Math 208: Discrete Mathematics

Department of Mathematics
The University of North Dakota

Published by UND MATH Department

http://arts-sciences.und.edu/math//

Fourth edition, First printing: July 2023

Contents

List of Figures

List of Tables

List of Algorithms

Introduction

Discrete math has become increasingly important in recent years, for a number of reasons: *The Art of Problem Solving* http://www.artofproblemsolving.com/articles/discrete-math

Discrete math is essential to college-level mathematics — and beyond.

Discrete math – together with calculus and abstract algebra – is one of the core components of mathematics at the undergraduate level. Students who learn a significant quantity of discrete math before entering college will be at a significant advantage when taking undergraduate-level math courses.

Discrete math is the mathematics of computing.

The mathematics of modern computer science is built almost entirely on discrete math, in particular combinatorics and graph theory. This means that in order to learn the fundamental algorithms used by computer programmers, students will need a solid background in these subjects. Indeed, at most universities, a undergraduate-level course in discrete mathematics is a required part of pursuing a computer science degree.

Discrete math is very much "real world" mathematics.

Many students' complaints about traditional high school math – algebra, geometry, trigonometry, and the like – is "What is this good for?" The somewhat abstract nature of these subjects often turn off students. By contrast, discrete math, in particular counting and probability, allows students – even at the middle school level – to very quickly explore non-trivial "real world" problems that are challenging and interesting.

Discrete math shows up on most middle and high school math contests.

Prominent math competitions such as MATHCOUNTS (at the middle school level) and the American Mathematics Competitions (at the high school level) feature discrete math questions as a significant portion of their contests. On harder high school contests, such as the AIME, the quantity of discrete math is even larger. Students that do not have a discrete math background will be at a significant disadvantage in these contests. In fact, one prominent MATHCOUNTS coach tells us that he spends nearly 50% of his preparation time with his students covering counting and probability topics, because of their importance in MATHCOUNTS contests.

Discrete math teaches mathematical reasoning and proof techniques.

Algebra is often taught as a series of formulas and algorithms for students to memorize (for example, the quadratic formula, solving systems of linear equations by substitution, etc.), and geometry is often taught as a series of ”definition-theorem-proof” exercises that are often done by rote (for example, the infamous ”two-column proof”). While undoubtedly the subject matter being taught is important, the material (as least at the introductory level) does not lend itself to a great deal of creative mathematical thinking. By contrast, with discrete mathematics, students will be thinking flexibly and creatively right out of the box. There are relatively few formulas to memorize; rather, there are a number of fundamental concepts to be mastered and applied in many different ways.

Discrete math is fun.

Many students, especially bright and motivated students, find algebra, geometry, and even calculus dull and uninspiring. Rarely is this the case with most discrete math topics. When we ask students what their favorite topic is, most respond either ”combinatorics” or ”number theory.” (When we ask them what their least favorite topic is, the overwhelming response is ”geometry.”) Simply put, most students find discrete math more fun than algebra or geometry.

Chapter 1

Logical Connectives and Compound Propositions

The basic objects in logic are **propositions**. A proposition is a statement which is either true (T) or false (F) but not both. For example in the language of mathematics $p : 3 + 3 = 6$ is a true proposition while $q : 2 + 3 = 6$ is a false proposition. *What do you want for lunch?* is a question, not a proposition. Likewise *Get lost!* is a command, not a proposition. The sentence *There are exactly* $10^{87} + 3$ *stars in the universe* is a proposition, despite the fact that no one knows its truth value. Here are two, more subtle, examples:

(1) *He is more than three feet tall* is not a proposition since, until we are told to whom *he* refers, the statement cannot be assigned a truth value. The mathematical sentence $x + 3 = 7$ is not a proposition for the same reason. In general, sentences containing variables are not propositions unless some information is supplied about the variables. More about that later however.

(2) *This sentence is false* is not a proposition. It seems to be both true and false. In fact if is T then it says it is F and if it is F then it says it is T. It can be dangerous using sentences that refer to themselves. Of course, using a knife can also be dangerous, but we do use knives safely when we are careful. Likewise, using self-referential sentences can be done safely if care is taken.

p	$\neg$	p
T	F	T
F	T	F

Table 1.1: Logical Negation

Sometimes a little common sense is required. For example *It is raining* is a proposition, but its truth value is not constant, and may be arguable. That is, someone might say *It is not raining*, *it is just drizzling*, or *Do you mean on Venus?* Feel free to ignore this sort of quibbling.

Simple propositions, such as *It is raining*, and *The streets are wet*, can be combined to create more complicated propositions such as *It is raining and the streets are not wet*. These sorts of involved propositions are called **compound propositions**. Compound propositions are built up from simple propositions using a number of **connectives** to join or modify the simple propositions. In the last example, the connectives are **and** which joins the two clauses, and **not**, which modifies the second clause.

It is important to keep in mind that since a compound proposition is, after all, a proposition, it must be classifiable as either true or false. That is, it must be possible to assign a truth value to any compound proposition. There are mutually agreed upon rules to allow the determination of exactly when a compound proposition is true and when it is false. Luckily these rules jive nicely with common sense (with one small exception), so they are easy to remember and understand.

The simplest logical connective is **negation**. In normal English sentences, this connective is indicated by appropriately inserting *not* in the statement, by preceding the statement with *it is not the case that*, or for mathematical statements, by using a slanted slash. For example, if p is the proposition $2+3=4$, then the negation of p is denoted by the symbol $\neg p$ and it is the proposition $2+3\neq 4$. In this case, p is false and $\neg p$ is true. If p is *It is raining*, then $\neg p$ is *It is not raining* or even the stilted sounding *It is not the case that it is raining*. The negation of a proposition p is the proposition whose truth value is the opposite of p in all cases. The behavior of $\neg p$ can be exhibited in a **truth table**. In each row of the truth table 1.1 we list a possible truth value of p and the corresponding truth value of $\neg p$.

The connective that corresponds to the word *and* is called **conjunction**. The conjunction of p with q is denoted by $p \wedge q$ and read as *p and q*. The conjunction of p with q is declared to be true exactly when both of p, q are true. It is false otherwise. This behavior is exhibited in the truth table 1.2.

p	q	p	$\wedge$	q
T	T	T	T	T
T	F	T	F	F
F	T	F	F	T
F	F	F	F	F

Table 1.2: Logical Conjunction

p	q	p	$\vee$	q	p	$\oplus$	q
T	T	T	T	T	T	F	T
T	F	T	T	F	T	T	F
F	T	F	T	T	F	T	T
F	F	F	F	F	F	F	F

Table 1.3: Logical or and xor

Four rows are required in this table since when p is true, q may be either true or false and when p is false it is possible for q to be either true or false. Since a truth value must be assigned to $p \wedge q$ in every possible case, one row in the truth table is needed for each of the four possibilities.

The logical connective **disjunction** corresponds to the word *or* of ordinary language. The disjunction of p with q is denoted by $p \vee q$, and read as p *or* q. The disjunction $p \vee q$ is true if at least one of p, q is true.

Disjunction is also called **inclusive-or**, since it includes the possibility that both component statements are true. In everyday language, there is a second use of *or* with a different meaning. For example, in the proposition *Your ticket wins a prize if its serial number contains a* 3 *or a* 5, the *or* would normally be interpreted in the inclusive sense (tickets that have both a 3 and 5 are still winners), but in the proposition *With dinner you get mashed potatoes or french fries*, the *or* is being used in the **exclusive-or** sense.

The exclusive-or is also called the **disjoint disjunction** of p with q and is denoted by $p \oplus q$. Read that as p *xor* q if it is necessary to say it in words. The value of $p \oplus q$ is true if exactly one of p, q is true. The exclusion of both being true is the difference between inclusive-or and exclusive-or. The truth table shown officially defines these two connectives. In a mathematical setting one usually assumes the inclusive-or is intended unless the exclusive sense is explicitly indicated.

The next two logical connectives correspond to the ordinary language phrases *If* $\cdots$, *then* $\cdots$ and the (rarely used in real life but common in mathematics) $\cdots$ *if and only if* $\cdots$.

p	q	p	$\rightarrow$	q
T	T	T	T	T
T	F	T	F	F
F	T	F	T	T
F	F	F	T	F

Table 1.4: Logical Implication

In mathematical discussions, ordinary English words are used in ways that usually correspond to the way we use words in normal conversation. The connectives **not**, **and**, **or** mean pretty much what would be expected. But the **implication**, denoted $p \rightarrow q$ and read as *If* p, *then* q can be a little mysterious at first. This is partly because when the *If* p, *then* q construction is used in everyday speech, there is an implied connection between the proposition p (called the **hypothesis**) and the proposition q (called the **conclusion**). For example, in the statement *If I study, then I will pass the test,* there is an assumed connection between studying and passing the test. However, in logic, the connective is going to be used to join any two propositions, with no relation necessary between the hypothesis and conclusion. What truth value should be assigned to such bizarre sentences as *If I study, then the moon is* $238,000$ *miles from earth*?

Is it true or false? Or maybe it is neither one? Well, that last option isn't too pleasant because that sentence is supposed to be a proposition, and to be a proposition it has to have truth value either T or F. So it is going to have to be classified as one or the other. In everyday conversation, the choice isn't likely to be too important whether it is classified it as either true or false in the case described. But an important part of mathematics is knowing when propositions are true and when they are false. The official choices are given in the truth table for $p \rightarrow q$. We can make sense of this with an example.

Example 1.1. *First consider the statement which Bill's dad makes to Bill: If you get an A in math, then I will buy you a new car. If Bill gets an A and his dad buys him a car, then dad's statement is true, and everyone is happy (that is the first row in the table). In the second row, Bill gets an A, and his dad doesn't come through. Then Bill's going to be rightfully upset since his father lied to him (dad made a false statement). In the last row of the table he can't complain if he doesn't get an A, and his dad doesn't buy him the car (so again dad made a true statement). Most people feel comfortable with those three rows. In the third row of the table, Bill doesn't get an A, and his dad buys him a car anyhow. This is the funny case. It seems that calling dad a liar in this case would be a little harsh on the old man. So it is declared that dad told the truth. Remember it this way: an implication is true unless the hypothesis is true and the conclusion is false.*

p	q	p	$\leftrightarrow$	q
T	T	T	T	T
T	F	T	F	F
F	T	F	F	T
F	F	F	T	F

Table 1.5: Logical biconditional

The **biconditional** is the logical connective corresponding to the phrase $\cdots$ *if and only if* $\cdots$. It is denoted by $p \Leftrightarrow q$, (read p *if and only if* q), and often more tersely written as p *iff* q. The biconditional is true when the two component propositions have the same truth value, and it is false when their truth values are different. Examine the truth table to see how this works.

The connectives described above combine at most two simple propositions. More complicated propositions can be formed by joining compound propositions with those connectives. For example, $p \wedge (\neg q)$, $(p \vee q) \rightarrow (q \wedge (\neg r))$, and $(p \rightarrow q) \Leftrightarrow ((\neg p) \vee q)$ are compound propositions, where parentheses have been used, just as in ordinary algebra, to avoid ambiguity. Such extended compound propositions really are propositions. That is, if the truth value of each component is known, it is possible to determine the truth value of the entire proposition. The necessary computations can be exhibited in a truth table.

Example 1.2. *Suppose that* p, q *and* r *are propositions. To construct a truth table for* $(p \wedge q) \rightarrow r$, *first notice that eight rows will be needed in the table to account for all the possible combinations of truth values of the simple component statements* p, q *and* r. *This is so since there are, as noted above, four rows needed to account for the choices for* p *and* q, *so there will be those four rows paired with* r *having truth value* T, *and four more with* r *having truth value* F, *for a total of* $4 + 4 = 8$. *In general, if there are* n *simple propositions in a compound statement, the truth table for the compound statement will have* 2^n *rows. The truth table for* $(p \wedge q) \rightarrow r$ *is given in Table 1.6, with an auxiliary column for* $p \wedge q$ *to serve as an aid for filling in the last column.*

Be careful about how propositions are grouped. For example, if truth tables for $p \wedge (q \rightarrow r)$ and $(p \wedge q) \rightarrow r$ are constructed, they turn out not to be the same in every row. Specifically if p is false, then $p \wedge q$ is false, and $(p \wedge q) \rightarrow r$ is true. Whereas when p is false $p \wedge (q \rightarrow r)$ is false. So writing $p \wedge q \rightarrow r$ is ambiguous.

Here are a few examples of translating between propositions expressed in ordinary language and propositions expressed in the language of logic.

Example 1.3. *Let* c *be the proposition It is cold and* s*: It is snowing, and* h*: I'm staying home.*

p	q	r	(p	$\wedge$	q)	$\rightarrow$	r
T	T	T	T	T	T	T	T
T	T	F	T	T	T	F	F
T	F	T	T	F	F	T	T
T	F	F	T	F	F	T	F
F	T	T	F	F	T	T	T
F	T	F	F	F	T	T	F
F	F	T	F	F	F	T	T
F	F	F	F	F	F	T	F

Table 1.6: Truth table for $(p \wedge q) \rightarrow r$

Then $(c \wedge s) \rightarrow h$ is the proposition If it is cold and snowing, then I'm staying home. While $(c \vee s) \rightarrow h$ is If it is either cold or snowing, then I'm staying home. Messier is $\neg(h \rightarrow c)$ which could be expressed as It is not the case that if I stay home, then it is cold, which is a little too convoluted for our minds to grasp quickly. Translating in the other direction, the proposition It is snowing and it is either cold or I'm staying home would be symbolized as $s \wedge (c \vee h)$. Notice the parentheses are needed in this last proposition since $(s \wedge c) \vee h$ does not capture the meaning of the ordinary language sentence, and $s \wedge c \vee h$ is ambiguous.

There is a connection between logical connectives and certain operations on bit strings. There are two **binary digits** (or **bits**): 0 and 1. A **bit string of length** n is any sequence of n bits. For example, 0010 is a bit sting of length four. Computers use bit strings to encode and manipulate information. Some bit string operations are really just disguised truth tables. Here is the connection: Since a bit can be one of two values, bits can be used to represent truth values. Let T correspond to 1, and F to 0. Then given two bits, logical connectives can be used to produce a new bit. For example $\neg 1 = 0$, and $1 \vee 1 = 1$. This can be extended to strings of bits of the same length by combining corresponding bit in the two strings. For example, $01011 \wedge 11010 = (0 \wedge 1)(1 \wedge 1)(0 \wedge 0)(1 \wedge 1)(1 \wedge 0) = 01010$.

Exercises

Exercise 1.1. *Determine which of the following sentences are propositions. Assume you are speaking the sentence.*

a) There are seven days in a week.

b) Get lost!

c) Pistachio is the best ice cream flavor.

d) All unicorns have four legs.

Exercise 1.2. *Construct truth tables for each of the following.*

a) $p \oplus \neg q$

b) $\neg(q \rightarrow p)$

c) $q \wedge \neg p$

d) $\neg q \vee p$

e) $p \rightarrow (\neg q \wedge r)$

Exercise 1.3. *Perform the indicated bit string operations. The bit strings are given in groups of four bits each for ease of reading.*

a) $(1101\ 0111 \oplus 1110\ 0010) \wedge 1100\ 1000$

b) $(1111\ 1010 \wedge 0111\ 0010) \vee 0101\ 0001$

c) $(1001\ 0010 \vee 0101\ 1101) \wedge (0110\ 0010 \vee 0111\ 0101)$

Exercise 1.4. *Let s be the proposition It is snowing and f be the proposition It is below freezing. Convert the following English sentences into statements using the symbols s, f and logical connectives.*

a) It is snowing and it is not below freezing.

b) It is below freezing and it is not snowing.

c) If it is not snowing, then it is not below freezing.

Exercise 1.5. *Let j be the proposition Jordan played and w be the proposition The Wizards won. Write the following propositions as English sentences.*

a) $\neg j \wedge w$

b) $j \rightarrow \neg w$

c) $w \vee j$

d) $w \rightarrow \neg j$

Exercise 1.6. *Let c be the proposition Sam plays chess, let b be Sam has the black pieces, and let w be Sam wins.*

a) *Translate into English:* $(c \wedge \neg b) \rightarrow w$.

b) *Translate into symbols: If Sam didn't win her chess game, then she played black.*

Problems

Problem 1.1. *Determine which of the following sentences are propositions*
Assume you are speaking the sentence.

a) *Today is Tuesday.*

b) *Why are you whining?*

c) *The Vikings are the worst team in professional sports.*

d) *This sentence has five words.*

e) *There is a black hole at the center of every galaxy.*

Problem 1.2. *Construct truth tables for each of the following.*

a) $\neg q \longrightarrow \neg p$.

b) $p \longrightarrow (q \wedge r)$.
(You will need eight rows for this one.)

Problem 1.3. *Perform the indicated bit string operations. The bit strings are given in groups of four bits each for ease of reading.*

a) $(1001\ 0101 \oplus 1010\ 0110) \wedge 1100\ 1000$

b) $(1110\ 1010 \wedge 0101\ 0010) \vee 0111\ 1001$

c) $(1111\ 0011 \vee 0111\ 0101) \wedge (0010\ 0010 \vee 0110\ 0100)$

Problem 1.4. *Let* s *be the proposition It is snowing and* f *be the proposition It is below freezing. Convert the following English sentences into statements using the symbols* s, f *and logical connectives.*

a) *It is snowing and it is below freezing.*

b) *If it is snowing, then it is below freezing.*

Chapter 2

Logical Equivalence

It is clear that the propositions *It is sunny and it is warm* and *It is warm and it is sunny* mean the same thing. More generally, for any propositions p, q, we see that $p \wedge q$ and $q \wedge p$ have the same meaning. To say it a little differently, for any choice of truth values for p and q, the propositions $p \wedge q$ and $q \wedge p$ have the same truth value. One more time: $p \wedge q$ and $q \wedge p$ have identical truth tables.

Two propositions with identical truth values are called **logically equivalent**. The expression $p \equiv q$ means p, q are logically equivalent.

Some logical equivalences are not as transparent as the example above. With a little thought it should be clear that *I am not taking math or I am not taking physics* means the same as *It's not the case that I am taking math and physics*. In symbols, $(\neg m) \vee (\neg p)$ means the same as $\neg(m \wedge p)$.

Example 2.1 (De Morgan)**.** *Prove that* $\neg(p \wedge q) \equiv (\neg p \vee \neg q)$ *using a truth table. We construct the truth table 2.1 in the order of precedence:* $\neg$ *before* $\wedge$ *or* $\vee$*, but the expresion in parentheses has highest precedence. We construct the table using additional columns for compound parts of the two expressions.*

It may be a little harder to believe $(p \to q) \equiv (\neg p \vee q)$, but checking a truth table shows they are in fact equivalent. Saying *If it is Monday, then I am tired* is identical to saying *It isn't Monday or I am tired.* You should construct a truth table to demonstrate their equivalence.

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

Table 2.1: De Morgan's Law

A proposition , $\mathbb{T}$, which is always true is called a **tautology**. A **contradiction** is a proposition, $\mathbb{F}$, which is always false. The prototype example of a tautology is $p \vee \neg p$, and for a contradiction, $p \wedge \neg p$. Notice that since $p \Leftrightarrow q$ is T exactly when p and q have the same truth value, two propositions p and q will be logically equivalent provided $p \Leftrightarrow q$ is a tautology.

Table 2.2 contains the most often used logical equivalences. These are well worth learning by sight and by name.

Equivalence	**Name**
$\neg(\neg p) \equiv p$	Double Negation
$p \wedge \mathbb{T} \equiv p$ $p \vee \mathbb{F} \equiv p$	Identity laws
$p \vee \mathbb{T} \equiv \mathbb{T}$ $p \wedge \mathbb{F} \equiv \mathbb{F}$	Domination laws
$p \vee p \equiv p$ $p \wedge p \equiv p$	Idempotent laws
$p \vee q \equiv q \vee p$ $p \wedge q \equiv q \wedge p$	Commutative laws
$(p \vee q) \vee r \equiv p \vee (q \vee r)$ $(\mathrm{p} \wedge \mathrm{q}) \wedge \mathrm{r} \equiv p \wedge (q \wedge r)$	Associative laws
$p \vee (q \wedge r) \equiv (p \vee q) \wedge (p \vee r)$ $p \wedge (q \vee r) \equiv (p \wedge q) \vee (p \wedge r)$	Distributive laws
$\neg(p \wedge q) \equiv (\neg p \vee \neg q)$ $\neg(p \vee q) \equiv (\neg p \wedge \neg q)$	De Morgan's laws
$p \vee \neg p \equiv \mathbb{T}$ $p \wedge \neg p \equiv \mathbb{F}$	Law of Excluded Middle Law of Contradiction
$p \to q \equiv \neg p \vee q$	Disjunctive form
$p \to q \equiv \neg q \to \neg p$ $\neg p \to \neg q \equiv q \to p$	Implication ≡ Contrapositive Inverse ≡ Converse

Table 2.2: Logical Equivalences

There are three propositions related to the basic **If . . . , then . . .** implication: $p \to q$. First

$\neg q \rightarrow \neg p$ is called the **contrapositive** of the implication. The **converse** of the implication is the proposition $q \rightarrow p$. Finally, the **inverse** of the implication is $\neg p \rightarrow \neg q$. Using a truth table, it is easy to check that an implication and its contrapositive are logically equivalent, as are the converse and the inverse. A common slip is to think the implication and its converse are logically equivalent. Checking a truth table shows that isn't so. The implication *If an integer ends with a* 2*, then it is even* is T, but its converse, *If an integer is even, then it ends with a* 2, is certainly F.

Five basic connectives have been given: $\neg, \wedge, \vee, \rightarrow, \Leftrightarrow$, but that is really just for convenience. It is possible to eliminate some of them using logical equivalences. For example, $p \Leftrightarrow q \equiv (p \rightarrow q) \wedge (q \rightarrow p)$ so there really is no need to explicitly use the biconditional. Likewise, $p \rightarrow q \equiv \neg p \vee q$, so the use of the implication can also be avoided. Finally, $p \wedge q \equiv \neg(\neg p \vee \neg q)$ so that there really is no need ever to use the connective $\wedge$. Every proposition made up of the five basic connectives can be rewritten using only $\neg$ and $\vee$ (probably with a great loss of clarity however).

The most often used standardization, or normalization, of logical propositions is the **disjunctive normal form (DNF)**, using only $\neg$ (negation), $\wedge$ (conjunction), and $\vee$ (disjunction). A propositional form is considered to be in DNF if and only if it is a disjunction of one or more conjunctions of one or more *literals* (a *literal* is a letter or a letter preceded by the negation symbol). For example, the following are all in disjunctive normal form:

- $p \wedge q$
- p
- $(a \wedge q) \vee r$
- $(p \wedge \neg q \wedge \neg r) \vee (\neg s \wedge t \wedge u)$

While, these are **not** in DNF:

- $\neg(p \vee q)$ *this is **not** the disjunction of literals.*
- $p \wedge (q \wedge (r \vee s))$ an or is embedded in a conjunction.

It is always possible the verify a logical equivalence via a truth table. But it also possible to verify equivalences by stringing together previously known equivalences. We provide two examples of this process.

Example 2.2. *Show* $\neg(p \vee (\neg p \wedge q)) \equiv \neg p \wedge \neg q$. *The plan is to start with the expression* $\neg(p \vee (\neg p \wedge q))$, *work through a sequence of equivalences ending up with* $\neg p \wedge \neg q$. *It's pretty much like proving identities in algebra or trigonometry.*

Proof.

$$
\begin{aligned}
\neg(p \vee (\neg p \wedge q)) &\equiv \neg p \wedge \neg(\neg p \wedge q) && \text{De Morgan's Law} \\
&\equiv \neg p \wedge (\neg(\neg p) \vee \neg q) && \text{De Morgan's Law} \\
&\equiv \neg p \wedge (p \vee \neg q) && \text{Double Negation Law} \\
&\equiv (\neg p \wedge p) \vee (\neg p \wedge \neg q) && \text{Distributive Law} \\
&\equiv (p \wedge \neg p) \vee (\neg p \wedge \neg q) && \text{Commutative Law} \\
&\equiv \mathbb{F} \vee (\neg p \wedge \neg q) && \text{Law of Contradiction} \\
&\equiv (\neg p \wedge \neg q) \vee \mathbb{F} && \text{Commutative Law} \\
&\equiv \neg p \wedge \neg q && \text{Identity Law}
\end{aligned}
$$

□

Example 2.3. *Show* $(p \wedge q) \to (p \vee q) \equiv \mathbb{T}$.

