$. The inductive hypothesis gives $r_{n-j} \geq f_{j+1} + f_j = f_{j+2}$. This completes the inductive proof of the claim and thus completes the proof of the Theorem. □

Exercises 4.2

1. Find a_1, a_2, a_3, a_4, and a_5 for each of the following recursively defined sequences.

(a) $a_1 = 2$ and $a_{n+1} = a_n + 3$ for $n \geq 1$.

(b) $a_1 = 2$ and $a_{n+1} = 3a_n$ for $n \geq 1$.

(c) $a_1 = 1$ and $a_{n+1} = 2^{a_n}$.

2. Give a recursive definition of the sequence $\{a_n\}$

(a) whose first few terms are $3, 7, 11, 15, 19, \ldots$.

(b) defined by the explicit formula $a_n = 3n - 2$.

3. Let $f_1, f_2, f_3, \ldots$ be the sequence of Fibonacci numbers.

(a) Prove $f_1^2 + f_2^2 + \cdots + f_n^2 = f_n f_{n+1}$.

(b) Prove $f_1 + f_3 + \cdots + f_{2n-1} = f_{2n}$.

(c) Prove $f_2 + f_4 + \cdots + f_{2n} = f_{2n+1} - 1$.

(d) Determine the number of divisions required to find $\gcd(f_n, f_{n+1})$.

4. Let $\{a_n\}$ be the sequence that is defined recursively by $a_0 = 1/2$ and $a_{n+1} = 2a_n - a_n^2$. Verify that

$$a_n = 1 - \left(\frac{1}{2}\right)^{2^n}$$

gives an explicit formula for a_n.