Proof.

$$
\begin{aligned}
(p \wedge q) \to (p \vee q) &\equiv \neg(p \wedge q) \vee (p \vee q) && \text{Disjunctive form} \\
&\equiv (\neg p \vee \neg q) \vee (p \vee q) && \text{De Morgan's Law} \\
&\equiv (p \vee \neg p) \vee (q \vee \neg q) && \text{Associative and Commutative Laws} \\
&\equiv \mathbb{T} \vee \mathbb{T} && \text{Commutative Law and Excluded Middle} \\
&\equiv \mathbb{T} && \text{Domination Law}
\end{aligned}
$$

□

Exercises

Exercise 2.1. *Use truth tables to verify each of the following equivalences:*

a) $(p \vee q) \vee r \equiv p \vee (q \vee r)$

b) $\neg p \wedge (p \vee q) \equiv \neg(q \rightarrow p)$

c) $p \vee (q \wedge r) \equiv (p \vee q) \wedge (p \vee r)$

Exercise 2.2. *Show that the statements are not logically equivalent.*

a) $p \wedge (q \rightarrow r) \not\equiv (p \wedge q) \rightarrow r$

b) $p \rightarrow q \not\equiv q \rightarrow p$

c) $p \rightarrow q \not\equiv \neg p \rightarrow \neg q$

Exercise 2.3. *Use truth tables to show that the following are tautologies.*

a) $[p \wedge (p \rightarrow q)] \rightarrow q$

b) $[(p \rightarrow q) \wedge (q \rightarrow r)] \rightarrow (p \rightarrow r)$

Exercise 2.4. *Consider the implication If it is Saturday, then I will mow the lawn.*

a) Write the converse of the implication.

b) Write the inverse of the implication.

c) Write the contrapositive of the implication.

Exercise 2.5. *Consider the three answers for exercise 2.4. One of the three is logically equivalent to the implication. Which one? Two of the three are not logically equivalent to the implication, but are logically equivalent to each other. Which two?*

Exercise 2.6. *The statements below are not tautologies. In each case, find an assignment of truth values to the literals, (that is, a letter or a letter preceded by the negation symbol), so the statement is false.*

a) $[(p \wedge q) \rightarrow r] \longleftrightarrow [(p \rightarrow r) \wedge (q \rightarrow r)]$

b) $[(p \wedge q) \vee r] \rightarrow [p \wedge (q \vee r)]$

Exercise 2.7. *Give a proof of* $(p \wedge \neg r) \rightarrow \neg q \equiv p \rightarrow (q \rightarrow r)$ *using the Fundamental Logical Equivalences, following the pattern of examples 2.2 and 2.3.*

Problems

Problem 2.1. *Use a truth table to show* $p \to q \equiv \neg p \lor q$.

Problem 2.2. *Use a truth table to show* $(p \land q) \to r \equiv p \to (q \to r)$.

Problem 2.3. *Use a truth table to show* $p \to q \equiv \neg q \to \neg p$.

Problem 2.4. *Consider the implication If I work, then I get paid.*

a) *Write the converse of the implication.*

b) *Write the inverse of the implication.*

c) *Write the contrapositive of the implication.*

d) *Write the inverse of the contrapositive of the implication.*

Problem 2.5. *Consider the four answers for exercise 2.4. Which of the are logically equivalent to the implication? Which of the four are not logically equivalent to the implication, (but are logically equivalent to each other)?*

Problem 2.6. *Use a truth table to show* $[p \to (q \to r)] \not\equiv [(p \to q) \to r]$.

Problem 2.7. *Use truth tables to show that the following are tautologies.*

a) $(p \land q) \to p$

b) $[(p \lor q) \to r] \to [(p \to r) \land (q \to r)]$

Problem 2.8. *Give a proof of* $\neg p \to (p \to q) \equiv \mathbb{T}$ *using the Fundamental Logical Equivalences, following the pattern of examples 2.2 and 2.3.*

Chapter 3

Predicates and Quantifiers

The sentence $x^2 - 2 = 0$ is not a proposition. It cannot be assigned a truth value unless some more information is supplied about the variable x. Such a statement is called a **predicate** or a **propositional function**.

Instead of using a single letter to denote a predicate, a symbol such as *S(x)* will be used to indicate the dependence of the sentence on a variable. Here are two more examples of predicates.

(1) *A(c)*: *Al drives a c*, and

(2) *B(x, y)*: *x is the brother of y*.

The second example is an instance of a **two-place predicate**.

With a given predicate, there is an associated set of objects which can be used in place of the variables. For example, in the predicate *S(x)*: $x^2 - 2 = 0$, it is understood that the x can be replaced by a number. Replacing x by, say, the word *blue* does not yield a meaningful sentence. For the predicate *A(c)* above, *c* can be replaced by, say, makes of cars (or maybe types of nails!). For *B(x, y)*, the x can be replaced by any human male, and the y by any human. The collection of possible replacements for a variable in a predicate is called the **domain of discourse** for that variable. Usually the domain of discourse is left for the reader to guess, but if the domain of discourse is something other than an obvious choice, the writer will mention the domain to be used.

A predicate is not a proposition, but it can be converted into a proposition. There are three ways to modify a predicate to change it into a proposition. Let's use $S(x)$: $x^2 - 2 = 0$ as an example.

The first way to change $S(x)$ to make it into a proposition is to assign a specific value from the variable's domain of discourse to the variable. For example, setting $x = 3$, gives the (false) proposition $S(3)$: $3^2 - 2 = 0$. On the other hand, setting $x = \sqrt{2}$ gives the (true) proposition $S(\sqrt{2}) : (\sqrt{2})^2 - 2 = 0$. The process of setting a variable equal to a specific object in its domain of discourse is called **instantiation**. Looking at the two-place predicate $B(x, y)$: x is the brother of y, we can instantiate both variables to get the (true - look up the Osmonds) proposition B(*Donny, Marie*): *Donny is the brother of Marie*. Notice that the sentence B(*Donny*, y): *Donny is the brother of* y has not been converted into a proposition since it cannot be assigned a truth value without some information about y. But it has been converted from a two-place predicate to a one-place predicate.

A second way to convert a predicate to a proposition is to precede the predicate with the phrase *There is an* x *such that*. For example, *There is an* x *such that* $S(x)$ would become *There is an* x *such that* $x^2 - 2 = 0$. This proposition is true if there is at least one choice of x in its domain of discourse for which the predicate becomes a true statement. The phrase *There is an* x *such that* is denoted in symbols by $\exists x$, so the proposition above would be written as $\exists x, S(x)$ or $\exists x(x^2 - 2 = 0)$. When trying to determine the truth value of the proposition $\exists x\, P(x)$, it is important to keep the domain of discourse for the variable in mind. For example, if the domain for x in $\exists x(x^2 - 2 = 0)$ is all integers, the proposition is false. But if its domain is all real numbers, the proposition is true. The phrase *There is an* x *such that* (or, in symbols, $\exists x$) is called **existential quantification**. In English it can also be read as *There exists* x or *For some* x.

The third and final way to convert a predicate into a proposition is by **universal quantification**. The phrase *For all* x is also rendered in English as *For each* x or *For every* x. The universal quantification of a predicate, $P(x)$, is obtained by preceding the predicate with the phrase *For all* x, producing the proposition *For all* x, $P(x)$, or, in symbols, $\forall x\, P(x)$. This proposition is true provided the predicate becomes a true proposition for every object in the variable's domain of discourse. Again, it is important to know the domain of discourse for the variable since the domain will have an effect on the truth value of the quantified proposition in general.

For multi-placed predicates, these three conversions can be mixed and matched. For example, using the obvious domains for the predicate $B(x, y)$: *x is the brother of y* here are some conversions into propositions:

(1) $B(\text{Donny}, \text{Marie})$ has both variables instantiated. The proposition is true.

(2) $\exists y\, B(\text{Donny}, y)$ is also a true proposition. It says *Donny* is somebody's brother. The first variable was instantiated, the second was existentially quantified.

(3) $\forall y\, B(\text{Donny}, y)$ says everyone has Donny for a brother, and that is false.

(4) $\forall x\, \exists y\, B(x, y)$ says every male is somebody's brother, and that is false.

(5) $\exists y\, \forall x\, B(\text{x}, \text{y})$ says there is a person for whom every male is a brother, and that is false.

(6) $\forall x\, B(x, x)$ says every male is his own brother, and that is false.

Translation between ordinary language and symbolic language can get a little tricky when quantified statements are involved. Here are a few more examples.

Example 3.1. *Let $P(x)$ be the predicate x* ***owns a Porsche****, and let $S(x)$ be the predicate x* ***speeds****. The domain of discourse for the variable in each predicate will be the collection of all drivers. The proposition $\exists x\, P(x)$ says* ***Someone owns a Porsche****. It could also be translated as* ***There is a person*** *x* ***such that*** *x* ***owns a Porsche****, but that sounds too stilted for ordinary conversation. A smooth translation is better. The proposition $\forall x(P(x) \rightarrow S(x))$ says* ***All Porsche owners speed.***

Translating in the other direction, the proposition ***No speeder owns a Porsche*** *could be expressed as $\forall x(S(x) \rightarrow \neg P(x))$.*

Example 3.2. *Here's a more complicated example: translate the proposition* ***Al knows only Bill*** *into symbolic form. Let's use $K(x, y)$ for the predicate x* ***knows*** *y. The translation would be $K(\text{Al, Bill}) \wedge \forall x\, (K(\text{Al}, x) \rightarrow (x = \text{Bill}))$.*

Example 3.3. *For one last example, let's translate* ***The sum of two even integers is even*** *into symbolic form. Let $E(x)$ be the predicate x* ***is even****. As with many statements in ordinary language, the proposition is phrased in a shorthand code that the reader is expected to unravel. As given, the statement doesn't seem to have any quantifiers, but they are implied. Before converting it to symbolic form, it might help to expand it to its more long winded version:* ***For every choice of two integers, if they are both even, then their sum is even****. Expressed this way, the translation to symbolic form is: $\forall x\, \forall y\, ((E(x) \wedge E(y)) \rightarrow E(x + y))$.*

Notice that if the domain of discourse consists of finitely many entries $a_1, \dots, a_n$, then $\forall x\, p(x) \equiv p(a_1) \vee p(a_2) \vee \ldots \vee p(a_n)$. So the quantifier $\forall$ can be expressed in terms of the logical connective $\wedge$. The existential quantifier and $\wedge$ are similarly linked: $\exists x\, p(x) \equiv p(a_1) \vee p(a_2) \vee \ldots \vee p(a_n)$.

From the associative and commutative laws of logic we see that we can rearrange any system of propositions which are linked only by ∧'s or linked only by ∨'s. For instance, consider the previous examples 3.1 – 3.3 with finite domains of discourse. Consequently any more generally quantified proposition of the form $\forall x \forall y\, p(x,y)$ is logically equivalent to $\forall y \forall x\, p(x,y)$. Similarly for statements which contain only existential quantifiers. But the distributive laws come into play when ∧'s and ∨'s are mixed. So care must be taken with predicates which contain both existential and universal quantifiers, as the following example shows.

Example 3.4. *Let $p(x,y)\colon x+y=0$ and let the domain of discourse be all real numbers for both x and y. The proposition $\forall y\, \exists x\, p(x,y)$ is true, since, for any given* y*, by setting (instantiating)* $x = -y$ *we convert $x+y=0$ to the true statement $(-y)+y=0$. In fact $(\forall y \in \mathbb{R})[(-y)+y=0]$ is a tautology. However the proposition $\exists x\, \forall y\, p(x,y)$ is false. If we set (instantiate) $y=1$, then $x+y=0$ implies $x=-1$. When we set $y=0$, we get $x=0$. Since $0 \neq -1$ there is no x which will work for all y, since it would have to work for the specific values of $y = 0$ and $y = 1$.*

To form the negation of quantified statements, we apply De Morgan's laws. This can be seen in case of a finite domain of discourse as follows:

$$\begin{aligned} \neg(\forall x\, p(x)) &\equiv \neg(p(a_1) \wedge p(a_2) \wedge \ldots \wedge p(a_n)) \\ &\equiv \neg p(a_1) \vee \neg p(a_2) \vee \cdots \vee \neg p(a_n) \\ &\equiv \exists x\, \neg p(x) \end{aligned}$$

In the same way, we have $\neg(\exists x\, p(x)) \equiv (\forall x\, \neg p(x))$.

Exercises

Exercise 3.1. *Let* $p(x)$: $2x \geq 4$*, for integers* x*. Determine the truth values of the following propositions.*

a) p(2)

b) $p(-3)$

c) $\forall x\,((x \leq 10) \rightarrow p(x))$

d) $\exists x\,\neg p(x)$

Exercise 3.2. *Let* $p(x, y)$ *be* x *has read* y*, where the domain of discourse for* x *is all students in this class, and the domain of discourse for* y *is all novels. Express the following propositions in English.*

a) $\forall x\, p(x,$ *War and Peace*$)$

b) $\exists x\,\neg p(x,$ *The Great Gatsby*$)$

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

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

Exercise 3.3. *Let* $F(x, y)$ *be the statement* x *can fool* y*, where the domain of discourse for both* x *and* y *is all people. Use quantifiers to express each of the following statements.*

a) *I can fool everyone.*

b) *George can't fool anybody.*

c) *No one can fool himself.*

d) *There is someone who can fool everybody.*

e) *There is someone everyone can fool.*

f) *Ralph can fool two different people.*

Exercise 3.4. *Negate each of the statements from exercise 3.2 in English.*

Exercise 3.5. *Negate each statement from exercise 3.3 in logical symbols. Of course, the easy answer would be to simply put* $\neg$ *in front of each statement. But use the principle given at the end of this chapter to move the negation across the quantifiers.*

Exercise 3.6. *Express symbolically: The product of an even integer and an odd integer is even.*

Exercise 3.7. *Express in words the meaning of*

$$\exists x\, P(x) \wedge \forall x\, \forall y\; ((P(x) \wedge P(y)) \rightarrow (x = y))\,.$$

Problems

Problem 3.1. *Let h be Ben is healthy., w: Ben is wealthy., and s: Ben is wise. Express the following in English:*

a) $h \wedge w$

b) $w \vee s$

c) $h \rightarrow (w \wedge s)$

d) $(h \rightarrow w) \wedge s$

Problem 3.2. *Let $S(x, y)$ be the predicate x has seen y where the domain of discourse for x is all students in this class and the domain of discourse for y is all movies. Express the following in logical symbols using quantifiers.*

a) *Every student in this class has seen The Avengers.*

b) *No student in this class has seen Life of Pi.*

c) *Hidden Figures has been seen by someone in this class.*

d) *Some students have seen every movie.*

e) *For each movie, there is at least one student in the class who has seen that movie.*

Problem 3.3. *Negate the propositions in 3.1 in English.*

Problem 3.4. *Negate the propositions in 3.2 in symbols. Note: An easy way to do this is to simply write $\neg$ in front of the answers in 3.2. Don't do that! Give the negation with no quantifiers coming after a negation symbol.*

Problem 3.5. *Negate the propositions in 3.2 in English. Note: An easy way to do this is to simply write It is not the case that .... in front of each proposition. Don't do that! Give the negation as a reasonably natural English sentence.*

Chapter 4

Rules of Inference

The heart of mathematics is proof. In this chapter, we give a careful description of what exactly constitutes a proof in the realm of propositional logic. Throughout the course various methods of proof will be demonstrated, including the particularly important style of proof called *induction*. It's important to keep in mind that all proofs, no matter what the subject matter might be, are based on the notion of a valid argument as described in this chapter, so the ideas presented here are fundamental to all of mathematics.

Imagine trying carefully to define what a proof is, and it quickly becomes clear just how difficult a task that is. So it shouldn't come as a surprise that the description takes on a somewhat technical looking aspect. But don't let all the symbols and abstract-looking notation be misleading. All these rules really boil down to plain old common sense when looked at correctly.

The usual form of a theorem in mathematics is: If a is true and b is true and c is true, etc., then s is true. The $a, b, c, \cdots$ are called the **hypotheses**, and the statement s is called the **conclusion**. For example, a mathematical theorem might be: if m is an even integer and n is an odd integer, then mn is an even integer. Here the hypotheses are *m is an even integer* and *n is and odd integer*, and the conclusion is *mn is an even integer*.

We begin by concerning ourselves with proofs from the realm of propositional logic rather than the sort of theorem mentioned above. We will be interested in arguments in which the **form** of the argument is the item of interest rather than the **content** of the statements in the argument.

For example, consider the simple argument: (1) *My car is either red or blue* and (2) *My car is not red*, and so (3) *My car is blue*. Here the hypotheses are (1) and (2), and the conclusion is (3). It should be clear that this is a **valid argument**. That means that if you agree that (1) and (2) are true, then you *must* accept that (3) is true as well.

Definition 4.1. An argument is called **valid** provided that if you agree that all the hypotheses are true, then you must accept the truth of the conclusion.

Now the content of that argument (in other words, the stuff about *my* and *cars* and *colors*) really have nothing to do with the validity of the argument. It is the **form** of the argument that makes it valid. The form of this argument is (1) $p \vee q$ and (2) $\neg p$, therefore (3) q. Any argument that has this form is valid, whether it talks about cars and colors or any other notions. For example, here is another argument of the very same form: (1) *I either read the book or just looked at the pictures* and (2) *I didn't read the book*, therefore (3) *I just looked at the pictures*.

Some arguments involve quantifiers. For instance, consider the classic example of a logical argument: (1) *All men are mortal* and (2) *Socrates is a man*, and so (3) *Socrates is mortal*. Here the hypotheses are the statements (1) and (2), and the conclusion is statement (3). If we let M(x) be x *is a man* and D(x) be x *is mortal* (with domain for x being everything!), then this argument could be symbolized as shown.

$$\begin{array}{l} \forall x(M(x) \rightarrow D(x)) \\ M(\text{Socrates}) \\ \hline \therefore D(\text{Socrates}) \end{array}$$

The general form of a proof that a logical argument is valid consists in assuming all the hypotheses have truth value T, and showing, by applying valid rules of logic, that the conclusion must also have truth value T.

Just what are the valid rules of logic that can be used in the course of the proof? They are called the Rules of Inference, and there are seven of them listed in table 4.1. Each rule of inference arises from a tautology, and actually there is no end to the rules of inference, since each new tautology can be used to provide a new rule of inference. But, in real life, people rely on only a few basic rules of inference, and the list provided in the table is plenty for all normal purposes.

It is important not to merely look on these rules as marks on the page, but rather to understand what each one says in words. For example, Modus Ponens corresponds to the common sense rule: if we are told *p is true*, and also *If p is true, then so is q*, then we would leap to the reasonable conclusion that

Name	Rule of Inference	
Modus Ponens	p and $p \to q$	$\therefore q$
Modus Tollens	$\neg q$ and $p \to q$	$\therefore \neg p$
Hypothetical Syllogism	$p \to q$ and $q \to r$	$\therefore p \to r$
Addition	p	$\therefore p \vee q$
Simplification	$p \wedge q$	$\therefore p$
Conjunction	p and q	$\therefore p \wedge q$
Disjunctive Syllogism	$p \vee q$ and $\neg p$	$\therefore q$

Table 4.1: Basic rules of inference

q is true. That is all Modus Ponens says. Similarly, for the rule of proof of Disjunctive Syllogism: knowing *Either p or q is true*, and *p is not true*, we would immediately conclude *q is true*. That's the rule we applied in the *car* example above. Translate the remaining six rules of inference into such common sense statements. Some may sound a little awkward, but they ought to all elicit an *of course that's right* feeling once understood. Without such an understanding, the rules seem like a jumble of mystical symbols, and building logical arguments will be pretty difficult.

What exactly goes into a logical argument? Suppose we want to prove (or show valid) an argument of the form *If a and b and c are true, then so is s.* One way that will always do the trick is to construct a truth table as in examples earlier in the course. We check the rows in the table where all the hypotheses are true, and make sure the conclusion is also true in those rows. That would complete the proof. In fact that is exactly the method used to justify the seven rules of inference given in the table. But building truth tables is certainly tedious business, and it certainly doesn't seem too much like the way we learned to do proofs in geometry, for example. An alternative is the construction of a logical argument which begins by assuming the hypotheses are all true and applies the basic rules of inferences from the table until the desired conclusion is shown to be true.

Here is an example of such a proof. Let's show that the argument displayed in figure 4.1 is valid.

$$\begin{array}{l} p \\ p \to q \\ s \vee r \\ \underline{r \to \neg q} \\ \therefore s \vee t \end{array}$$

Figure 4.1: A logical argument

Argument:

p
$p \to q$
$s \vee r$
$r \to \neg q$
$\therefore s \vee t$

Proof:

(1)	p	hypothesis
(2)	$p \to q$	hypothesis
(3)	q	Modus Ponens (1) and (2)
(4)	$r \to \neg q$	hypothesis
(5)	$q \to \neg r$	logical equivalent of (4)
(6)	$\neg r$	Modus Ponens (3) and (5)
(7)	$s \vee r$	hypothesis
(8)	$r \vee s$	logical equivalence of (7)
(9)	s	Disjunctive Syllogism (6) and (8)
(10)	$s \vee t$	Addition

Table 4.2: Proof of the validity of an argument

Each step in the argument will be justified in some way, either

(1) as a hypothesis (and hence assumed to have truth value T), or

(2) as a consequence of previous steps and some rule of inference from the table, or

(3) as a statement logically equivalent to a previous statement in the proof.

Finally the last statement in the proof will be the desired conclusion. Of course, we could prove the argument valid by constructing a 32 row truth table instead! Well, actually we wouldn't need all 32 rows, but it would be pretty tedious in any case.

Such proofs can be viewed as games in which the hypotheses serve as the starting position in a game, the goal is to reach the conclusion as the final position in the game, and the rules of inference (and logical equivalences) specify the legal moves. Following this outline, we can be sure every step in the proof is a true statement, and, in particular, the desired conclusion is true, as we hoped to show.

One step more complicated than the last example are arguments that are presented in words rather than symbols. In such a case, it is necessary to first convert from a verbal argument to a symbolic argument, and then check the argument to see if it is valid. For example, consider the argument: *Tom is a cat. If Tom is a cat, then Tom likes fish. Either Tweety is a bird or Fido is a dog. If Fido is a dog, then Tom does not like fish. So, either Tweety is a bird or I'm a monkey's uncle.* Just reading this argument, it is difficult to decide if it is valid or not. It's just a little too confusing to process. But it is valid, and in fact it is the very same argument as given above. Let p be *Tom is a cat*, let q be *Tom likes fish*, let s be *Tweety is a bird*, let r be *Fido is a dog*, and let t be *I'm a*

Name	Instantiation Rules
Universal Instantiation	$\forall x P(x)\ \therefore\ P(c)$ if c is in the domain of x
Existential Instantiation	$\exists x P(x)\ \therefore\ P(c)$ for some c in the domain of x

Name	Generalization Rules
Universal Generalization	$P(c)$ for **arbitrary** c in the domain of x $\therefore\ \forall x\, P(x)$
Existential Generalization	$P(c)$ for some c in the domain of x $\therefore\ \exists x\, P(x)$

Table 4.3: Quantification rules

monkey's uncle. Expressing the statements in the argument in terms of p, q, r, s, t produces exactly the symbolic argument proved above.

Some logical arguments have a convincing ring to them but are nevertheless invalid. The classic example is an argument of the form *If it is snowing, then it is winter. It is winter. So it must be snowing.* A moment's thought is all that is needed to be convinced the conclusion does not follow from the two hypotheses. Indeed, there are many winter days when it does not snow. The error being made is called the **fallacy of affirming the conclusion**. In symbols, the argument is claiming that $[(p \to q) \wedge q] \to p$ is a tautology, but in fact, checking a truth table shows that it is not a tautology. Fallacies arise when statements that are not tautologies are treated as if they were tautologies.

Logical arguments involving propositions using quantifiers require a few more rules of inference. As before, these rules really amount to no more than a formal way to express common sense. For instance, if the proposition $\forall x\, P(x)$ is true, then certainly for every object c in the universe of discourse, $P(c)$ is true. After all, if the statement $P(x)$ is true for every possible choice of x, then, in particular, it is true when $x = c$. The other three rules of inference for quantified statements are just as obvious. All four quantification rules appear in table 4.3.

Example 4.2. *Let's analyze the following (fictitious, but obviously valid) argument to see how these rules of inference are used. All books written by Sartre are hard to understand. Sartre wrote a book about kites. So, there is a book about kites that is hard to understand. Let's use to following predicates to symbolize the argument:*

(1) $S(x)$: x was written by Sartre.

(2) $H(x)$: x is hard to understand.

(3) $K(x)$: x is about kites.

The domain for x in each case is all books. In symbolic form, the argument and a proof are

Argument: $\forall x(S(x) \rightarrow H(x))$

$\underline{\exists x(S(x) \wedge K(x))}$

$\therefore \exists x(K(x) \wedge H(x))$

Proof:

1) $\exists x(S(x) \wedge K(x))$	hypothesis
2) $S(c) \wedge K(c)$ for some c	Existential Instantiation (1)
3) $S(c)$	Simplification (2)
4) $\forall x(S(x) \rightarrow H(x))$	hypothesis
5) $S(c) \rightarrow H(c)$	Universal Instantiation (4)
6) $H(c)$	Modus Ponens (3) and (5)
7) $K(c) \wedge S(c)$	logical equivalence (2)
8) $K(c)$	Simplification (7)
9) $K(c) \wedge H(c)$	Conjunction (8) and (6)
10) $\exists x(K(x) \wedge H(x))$	Existential Generalization (9)

Exercises

Exercise 4.1. *Show* $p \vee q$ *and* $\neg p \vee r, \therefore q \vee r$ *is a valid rule of inference. It is called* ***Resolution***.

Exercise 4.2. *Prove the following argument is valid. All Porsche owners are speeders. No owners of sedans buy premium fuel. Car owners that do not buy premium fuel never speed. So Porsche owners do not own sedans. Use all car owners as the domain of discourse.*

Exercise 4.3. *Prove the following symbolic argument is valid.*

$$\begin{array}{l} \neg p \rightarrow (r \wedge \neg s) \\ t \rightarrow s \\ u \rightarrow \neg p \\ \neg w \\ \underline{u \vee w \qquad\qquad} \\ \therefore \neg t \vee w \end{array}$$

Problems

Problem 4.1. *Show that* $p \rightarrow q$ *and* $\neg p, \quad \therefore \neg q$ *is not a valid rule of inference. It is called the* ***Fallacy of denying the hypothesis***.

Problem 4.2. *Prove the following symbolic argument is valid.*

$$\begin{array}{l} \neg p \wedge q \\ r \rightarrow p \\ \neg r \rightarrow s \\ \underline{s \rightarrow t \quad} \\ \therefore t \end{array}$$

Problem 4.3. *Prove the following symbolic argument is valid.*

$$
\begin{array}{l}
p \vee q \\
q \rightarrow r \\
(p \wedge s) \rightarrow t \\
\neg r \\
\underline{\neg q \rightarrow (p \wedge s)} \\
\therefore t
\end{array}
$$

Problem 4.4. *Prove the following symbolic argument is valid.*

$$
\begin{array}{l}
(\neg p \vee q) \rightarrow r \\
s \vee \neg q \\
\neg t \\
p \rightarrow t \\
\underline{(\neg p \wedge r) \rightarrow \neg s} \\
\therefore \neg q
\end{array}
$$

Problem 4.5. *Express the following argument is symbolic form and prove the argument is valid. If Ralph doesn't do his homework or he doesn't feel sick, then he will go to the party and he will stay up late. If he goes to the party, he will eat too much. He didn't eat too much. So Ralph did his homework.*

Problem 4.6. *In problem 4.5, show that you can logically deduce that Ralph felt sick.*

Problem 4.7. *In problem 4.5, can you logically deduce that Ralph stayed up late?*

Problem 4.8. *Prove the following symbolic argument.*

$$
\begin{array}{l}
\exists x(A(x) \wedge \neg B(x)) \\
\underline{\forall x(A(x) \rightarrow C(x))} \\
\therefore \exists x(C(x) \wedge \neg B(x))
\end{array}
$$

Chapter 5

Sets: Basic Definitions

A **set** is a collection of objects. Often, but not always, sets are denoted by capital letters such as $A, B, \cdots$ and the objects that make up a set, called its **elements**, are denoted by lowercase letters. Write $x \in A$ to mean that the object x is an element of A. If the object x is not an element of A, write $x \notin A$.

Two sets A and B are **equal**, written $A = B$ provided A and B comprise exactly the same elements. Another way to say the same thing: $A = B$ provided $\forall x\,(x \in A \iff x \in B)$.

There are a number of ways to specify a given set. We consider two of them.

One way to describe a set is to list its elements. This is called the **roster method**. Braces are used to signify when the list begins and where it ends, and commas are used to separate elements. For instance, $A = \{1, 2, 3, 4, 5\}$ is the set of positive whole numbers between 1 and 5 inclusive. It is important to note that the order in which elements are listed is immaterial. For example, $\{1, 2\} = \{2, 1\}$ since $x \in \{1, 2\}$ and $x \in \{2, 1\}$ are both true for $x = 1$ and $x = 2$ and false for all other choices of x. Thus $x \in \{1, 2\}$ and $x \in \{2, 1\}$ always have the same truth value, and that means $\forall x\,(x \in \{1, 2\} \Leftrightarrow x \in \{2, 1\})$ is true. According to the definition of equality given above, it follows that $\{1, 2\} = \{2, 1\}$. The same sort of reasoning shows that repetitions in the list of elements of a set can be ignored. For example $\{1, 2, 3, 2, 4, 1, 2, 3, 2\} = \{1, 2, 3, 4\}$. There is no point in listing an element of a set more than once.

The roster method has certain drawbacks. For example we probably don't want to list all of the elements in the set of positive integers between 1 and 99 inclusive. One option is to use an **ellipsis**. The idea is that we list elements until a pattern is established, and then replace the missing elements with ... (which is the ellipsis). So $\{1, 2, 3, 4, \ldots, 99\}$ would describe our set.

The use of an ellipsis has one pitfall. It is **hoped** that whoever is reading the list will be able to guess the proper pattern and apply it to fill in the gap.

Another method to specify a set is via the use of **set-builder notation**. A set can be described in set-builder notation as $A = \{x | p(x)\}$. Here we read A is the set of all objects x for which the predicate $p(x)$ is true. So $\{1, 2, 3, 4, \ldots, 99\}$ becomes $\{x | x$ is a whole number and $1 \leq x \leq 99\}$.

Certain sets occur often enough that we have special notation for them.

$\mathbb{N} = \{x | x$ is a non-negative whole number$\} = \{0, 1, 2, \ldots\}$, **the natural numbers.**

$\mathbb{Z} = \{x | x$ is a whole number$\} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$, **the integers.**

$\mathbb{Q} = \{x | x = p/q, p, q$ integers and $q \neq 0\}$ **the rational numbers.**

$\mathbb{R} = \{x | x$ is a real number$\}$, **the real numbers.**

$\mathbb{C} = \{x | x = a + ib, a, b \in \mathbb{R}, i^2 = -1\}$ **the complex numbers.**

In addition to the above sets, there is a set with no elements, written as $\emptyset$ (also written using the roster style as { }), and called the **empty set**. This set can be described using set builder style in many different ways. For example, $\{x \in \mathbb{R} | x^2 = -2\} = \emptyset$. In fact, if P(x) is any predicate which is always false, then $\{x \mid P(x)\} = \emptyset$. There are two easy slips to make involving the empty set. First, don't write $\emptyset = 0$ (the idea being that both $\emptyset$ and 0 represent *nothing*). That is not correct since $\emptyset$ is a set, and 0 is a number, and it's not fair to compare two different types of objects. The other error is thinking $\emptyset = \{\emptyset\}$. This cannot be correct since the right-hand set has an element, but the left-hand set does not.

At the other extreme from the empty set is the **universal set**, denoted $\mathcal{U}$. The universal set consists of all objects under consideration in any particular discussion. For example, if the topic du jour is basic arithmetic then the universal set would be the set of all integers. Usually the universal set is left for the reader to guess. If the choice of the universal set is not an obvious one, it will be pointed out explicitly.

The set A is a **subset** of the set B, written as $A \subseteq B$, in case $\forall x(x \in A \rightarrow x \in B)$ is true. In plain English, $A \subseteq B$ if every element of A also is an element of B. For example, $\{1,2,3\} \subseteq \{1,2,3,4,5\}$. On the other hand, $\{0,1,2,3\} \not\subset \{1,2,3,4,5\}$ since 0 is an element of the left-hand set but not of the right-hand set. The meaning of $A \not\subset B$ can be expressed in symbols using De Morgan's law:

$$A \not\subset B \iff \neg(\forall x(x \in A \rightarrow x \in B)) \tag{5.1}$$

$$\equiv \exists x \neg(x \in A \rightarrow x \in B) \tag{5.2}$$

$$\equiv \exists x \neg(\neg(x \in A) \vee x \in B) \tag{5.3}$$

$$\equiv \exists x(x \in A \wedge x \notin B) \tag{5.4}$$

and the last line says $A \not\subset B$ if there is at least one element of A that is not an element of B.

The empty set is a subset of every set. To check that, suppose A is any set, and let's check to make sure $\forall x(x \in \emptyset \rightarrow x \in A)$ is true. But it is since for any x, the hypothesis of $x \in \emptyset \longrightarrow x \in A$ is F, and so the implication is T. So $\emptyset \subseteq A$. Another way to same the same thing is to notice that to claim $\emptyset \not\subset A$ is the same as claiming there is at least one element of $\emptyset$ that is not an element of A, but that is ridiculous, since $\emptyset$ has no elements at all.

To say that $A = B$ is the same as saying every element of A is also an element of B and every element of B is also an element of A. In other words, $A = B \iff (A \subseteq B \wedge B \subseteq A)$, and this indicates the method by which the common task of showing two sets are equal is carried out: to show two sets are equal, show that each is a subset of the other.

If $A \subseteq B$, and $A \neq B$, A is a **proper subset** of B, denoted by $A \subset B$, or $A \subsetneq B$. In words, $A \subset B$ means every element of A is also an element of B and there is at least one element of B that is not an element of A. For example $\{1,2\} \subset \{1,2,3,4,5\}$, and $\emptyset \subset \{1\}$.

A set is **finite** if the number of distinct elements in the set is a non-negative integer. In this case we call the number of distinct elements in the set its **cardinality** and denote this natural number by $|A|$. For example, $|\{1,3,5\}| = 3$ and $|\emptyset| = 0$, $|\{\emptyset\}| = 1$, and $|\{\emptyset, \{a, b, c\}, \{X, Y\}\}| = 3$. A set, such as $\mathbb{Z}$, which is not finite, is **infinite**.

Given a set A the **power set of** A, denoted $P(A)$, is the set of all subsets of A. For example if $A = \{1,2\}$, then $P(A) = \{\emptyset, \{1\}, \{2\}, \{1,2\}\}$. It is not hard to see that if $|A| = n$, then $|P(A)| = 2^n$.

Exercises

Exercise 5.1. *List the members of the following sets.*

a) $\{x \in \mathbb{Z} | 3 \leq x^3 < 100\}$

b) $\{x \in \mathbb{R} | 2x^2 = 50\}$

c) $\{x \in \mathbb{N} | 7 > x \geq 4\}$

Exercise 5.2. *Use set-builder notation to give a description of each set.*

a) $\{-5, 0, 5, 10, 15\}$

b) $\{0, 1, 2, 3, 4\}$

c) The interval of real numbers: $[\pi, 4)$

Exercise 5.3. *Determine the cardinality of the sets in exercises 5.1 and 5.2.*

Exercise 5.4. *Is the proposition Every element of the empty set has three toes true or false? Explain your answer!*

Exercise 5.5. *Determine the power set of* $\{1, \{2\}\}$.

Exercise 5.6. *True or False:* $\{1, 2, 3, 4, 5\} = \{5, 2, 3, 1, 2, 1, 5, 4, 3, 2, 1\}$.

Exercise 5.7. *True or False: The set of even integers is a subset of the set of integers that are multiples of four.*

Problems

Problem 5.1. *List the members of the following sets.*

a) $\{x \in \mathbb{N} | 3 \leq x^2 < 100\}$

b) $\{x \in \mathbb{Z} | 3 \leq x^2 < 100\}$

c) $\{x \in \mathbb{R} | 0 < x \leq 5\}$

d) $\{x \in \mathbb{N} | x^2 < 0\}$

e) $\{1, \{1\}, \{1, 2\}\}$

Problem 5.2. *Determine the power set of* $\{0, 1, 2\}$.

Problem 5.3. *Determine the truth value of the following propositions:*

a) Every element of the empty set is positive.

b) Some element of the empty set is positive.

c) $0 \in \emptyset$.

d) $0 = \emptyset$.

Problem 5.4. *Determine the truth value of the following propositions:*

a) For any set A, $A \subseteq P(A)$.

b) For any set A, $A \in P(A)$.

c) For any finite set A, *the cardinality of* $P(A)$ *is greater than the cardinality of* A.

Problem 5.5. *Determine the cardinality of the following sets:*

a) The empty set.

b) The power set of the empty set.

c) The power set of the power set of the empty set.

d) The power set of $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$.

e) $\{1, \{1\}, \{1, 2\}\}$

Problem 5.6. *True or False:*

a) If $n \in \mathbb{N}$, *then* $n + 1 \in \mathbb{N}$.

b) If $n \in \mathbb{N}$, *then* $n - 1 \in \mathbb{N}$.

Chapter 6

Set Operations

There are several ways of combining sets to produce new sets.

The **intersection** of A with B denoted $A \cap B$ is defined as $\{x | x \in A \wedge x \in B\}$. For example $\{1, 2, 3, 4, 5\} \cap \{1, 3, 5, 7, 9\} = \{1, 3, 5\}$. So the intersection of two sets consists of the objects which are in both sets simultaneously. Two sets are **disjoint** if $A \cap B = \emptyset$.

Set operations can be visualized using **Venn diagrams**. A circle (or other closed curve) is drawn to represent a set. The points inside the circle are used to stand for the elements of the set. To represent the set operation of intersection, two such circles are drawn with an overlap to indicate the two sets may share some elements. In the Venn diagram, figure 6.1, the shaded area represents the intersection of A and B.

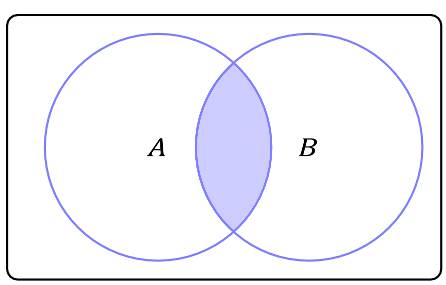


Figure 6.1: Venn diagram for $A \cap B$

The **union** of A with B denoted $A \cup B$ is $\{x | x \in A \vee x \in B\}$. In words, $A \cup B$ consists of those elements that appear in at least one of A and B. So

$$\{1, 2, 3, 4, 5\} \cup \{1, 3, 5, 7, 9\} = \{1, 2, 3, 4, 5, 7, 9\},$$

for example. The Venn Diagram 6.2 represents the union of A and B.

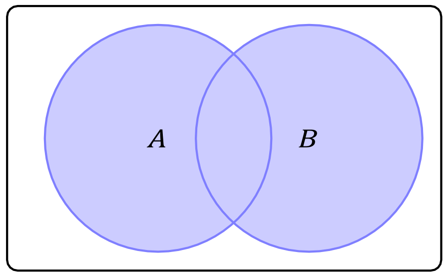


Figure 6.2: Venn diagram for $A \cup B$

The **symmetric difference** of A and B is defined to be $A \oplus B = \{x | x \in A \oplus x \in B\}$. So $A \oplus B$ consists of those elements which appear in exactly one of A and B. For example

$$\{1, 2, 3, 4, 5\} \oplus \{1, 3, 5, 7, 9\} = \{2, 4, 7, 9\}.$$

The Venn diagram for the symmetric difference is figure 6.3.

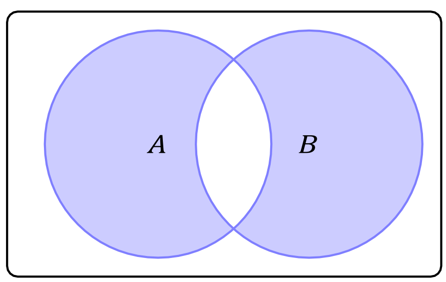


Figure 6.3: Venn diagram for $A \oplus B$

The **complement of** B **relative to** A, denoted $A - B$ is $\{x | x \in A \wedge x \notin B\}$. So $\{1, 2, 3, 4, 5\} - \{1, 3, 5, 7, 9\} = \{2, 4\}$. Figure 6.4 is the corresponding Venn diagram.

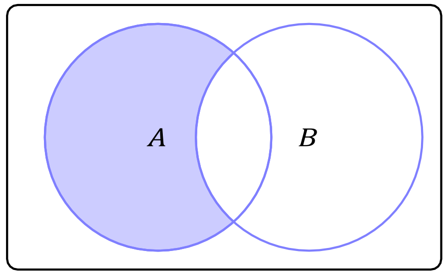


Figure 6.4: Venn diagram for $A - B$

When $\mathcal{U}$ is a universal set, we denote $\mathcal{U} - A$ by $\bar{A}$ and call it the **complement** of A. The Venn diagram for is in figure 6.5. If $\mathcal{U} = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$, then $\overline{\{0,1,2,3,4\}} = \{5,6,7,8,9\}$. The universal set matters here. If $\mathcal{U} = \{x \in \mathbb{N} | x \leq 100\}$, then $\overline{\{0,1,2,3,4\}} = \{5,6,7, \ldots ,100\}$.

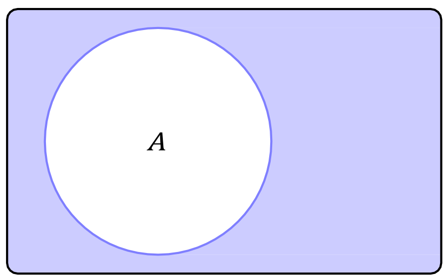


Figure 6.5: Venn diagram for $\bar{A} = \mathcal{U} - A$

There is a close connection between many set operations and the logical connectives of Chapter 1. The intersection operation is related to conjunction, union is related to disjunction, and complementation is related to negation. It is not surprising then that the various laws of logic, such as the associative, commutative, and distributive laws carry over to analogous laws for the set operations. Table 6.1 exhibits some of these properties of these set operations.

These can be verified by using **membership tables** which are the analogs of truth tables used to verify the logical equivalence of propositions. For a set A either an element under consideration is in A or it is not. These binary possibilities are kept track of using 1 if $x \in A$ and 0 if $x \notin A$ and then performing related bit string operations.

IDENTITY	**NAME**
$\overline{(\overline{\boldsymbol{A}})} = \boldsymbol{A}$	Double Negation
$\boldsymbol{A} \cap \boldsymbol{\mathcal{U}} = \boldsymbol{A}$	
$\boldsymbol{A} \cup \emptyset = \boldsymbol{A}$	Identity Laws
$\boldsymbol{A} \cap \emptyset = \emptyset$	
$\boldsymbol{A} \cup \boldsymbol{\mathcal{U}} = \boldsymbol{\mathcal{U}}$	Domination Laws
$\boldsymbol{A} \cap \boldsymbol{A} = \boldsymbol{A}$	
$\boldsymbol{A} \cup \boldsymbol{A} = \boldsymbol{A}$	Idempotent Laws
$\boldsymbol{A} \cup \boldsymbol{B} = \boldsymbol{B} \cup \boldsymbol{A}$	
$\boldsymbol{A} \cap \boldsymbol{B} = \boldsymbol{B} \cap \boldsymbol{A}$	Commutative Laws
$(\boldsymbol{A} \cup \boldsymbol{B}) \cup \boldsymbol{C} = \boldsymbol{A} \cup (\boldsymbol{B} \cup \boldsymbol{C})$	
$\boldsymbol{A} \cap (\boldsymbol{B} \cap \boldsymbol{C}) = \boldsymbol{A} \cap (\boldsymbol{B} \cap \boldsymbol{C})$	Associative Laws
$\boldsymbol{A} \cap (\boldsymbol{B} \cup \boldsymbol{C}) = (\boldsymbol{A} \cap \boldsymbol{B}) \cup (\boldsymbol{A} \cap \boldsymbol{C})$	
$\boldsymbol{A} \cup (\boldsymbol{B} \cap \boldsymbol{C}) = (\boldsymbol{A} \cup \boldsymbol{B}) \cap (\boldsymbol{A} \cup \boldsymbol{C})$	Distributive Laws
$\overline{(\boldsymbol{A} \cap \boldsymbol{B})} = \overline{\boldsymbol{A}} \cup \overline{\boldsymbol{B}}$	
$\overline{(\boldsymbol{A} \cup \boldsymbol{B})} = \overline{\boldsymbol{A}} \cap \overline{\boldsymbol{B}}$	De Morgan's Laws
$\boldsymbol{A} \cup \overline{\boldsymbol{A}} = \boldsymbol{\mathcal{U}}$	Law of Excluded Middle
$\boldsymbol{A} \cap \overline{\boldsymbol{A}} = \emptyset$	Law of Contradiction

Table 6.1: Laws of Set Theory

Example 6.1. *Verify the De Morgan's law given by* $\overline{A \cap B} = \bar{A} \cup \bar{B}$.

The meaning of the first row of the table is that if $x \in A$ *and* $x \in B$, *then* $x \notin A \cap B$, *as indicated by the* 0 *in the first row, fourth column, and also not in* $A \cup B$ *as indicated by the* 0 *in the first row, last column. Since the columns for* $\overline{A \cap B}$ *and* $\bar{A} \cup \bar{B}$ *are identical, it follows that* $\overline{A \cap B} = \bar{A} \cup \bar{B}$ *as promised.*

Just as compound propositions can be analyzed using truth tables, more complicated combinations of sets can be handled using membership tables. For example, using a membership table, it is

A	B	$A \cap B$	$\overline{(A \cap B)}$	$\overline{A}$	$\overline{B}$	$\overline{A} \cup \overline{B}$
1	1	1	0	0	0	0
1	0	0	1	0	1	1
0	1	0	1	1	0	1
0	0	0	1	1	1	1

easy to verify that $\overline{A \cup (B \cap C)} = \bar{A} \cap (\bar{B} \cup \bar{C})$. But, just as with propositions, it is usually more enlightening to verify such equalities by applying the few basic laws of set theory listed above.

Example 6.2. *Let's prove* $\overline{A \cup (B \cap C)} = \bar{A} \cap (\bar{B} \cup \bar{C})$

Proof. The proof is just two applications of De Morgan's laws:

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

□

There is a correspondence between set operations of finite sets and bit string operations. Let $\mathcal{U} = \{u_1, u_2, \ldots, u_n\}$ be a finite universal set with distinct elements listed in a specific order. Notice the universal set is **ordered**. We may write it as an n −tuple: $\mathcal{U} = (u_1, u_2, \ldots, u_n)$. For a set A under consideration, we have $A \subseteq \mathcal{U}$. By the law of excluded middle, for each $u_j \in \mathcal{U}$, either $u_j \in A$ or $u_j \notin A$. We define a binary string of length n, called the **characteristic vector** of A, denoted $\chi(A)$, by setting the jth bit of $\chi(A)$ to be 1 if $u_j \in A$ and 0 if $u_j \notin A$. For example if $\mathcal{U} = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$, and $A = \{1, 3, 4, 5, 8\}$, then $\chi(A) = 101110010$.

An interesting side-effect is that for example $\chi(A \cap B) = \chi(A) \wedge \chi(B)$, $\chi(A \cup B) = \chi(A) \vee \chi(B)$ and $\chi(\bar{A}) = \neg\chi(A)$. As a function, we say that χ maps intersection to conjunction.

Since every proposition can be expressed using ∧, ∨ and ¬, if we represent sets by their characteristic vectors, we can get a machine to perform set operations as logical operations on bit strings. This is the method programmers use to manipulate sets in computer memory.

The order in which elements of a set are listed does not matter. But there are times when order is important. For example, in a horse race, knowing the order in which the horses cross the finish line is more interesting than simply knowing which horses were in the race. There is a familiar way, introduced in algebra, of indicating order is important: ordered pairs. Ordered pairs of numbers are used to specify points in the Euclidean plane when graphing functions. For instance, when graphing $y = 2x + 1$, setting x = 3 gives y = 7, and so the ordered pair (3, 7) will indicate one of the points on the graph.

In this course, ordered pairs of any sorts of objects, not just numbers, will be of interest. An **ordered pair** is a collection of two objects (which might both be the same) with one specified as first (the first coordinate) and the other as second (the second coordinate). The ordered pair with a

specified as first and b as second is written (as usual) (a, b). The most important feature of ordered pairs is that $(a, b) = (c, d)$ *if and only if* $a = c$ *and* $b = d$. In words, two ordered pairs are equal provided they match in both coordinates. So $(1, 2) \neq (2, 1)$.

More generally, an **ordered n −tuple** $(a_1, a_2, \dots, a_n)$ is the ordered collection with a_1 as its first coordinate, a_2 as its second coordinate, and so on. Two ordered n −tuples are equal provided they match in every coordinate.

The last operation to be considered for combining sets is the **Cartesian product** of two sets A and B. It is defined by $A \times B = \{(a, b) | a \in A \wedge b \in B\}$. In other words, $A \times B$ comprises all ordered pairs that can be formed taking the first coordinate from A and the second coordinate from B. So if $A = \{1, 2\}$, and $B = \{\alpha, \beta\}$, then $A \times B = \{(1, \alpha), (2, \alpha), (1, \beta), (2, \beta)\}$. Notice that in this case $A \times B \neq B \times A$ since, for example, $(1, \alpha) \in A \times B$, but $(1, \alpha) \notin B \times A$.

A special case occurs when $A = B$. In this case we denote the Cartesian product of A with itself by A^2. The familiar example $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$ is called the Euclidean plane or the Cartesian plane.

More generally given sets $A_1, \dots, A_n$ the Cartesian product of these sets is written as $A_1 \times A_2 \times \dots \times A_n = \{(a_1, a_2, \dots, a_n) | a_i \in A_i, 1 \leq i \leq n\}$. Also A^n denotes the Cartesian product of A with itself n times.

In order to avoid the use of an ellipsis we also denote the Cartesian product of $A_1, \dots, A_n$ as $\prod_{k=1}^{n} A_k$

The variable k is called the **index** of the product. Most often the index is a whole number. Unless we are told otherwise we start with $k = 1$ and increment k by 1 successively until we reach n. So if we are given A_1, A_2, A_3, A_4 and A_5, then $\prod_{k=1}^{5} A_k = A_1 \times A_2 \times A_3 \times A_4 \times A_5$.

Exercises

Exercise 6.1. *Let* $A = \{2, 3, 4, 5, 6, 7, 8\}$ *and* $B = \{1, 2, 4, 6, 7, 8, 9\}$. *Find*

a) $A \cap B$

b) $A \cup B$

c) $A - B$

d) $B - A$

Exercise 6.2. *Determine the sets* A *and* B, *if* $A - B = \{1, 2, 7, 8\}$, $B - A = \{3, 4, 10\}$ *and* $A \cap B = \{5, 6, 9\}$.

Exercise 6.3. *Use membership tables to show that* $A \oplus B = (A \cup B) - (A \cap B)$.

Exercise 6.4. *Verify* $A \oplus B = (A \cup B) - (A \cap B)$ *using Venn diagrams.*

Exercise 6.5. *Verify* $A \cup (A \cap B) = A$ *using the rules of set algebra.*

Exercise 6.6. *Let* $A = \{1, 2, 3, 4\}, B = \{a, b, c\}, C = \{\alpha, \beta\}$, *and* $D = \{7, 8, 9\}$. *Write out the following Cartesian products.*

a) $A \times B$

b) $B \times A$

c) $C \times B \times D$

Exercise 6.7. *What can you conclude about* A *and* B *if* $A \times B = B \times A$.

Exercise 6.8. *If* $\mathcal{U} = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$, *determine* $\chi(\{1, 2, 4, 8\})$.

Exercise 6.9. *Let* $A = \{1, 2, 3\} \times \{1, 2, 3, 4\}$. *List the elements in the set* $B = \{(s, t) \in A \mid s \geq t\}$.

Problems

Problem 6.1. *Let* $A = \{1, 2, 3, 5, 6, 7, 9\}$ *and* $B = \{1, 3, 4, 6, 8, 9\}$. *Find*

a) $A \cap B$

b) $A \cup B$

c) $A - B$

d) $B - A$

Problem 6.2. *Use the rules of set algebra to verify* $A \oplus B = (A \cup B) - (A \cap B)$.

Problem 6.3. *Let* $A = \{1, 2, 3\} \times \{1, 2, 3, 4\}$. *List the elements of the set* $B = \{(s, t) \in A \mid s < t\}$.

Problem 6.4. *Let* $A = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\} \times \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$.

a) *What is the cardinality of* A?

b) *What is the cardinality of* $A \times A$?

c) *What is the cardinality of* $B = \{s \mid (s, s^2) \in A\}$?

Problem 6.5. *True or False:*

a) *For all sets* A, B, C, $A \times (B \cap C) = (A \times B) \cap (A \times C)$.

b) *For all sets* A, B, C, $A \times (B \cup C) = (A \times B) \cup (A \times C)$.

c) *For all sets* A, B, C, D, $(A \cap B) \times (C \cap D) = (A \times C) \cap (B \times D)$.

d) *For all sets* A, B, C, D, $(A \cup B) \times (C \cup D) = (A \times C) \cup (B \times D)$.

Problem 6.6. *Let* $\mathcal{U} = \{1, 2, 3, 4, 5, 6, 7, 8\}$. *Find* $\chi(\{1, 2, 3, 4\})$.

Chapter 7

Styles of Proof

Earlier, we practiced proving the validity of logical arguments, both with and without quantifiers. The technique introduced there is one of the main tools for constructing proofs in a more general setting. In this chapter, various common styles of proof in mathematics are described. Recognizing these styles of proof will make both reading and constructing proofs a little less onerous. The example proofs in this chapter will use some familiar facts about integers, which we will prove in a later chapter.

As mentioned before, the typical form of the statement of a theorem is: *if a and b and c and $\cdots$, then d*. The propositions $a,b,c,\cdots$ are called the hypotheses, and the proposition d is called the conclusion. The goal of the proof is to show that $(a \wedge b \wedge c \wedge \cdots) \rightarrow d$ is a true proposition. In the case of propositional logic, the only thing that matters is the *form* of a logical argument, not the particular propositions that are involved. That means the proof can always be given in the form of a truth table. In areas outside of propositional logic that is no longer possible. Now the content of the propositions must be considered. In other words, what the words mean, and not merely how they are strung together, becomes important.

Suppose we want to prove an implication **Theorem:** *If p, then q*. In other words, we want to show $p \rightarrow q$ is true. There are two possibilities: Either p is false, in which case $p \rightarrow q$ is automatically true, or p is true. In this second case, we need to show that q is true as well to conclude $p \rightarrow q$ is true. In other words, to show $p \rightarrow q$ is true, we can begin by assuming p is true, and then give an *argument* that q must be true as well. The outline of such a proof will look like:

Proof.	
Step 1)	Reason 1
Step 2)	Reason 2
.	.
Step l)	Reason l $\square$

The symbol $\square$ in the last line informs the reader that the proof is finished. Every step in the proof must be a true proposition, and since the goal is to conclude q is true, the proposition q will be the last step in the proof. **There are only four acceptable reasons** that can be invoked to justify a step in a proof. Each step can be: (1) a *hypothesis* (and so assumed to be true), (2) an application of a *definition,* (3) a *known fact* proved previously, and so known to be true, or (4) a consequence of applying a *rule of inference or a logical equivalence* to earlier steps in the proof. The only difference between these sorts of formal proofs and the proofs of logical arguments we practiced earlier is the inclusion of definitions as a justification of a step.

Before giving a few examples, there is one more point to consider. Most theorems in mathematics involve variables in some way, along with either universal or existential quantifiers. But, in the case of universal quantifiers, tradition dictates that the mention of the quantifier is often suppressed, and left for the reader to fill in. For example consider: **Theorem:** *If n is an even integer, then n^2 is an even integer.* The statement is really shorthand for **Theorem:** *For every $n \in \mathbb{Z}$, if n is even, then n^2 is even.* If we let $E(n)$ be the predicate *n is even* with universe of discourse $\mathbb{Z}$, the theorem becomes **Theorem:** $\forall n(E(n) \rightarrow E(n^2))$. The truth of such a universally quantified statement can be accomplished with an application of the rule of universal generalization. In other words, we prove that for an arbitrary $n \in \mathbb{Z}$, the proposition $E(n) \rightarrow E(n^2)$ is true.

Theorem 7.1. *If n is an even integer, then n^2 is an even integer.*

Proof.

1) n is an even integer	hypothesis
2) $n = 2k$ for an integer k	definition of even
3) $n^2 = 4k^2$	algebra fact
4) $n^2 = 2(2k^2)$	algebra fact
5) n^2 is even	definition of even $\square$

Usually proofs are not presented in the dry step-wise style of the last example. Instead, a more narrative style is used. So the above proof could go as follows:

Theorem 7.2. *If n is an even integer, then n^2 is an even integer.*

Proof. Suppose n is an even integer. That means $n = 2k$ for some integer k. Squaring both sides gives $n^2 = (2k)^2 = 4k^2 = 2(2k^2)$ which shows n^2 is even. □

All the ingredients of the step-wise proof are present in the narrative form, but this second form is a little more reader friendly. For example, we can include a few comments, such as *squaring both sides gives* to help the reader figure out what is happening.

The method of proof given above is called **direct proof**. The characteristic feature of a direct proof is that in the course of the proof, the hypotheses appear as steps, and the last step in the proof is the conclusion of the theorem.

Here is one more example of a direct proof.

Theorem 7.3. *If n and m are odd integers, then $n+m$ is even.*

Proof. Suppose m and n are odd integers. That means $m = 2j+1$ for some integer j, and $n = 2k+1$ for some integer k. Adding gives $m + n = (2j + 1) + (2k + 1) = 2j + 2k + 2 = 2(j + k + 1)$, and so we see $m + n$ is even. □

There are situations where a direct proof is not very convenient for one reason or another. There are several other styles of proof, each based on some logical equivalence.

For example, since $p \to q \equiv \neg q \to \neg p$, we can prove the

Theorem 7.4. $p \to q$

by instead giving a proof of

Theorem 7.5. $\neg q \to \neg p$.

In other words, we replace the requested implication with its contrapositive and prove that instead. This method of proof is called **indirect proof**. Here's an example.

Theorem 7.6. *If m^2 is an even integer, then m is an even integer.*

Proof. Suppose m is not even. Then m is odd. So $m = 2k + 1$ for some integer k. Squaring both sides of that equation gives $m^2 = (2k + 1)^2 = 4k^2 + 4k + 1 = 2(2k^2 + 2k) + 1$, which shows m^2 is not even. □

Notice that we gave a *direct proof* of the equivalent theorem: *If m is not an even integer, then m^2 is not an even integer.*

Another alternative to a direct proof is **proof by contradiction**. In this method the plan is to replace the requested **Theorem:** r (where r can be any simple or compound proposition) with **Theorem:** $\neg r \to \mathbb{F}$, where $\mathbb{F}$ is any proposition known to be false. The reason proof by contradiction is a valid form of proof is that $\neg r \to \mathbb{F} \equiv r$, so that showing $\neg r \to \mathbb{F}$ is true is identical to showing r is true. Proofs by contradiction can be a bit more difficult to discover than direct or indirect proofs. The reason is that in those two types of proof, we know exactly what the last line of our proof will be. We know where we want to get to. But in a proof by contradiction, we only know that we want to end up with some (any) proposition known to be false. Typically, when writing a proof by contradiction, we experiment, trying various logical arguments, hoping to stumble across some false proposition, and so conclude the proof. For example, consider the following.

Theorem 7.7. *$\sqrt{2}$ is irrational*.

The plan is to replace the requested theorem with

Theorem 7.8. *If $\sqrt{2}$ is rational, then* $\mathbb{F}$ (some fact known to be false).

And now, we may give a direct proof of this replacement theorem:

Proof. Suppose that $\sqrt{2}$ is rational. Then there exist integers m and n with $n \neq 0$ so that $\sqrt{2} = m/n$, with m/n in lowest terms. Squaring both sides gives $2 = m^2/n^2$. Thus $m^2 = 2n^2$ and so m^2 is even. Therefore m is even. So $m = 2k$. Substituting $2k$ for m in $m^2 = 2n^2$ shows $(2k)^2 = 4k^2 = 2n^2$. Which means that $n^2 = 2k^2$. Therefore n^2 is even, which means that n is even. Now since both m and n are even, they have 2 as a common factor. Therefore m/n is in lowest terms and it is not in lowest terms. →← □

The symbol $\rightarrow\leftarrow$ (two arrows crashing into each other head on) denotes that we have reached a *fallacy* ($\mathbb{F}$), a statement known to be false. It usually marks the end of a proof by contradiction.

In the next example, we will prove a proposition of the form $p \to q$ by contradiction. The theorem is about real numbers x and y.

Theorem 7.9. *If* $0 < x < y$, *then* $\sqrt{x} < \sqrt{y}$.

Think of the statement of the theorem in the form $p \to q$. The plan is to replace the requested theorem with

Theorem 7.10. $\neg(p \to q) \to \mathbb{F}$.

But $\neg(p \to q) \equiv \neg(\neg p \vee q) \equiv p \wedge \neg q$. So we will actually prove $(p \wedge \neg q) \to \mathbb{F}$. In other words, we will prove (directly)

Theorem 7.11. *If* $0 < x < y$ and $\sqrt{x} \geq \sqrt{y}$, *then* $\mathbb{F}$ *(some fallacy).*

Proof. Suppose $0 < x < y$ and $\sqrt{x} \geq \sqrt{y}$. Since $\sqrt{x} > 0, \sqrt{x}\sqrt{x} \geq \sqrt{x}\sqrt{y}$, which is the same as $x \geq \sqrt{xy}$. Also since $\sqrt{y} > 0, \sqrt{y}\sqrt{x} \geq \sqrt{y}\sqrt{y}$, which is the same as $\sqrt{xy} \geq y$. Putting $x \geq \sqrt{xy}$ and $\sqrt{xy} \geq y$ together, we conclude that $x \geq y$. Thus $x < y$ and $x \geq y$. $\rightarrow\leftarrow$ □

The only other common style of proof is **proof by cases**. Let's first look at the justification for this proof technique. Suppose we are asked to prove

Theorem 7.12 (Theorem X)**.** $p \to q$.

We *dream up* some propositions, r and s, and replace the requested theorem with three theorems:

Theorem 7.13 (Theorem XS)**.** *(1)* $p \to (r \vee s)$,

(2) $r \to q$, *and*

(3) $s \to q$.

The propositions r, s we dream up are called the *cases*. There can be any number of cases. If we dream up three cases, then we would have four theorems to prove, and so on. The hope is that the proofs of these replacement theorems will be much easier than a proof of the original theorem. This is the *divide and conquer* approach to a proof.

The reason proof by cases is a valid proof technique is that

$$[(p \to (r \vee s)) \wedge (r \to q) \wedge (s \to q)] \to (p \to q)$$

is a tautology - which you should be able to verify. Proof by cases, as with proof by contradiction, is generally a little trickier than direct and indirect proofs. In a proof by contradiction, we are not sure exactly what we are shooting for. We just hope some contradiction will pop up. For a proof by cases, we have to dream up the cases to use, and it can be difficult at times to dream up good cases.

Theorem 7.14. *For any integer n, $|n| \geq n$.*

Proof. Suppose n is an integer. There are two cases: Either (1): $n > 0$, or (2): $n \leq 0$. (This has the form $p \to (r \vee s)$ of (1) in Theorem 7.)

Case 1: We need to show *If $n > 0$, then $|n| \geq n$.* (We will do this with a direct proof.) Suppose $n > 0$. Then $|n| = n$. Thus $|n| \geq n$ is true.

Case 2: We need to show *If $n \leq 0$, then $|n| \geq n$.* (We will again use a direct proof.) Suppose $n \leq 0$. Now $0 \leq |n|$. Thus, $n \leq |n|$.

So, in any case, $n \leq |n|$ is true, and that proves the theorem. □

A proof of a statement of the form $\exists x P(x)$ is called an **existence proof**. The proof may be **constructive**, meaning that the proof provides a specific example of, or at least an explicit recipe for finding, an x so that $P(x)$ is true; or the proof may be **non-constructive**, meaning that it establishes the existence of x without giving a method of actually producing an example of an x for which $P(x)$ is true.

To give examples of each type of existence proof, let's use a familiar fact (which will be proved a little later in the course): There are infinitely many primes. Recall that a prime is an integer greater than 1 whose only positive divisors are 1 and itself. The next two theorems are contrived, but they demonstrate the ideas of constructive and nonconstructive proofs.

Theorem 7.15. *There is a prime with more than two digits.*

Proof. Checking shows that 101 has no positive divisors besides 1 and itself. Also, 101 has more than two digits. So we have produced an example of a prime with more than two digits. □

That is a constructive proof of the theorem. Now, here is a non-constructive proof of a similar theorem.

Theorem 7.16. *There is a prime with more than one billion digits.*

Proof. Since there are infinitely many primes, they cannot all have one billion or fewer digits. So there must some primes with more than one billion digits. □

Finally, suppose we are asked to prove a theorem of the form $\forall x\, P(x)$, and for one reason or another we come to believe the proposition is not true. The proposition can be shown to be false by exhibiting a specific element from the domain of x for which $P(x)$ is false. Such an example is called a **counterexample** to the theorem. Let's look at a specific instance of the counterexample technique.

Theorem 7.17 (not really!). *For all positive integers* $n, n^2 - n + 41$ *is prime*

Counterexample 7.18. *To disprove the theorem, we explicitly specify a positive integer* n *such that* $n^2 - n + 41$ *is not prime. In fact, when* $n = 41$, *the expression is not a prime since clearly* $41^2 - 41 + 41 = 41^2$ *is divisible by* 41. *So,* $n = 41$ *is a counterexample to the proposition.*

An interesting fact about this example is that $n = 41$ is the smallest counterexample. For $n = 1, 2, \cdots 40$, it turns out that $n^2 - n + 41$ is a prime! This examples shows the danger of checking a theorem of the form $\forall x\, P(x)$ for a few (or a few billion!) values of x, finding $P(x)$ true for those cases, and concluding it is true for every possible value of x.

For the purpose of these exercises and problems, feel free to use familiar facts and definitions about integers. For example: Recall, an integer n is even if $n = 2k$ for some integer k. And an integer n is odd if $n = 2k + 1$ for some integer k.

Exercises

Exercise 7.1. *Give a direct proof that the sum of two even integers is even.*

Exercise 7.2. *Give an indirect proof that if the square of the integer* n *is odd, then* n *is odd.*

Exercise 7.3. *Give a proof by contradiction that the sum of a rational number and an irrational number is irrational.*

Exercise 7.4. *Give a proof by contradiction that if* $5n - 1$ *is odd, then* n *is even.*

Exercise 7.5. *Give a counterexample to the proposition Every positive integer that ends with a* 7 *is a prime.*

Problems

Problem 7.1. *Give a direct proof that the sum of an even integer and an odd integer is odd.*

Hint: Start by letting m be an even integer and letting n be an odd integer. That means $m = 2k$ for some integer k and $n = 2j + 1$ for some integer j. You are interested in $m + n$, so add them up and see what you get. Why is the thing you get an odd integer (think about the definition of odd)?

Problem 7.2. *Give a direct proof that the sum of two odd integers is even.*

Problem 7.3. *Give an indirect proof that if n^3 is even, then n is even. Hint: Study the solution of a similar statement in the sample exercises for this lesson.*

Problem 7.4. *Give a proof by contradiction that if $3n + 2$ is odd, then n is odd.*

Hint: This is the problem in this set that gives the most grief. Study the section in the notes where the mechanics of proving a statement of the form If P, then Q by contradiction is discussed. Be sure you understand why the first line of the proof should be something like Suppose $3n + 2$ is odd **and** *n is even.*

Problem 7.5. *Give an example of a predicate $P(n)$ about positive integers n, such that $P(n)$ is true for every positive integer from 1 to one billion, but which is never-the-less not true for all positive integers. (Hint: there is a really simple choice possible for the predicate $P(n)$.)*

Problem 7.6. *The* ***maximum*** *of two numbers, a and b is a provided $a \geq b$. Notation: $\max(a, b) = a$. The* ***minimum*** *of a and b is a provided $a \leq b$. Notation: $\min(a, b) = a$. Examples: $\max(2, 3) = 3$, $\max(5, 0) = 5$, $\min(2, 3) = 2$, $\min(5, 0) = 0$, $\max(4, 4) = \min(4, 4) = 4$.*
Give a proof by cases that for any numbers s, t,

$$\min(s, t) + \max(s, t) = s + t.$$

Problem 7.7. *Give a proof by cases that for integers m, n, we have $|mn| = |m||n|$. Hint: Consider four cases: (1) $m \geq 0$ and $n \geq 0$, (2) $m \geq 0$ and $n < 0$, (3) $m < 0$ and $n \geq 0$, and (4) $m < 0$ and $n < 0$.*

Chapter 8

Relations

Two-place predicates, such as $B(x, y)$: *x is the brother of y*, play a central role in mathematics. Such predicates can be used to describe many basic concepts. As examples, consider the predicates given verbally:

(1) $G(x, y)$: *x is greater than or equal to y* which compares the magnitudes of two values.

(2) $P(x, y)$: *x has the same parity as y* which compares the parity of two integers.

(3) $S(x, y)$: *x has square equal to y* which relates a value to its square.

Two-place predicates are called **relations**, probably because of examples such as the *brother of* given above. To be a little more complete about it, if $P(x, y)$ is a two-place predicate, and the domain of discourse for x is the set A, and the domain of discourse for y is the set B, then P is called a **relation from A to B**. When working with relations, some new vocabulary is used. The set A (the domain of discourse for the first variable) is called the **domain** of the relation, and the set B (the domain of discourse for the second variable) is called the **codomain** of the relation.

There are several different ways to specify a relation. One way is to give a verbal description as in the examples above. As one more example of a verbal description of a relation, consider

$E(x, y)$: *The word x ends with the letter y.* Here the domain will be words in English, and the codomain will the the twenty-six letters of the alphabet. We say the ordered pair (cat, t) **satisfies** the relation E, but that (dog, w) does not.

When dealing with abstract relations, a verbal description is not always convenient. An alternate method is to tell what the domain and codomain are to be, and then simply list the ordered pairs which will satisfy the relation. For example, if $A = \{1, 2, 3, 4\}$ and $B = \{a, b, c, d\}$, then one of many possible relations from A to B would be $\{(1, b), (2, c), (4, c)\}$. If we name this relation R, we will write $R = \{(1, b), (2, c), (4, c)\}$. It would be tough to think of a natural verbal description of R.

When thinking of a relation, R, as a set of ordered pairs, it is common to write aRb in place of $(a, b) \in R$. For example, using the relation G defined above, we can convey the fact that the pair $(3, 2)$ satisfies the relation by writing any one of the following: (1) $G(3, 2)$ is true, (2) $(3, 2) \in G$, or (3) $3G2$. The third choice is the preferred one when discussing relations abstractly.

Sometimes the ordered pair representation of a relation can be a bit cumbersome compared to the verbal description. Think about the ordered pair form of the relation E given above: $E = \{(\text{cat, t}), (\text{dog, g}), (\text{antidisestablishmentarianism, m}), \cdots\}$.

Another way represent a relation is with a **graph**. Here, a graph is a diagram made up of dots, called **vertices**, some of which are joined by lines, called **edges**. To draw a graph of a relation R from A to B, make a column of dots, one for each element of A, and label the dots with the names of those elements. Then, to the right of A's column make a column of dots for the elements of B. Then connect the vertex labelled $a \in A$ to a vertex $b \in B$ with an edge provided $(a, b) \in R$. The diagram is called the **bipartite graph representation** of R.

Example 8.1. *Let* $A = \{1, 2, 3, 4\}$, $B = \{a, b, c, d\}$, *and let* $R = \{(1, a), (2, b), (3, c), (3, d), (4, d)\}$. *Then the bipartite graph which represents* R *is given in figure 8.1.*

The choices made about the ordering and the placement of the vertices for the elements of A and B may make a difference in the appearance of the graph, but all such graphs are considered equivalent. Also, edges can be curved lines. All that matters is that such diagrams convey graphically the same information as R given as a set of ordered pairs.

It is common to have the domain and the codomain of a relation be the same set. If R is a relation from A to A, then we will say R is a **relation on** A. In this case there is a shorthand way of representing the relation by using a **digraph**. The word digraph is shorthand for *directed graph* meaning the edges have a direction indicated by an arrowhead. Each element of A is used to label

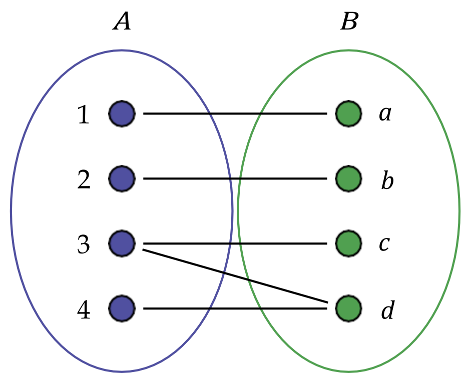


Figure 8.1: Example bipartite graph

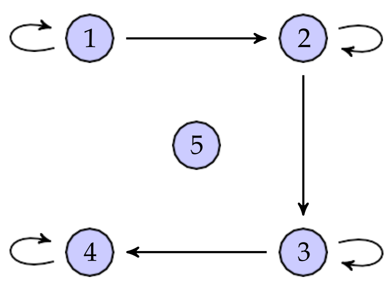


Figure 8.2: Example digraph

a single point. An arrow connects the vertex labelled s to the one labelled t provided $(s, t) \in R$. An edge of the form (s, s) is called a **loop**.

Example 8.2. *Let* $A = \{1, 2, 3, 4, 5\}$ *and*
$R = \{(1, 1), (1, 2), (2, 2), (2, 3), (3, 3), (3, 4), (4, 4)\}$. *Then a digraph for* R *is shown in figure 8.2*

Again it is true that a different placement of the vertices may yield a different-looking, but equivalent, digraph.

The last method for representing a relation is by using a $0-1$ matrix. This method is particularly handy for encoding a relation in computer memory. An $m \times n$ **matrix** is a rectangular array with m rows and n columns. Matrices are usually denoted by capital English letters. The entries of a matrix, usually denoted by lowercase English letters, are indexed by row and column. Either a_{ij} or $a_{i,j}$ stands for the entry in a matrix in the ith row and jth column. A $0-1$ **matrix** is one all of whose entries are 0 or 1. Given two finite sets A and B with m and n elements respectively, we may use the elements of A (in some fixed order) to index the rows of an $m \times n, 0-1$ matrix, and use the elements of B to index the columns. So for a relation R from A to B, there is a matrix of R, M_R with respect to the orderings of A and B which represents R. The entry of M_R in the row labelled by a and column labelled by b is 1 if aRb and 0 otherwise. This is exactly like using characteristic vectors to represent subsets of $A \times B$, except that the vectors are cut into n chunks of size m.

Example 8.3. *Let* A = {1, 2, 3, 4} *and* $B = \{a, b, c, d\}$ *as before, and consider the relation* R = {(1, *a*), (1, *b*), (2, *c*), (4, *c*), (4, *a*)}. *Then a* 0−1 *matrix which represents* R *using the natural orderings of* A *and* B *is Note: This matrix may change appearance if* A *or* B *is listed in a different order.*

$$M_R = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

Since relations can be thought of as sets of ordered pairs, it makes sense to ask if one relation is a subset of another. Also, set operations such as union and intersection can be carried out with relations.

These notions can be expressed in terms of the matrices that represent the relations. Bit-wise operations on 0-1 matrices are defined in the obvious way. Then $M_{R\cup S} = M_R \vee M_S$, and $M_{R\cap S} = M_R \wedge M_S$. Also, for two 0-1 matrices of the same size $M \leq N$ means that wherever N has a 0 entry, the corresponding entry in M is also 0. Then $R \subseteq S$ means the same as $M_R \leq M_S$.

There are two new operations possible with relations.

First, if R is a relation from A to B, then by reversing all the ordered pairs in R, we get a new relation, denoted R^{-1}, called the **inverse** of R. In other words, R^{-1} is the relation from B to A given by $R^{-1} = \{(b, a) | (a, b) \in R\}$. A bipartite graph for R^{-1} can be obtained from a bipartite graph for R simply by interchanging the two columns of vertices with their attached edges (or, by rotating the diagram 180°).

If the matrix for R is M_R, then the matrix for R^{-1} is produced by taking the columns of $M_{R^{-1}}$ to be the rows of M_R. A matrix obtained by changing the rows of M into columns is called the **transpose** of M, and written as M^T . So, in symbols, if M is a matrix for R, then M^T is a matrix for R^{-1}.

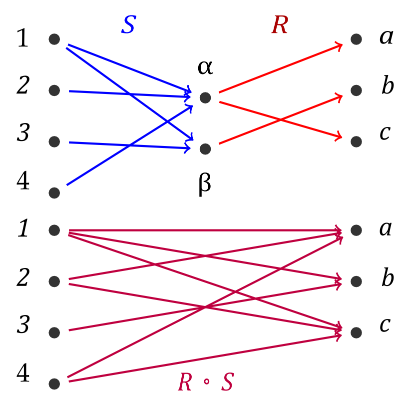


Figure 8.3: Composing relations: $R \circ S$

The second operation with relations concerns the situation when S is a relation from A to B and R is a relation from B to C. In such a case, we can form the **composition of** S **by** R which is denoted $R \circ S$. The composition is defined as

$$R \circ S = \{(a, c) | a \in A, c \in C \text{ and } \exists b \in B, \text{ such that } (a, b) \in S \text{ and } (b, c) \in R\}.$$

Example 8.4. *Let* $A = \{1, 2, 3, 4\}, B = \{\alpha, \beta\}$ *and* $C = \{a, b, c\}$. *Further let* $S = \{(1, \alpha), (1, \beta), (2, \alpha), (3, \beta), (4, \alpha)\}$ *and* $R = \{(\alpha, a), (\alpha, c), (\beta, b)\}$.

Since $(1, \alpha) \in S$ *and* $(\alpha, a) \in R$, *it follows that* $(1, a) \in R \circ S$. *Likewise, since* $(2, \alpha) \in S$ *and* $(\alpha, c) \in R$, *it follows that* $(2, c) \in R \circ S$. *Continuing in that fashion shows that*

$$R \circ S = \{(1, a), (1, b), (1, c), (2, a), (2, c), (3, b), (4, a), (4, c)\}.$$

The composition can also be determined by using the bipartite graphs of the relations. Make a column of vertices for A *labelled* $1, 2, 3, 4$, *then to the right a column of points for* B *labelled* α, β, *then again to the right a column of points for* C *labelled* a, b, c. *Draw in the edges as usual for* R *and* S. *(See figure 8.3.) Then a pair* (x, y) *will be in* $R \circ S$ *provided there is a two edge path from* x *to* y.

From the picture it is instantly clear that, for example, $(1, c) \in R \circ S$.

In terms of $0{-}1$ *matrices if* M_S *is the* $m \times k$ *matrix of* S *with respect to the given orderings of* A *and* B, *and if* M_R *is the* $k \times n$ *matrix of* R *with respect to the given orderings of* B *and* C, *then whenever the* i, l *entry of* S *and* l, j *entry of* R *are both* 1, *then* $(a_i, c_j) \in R \circ S$.

This example motivates the definition of the **Boolean product** of M_S and M_R as the corresponding matrix $M_{R \circ S}$ of the composition. More rigorously when M is an $m \times k$ $0{-}1$ matrix and N is an $k \times n$ $0{-}1$ matrix, $M \odot N$ is the $m \times n$ $0{-}1$ matrix whose i, j entry is $(m_{i1} \wedge n_{1j}) \vee (m_{i2} \wedge n_{2j}) \vee \ldots \vee m_{ik} \wedge n_{kj})$. This looks worse than it is. It achieves the desired result. The Boolean product is computed the same way as the ordinary matrix product where multiplication and addition have been replaced with and and or, respectively.

For the relations in the example above example

$$M_{R \circ S} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \odot \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} = M_S \odot M_R$$

Exercises

Exercise 8.1. *Let $A = \{a, b, c, d\}$ and*
$R = \{(a,a), (a,c), (b,b), (b,d), (c,a), (c,c), (d,b), (d,d)\}$ be a relation on A. Draw a digraph which represents R. Find the matrix which represents R ***with respect to the ordering*** (d, c, a, b)**.**

Exercise 8.2. *The matrix of a relation S from $\{1,2,3,4,5\}$ to $\{a, b, c, d\}$ with respect to the given orderings is displayed below. Represent S as a bipartite graph, and as a set of ordered pairs.*

$$M_S = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \end{bmatrix}$$

Exercise 8.3. *Find the composition $R \circ S$ (where R and S are defined in exercises 8.1 and 8.2) as a set of ordered pairs. Use the Boolean product to find $M_{R \circ S}$ with respect to the natural orderings.*

N.B. The natural ordering for R is ***not*** *the ordering used in exercise 8.1.*

Exercise 8.4. *Let $B = \{1,2,3,4,5,6\}$ and let*

$$\begin{aligned} R_1 &= \{(1,2),(1,3),(1,5),(2,1),(2,2),(2,4),(3,3),(3,4), \\ &\quad (4,1),(4,5),(5,5),(6,6)\} \text{ and} \\ R_2 &= \{(1,2),(1,6),(2,1),(2,2),(2,3),(2,5),(3,1),(3,3),(3,6), \\ &\quad (4,2),(4,3),(4,4),(5,1),(5,5),(5,6),(6,2),(6,3),(6,6)\}. \end{aligned}$$

(a) *Find $R_1 \cup R_2$, $R_1 \cap R_2$, and $R_1 \oplus R_2$.*

(b) *With respect to the given ordering of B find the matrix of each relation in part (a)*

Problems

Problem 8.1. *Let* $A = \{a, b, c, d\}$ *and* $R = \{(a, a), (a, c), (b, d), (c, a), (c, c), (d, b)\}$ *be a relation on* A. *Draw a digraph which represents* R. *Draw the bipartite graph which represents* R.

Problem 8.2. *Let* $A = \{a, b, c, d\}$ *and* $R = \{(a, a), (a, c), (b, d), (c, a), (c, c), (d, b)\}$ *be a relation on* A. *What is the inverse of* R?

Problem 8.3. *Find the composition,* $R \circ S$, *where* $S = \{(1, a), (4, a), (5, b), (2, c), (5, c), (3, d)\}$ *and* $R = \{(a, x), (a, y), (b, x), (c, z), (d, z)\}$, *as a set of ordered pairs.*

Problem 8.4. *Let* $R_1 = \{(1, 2), (1, 3), (1, 5), (2, 1), (6, 6)\}$ *and* $R_2 = \{(1, 2), (1, 6), (3, 6), (4, 2), (5, 6), (6, 2), (6, 3)\}$. *Find* $R_1 \cup R_2$ *and* $R_1 \cap R_2$.

Problem 8.5. *Let* L *be the relation less than on the set of integers. Examples* $3\,L\,7$ *and* $-8\,L\,0$ *are true, but* $5\,L\,2$ *and* $6\,L\,6$ *are false. How would describe the relation* L^{-1}?

Problem 8.6. *True or False: For any relation* R, $(R^{-1})^{-1} = R$. *Explain your answer.*

Problem 8.7. *Are there relations* R *for which* $R = R^{-1}$? *If not, explain why it is not possible. If so, give an example of such a relation.*

Problem 8.8. *Let* R *be a relation on a set* A, *and let* R^{-1} *be its inverse. Prove that if* $(a, b) \in R \circ R^{-1}$, *then* $(b, a) \in R \circ R^{-1}$.

Problem 8.9. *Let* A *and* B *be two sets. Explain why the empty set,* $\emptyset$, *is a relation from* A *to* B.

Problem 8.10. *Let* S *be a relation from* A *to* B, *and let* R *be a relation from* B *to* C. *Prove* $(R \circ S)^{-1} = S^{-1} \circ R^{-1}$.

Chapter 9

Properties of Relations

There are several conditions that can be imposed on a relation R on a set A that make it useful. These requirements distinguish those relations which are interesting for some reason from the garden variety junk, which is, let's face it, what most relations are.

A relation R on A is **reflexive** provided $\forall a \in A, aRa$. In plain English, a relation is reflexive if every element of its domain is related to itself. The relation $B(x, y)$: x **is the brother of** y is not reflexive since no person is his own brother. On the other hand, the relation $S(m, n)$: $m + n$ **is even**. is a reflexive relation on the set of integers since, for any integer m, $m + m = 2m$ is even.

In symbolic logic a relation R on a set A is reflexive means $(\forall a \in A)[aRa]$

It is easy to spot a reflexive relation from its digraph: there is a loop at every vertex. Also, a reflexive relation can be spotted quickly from its matrix. First, let's agree that when the matrix of a relation on a set A is written down, the same ordering of the elements of A is used for both the row and column designators. For a reflexive relation, the entries on the **main diagonal** of its matrix will all be 1's. The main diagonal of a square matrix runs from the upper left corner to the lower right corner.

The flip side of the coin from reflexive is irreflexive. A relation R on A is **irreflexive** in case $a \not R a$ for all $a \in A$. In other words, no element of A is related to itself. The *brother of* relation is irreflexive. The digraph of an irreflexive relation contains no loops, and its matrix has all 0's on the main diagonal.

In symbols a relation R on a set A is irreflexive means $(\forall a \in A)[a \not R a] \equiv \neg(\exists a \in A)[aRa]$.

Actually, that discussion was a little careless. To see why, consider the relation

$S(x, y)$: **the square of** x is bigger than or equal to y.

Is this relation reflexive? The answer is: we can't tell. The answer depends on the domain of the relation, and we haven't been told what that is to be. For example, if the domain is the set $\mathbb{N}$ of natural numbers, then the relation is reflexive, since $n^2 \geq n$ for all $n \in \mathbb{N}$. However, if the domain is the set $\mathbb{R}$ of all real numbers, the relation is not reflexive. In fact a counterexample to the claim that S is reflexive on $\mathbb{R}$ is the number $1/2$ since $(1/2)^2 = 1/4$ and $1/4 < 1/2$. The lesson to be learned from this example is that the question of whether a relation is reflexive cannot be answered until the domain has been specified. The same is true for the irreflexive condition and the other conditions defined below. Always be sure you know the domain before trying to determine which properties a relation satisfies.

A relation R on A is **symmetric** provided $(a, b) \in R \rightarrow (b, a) \in R$. Another way to say the same thing: R is symmetric provided $R = R^{-1}$. In words, R is symmetric provided that whenever a is related to b, then b is related to a. Any digraph representing a symmetric relation R will have a return edge for every non-loop. Think of this as saying the graph has no one-way streets. The matrix M of a symmetric relation satisfies $M = M^T$. In this case M is symmetric about its main diagonal in the usual geometric sense of symmetry. The $B(x, y)$: x **is the brother of** y relation mentioned before is not symmetric if the domain is taken to be all people since, for example, Donny B Marie, but Marie $\not B$ Donny. On the other hand, if we take the domain to be all (human) males, then B is symmetric.

In symbolic logic a relation R on a set A is symmetric means $(\forall a, b \in A)[aRb \rightarrow bRa]$.

A relation R on A is **antisymmetric** if whenever $(a, b) \in R$ and $(b, a) \in R$, then $a = b$. In other words, the only objects that are each related to the other are objects that are the same. For example, the usual $\leq$ relation for the integers is antisymmetric since if $m \leq n$ and $n \leq m$, then $n = m$. A digraph representing an antisymmetric relation will have all streets one-way except loops. If M is a matrix for R, then whenever $a_{i,j} = 1$ and $i \neq j$, $a_{j,i} = 0$.

In symbolic logic a relation R on A is antisymmetric means $(\forall a, b \in A)[(aRb) \wedge (bRa) \rightarrow (a = b)]$.

A relation R on A is **transitive** if whenever $(a, b) \in R$ and $(b, c) \in R$, then $(a, c) \in R$. This can also be expressed by saying $R \circ R \subseteq R$. In a digraph for a transitive relation whenever we have a directed path of length two from a to c through b, we must also have a direct link from a to c. This means that any digraph of a transitive relation has lots of triangles. This includes degenerate triangles where a, b and c are not distinct. A matrix M of a transitive relation satisfies $M \odot M \leq M$. The relation $\leq$ on $\mathbb{N}$ is transitive, since from $k \leq m$ and $m \leq n$, we can conclude $k \leq n$.

In symbolic logic a relation R on A is transitive means $(\forall a, b, c \in A)[(aRb) \wedge (bRc) \rightarrow (aRc)]$.

Example 9.1.
Define a relation, N on the set of all living people by the rule $a\,N\,b$ if and only if a, b live within one mile of each other. This relation is reflexive since every person lives within a mile of himself. It is not irreflexive since I live within a mile of myself. It is symmetric since if a lives within a mile of b, then b lives within a mile of a. It is not antisymmetric since Mr. and Mrs. Smith live within a mile of each other, but they are not the same person. It is not transitive: to see why, think of the following situation (which surely exists somewhere in the world!): there is a straight road of length 1.5 miles. Say Al lives at one end of the road, Cal lives at the other end, and Sal lives half way between Al and Cal. Then $Al\,N$ Sal and $Sal\,N$ Cal, but not $Al\,N$ Cal.

Example 9.2. *Let $\mathrm{A} = \mathbb{R}$ and define aRb iff $a \leq b$, then R is a reflexive, transitive, antisymmetric relation. Because of this example, any relation on a set that is reflexive, antisymmetic, and transitive is called an* ***ordering*** *relation. The subset relation on any collection of sets is another ordering relation.*

Example 9.3. *Let $\mathrm{A} = \mathbb{R}$ and define aRb iff $a < b$. Then R is irreflexive, and transitive.*

Example 9.4. *If $A = \{1, 2, 3, 4, 5, 6\}$ then*

$$R = \{(1,1), (2,2), (3,3), (4,4), (5,5), (6,6), (1,3), (3,1), (1,5), \\ (5,1), (2,4), (4,2), (2,6), (6,2), (3,5), (5,3), (4,6), (6,4)\}$$

is reflexive, symmetric, and transitive. In artificial examples such as this one, it can be a tedious chore checking that the relation is transitive.

Example 9.5. *If $A = \{1, 2, 3, 4\}$ and $R = \{(1,1), (1,2), (2,3), (1,3), (3,4), (2,4), (4,1)\}$ then R is not reflexive, not irreflexive, not symmetric, and not transitive but it is antisymmetric.*

Exercises

Exercise 9.1. *Define a relation on* $\{1, 2, 3\}$ *which is both symmetric and antisymmetric.*

Exercise 9.2. *Define a relation on* $\{1, 2, 3, 4\}$ *by*

$$R = \{(1, 2), (2, 1), (2, 3), (3, 2), (3, 4), (4, 3)\}.$$

For each of the five properties of a relation defined in this chapter (reflexive, irreflexive, symmetric, antisymmetric, and transitive) either show R *satisfies the property, or explain why it does not.*

Exercise 9.3. *Each matrix below specifies a relation* R *on* $\{1, 2, 3, 4, 5, 6\}$ *with respect to the given ordering* $1, 2, 3, 4, 5, 6$.

For each of the five properties of a relation defined in this chapter (reflexive, irreflexive, symmetric, antisymmetric, and transitive) either show R *satisfies the property, or explain why it does not.*

$$a)\begin{bmatrix}1&1&1&0&0&0\\1&1&1&0&0&0\\1&1&1&0&0&0\\0&0&0&1&1&1\\0&0&0&1&1&1\\0&0&0&1&1&1\end{bmatrix}\quad b)\begin{bmatrix}1&1&1&1&1&1\\0&0&0&0&0&1\\0&0&0&0&0&1\\0&0&0&0&0&1\\0&0&0&0&0&1\\0&0&0&0&0&1\end{bmatrix}$$

$$c)\begin{bmatrix}1&0&0&0&0&1\\0&1&0&0&1&0\\0&0&1&1&0&0\\0&0&1&1&0&0\\0&1&0&0&1&0\\1&0&0&0&0&0\end{bmatrix}\quad d)\begin{bmatrix}1&0&0&0&0&1\\0&1&0&0&1&0\\0&0&1&1&0&0\\0&0&1&1&0&0\\0&1&0&0&1&0\\1&0&0&0&0&1\end{bmatrix}$$

Exercise 9.4. *Define the relation* $C(A, B) : |A| \leq |B|$, *where the domains for* A *and* B *are all subsets of* $\mathbb{Z}$.

For each of the five properties of a relation defined in this chapter (reflexive, irreflexive, symmetric, antisymmetric, and transitive) either show C *satisfies the property, or explain why it does not.*

Exercise 9.5. *Explain why* $\emptyset$ *is a relation on any set.*

Exercise 9.6. *Define the relation $M(A, B) : |A \cap B| = 1$ (or, in plain English, A and B have exactly one element in common), where the domains for A and B are all subsets of $\mathbb{Z}$. A few examples:*

- $\{5, 10\}\ M\ \{1, 2, 3, 4, 5, 6\}$ *is true since the sets* $\{5, 10\}$ *and* $\{1, 2, 3, 4, 5, 6\}$ *have exactly one element in common (namely* 5*).*
- $\{1, 2, 3\}\ M\ \{6, 7, 8, 9\}$ *is false since* $\{1, 2, 3\}$ *and* $\{6, 7, 8, 9\}$ *have no elements in common.*
- $\{1, 2, 3, 4\}\ M\ \{2, 4, 6, 8\}$ *is false since* $\{1, 2, 3, 4\}$ *and* $\{2, 4, 6, 8\}$ *have more than one element in common.*
- $\{n | n \in \mathbb{Z}\ and\ n \leq 0\}\ \mathrm{M}\ \{n | n \in \mathbb{Z}\ and\ n \geq 0\}$ *is true since* $\{n | n \in \mathbb{Z}\ and\ n \leq 0\}$ *and* $\{n | n \in Z\ and\ n \geq 0\}$ *have exactly one element in common (namely* 0*).*

For each of the five properties of a relation defined in this chapter (reflexive, irreflexive, symmetric, antisymmetric, and transitive) either show M satisfies the property or explain why it does not.

Problems

Problem 9.1. *Let R be the relation $\{(1,1)\}$ on the set $A = \{1,2\}$. For each of the five properties of a relation defined in this chapter (reflexive, irreflexive, symmetric, antisymmetric, and transitive) either show R satisfies the property, or explain why it does not.*

Problem 9.2. *Let S be the relation $\{(1,1),(1,2),(1,3),(2,3)\}$ on the set $A = \{1,2,3\}$. For each of the five properties of a relation defined in this chapter (reflexive, irreflexive, symmetric, antisymmetric, and transitive) either show S satisfies the property, or explain why it does not.*

Problem 9.3. *Let A be the relation on the set $\mathbb{Z}$ of all integers defined by sAt iff $|s| \leq |t|$. For each of the five properties of a relation defined in this chapter (reflexive, irreflexive, symmetric, antisymmetric, and transitive) either show A satisfies the property, or explain why it does not.*

Problem 9.4. *Let D be the relation on the natural numbers defined by the rule mDn if and only if m does not equal n. Examples: $5D7$ is true and $4D4$ is false. For each of the five properties of a relation defined in this chapter (reflexive, irreflexive, symmetric, antisymmetric, and transitive) either show D satisfies the property, or explain why it does not.*

Problem 9.5. *Let R be the relation $\{(1,2),(2,3),(3,4)\}$ on the set $A = \{1,2,3\}$. The relation R is not transitive on A. What is the fewest number of ordered pairs that need to be added to R so it becomes a transitive relation on A?*

Problem 9.6. *Give a counterexample to the claim that a relation R on a set A that is both symmetric and transitive must be reflexive. Hint: There is a very simple example!*

Problem 9.7. *Define the relation $M(A,B) : A \cap B = \emptyset$, where the domains for A and B are all subsets of $\mathbb{Z}$.*

For each of the five properties of a relation defined in this chapter (reflexive, irreflexive, symmetric, antisymmetric, and transitive) either show M satisfies the property, or explain why it does not.

Problem 9.8.

(a) Let $A = \{1\}$, and consider the empty relation, $\emptyset$, on A. For each of the five properties of a relation defined in this chapter (reflexive, irreflexive, symmetric, antisymmetric, and transitive) either show $\emptyset$ satisfies the property, or explain why it does not.

(b) Same question as (a), but now with $A = \emptyset$.

Chapter 10

Equivalence Relations

Relations capture the essence of many different mathematical concepts. In this chapter, we will show how to put the idea of *are the same kind* in terms of a special type of relation.

Before considering the formal concept of *same kind* let's look at a few simple examples. Consider the question, posed about an ordinary deck of 52 cards: *How many different kinds of cards are there?* One possible answer is: *There are* 52 *kinds of cards,* since all the cards are different. But another possible answer in certain circumstances is: *There are four kinds of cards* (namely clubs, diamonds, hearts, and spades). Another possible answer is: *There are two kinds of cards, red and black.* Still another answer is: *There are* 13 *kinds of cards: aces, twos, threes,* $\cdots$*, jacks, queens, and kings.* Another answer, for the purpose of many card games is: *There are ten kinds of cards, aces, twos, threes, up to nines, while tens, jacks, queens, and kings are all considered to be the same value (usually called 10).* You can certainly think of many other ways to split the deck into a number of different kinds.

Whenever the idea of *same kind* is used, some properties of the objects being considered are deemed important and others are ignored. For instance, when we think of the the deck of cards made of the 13 different ranks, ace through king, we are agreeing the the suit of the card is irrelevant. So the jack of hearts and the jack of clubs are taken to be the same for what ever purposes we have in mind.

The mathematical term for *same kind* is **equivalent**. There are three basic properties always associated with the idea of equivalence.

(1) *Reflexive*: Every object is equivalent to itself.

(2) *Symmetric*: If object a is equivalent to object b, then b is also equivalent to a.

(3) *Transitive*: If a is equivalent to b and b is equivalent to c, then a is equivalent to c.

To put the idea of equivalence in the context of a relation, suppose we have a set A of objects, and a rule for deciding when two objects in A are the same kind (equivalent) for some purpose. Then we can define a relation E on the set A by the rule that the pair (s, t) of elements of A is in the relation E if and only if s and t are the same kind. For example, consider again the deck of cards, with two cards considered to be the same if they have the same rank. Then a few of the pairs in the relation E would be (ace hearts, ace spades), (three diamonds, three clubs), (three clubs, three diamonds), (three diamonds, three diamonds), (king diamonds, king clubs), and so on.

Using the terminology of the previous chapter, this relation E, and in fact any relation that corresponds to notion of equivalence, will be reflexive, symmetric, and transitive. For that reason, any reflexive, symmetric, transitive relation on a set A is called an **equivalence relation** on A.

Suppose E is an equivalence relation on a set A and that x is one particular element of A. The **equivalence class of** x is the set of all the things in A that are equivalent to x. The symbol used for the equivalence class of x is $[x]$, so the definition can be written in symbols as $[x] = \{y \in A | y\, E\, x\}$.

For instance, think once more about the deck of cards with the equivalence relation *having the same rank*. The equivalence class of the two of spades would be the set $[2\spadesuit] = \{2\clubsuit, 2\diamondsuit, 2\heartsuit, 2\spadesuit\}$. That would also be the equivalence class of the two of diamonds. On the other hand, if the equivalence relation we are using for the deck is *having the same suit*, then the equivalence class of the two of spades would be

$$[2\spadesuit] = \{A\spadesuit, 2\spadesuit, 3\spadesuit, 4\spadesuit, 5\spadesuit, 6\spadesuit, 7\spadesuit, 8\spadesuit, 9\spadesuit, 10\spadesuit, J\spadesuit, Q\spadesuit, K\spadesuit\}.$$

The most important fact about the collection of different equivalence classes for an equivalence relation on a set A is that they split the set A into separate pieces. In fancier words, they **partition** the set A. For example, the equivalence relation of having the same rank splits a deck of cards into 13 different equivalence classes. In a sense, when using this equivalence relation, there are only 13 different objects, four of each kind.

Here are a few more examples of equivalence relations.

Example 10.1. *Define R on $\mathbb{N}$ by aRb if and only if $a = b$. In other words, equality is an equivalence relation. If fact, this example explains the choice of name for such relations.*

Example 10.2. *Let A be the set of logical propositions and define R on A by pRq iff $p \equiv q$.*

Example 10.3. *Let A be the set of people in the world and define R on A by aRb iff a and b are the same age in years.*

Example 10.4. *Let $A = \{1, 2, 3, 4, 5, 6\}$ and R be the relation on A with the matrix from exercise 3. part a) of chapter 9.*

Example 10.5. *Define P on $\mathbb{Z}$ by $a\,P\,b$ if and only if a and b are both even, or both odd. We say a and b have the same parity.*

For the equivalence relation *has the same rank* on a set of cards in a 52 card deck, there are 13 different equivalence classes. One of the classes contains all the aces, another contains all the 2's, and so on.

Example 10.6. *For the equivalence relation from example 10.5, the equivalence class of 2 is the set of all even integers.*

$$[2] = \{n \mid 2\,P\,n\} = \{n \mid 2 \text{ has the same parity as } n\}$$
$$= \{n \mid n \text{ is even}\} = \{\dots, -4, -2, 0, 2, 4, \dots\}$$

In this example, there are two different equivalence classes, the one comprising all the even integers, and the other comprising all the odd integers. As far as parity is concerned, -1232215 and 171717 are the same.

Suppose E is an equivalence relation on A. The most important fact about equivalence classes is that every element of A belongs to exactly one equivalence class. Let's prove that.

Theorem 10.7. *Let E be an equivalence relation on a set A, and let $a \in A$. Then there is exactly one equivalence class to which a belongs.*

Proof. Let E be an equivalence relation on a set A, and suppose $a \in A$. Since E is reflexive, $a\,E\,a$, and so $a \in [a]$ is true. That proves that a is in at least one equivalence class. To complete the proof, we need to show that if $a \in [b]$ then $[b] = [a]$.

Now, stop and think: Here is what we know:

(1) E is an equivalence relation on A,

(2) $a \in [b]$, and

(3) the definition of equivalence class.

Using those three pieces of information, we need to show the two sets $[a]$ and $[b]$ are equal. Now, to show two sets are equal, we show they have the same elements. In other words, we want to prove

(1) If $c \in [a]$, then $c \in [b]$, and

(2) If $c \in [b]$, then $c \in [a]$.

Let's give a direct proof of (2).

Suppose $c \in [b]$. Then, according to the definition of $[b]$, $c\,E\,b$. The goal is to end up with $c \in [a]$. Now, we know $a \in [b]$, and that means $a\,E\,b$. Since E is symmetric and $a\,E\,b$, it follows that $b\,E\,a$. Now we have $c\,E\,b$ and $b\,E\,a$. Since E is transitive, we can conclude $c\,E\,a$, which means $c \in [a]$ as we hoped to show. That proves (2). □

For homework, you will complete the proof of this theorem by doing part (1).

Definition 10.8. A **partition** of a set A is a collection of nonempty, pairwise disjoint subsets of A, so that A is the union of the subsets in the collection. So for example {{1, 2, 3} {4, 5, 6}} is a partition of $\{1, 2, 3, 4, 5, 6\}$. The subsets forming a partition are called the **parts of the partition**.

So to express the meaning of theorem 10.7 above in different words: The different equivalence classes of an equivalence relation on a set partition the set into nonempty disjoint pieces. More briefly: the equivalence classes of E **partition** A.

The fact that an equivalence relation partitions the underlying set is reflected in the digraph of an equivalence relation. If we pick an equivalence class $[a]$ of an equivalence relation E on a finite set A and we pick $b \in [a]$, then $b\,E\,c$ for all $c \in [a]$. This is true since $a\,E\,b$ implies $b\,E\,a$ and if $a\,E\,c$, then transitivity fills in $b\,E\,c$. So in any digraph for E every vertex of $[a]$ is connected to every other vertex in $[a]$ (including itself) by a directed edge. Also no vertex in $[a]$ is connected to any vertex in $A - [a]$. So the digraph of E consists of separate components, one for each distinct equivalence class, where each component contains every possible directed edge.

In terms of a matrix representation of an equivalence relation E on a finite set A of size n, let the distinct equivalence classes have size $k_1, k_2, \dots k_r$, where $k_1 + \cdots + k_r = n$. Next list the elements of A as $a_{11}, \dots, a_{k_1 1}, a_{12}, \dots, a_{k_2 2}, \dots, a_{1r}, \dots, a_{k_r r}$ where the ith equivalence class is $\{a_{1i}, \dots, a_{k_i i}\}$. Then the matrix for R with respect to this ordering is of the form

$$\begin{bmatrix} J_{k_1} & 0 & 0 & \dots & 0 \\ 0 & J_{k_2} & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & J_{k_r-1} & 0 \\ 0 & \cdots & 0 & 0 & J_{k_r} \end{bmatrix}$$

where J_{k_m} is the all 1's matrix of size $k_m \times k_m$. Conversely if the digraph of a relation can be drawn to take the above form, or if it has a matrix representation of the above form, then it is an equivalence relation and therefore reflexive, symmetric, and transitive.

Exercises

Exercise 10.1. *Let* $A = \{0,1,2\}$. *Let* $R = \{(0,0),(1,1),(2,2),(0,1),(1,0)\}$. *Is* R *an equivalence relation on* A? *If it is, what are the equivalence classes?*

Exercise 10.2. *Let* $A = \{0,1,2,3\}$. *Let* $R = \{(0,0),(1,1),(2,2),(0,1),(1,0)\}$. *Is* R *an equivalence relation on* A? *If it is, what are the equivalence classes?*

Exercise 10.3. *Let* $A = \{0,1,2\}$. *Let* $R = \{(0,0),(1,1),(2,2),(0,1)\}$. *Is* R *an equivalence relation on* A? *If it is, what are the equivalence classes?*

Exercise 10.4. *Let* $A = \{0,1,2\}$. *Let* $R = \{(0,0),(1,1),(2,2),(0,1),(1,0),(1,2),(2,1)\}$. *Is* R *an equivalence relation on* A? *If it is, what are the equivalence classes?*

Exercise 10.5. *True or False: The relation* $R = \{(1,1),(2,2)\}$ *on* $A = \{1,2\}$ *is both symmetric and antisymmetric.*

Exercise 10.6. *The relation* S *is defined on the set* $\mathbb{Z}$ *of all integers by the rule* $m\,S\,n$ *if and only if* $m^2 = n^2$. *Is* S *an equivalence relation on* $\mathbb{Z}$? *If it is, what are the equivalence classes of* S?

Exercise 10.7. *Let* L *be the collection of all straight lines in the plane. Four examples of elements in* L: $x+y = 0$, $2x\text{-}y = 5$, $x = 7$, $y = 0$. *A relation* C *on* L *is defined by the rule* $l_1\,C\,l_2$ *provided the lines* l_1 *and* l_2 *have at least one point in common. (The letter* C *should remind us of cross, and, loosely speaking, two lines are related if they cross each other. We will have to agree that a line crosses itself.) Is* C *an equivalence relation on* L? *If it is, what are the equivalence classes of* C?

Exercise 10.8. *Let* R *be a relation on a non-empty set* A *that is both symmetric, transitive. And, suppose that for each* $a \in A$, aRb *for at least one* $b \in A$. *Prove that* R *is reflexive, hence, an equivalence relation.*

Exercise 10.9. *Let* E *be an equivalence relation on a set* A, *and let* $a, b \in A$. *Prove that either* $[a] \cap [b] = \emptyset$ *or else* $[a] = [b]$.

Exercise 10.10. *Let* $A = \{1,2,3,4,5,6,7,8\}$. *Form a partition of* A *using* $\{1,2,4\}, \{3,5,7\}$, *and* $\{6,8\}$. *These are the equivalence classes for an equivalence relation,* E, *on* A.

a) Draw a digraph of E.

b) Determine a $0-1$ *matrix of* E.

Exercise 10.11. *Let* $A = \{a, b, c, d, e, f, g, h\}$. *Determine if each matrix represents an equivalence relation on* A. *If the matrix represents an equivalence relation find the equivalence classes. The natural ordering of the elements,* a, b, c, d, e, f, g, h, *is used to define the matrices.*

$$(a)\begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \qquad (b)\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

Exercise 10.12. *Complete the proof of theorem 10.7 on page 67 by proving part (1).*

Problems

Problem 10.1. *Let A be the set of people alive on earth. For each relation defined below, determine if it is an equivalence relation on A. If it is, describe the equivalence classes. If it is not, determine which properties of an equivalence relation fail.*

a) $a\,H\,b \Leftrightarrow$ *a and b are the same height.*

b) $a\,G\,b \Leftrightarrow$ *a and b have a common grandparent.*

c) $a\,L\,b \Leftrightarrow$ *a and b have the same last name.*

d) $a\,N\,b \Leftrightarrow$ *a and b have a name (first name or last name) in common.*

e) $a\,W\,b \Leftrightarrow$ *a and b were born less than a day apart.*

Problem 10.2. *Let L be the collection of all straight lines in the plane. Four examples of elements in L: $x + y = 0$, $2x\text{-}y = 5, x = 7, y = 0$. A relation P on L is defined by the rule $l_1\, P\, l_2$ provided the lines l_1 and l_2 are parallel. Is P an equivalence relation on L? If it is, what are the equivalence classes of P?*

Problem 10.3. *Consider the relation $S(x, y)$: x* ***is a sibling of*** *y on the set, A, of people alive on earth. Is S reflexive? Is S symmetric? Is S transitive? (To be precise, siblings will mean two different people with the same two parents. Don't consider half-siblings for this problem.)*

Problem 10.4. *The relation $R = \{(a, a), (a, b)\}$ is not an equivalence relation on the set $A = \{a, b, c\}$. What is the fewest number of ordered pairs that need to be added to R so the result is an equivalence relation on A?*

Problem 10.5. *Let A be the set of all ordered pairs of positive integers. Some members of A are $(3,6), (7,7), (11,4), (1,2981)$. A relation R on A is defined by the rule $(a, b)R(c, d)$ if and only if $ad = bc$. For example $(3,5)R(6,10)$ is true since $(3)(10) = (5)(6)$.*

a) *Explain why R is an equivalence relation on A.*

b) *List four ordered pairs in the equivalence class of $(2,3)$.*

Problem 10.6. *Let $A = \{1,2,3,4,5,6\}$. Form a partition of A using $\{1,2\}, \{3,4,5\}$, and $\{6\}$. These are the equivalence classes for an equivalence relation, E, on A. Draw the* **digraph** *of E.*

Problem 10.7. *Let $A = \{1,2,3\}$. The relation $E = \{(1,1),(2,2),(3,3),(2,3),(3,2)\}$ is an equivalence relation on A. The relation $F = \{(1,1),(2,2),(3,3),(1,2),(2,1)\}$ is another equivalence relation on A. Compute the composition $F \circ E$. Is $F \circ E$ and equivalence relation on A?*

Chapter 11

Functions and Their Properties

In algebra, functions are thought of as formulas such as $f(x) = x^2$ where x is any real number. This formula gives a rule that describes how to determine one number if we are handed some number x. So, for example, if we are handed $x = 2$, the function f says that determines the value 4, and if we are handed 0, f says that determines 0. There is one condition that, by mutual agreement, such a function rule must obey to earn the title function: the rule must always determine *exactly one value* for each (reasonable) value it is handed. Of course, for the example above, $x = blue$ isn't a reasonable choice for x, so f doesn't determine a value associated with *blue*. The **domain** of this function is all real numbers.

Instead of thinking of a function as a formula, we could think of a function as any rule which determines exactly one value for every element of a set A. For example, suppose W is the set of all words in English, and consider the rule, I, which associates with each word, w, the first letter of w. Then $I(cat) = c, I(dog) = d, I(a) = a,$ and so on. Notice that for each word w, I always determines exactly one value, so it meets the requirement of a function mentioned above. Notice that for the same set of all English words, the rule $T(w)$ *is the third letter of the word* w is not a function since, for example, $T(be)$ has no value.

Here is the semi-formal definition of a function: A **function** from the set A to the set B is any rule which describes how to determine exactly one element of B for each element of A. The set A is called the **domain** of f, and the set B is called the **codomain** of f. The notation $f: A \to B$ means f is a function from A to B.

x	$f(x)$
1	a
2	a
3	c
4	b
5	d
6	e

Table 11.1: A simple function

There are cases where it is not convenient to describe a function with words or formulas. In such cases, it is often possible to simply make a table listing the members of the domain along with the associated member of the codomain.

Example 11.1. *Let $A = \{1, 2, 3, 4, 5, 6\}$, $B = \{a, b, c, d, e\}$ and let $f : A \to B$ be specified by table 11.1*

It is hard to imagine a verbal description that would act like f, but the table says it all. It is traditional to write such tables in a more compact form as

$$f = \{(1, a), (2, a), (3, c), (4, b), (5, d), (6, e)\}.$$

The last result in example 11.1 looks like a relation, and that leads to the modern definition of a function:

Definition 11.2. A function, f, with domain A and codomain B is a relation from A to B (hence $f \subseteq A \times B$) such that each element of A is the first coordinate of exactly one ordered pair in f.

That completes the evolution of the concept of function from formula, through rule, to set of ordered pairs. When dealing with functions, it is traditional to write $b = f(a)$ instead of $(a, b) \in f$.

In algebra and calculus, the functions of interest have a domain and a codomain consisting of sets of real numbers, $A, B \subseteq \mathbb{R}$. The *graph* of f is the set of ordered pairs in the Cartesian plane of the form $(x, f(x))$. Normally in this case, the output of the function f is determined by some formula. For example, $f(x) = x^2$.

We can spot a function in this case by the **vertical line test**. A relation from a subset A of $\mathbb{R}$ to another subset of $\mathbb{R}$ is a function if every vertical line of the form $x = a$, where $a \in A$ intersects the graph of f exactly once.

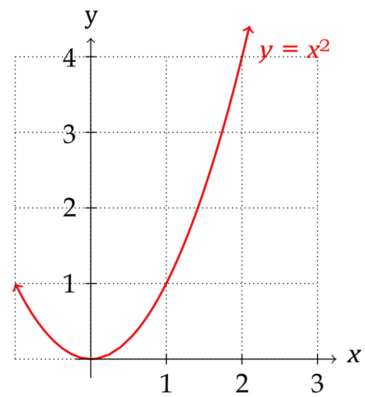


Figure 11.1: Graph of $y = x^2$

In discrete mathematics, most functions of interest have a domain and codomain some finite sets, or, perhaps a domain or codomain consisting of integers. Such domains and codomains are said to be **discrete**.

When $f: A \to B$ is a function and both A and B are finite, then since f is a relation, we can represent f either as a $0-1$ matrix or a bipartite graph. If M is a $0-1$ matrix which represents a function, then since every element of A occurs as the first entry in exactly one ordered pair in f, it must be that every row of M has exactly one 1 in it. So it is easy to distinguish which relations are functions, and which are not from the matrix for the relation. This is the discrete analog of the vertical line test, (but notice that rows are horizontal).

Example 11.3. *Again, let's consider the function defined, as in example 11.1, by f is from $A = \{1, 2, 3, 4, 5, 6\}$ to $B = \{a, b, c, d, e\}$ given by the relation $f = \{(1, a), (2, a), (3, c), (4, b), (5, d), (6, e)\}$.*

If we take the given orderings of A and B, then the $0{-}1$ matrix representing the function f appears in figure 11.2.

Notice that in matrix form the number of 1's in a column coincides with the number of occurrences of the column label as output of the function. So the sum of all entries in a given column equals the number of times the element labeling that column is an output of the function.

When a function from A to B is represented as a bipartite graph, every vertex of A is connected to exactly one element of B.

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Figure 11.2: A function in 0−1 matrix form

In the case of a function whose domain is a subset of $\mathbb{R}$, the number of times that the horizontal line $y = b$ intersects the graph of f, is the number of inputs from A for which the function value is b. Notice that these criteria are twisted again. In the finite case we are now considering vertical information, and in the other case we are considering horizontal information. In either case, these criteria will help us determine which of several special properties a function either has or lacks.

We say that a function $f: A \to B$ is **one-to-one** provided $f(s) = f(t)$ implies $s = t$. A more sophisticated word for one-to-one is **injective**. The definition can also be expressed in the contrapositive as: f is one--to-one provided $s \neq t$ implies $f(s) \neq f(t)$. But the definition is even easier to understand in words: a function is one-to-one provided different inputs always result in different outputs. As an example, consider the function $f: \mathbb{R} \to \mathbb{R}$ given by the formula $f(x) = x^2$. This function is not one-to-one since both inputs 2 and −2 are associated with the same output: $f(2) = 2^2 = 4$ and $f(-2) = (-2)^2 = 4$.

Example 11.4. *Proving a function is one-to-one can be a chore. Let $f: \mathbb{R} \to \mathbb{R}$ be given by $f(x) = x^3 - 2$. Let's prove f is one-to-one.*

Proof. Suppose $f(s) = f(t)$, then $s^3 - 2 = t^3 - 2$. Thus $s^3 = t^3$. So $s^3 - t^3 = 0$. Now $s^3 - t^3 = (s - t)(s^2 + st + t^2)$ implies $s - t = 0$ or $s^2 + st + t^2 = 0$. The first case leads to $s = t$. Using the quadratic formula, the second case leads to $s = -t + \frac{\sqrt{t^2 - 4t^2}}{2}$. Since s has to be a real number, the expression under the radical cannot be negative. The only other option is that it is zero, and that means $t = 0$. Of course if $t = 0$ this leads to $s = 0 = t$. So, in any case, $s = t$. □

The one-to-one property is very easy to spot from either the matrix or the bipartite graph of a function. When $f: A \to B$ is one-to-one, and $|A| = m$ and $|B| = n$ for some $m, n \in \mathbb{N} - \{0\}$, then when f is represented by a 0−1 matrix M, there can be no more than one 1 in any column. So the column sums of any $0 - 1$ matrix representing a one-to-one function are all less than or equal to 1.

Since every row sum of M is 1 and there are m rows, we must have $m \leq n$. The bipartite graph of a one-to-one function can be recognized by the feature that no vertex of the codomain has more than one edge leading to it.

We say that a function $f: A \to B$ is **onto**, or **surjective**, if every element of B equals $f(a)$ for some $a \in A$. Consequently any matrix representing an onto function has each column sum at least one, and thus $m \geq n$. In terms of bipartite graphs, for an onto function, every element of the codomain has at least one edge leading to it.

As an example, consider again the function L from all English words to the set of letters of the alphabet defined by the rule $L(w)$ is the last letter of the word w. This function is not one-to-one since, for example, $L(cat) = L(mutt)$, so two different members of the domain of L are associated with the same member (namely t) of the codomain. However, L is onto. We could prove that by making a list of twenty-six words, one ending with a, one ending with b, $\cdots$, one ending with z. (Only the letters j and q might take more than a moment's thought.)

A function $f: A \to B$ which is both one-to-one and onto is called **bijective**. In the matrix of a bijection, every column has exactly one 1 and every row has exactly one 1. So the number of rows must equal the number of columns. In other words, if there is a bijection $f: A \to B$, where A is a finite set, then A and B have the same number of elements. In such a case we will say the sets have the same **cardinality** or that they are **equinumerous**, and write that as $|A| = |B|$. The general definition (whether A and B are finite or not) is:

Definition 11.5. A and B are *equinumerous* provided there exists a bijection from A to B.

Notice that for finite sets with the same number of elements, A, B, any one-to-one function must be onto and vice versa. This is not true for infinite sets. For example the function $f: \mathbb{Z} \to \mathbb{Z}$ by $f(m) = 2m$ is one-to-one, but not onto, since $f(n) = 1$ is impossible for any n. On the other hand, the function $g: \mathbb{Z} \to \mathbb{Z}$ given by the rule $g(n)$ is the smallest integer that is greater than or equal to $n/2$ is onto, but not one-to-one. As examples $g(6) = 2$ and $g(-5) = -2$. This function is onto since clearly $g(2n) = n$ for any integer n, so every element of the codomain has at least one edge leading to it. But $g(1) = g(2) = 1$, so g is not one-to-one.

Since functions are relations, the **composition** of a function g: $A \to B$ by a function f: $B \to C$, makes sense. As usual, this is written as $f \circ g$: $A \to C$, but that's a little presumptive since it seems to assume that $f \circ g$ really is a function.

Theorem 11.6. *If $g: A \to B$ and $f: B \to C$, then $f \circ g$ is a function.*

Proof. We need to show that for each $a \in A$ there is exactly one $c \in C$ such that $(a, c) \in f \circ g$. So suppose $a \in A$. since $g: A \to B$, there is some $b \in B$ with $(a, b) \in g$. Since $f: B \to C$, there is a $c \in C$ such that $(b, c) \in f$. So, by the definition of composition, $(a, c) \in f \circ g$. That proves there is at least one $c \in C$ with $(a, c) \in f \circ g$. To complete the proof, we need to show that there is only one element of C that $f \circ g$ pairs up with a. So, suppose that (a, c) and (a, d) are both in $f \circ g$. We need to show $c = d$. Since (a, c) and (a, d) are both in $f \circ g$, there must be elements $s, t \in B$ such that $(a, s) \in g$ and $(s, c) \in f$, and also $(a, t) \in g$ and $(t, d) \in f$. Now, since g is a function, and both (a, s) and (a, t) are in g, we can conclude $s = t$. So when we write $(t, d) \in f$, we might as well write $(s, d) \in f$. So we know (s, c) and (s, d) are both in f. As f is a function, we can conclude $c = d$. □

If $g: A \to B$ and $f: B \to C$, and $(a, b) \in g$ and $(b, c) \in f$, then $(a, c) \in f \circ g$. Another way to write that is $g(a) = b$ and $f(b) = c$. So $c = f(b) = f(g(a))$. That last expression looks like the familiar formula for the composition of functions found in algebra texts: $(f \circ g)(x) = f(g(x))$.

When $f: A \to B$ is a function, we can form the relation f^{-1} from B to A. But f^{-1} might not be a function. For example, suppose $f: \{a, b\} \to \{1, 2\}$ is $f = \{(a, 1), (b, 1)\}$. Then $f^{-1} = \{(1, a), (1, b)\}$, definitely not a function.

If in fact f^{-1} is a function, then for all $a \in A$ with $b = f(a)$, we have $f^{-1}(b) = a$ so $f^{-1} \circ f(a) = f^{-1}(f(a)) = f^{-1}(b) = a, \forall a \in A$. Similarly $(f \circ f^{-1})(b) = b, \forall b \in B$. In this case we say f is **invertible**. Another way to say the same thing: the inverse of a function $f: A \to B$ is a function $g: B \to A$ which undoes the operation of f. As a particular example, consider the function $f: \mathbb{Z} \to \mathbb{Z}$ given by the formula $f(n) = n + 3$. In words, f is the **add** 3 function. The operation which undoes the effect of f is clearly the **subtract** 3 function. That is, $f^{-1}(n) = n - 3$.

For any set S define $1_S: S \to S$ by $1_S(x) = x$ for every $x \in S$. In other words $1_S = \{(x, x) | x \in S\}$. The function 1_S is called the **identity function on** S. So the computations show $f^{-1} \circ f = 1_A$ and $f \circ f^{-1} = 1_B$.

Theorem 11.7. *A function $f: A \to B$ is invertible iff f is bijective.*

Proof. First suppose that $f: A \to B$ is invertible. Then $f^{-1}: B \to A$ exists. If $f(a_1) = f(a_2)$, then since f^{-1} is a function, $a_1 = f^{-1}(f(a_1)) = f^{-1}(f(a_2)) = a_2$. Thus f is one-to-one. Also if $b \in B$ with $f^{-1}(b) = a$, then $f(a) = f(f^{-1}(b)) = b$. So f is onto. Since f is one-to-one and onto, f is bijective.

Now suppose that f is bijective and let $b \in B$. Since f is onto, we have some $a \in A$ with $f(a) = b$. If $e \in A$ with $f(e) = b$, then $e = a$ since f is one-to-one. Thus b is the first entry in exactly one ordered pair in the inverse relation f^{-1}. Whence, f^{-1} is a function. □

Do not make the error of confusing inverses and reciprocals when dealing with functions. The **reciprocal** of $f: \mathbb{Z} \to \mathbb{Z}$ given by the formula $f(n) = n + 3$ is $1f(n) = 1(n + 3)$ which is not the inverse function for f. For example $f(0) = 3$, but the reciprocal of f does not convert 3 back into 0, instead the reciprocal associates $1/6$ with 3.

In fact, there are other problems with the reciprocal: it doesn't even make sense when $n = -3$ since that would give a division by 0, which is undefined. So, be very careful when working with functions not to confuse the words reciprocal and inverse. They are entirely different things.

The characteristic vector (see section 6) of a set may be used to define a special 0-1 function representing the given set.

Example 11.8. *Let $\mathcal{U}$ be a finite universal set with n elements ordered $u_1, \dots, u_n$. Let B_n denote all binary strings of length n. The characteristic function $\chi: P(\mathcal{U}) \to B_n$, which takes a subset A to its characteristic vector is bijective. Thus there is no danger of miscalculation. We can either manipulate subsets of $\mathcal{U}$ using set operations and then represent the result as a binary vector or we can represent the subsets as binary vectors and manipulate the vectors with appropriate bit string operations. We'll get exactly the same answer either way.*

The process in example 11.8 allows us therefore to translate any set theory problem with finite sets into the world of 0's and 1's. This is the essence of computer science.

Exercises

Exercise 11.1. *Recall that* $\mathbb{R}$ *is the set of all real numbers. In each case, give an example of a function* $f\colon \mathbb{R} \to \mathbb{R}$ *with the indicated properties, or explain why no such function exists.*

(a) f *is bijective, but* f *is not the identity function* $f(x) = x$.

(b) f *is neither one-to-one nor onto.*

(c) f *is one-to-one, but not onto.*

(d) f *is onto, but not one-to-one.*

Exercise 11.2. *Let* $\mathrm{A} = \{1, 2, 3, 4, 5\}$ *and* $\mathrm{B} = \{\mathrm{a}, \mathrm{b}, \mathrm{c}, \mathrm{d}, \mathrm{e}, \mathrm{f}\}$. *In each case, give an example of a function* $\mathrm{f} : \mathrm{A} \to \mathrm{B}$ *with the indicated properties, or explain why no such function exists.*

(a) $f\colon A \to B, f$ *is one-to-one.*

(b) $g\colon B \to A, g$ *is one-to-one.*

(c) $f\colon A \to B, f$ *is onto.*

(d) $g\colon B \to A, g$ *is onto.*

Exercise 11.3. *Prove or give a counterexample: If* E *is an equivalence relation on a set* A, *then* $E \circ E$ *is an equivalence relation on* A.

Exercise 11.4. *Suppose* $g\colon A \to B$ *and* $f\colon B \to C$ *are both one-to-one. Prove* $f \circ g$ *is one-to-one.*

Problems

Problem 11.1. *Let* A = $\{1, 2, 3, 4, 5, 6\}$. *In each case, give an example of a function* $f : A \to A$ *with the indicated properties, or explain why no such function exists.*

(a) f *is bijective, but is not the identity function* $f(x) = x$.

(b) f *is neither one-to-one nor onto.*

(c) f *is one-to-one, but not onto.*

(d) f *is onto, but not one-to-one.*

Problem 11.2. *Repeat problem 11.1 with the set* $A = \mathbb{N}$.

Problem 11.3. *Repeat problem 11.1 with the set* $A = \mathbb{Z}$.

Chapter 12

Special Functions

Certain functions arise frequently in discrete mathematics. Here is a catalog of some important ones.

To begin with, the **floor function** is a function from $\mathbb{R}$ to $\mathbb{Z}$ which assigns to each real number x, the largest integer which is less than or equal to x. We denote the floor function by $\lfloor x \rfloor$. So $\lfloor x \rfloor = n$ means $n \in \mathbb{Z}$ and $n \leq x < n + 1$. For example, $\lfloor 4.2 \rfloor = 4$, and $\lfloor 7 \rfloor = 7$. Notice that for any integer $n, \lfloor n \rfloor = n$. Be a little careful with negatives: $\lfloor \pi \rfloor = 3$, but $\lfloor -\pi \rfloor = -4$. A dual function is denoted $\lceil x \rceil$, where $\lceil x \rceil = n$ means $n \in \mathbb{Z}$ and $n \geq x > n - 1$. This is the **ceiling function**. For example, $\lceil 4.2 \rceil = 5$ and $\lceil -4.2 \rceil = -4$.

The graph (in the college algebra sense!) of the floor function appears in figure 12.1.

The **fractional part** of a number $x \geq 0$ is denoted $frac(x)$ and equals $x - \lfloor x \rfloor$. For numbers $x \geq 0$, the fractional part of x is just what would be expected: the stuff following the decimal point.

The definition above is the Mathematica and Wolfram/Alpha definition of the fractional part. The Graham definition extends the domain: $frac(x) = x - \lfloor x \rfloor$, for all x.

For example, $frac(5.2) = 5.2 - \lfloor 5.2 \rfloor = 5.2 - 5 = 0.2$. When x is negative its fractional part is

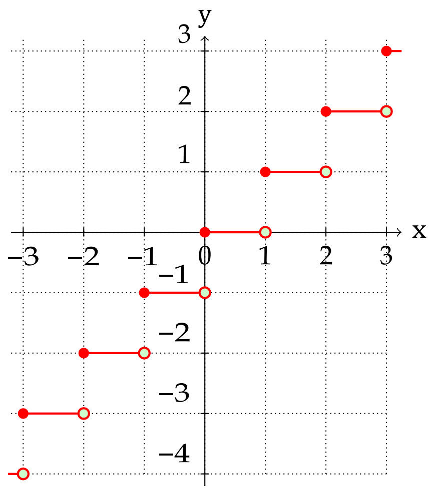


Figure 12.1: Floor function

defined to be $frac(x) = x - \lfloor x \rfloor$. Hence, we have

$$frac(x) = \begin{cases} x - \lfloor x \rfloor, x \geq 0, \\ x - \lceil x \rceil, x < 0. \end{cases}$$

For example, $frac(-5.2) = -5.2 - \lceil -5.2 \rceil = -5.2 - (-5) = -0.2$ In plain English, to determine the fractional part of a number x, take the stuff after the decimal point and keep the sign of the number. The graph of the fractional part function is shown in figure 12.2.

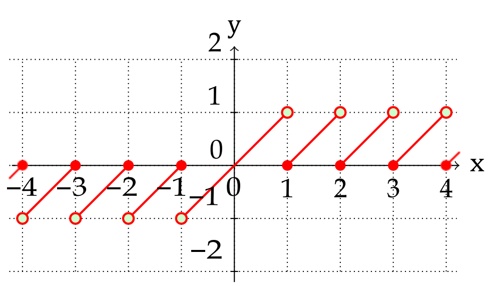


Figure 12.2: Fractional part function

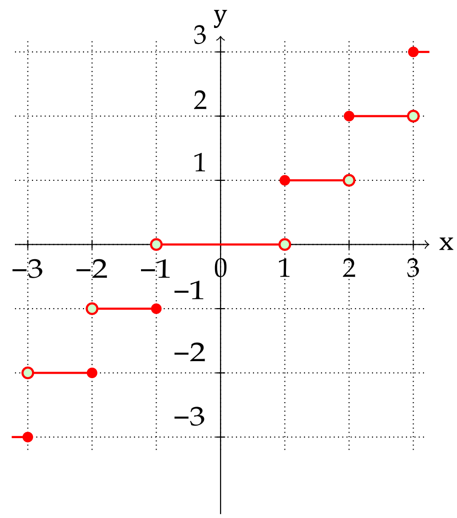


Figure 12.3: Integral part function

For any real number x its **integral part** is defined to be $x - frac(x)$.

The integral part can equivalently be defined by

$$[x] = \begin{cases} \lfloor x \rfloor, x \geq 0, \\ \lceil x \rceil, x < 0. \end{cases}$$

The integral part of x is denoted by $[x]$, or, sometimes, by $int(x)$. In words, the integral part of x is found by discarding everything following the decimal (at least if we agree not to end decimals with an infinite string of 9's such as 2.9999...). The graph of the integral part function is displayed in figure 12.3.

The **power functions** are familiar from college algebra. They are the functions of the form $f(x) = x^2$, $f(x) = x^3, f(x) = x^4$ and so on. By extension, $f(x) = x^a$, where a is any constant greater than or equal to 1 will be called a power function.

For any set X, the unit power function $1_X(x) = x$ for all $x \in X$ is called the **identity** function.

Exchanging the roles of the variable and the constant in the power functions leads to a whole class of nice functions, those of the form $f: \mathbb{R} \to \mathbb{R}$, where $f(x) = a^x$, and $0 < a$. Such a function

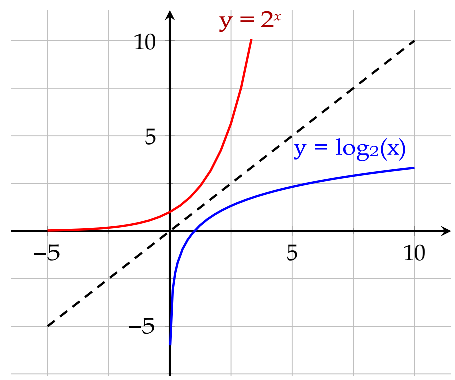


Figure 12.4: 2^x and $\log_2 x$ functions

f is called the **base a exponential function**. The function is not interesting when $a = 1$. Also if $0 < b < 1$, then the function $g(x) = b^x = 1/f(x)$, where $f(x) = a^x$ and $a = 1/b > 1$. So we may focus on $a > 1$. In fact the most important values for a are 2, e and 10.

The number $e \approx 2.718281828459...$ is called the **natural base**, but that story belongs to calculus. Base 2 is the usual base for computer science. Engineers are most interested in base 10, while mathematicians often use the natural exponential function, e^x.

By graphing the function $f: \mathbb{R} \to (0, \infty)$ defined by $y = f(x) = e^x$ we can see that it is bijective. We denote the inverse function $f^{-1}(x)$ by $\ln x$ and call it the **natural log function**. Since these are inverse functions we have

$$e^{\ln a} = a, \forall a > 0 \text{ and } \ln(e^b) = b, \forall b \in \mathbb{R}.$$

As a consequence $a^x = \left(e^{\ln a}\right)^x = e^{(\ln a)\cdot x} = e^{x \ln a}$ is determined as the composition of $y = (\ln(a))\, x$ by the natural exponential function. So every exponential function is invertible with inverse denoted as $\log_a x$, the **base a logarithmic function**. Besides the natural log, $\ln x$, we often write $\lg x$ for the base 2 logarithmic function, and $\log x$ with no subscript to denote the base 10 logarithmic function.

The basic facts needed for manipulating exponential and logarithmic functions are the laws of exponents.

Theorem 12.1 (Laws of Exponents)**.** *For $a, b, c \in \mathbb{R}$, $a^{b+c} = a^b a^c$ and $a^c b^c = (ab)^c$.*

From the laws of exponents, we can derive the

Theorem 12.2 (Laws of Logarithms)**.** *For* $a, b, c > 0, \log_a bc = \log_a b + \log_a c$, *and* $\log_a(b^c) = c \log_a b$.

Proof. We rely on the fact that all exponential and logarithmic functions are one-to-one. Hence, we have that

$$a^{\log_a bc} = bc = a^{\log_a b} a^{\log_a c} = a^{\log_a b + \log_a c}$$

implies

$$\log_a bc = \log_a b + \log_a c.$$

Similarly, the second identity follows from

$$a^{\log_a b^c} = b^c = \left(a^{\log_a b}\right)^c = a^{c(\log_a b)}.$$

□

Notice that $\log_a(1/b) = \log_a(b^{-1}) = -\log_a b$.

Calculators typically have buttons for logs base e and base 10. If $\log_a b$ is needed for a base different from e and 10, it can be computed in a roundabout way. Suppose we need to find $c = \log_a b$. In other words, we need the number c such that $a^c = b$. Taking the ln of both sides of that equation we get

$$\begin{aligned} a^c &= b \\ \ln a^c &= \ln b \\ c \ln a &= \ln b \\ c &= \frac{\ln b}{\ln a} \end{aligned}$$

Hence, we have the general relation between logarithms as follows.

Corollary 12.3. *So we have* $\log_a x = \ln x \,/\, \ln a$.

Example 12.4. *For example, we see that* $\log_2 100 = \ln 100 \,/\, \ln 2 \approx 6.643856$.

Exercises

Exercise 12.1. *In words,* $\lfloor x \rfloor$ *is the largest integer less than or equal to* x. ***Complete the sentence***: In words, $\lceil x \rceil$ is the smallest ....

Exercise 12.2. *Draw a (college algebra) graph of* $f(x) = \lceil x \rceil$.

Exercise 12.3. *Draw a (college algebra) graph of* $f(x) = \lfloor 2x - 1 \rfloor$.

Exercise 12.4. *Draw a (college algebra) graph of* $f(x) = 2\lfloor x - 1 \rfloor$.

Exercise 12.5. *Let* $(x) = 18x$ *and let* $g(x) = x^3/2$. *Sketch the graphs of* f, g *for* $x \geq 1$ *on the same set of axes. Notice that the graph* g *is lower than the graph of* f *when* $x = 1$, *but it is above the graph of* f *when* $x = 9$. *Where does* g *cross the graph of* f *(in other words, where does* g *catch up with* f*)?*

Exercise 12.6. *Let* $f(x) = 4x^5$ *and let* $g(x) = 2^x$. *For values of* $x \geq 1$ *it appears that the graph of* g *is lower than the graph of* f . *Does* g *ever catch up with* f, *or does* f *always stay ahead?*

Exercise 12.7. *The* x^y *button on your calculator is broken. Show how you can approximate* $2^{\sqrt{2}}$ *with your calculator anyhow.*

Problems

Problem 12.1. *Write* $(5/2)\ln 5 - 4\ln 3$ *as a single logarithm.*

Problem 12.2. *Draw the college algebra style graph of* $f(x) = e^{x+3} - 1$.

Problem 12.3. *Let* $f(x) = 2x^3$ *and* $g(x) = 3^x$. *Notice that* $f(1) < g(1)$ *and* $f(2) > g(2)$. *Does* g *ever catch up with* f *again, or does* f *always stay ahead of* g*?*

Problem 12.4. *Write* $\log 1 + \log 2 + \log 3 + \log 4 + \log 5 + \log 6$ *as a single logarithm.*

Problem 12.5. *Write* $\sum_{k=1}^{n} \log k$ *as a single logarithm.*

Chapter 13

Sequences and Summation

A **sequence** is a list of numbers in a specific order. For example, the positive integers 1, 2, 3,... is a sequence, as is the list 4, 3, 3, 5, 4, 4, 3, 5, 5, 4 of the number of letters in the English words of the ten digits in order *zero, one, ... , nine*. Actually, the first is an example of an infinite sequence, the second is a finite sequence. The first sequence goes on forever; there is no last number. The second sequence eventually comes to a stop. In fact the second sequence has only ten items. A **term** of a sequence is one of the numbers that appears in the sequence. The first term is the first number in the list, the second term is the second number in the list, and so on.

A more general way to think of a sequence is as a function from some subset of $\mathbb{Z}$ having a least member (in most cases either $\{0, 1, 2, \ldots\}$ or $\{1, 2, \ldots\}$) with codomain some *arbitrary* set.

Computer science texts use the former and elementary math application texts use the later. Mathematicians use any such well-ordered domain set.

In most mathematics courses the codomain will be a set of numbers, but that isn't necessary. For example, consider the finite sequence of initial letters of the words in the previous paragraph: a, s, i, a, l, o, n, ..., a, s, o. If the letter L is used to denote the function that forms this sequence, then $L(1) = \text{a}$, $L(2) = \text{s}$, and so on.

The examples of sequences given so far were described in words, but there are other ways to tell what objects appear in the sequence. One way is with a formula. For example, let $s(n) = n^2$, for n = 1, 2, 3, ... As the values 1, 2, 3 and so on are plugged into $s(n)$ in succession, the infinite sequence 1, 4, 9, 16, 25, 36, ... is built up. It is traditional to write s_n (or t_n, etc) instead of $s(n)$ when describing the terms of a sequence, so the formula above would usually be seen as $s_n = n^2$. Read that as *s sub n equals* n^2. When written this way, the n in the s_n is called a *subscript* or *index*. The subscript of s_{173} is 173.

Example 13.1. *What is the 50th term of the sequence defined by the formula* $s_j = \frac{j+1}{j+2}$, where j =1,2,3,...? *We see that*

$$s_{50} = \frac{51}{52}.$$

Example 13.2. *What is the 50th term of the sequence defined by the formula* $t_k = \frac{k+1}{k+2}$, where k = 0,1,2,3,...? *We see that*

$$t_{49} = \frac{50}{51}.$$

A sequence can also be specified by listing an initial portion of the sequence and trust the reader to successfully perform the mind reading trick of guessing how the sequence is to continue based on the pattern suggested by those initial terms. For example, consider the sequence 7, 10, 13, 16, 19, 22, ... The symbol ... means *and so on*. In other words, you *should* be able to figure out the way the sequence will continue. This method of specifying a sequence is dangerous of course. For instance, the number of terms sufficient for one person to spot the pattern might not be enough for another person. Also, maybe there are several different *obvious* ways to continue the pattern

Example 13.3. *What is the next term in the sequence* 1, 3, 5, 7 ...? *One possible answer is* 9, *since it looks like we are listing the positive odd integers in increasing order. But another possible answer is* 8: *maybe we are listing each positive integer with an* **e** *in its name. You can probably think of other ways to continue the sequence.*

In fact, for any finite list of initial terms, there are always infinitely many more or less natural ways to continue the sequence. A reason can always be provided for absolutely any number to be the next in the sequence. However, there will typically be only one or two *obvious* simple choices for continuing a sequence after five or six terms.

The simple pattern suggested by the initial terms 7, 10, 13, 16, 19, 22, ... is that the sequence begins with a 7, and each term is produced by adding 3 to the previous term. This is an important type of sequence. The general form is $s_1 = a$ (a is just some specific number), and, from the second term on, each new term is produced by adding d to the previous term (where d is some fixed number).

In the last example, $a = 7$ and $d = 3$. A sequence of this form is called an **arithmetic sequence**. The number d is called the **common difference**, which makes sense since d is the difference of any two consecutive terms of the sequence. It is possible to write down a formula for s_n in this case. After all, to compute s_n we start with the number a, and begin adding d's to it. Adding one d gives $s_2 = a + d$, adding two d's gives $s_3 = a + 2d$, and so on. For s_n we will add $n - 1$ d's to the a, and so we see $s_n = a + (n - 1)d$. In the numerical example above, the 5^{th} term of the sequence ought to be $s_5 = 7 + 4 \cdot 3 = 19$, and sure enough it is. The 407^{th} term of the sequence is $s_{407} = 7 + 406 \cdot 3 = 1225$.

Example 13.4. *The 1^{st} term of an arithmetic sequence is 11 and the 8^{th} term is 81. What is a formula for the n^{th} term?*

We know $a_1 = 11$ and $a_8 = 81$. Since $a_8 = a_1 + 7d$, where d is the common difference, we get the equation $81 = 11 + 7d$. So $d = 10$. We can now write down a formula for the terms of this sequence: $a_n = 11 + (n - 1)10 = 1 + 10n$. Checking, we see this formula does give the required values for a_1 and a_8.

For an arithmetic sequence we added the same quantity to get from one term of the sequence to the next. If instead of adding we multiply each term by the same thing to produce the next term the result is called a **geometric sequence**.

Example 13.5. *Let $s_1 = 2$, and suppose we multiply by 3 to get from one term to the next. The sequence we build now looks like $2, 6, 18, 54, 162, \cdots$, each term being 3 times as large as the previous term.*

In general, if $s_1 = a$, and, for $n \geq 1$, each new term is r times the preceding term, then the formula for the n^{th} term of the sequence is $s_n = ar^{n-1}$, which is reasoned out just as for the formula for the arithmetic sequence above. The quantity r in the geometric sequence is called the **common ratio** since it is the ratio of any term in the sequence to its predecessor (assuming $r \neq 0$ at any rate).

A sequence of numbers is an ordered list of numbers. A **summation** (or just **sum**) is a sequence of numbers added up. A sum with n terms (that is, with n numbers added up) will be denoted by S_n typically. Thus if we were dealing with sequence $1, 3, 5, 7, ..., 2n - 1, ...$, then $S_3 = 1 + 3 + 5$, and $S_n = 1 + 3 + 5 + \cdots + (2n - 1)$. For the arithmetic sequence $a, a + d, a + 2d, a + 3d, ...$, we see $S_n = a + (a + d) + (a + 2d) + \cdots + (a + (n - 1)d)$.

It gets a little awkward writing out such extended sums and so a compact way to indicate a sum, called **summation notation**, is introduced. For the sum of the first 3 odd positive integers above we would write $\sum_{j=1}^{3}(2j-1)$. The Greek letter sigma (Σ) is supposed to be reminiscent of the word summation. The j is called the **index of summation** and the number on the bottom of the Σ specifies the starting value of j while the number above the Σ gives the ending value of j. The idea is that we replace j in turn by 1, 2 and 3, in each case computing the value of the expression following the Σ, and then add up the terms produced. In this example, when $j = 1$, $2j - 1 = 1$, when $j = 2$, $2j - 1 = 3$ and finally, when $j = 3$, $2j - 1 = 5$. We've reached the stopping value, so we have $\sum_{j=1}^{3}(2j-1) = 1 + 3 + 5 = 9$.

Notice that the index of summation takes only integer values. If it starts at 6, then next it is replaced by 7, and so on. If it starts at −11, then next it is replaced by −10, and then by −9, and so on.

The symbol used for the index of summation does not have to be j. Other traditional choices for the index of summation are i, k, m and n. So for example,

$$\sum_{j=0}^{4}(j^2+2) = 2+3+6+11+18 = \sum_{i=0}^{4}(i^2+2) = \sum_{m=0}^{4}(m^2+2)$$

and so on. Even though a different index letter is used, the formulas produce the same sequence of numbers to be added up in each case, so the sums are the same.

Also, the starting and ending points can for the index can be changed without changing the value of the sum provided care is taken to change the formula appropriately. Notice that

$$\sum_{k=1}^{3}(3k-1) = \sum_{k=0}^{2}(3k+2)$$

In fact, if the terms are written out, we see

$$\sum_{k=1}^{3}(3k-1) = 2+5+8$$

and

$$\sum_{k=0}^{2}(3k+2) = 2+5+8.$$

Example 13.6. *We see that*

$$\sum_{m=-1}^{5} 2^m = 2^{-1} + 2^0 + 2^1 + 2^2 + 2^3 + 2^4 + 2^5 = \frac{127}{2}.$$

Example 13.7. *We find that*

$$\sum_{n=3}^{6} 2 = 2 + 2 + 2 + 2 = 8.$$

There are two important formulas for finding sums that are worth remembering. The first is the sum of the first n terms of an arithmetic sequence.

$$S_n = a + (a + d) + (a + 2d) + \cdots + (a + (n-1)d).$$

Here is a clever trick that can be used to find a simple formula for the quantity S_n: the list of numbers is added up twice, once from left to right, the second time from right to left. When the terms are paired up, it is clear the sum is $2S_n = n[a + (a + (n-1)d)]$. A diagram will make the idea clearer:

$$\begin{array}{cccc} a & +(a+d) \quad + (a+2d) & + \cdots + & (a+(n-1)d) \\ +(a+(n-1)d) & +(a+(n-2)d) + (a+(n-3)d) & + \cdots + & a \\ \hline (2a+(n-1)d) & +(2a+(n-1)d) + (2a+(n-1)d) & + \cdots + & (2a+(n-1)d) \end{array}$$

The bottom row contains n terms, each equal to $2a + (n-1)d$, and so $2S_n = n\,[2a + (n-1)d)]$. Dividing by 2 gives the important formula, for $n = 1, 2, 3, \ldots$,

$$S_n = n\left(\frac{2a + (n-1)d}{2}\right) = n\left(\frac{a + (a + (n-1)d)}{2}\right) \tag{13.1}$$

An easy way to remember the formula is to think of the quantity in the parentheses as the average of the first and last terms to be added, and the coefficient, n, as the number of terms to be added.

Example 13.8. *The sum of the first* 20 *terms of the arithmetic sequence* $5, 9, 13, \cdots$ *is found to be*

$$s_{20} = 20\left(\frac{5 + 81}{2}\right) = 860.$$

For a geometric sequence, a little algebra produces a formula for the sum of the first n terms of the sequence. The resulting formula for $S_n = a + ar + ar^2 + \cdots + ar^{n-1}$, is

$$S_n = \frac{a - ar^n}{1 - r} = a\left(\frac{1 - r^n}{1 - r}\right), \text{ if } r \neq 1.$$

Example 13.9. *The sum of the first ten terms of the geometric sequence* 2,2/3,2/9, ... *would be*

$$S_{10} = \frac{2 - 2\left(\frac{1}{3}\right)^{10}}{1 - \left(\frac{1}{3}\right)}$$

The expression for S_{10} *can be simplified as*

$$S_{10} = \frac{2 - 2\left(\frac{1}{3}\right)^{10}}{1 - \left(\frac{1}{3}\right)} = 2\left(\frac{1 - \left(\frac{1}{3}\right)^{10}}{1 - \left(\frac{1}{3}\right)}\right) = 2\left(\frac{1 - \left(\frac{1}{3}\right)^{10}}{\frac{2}{3}}\right) = 3\left(1 - \frac{1}{3^{10}}\right) = 3 - \frac{1}{3^9}.$$

Notice that the numerator in this case is the difference of the first term we have to add in and the term immediately following the last term we have to add in.

Here is the algebra that shows the geometric sum formula is correct.

Let $S_n = a + ar + ar^2 + \cdots + ar^{n-1}$. Multiply both sides of that equation by r to get

$$rS_n = r(a + ar + ar^2 + \cdots + ar^{n-1}) = ar + ar^2 + ar^3 + \cdots + ar^{n-1} + ar^n$$

Now subtract, and observe that most terms will cancel:

$$\begin{aligned} S_n - rS_n &= (a + ar + ar^2 + \cdots + ar^{n-1}) - (ar + ar^2 + ar^3 + \cdots + ar^{n-1} + ar^n) \\ &= a + (ar + ar^2 + ar^3 + \cdots + ar^{n-1}) - (ar + ar^2 + ar^3 + \cdots + ar^{n-1}) - ar^n \\ &= a - ar^n \end{aligned}$$

So $S_n(1 - r) = a - ar^n$. Assuming $r \neq 1$, we can divide both sides of that equation by $1 - r$, producing the promised formula:

$$S_n = \frac{a - ar^n}{1 - r} = a\left(\frac{1 - r^n}{1 - r}\right), \text{ if } r \neq 1. \tag{13.2}$$

You should be able to find a formula for S_n when $r = 1$.

Exercises

Exercise 13.1. *Guess the next term in the sequence* $1, 2, 4, 5, 7, 8, \cdots$. *What's another possible answer?*

Exercise 13.2. *What is the* 100^{th} *term of the arithmetic sequence with initial term* 2 *and common difference* 6*?*

Exercise 13.3. *The* 10^{th} *term of an arithmetic sequence is* -4 *and the* 16^{th} *term is* 47. *What is the* 11^{th} *term?*

Exercise 13.4. *What is the* 5^{th} *term of the geometric sequence with initial term* 6 *and common ratio* 2*?*

Exercise 13.5. *The first two terms of a geometric sequence are* $g_1 = 5$ *and* $g_2 = -11$. *What is* g_5*?*

Exercise 13.6. *Which sequences are both a geometric sequence and also an arithmetic sequence?*

Exercise 13.7. *Evaluate* $\sum_{j=1}^{4}(j^2 + 1)$.

Exercise 13.8. *Evaluate* $\sum_{k=-2}^{4}(2k - 3)$.

Exercise 13.9. *What is the sum of the first* 100 *terms of the arithmetic sequence with initial term* 2 *and common difference* 6*?*

Exercise 13.10. *What is the sum of the first five terms of the geometric sequence with initial term* 6 *and common ratio* 2*?*

Exercise 13.11. *Evaluate* $\sum_{i=0}^{4}\left(-\frac{3}{2}\right)^i$.

Exercise 13.12. *Express in summation notation:* $\frac{1}{2} + \frac{1}{4} + \frac{1}{6} + \cdots + \frac{1}{2n}$ *is the sum of the reciprocals of the first* n *even positive integers.*

Problems

Problem 13.1. *Guess the next term in the sequence* $1, 3, 5, 7, 8, 9 \cdots$. *What's another possible answer?*

Problem 13.2. *Guess the next term in the sequence* $1, 2, 2, 3, 2, 4, 2, 4, 3 \cdots$.

Problem 13.3. *A sequence begins* $1, 3, 9, 15$. *Could it be an arithmetic sequence? Could it be a geometric sequence?*

Problem 13.4. *What is the* 20^{th} *term of the arithmetic sequence with initial term* 4 *and common difference* 5?

Problem 13.5. *The* 8^{th} *term of an arithmetic sequence is* 20 *and the* 12^{th} *term is* 40. *What is the* 25^{th} *term?*

Problem 13.6. *What is the* 7^{th} *term of the geometric sequence with initial term* 3 *and common ratio* 4?

Problem 13.7. *Two terms of a geometric sequence are* $g_3 = 2$ *and* $g_5 = 72$. *There are two possible values for* g_4. *What are those two values?*

Problem 13.8. *A geometric sequence has initial term* 3, *and common ratio* 7. *Determine the smallest value of* n *so that the* n^{th} *term of the sequence is more that one million.*

Problem 13.9. *Evaluate* $\sum_{j=1}^{4}(j+1)^2$.

Problem 13.10. *Evaluate* $\sum_{k=-2}^{4}(2k+3)$.

Problem 13.11. *What is the sum of the first* 100 *terms of the arithmetic sequence with initial term* 2 *and common difference* 6?

Problem 13.12. *What is the sum of the first four terms of the geometric sequence with initial term* 3 *and common ratio* -2?

Problem 13.13. *What is the sum of the first four thousand terms of the geometric sequence with initial term* 3 *and common ratio* -1?

Problem 13.14. *Evaluate* $\sum_{i=0}^{4}\left(\frac{3}{2}\right)^i$.

Problem 13.15. *Express in summation notation:* $\frac{1}{1} + \frac{1}{3} + \frac{1}{5} + \cdots + \frac{1}{2n-1}$ *is the sum of the reciprocals of the first* n *odd positive integers.*

Chapter 14

Recursively Defined Sequences

Besides specifying the terms of a sequence with a formula, such as $a_n = n^2$, an alternative is to give an initial term, usually something like b_1, (or the first few terms, b_1, b_2,b_3 ...) of a sequence, and then give a rule for building new terms from old ones. In this case, we say the sequence has been defined **recursively**.

Example 14.1. *For example, suppose* $b_1 = 1$, *and for* $n > 1$, $b_n = 2b_{n-1}$. *Then the* 1^{st} *term of the sequence will be* $b_1 = 1$ *of course. To determine* b_2, *we apply the rule* $b_2 = 2b_{2-1} = 2b_1 = 2 \cdot 1 = 2$. *Next, applying the rule again,* $b_3 = 2b_{3-1} = 2b_2 = 2 \cdot 2 = 4$. *Next* $b_4 = 2b_3 = 8$. *Continuing in this fashion, we can form as many terms of the sequence as we wish:* $1, 2, 4, 8, 16, 32, \cdots$. *In this case, it is easy to guess a formula for the terms of the sequence:* $b_n = 2^{n-1}$.

In general, to define a sequence recursively, (1) we first give one or more initial terms (this information is called the **initial condition(s)** for the sequence), and then (2) we give a rule for forming new terms from previous terms (this rule is called the **recursive formula**).

Example 14.2. *Consider the sequence defined recursively by* $a_1 = 0$, *and, for* $n \geq 2$, $a_n = 2a_{n-1} + 1$. *The five terms of this sequence are*

$$0, \quad 2 \cdot 0 + 1 = 1, \quad 2 \cdot 1 + 1 = 3, \quad 2 \cdot 3 + 1 = 7, \quad 2 \cdot 7 + 1 = 15 \quad \cdots$$

In words, we can describe this sequence by saying the initial term is 0 *and each new term is one* *more than twice the previous term. Again, it is easy to guess a formula that produces the terms of this sequence:* $a_n = 2^{n-1} - 1$.

Such a formula for the terms of a sequence is called a **closed form formula** to distinguish it from a recursive formula.

There is one big advantage to knowing a closed form formula for a sequence. In example 14.2 above, the closed form formula for the sequence tells us immediately that $a_{101} = 2^{100} - 1$, but using the recursive formula to calculate a_{101} means we have to calculate in turn $a_1, a_2, \cdots a_{100}$, making 100 computations. The closed form formula allows us to jump directly to the term we are interested in. The recursive formula forces us to compute 99 additional terms we don't care about in order to get to the one we want. With such a major drawback why even introduce recursively defined sequences at all? The answer is that there are many naturally occurring sequences that have simple recursive definitions but have no reasonable closed form formula, or even no closed form formula at all in terms of familiar operations. In such cases, a recursive definition is better than nothing.

There are methods for determining closed form formulas for some special types of recursively defined sequences. Such techniques are studied later in chapter 35. For now we are only interested in understanding recursive definitions, and determining some closed form formulas by the method of *pattern recognition* (aka *guessing*).

The most famous recursively defined sequence is due to Fibonacci. There are two initial conditions: $f_0 = 0$ and $f_1 = 1$. The index starts at **zero**, by tradition. The recursive rule is, for $n \geq 2$, $f_n = f_{n-1} + f_{n-2}$. In words, each new term is the sum of the two terms that precede it. So, the **Fibonacci sequence** begins

$$0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, \cdots$$

There is a closed form formula for the Fibonacci Sequence, but it is not at all easy to guess:

$$f_n = \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^n - \frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^n, n \geq 0.$$

For a positive integer n, the symbol $n!$ is read ***n* factorial** and it is defined to be the product of all the positive integers from 1 to n. For example, $5! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 120$. In order to make many formulas work out nicely, the value of $0!$ is defined to be 1.

A recursive formula can be given for $n!$. The initial term is $0! = 1$, and the recursive rule is, for $n \geq 1$, $n! = n[(n-1)!]$. Hence, the first few factorial values are:

$$
\begin{aligned}
1! &= 1[0!] = 1 \cdot 1 = 1,\\
2! &= 2[1!] = 2 \cdot 1 = 2,\\
3! &= 3[2!] = 3 \cdot 2 = 6,\\
4! &= 4[3!] = 4 \cdot 6 = 24,\\
&\ \ \vdots
\end{aligned}
$$

We sometimes write a *general* formula for the factorial as

$$n! = 1 \cdot 2 \cdot 3 \cdot 4 \cdots n, \text{ for } n > 0.$$

The sequence of factorials grows very quickly. Here are the first few terms:

1, 2, 6, 24, 120, 720, 5040, 40320, 362880, 3628800, 39916800, 479001600, 6227020800, $\cdots$

Consider the terms of an arithmetic sequence with initial term a and common difference d:

$$a, \ (a+d), \ (a+2d), \cdots, \ (a+(n-1)d), \cdots .$$

These terms may clearly be found by adding d to the current term to get the next. That is, the arithmetic sequence may be defined recursively as (1) $a_1 = a$, and (2) for $n \geq 2$, $a_n = a_{n-1} + d$.

Exercises

Exercise 14.1. *List the first five terms of the sequence defined recursively by* $a_1 = 3$*, and, for* $n \geq 2$*,* $a_n = a_{n-1}(2 + a_{n-1})$.

Exercise 14.2. *List the first seven terms of the sequence defined recursively by* $a_0 = 1$*,* $a_1 = 1$*, and, for* $n \geq 2$*,* $a_n = 1 + a_{n-1}a_{n-2}$.

Exercise 14.3. *List the first ten terms of the sequence defined recursively by* $a_0 = 1$*, and, for* $n \geq 1$*,* $a_n = 1 + a_{\lfloor n/2 \rfloor}$.

Exercise 14.4. *List the first ten terms of the sequence defined recursively by* $a_0 = 1$*, and for* $n \geq 1$*,* $a_n = 2n - a_{n-1} - 1$*, and guess a closed form formula for* a_n.

Exercise 14.5. *The first few terms of a sequence are*

1, 11, 21, 1211, 111221, 312211, 13112221, 1113213211.

There is an easy recursive rule for building the terms of this sequence. Guess the next term.

Exercise 14.6. *Let* d *be a fixed real number. For a positive integer* n*, the symbol* nd *means the sum of* n d*'s. Give a recursive definition of* nd *analogous to the definition of* $n!$ *given in this chapter.*

Problems

Problem 14.1. *List the first five terms of the sequence defined recursively by* $a_1 = 2$*, and, for* $n \geq 2$*,* $a_n = a_{n-1}^2 - 1$.

Problem 14.2. *List the first five terms of the sequence defined recursively by* $a_1 = 2$*, and, for* $n \geq 2$*,* $a_n = 3a_{n-1} + 2$*. Guess a closed form formula. Hint: This is a lot like example 14.2. for the sequence.*

Problem 14.3. *List the first five terms of the sequence with initial terms* $u_0 = 2$ *and* $u_1 = 5$*, and, for* $n \geq 2$*,* $u_n = 5u_{n-1} - 6u_{n-2}$*. Guess a closed form formula for the sequence. Hint: The terms are simple combinations of powers of* 2 *and powers of* 3*.*

Problem 14.4. *Let* r *be a fixed real number different from* 0*. For a positive integer* n*, the symbol* r^n *means the product of* n r*'s. For convenience,* r^0 *is defined to be* 1*. Give a recursive definition of* r^n *analogous to the definition of* $n!$ *given in this chapter.*

Problem 14.5. *Give a recursive definition of the geometric sequence with initial term* 3 *and common ratio* 2*.*

Problem 14.6. *Generalize problem 5: give a recursive definition of the geometric sequence with initial term* a *and common ratio* r*.*

Chapter 15

Recursively Defined Sets

Two different ways of defining a set have been discussed. We can describe a set by the roster method, listing all the elements that are to be members of the set, or we can describe a set using set-builder notation by giving a predicate that the elements of the set are to satisfy. Here we consider defining sets in another natural way: recursion.

Recursive definitions can also be used to build sets of objects. The spirit is the same as for recursively defined sequences: give some initial conditions and a rule for building new objects from ones already known.

Example 15.1. *For instance, here is a way to recursively define the set of positive even integers, E. First the initial condition: $2 \in E$. Next the recursive portion of the definition: If $x \in E$, then $x + 2 \in E$. Here is what we can deduce using these two rules. First of course, we see $2 \in E$ since that is the given initial condition. Next, since we know $2 \in E$, the recursive portion of the definition, with* x *being played by* 2*, says $2 + 2 \in E$, so that now we know $4 \in E$. Since $4 \in E$, the recursive portion of the definition, with x now being played by 4, says $4 + 2 \in E$, so that now we know $6 \in E$. Continuing in this way, it gets easy to believe that E really is the set of positive even integers.*

Actually, there is a little more to do with example 15.1. The claim is that E consists of exactly all the positive even integers. In other words, we also need to make sure that no other things appear in E besides the positive even integers. Could 312211 somehow have slithered into the set E? To verify that such a thing does not happen, we need one more fact about recursively defined sets. The only elements that appear in a set defined recursively are those that make it on the basis of either the initial condition or the recursive portion of the definition. No elements of the set appear, as if by magic, from nowhere.

In this case, it is easy to see that no odd integers sneak into the set. For if so, there would be a smallest odd integer in the set and the only way it could be elected to the set is if the integer two less than it were in the set. But that would mean a yet smaller odd integer would be in the set, a contradiction. We won't go into that sort of detail for the following examples in general. We'll just consider the topic at the intuitive level only.

Example 15.2. *Give a recursive definition of the set, S, of all non-negative integer powers of* 2*.*

Initial condition: $1 \in S$*. Recursive rule: If* $x \in S$*, then* $2x \in S$*. Applying the initial condition and then the recursive rule repeatedly gives the elements:*

$$1 \qquad 2 \cdot 1 = 2 \qquad 2 \cdot 2 = 4 \qquad 2 \cdot 4 = 8 \qquad 2 \cdot 8 = 16$$

and so on, and that looks like the set of nonnegative powers of 2*.*

Example 15.3. *A set, S, is defined recursively by*

(1) (initial conditions) $1 \in S$ *and* $2 \in S$*, and*

(2) (recursive rule) If $x \in \mathrm{S}$*, then* $x + 3 \in S$*. Describe the integers in* S*.*

The plan is to use the initial conditions and the recursive rule to build elements of S *until we can guess a description of the integers in* S*. From the initial conditions we know* $1 \in S$ *and* $2 \in S$*. Applying the recursive rule to each of those we get* $4, 5 \in S$*, and using the recursive rule on those gives* $7, 8 \in S$*, and so on.*

So we get $S = \{1, 2, 4, 5, 7, 8, 10, 11, \cdots\}$ *and it's apparent that* S *consists of of the positive integers that are not multiples of* 3*.*

Recursively defined sets appear in certain computer science courses where they are used to describe sets of strings. To form a string, we begin with an **alphabet** which is a set of symbols, traditionally denoted by Σ. For example $\Sigma = \{a, b, c\}$ is an alphabet of three symbols, and $\Sigma = \{!, @, \#, \$, \%, \&, \mathrm{X}, 5\}$ is an alphabet of eight symbols. A **string** over the alphabet Σ is any finite sequence of symbols from the alphabet. For example $aaba$ is a string of **length** four over the alphabet $\Sigma = \{a, b, c\}$, and !!5X$$5@@ is a length nine string over $\Sigma = \{!, @, \#, \$, \%, \&, X, 5\}$. There is a special string over any alphabet denoted by λ called the **empty string**. It contains no symbols, and has length 0.

Example 15.4. *A set, S, of strings over the alphabet $\Sigma = \{a, b\}$ is given recursively by (1) $\lambda \in S$, and (2) If $x \in S$, then $axb \in S$. Describe the strings in S.*

The notation axb means write down the string a followed by the string x followed by the string b. So if $x = aaba$ then $axb = aaabab$. Let's experiment with the recursive rule a bit, and then guess a description for the strings in S. Starting with the initial condition we see $\lambda \in S$. Applying the recursive rule to λ gives $a\lambda b = ab \in S$. Applying the recursive rule to ab gives $aabb \in S$, and applying the recursive rule to $aabb$ shows $aaabbb \in S$. It's easy to guess the nature of the strings in S: Any finite string of a's followed by the same number of b's.

Example 15.5. *Give a recursive definition of the set S of strings over $\Sigma = \{a, b, c\}$ which do not contain adjacent a's. For example $ccabbbabba$ is acceptable, but $abcbaabaca$ is not.*

For the initial conditions we will use (1) $\lambda \in S$, and $a \in S$. If we have a string with no adjacent a's, we can extend it by adding b or c to either end. But we'll need to be careful when adding more a's. For the recursive rule we will use (2) if $x \in S$, then $bx, xb, cx, xc \in S$ and $abx, xba, acx, xca \in S$.

Notice how the string a had to be put into S in the initial conditions since the recursive rule won't allow us to form that string from λ.

Here is different answer to the same question. It's a little harder to dream up, but the rules are much cleaner. The idea is that if we take two strings with no adjacent a's, we can put them together and be sure to get a new string with no adjacent a's provided we stick either b or c between them. So, we can define the set recursively by (1) $\lambda \in S$ and $a \in S$, and (2) if $x, y \in S$, then $xby, xcy \in S$.

Example 15.6. *Give a recursive definition of the set S of strings over $\Sigma = \{a, b\}$ which contain more a's than b's.*

The idea is that we can build longer strings from smaller ones by (1) sticking two such strings together, or (2) sticking two such strings together along with a b before the first one, between the two strings, or after the last one. That leads to the following recursive definition: (1) $a \in S$ and (2) if $x, y \in S$ then $xy, bxy, xby, xyb \in S$. That looks a little weird since in the recursive rule we added b, but since x and y each have more a's than b's, the two together will have a least two more a's than b's, so it's safe to add b in the recursive rule.

Starting with the initial condition, and then applying the recursive rule repeatedly, we form the following elements of S:

$$a, aa, baa, aba, aab, aaa, baaa, abaa, aaba, baaa, \cdots$$

Example 15.7. *A set, S, of strings over the alphabet $\Sigma = \{a, b\}$ is defined recursively by the rules (1) $a \in S$, and (2) if $x \in S$, then $xbx \in S$. Describe the strings in S.*

Experimenting we find the following elements of S:

$$a, aba, abababa, abababababababa, \cdots$$

It looks like S is the set of strings beginning with a *followed by a certain number of ba's. If we look at the number of ba's in each string, we can see a pattern:* $0, 1, 3, 7, 15, 31, \cdots$*, which we recognize as being the numbers that are one less than the positive integer powers of* 2 $(1, 2, 4, 8, 16, 32, \cdots)$. *So it appears S is the set of strings which consisting of* a *followed by* $2^n - 1$ *pairs ab for some integer* $n \geq 0$.

Exercises

Exercise 15.1. *The set* S *is described recursively by (1)* $1 \in S$, *and (2) if* $n \in S$, *then* $n+1 \in S$. *To what familiar set is* S *equal?*

Exercise 15.2. *Give a recursive definition of the set of positive integers that end with the digits* 17.

Exercise 15.3. *Give a recursive definition of the set of positive integers that are not multiples of* 4.

Exercise 15.4. *Describe the strings in the set* S *of strings over the alphabet* $\Sigma = \{a, b, c\}$ *defined recursively by (1)* $\lambda \in S$ *and (2) if* $x \in S$, *then* $axbc \in S$.

Exercise 15.5. *Describe the strings in the set* S *of strings over the alphabet* $\Sigma = \{a, b, c\}$ *defined recursively by (1)* $c \in S$ *and (2) if* $x \in S$ *then* $ax \in S$ *and* $bx \in S$ *and* $xc \in S$.

Exercise 15.6. *A* **palindrome** *is a string that reads the same in both directions. For example, a classic palindrome with length 21 is: A man, a plan, a canal: panama. For another example,* $aabaa$ *is a palindrome of length five and* $babccbab$ *is a palindrome of length eight. The empty string is also a palindrome. Give a recursive definition of the set of palindromes over the alphabet* $\Sigma = \{a, b, c\}$.

Problems

Problem 15.1. *A set S of integers is defined recursively by the rules: (1) $1 \in S$, and (2) If $n \in S$, then $2n+1 \in S$.*

(1) Is $15 \in S$? Explain your answer.

(2) Is $65 \in S$? Explain your answer.

Problem 15.2. *A set of integers is defined recursively by the rules (1) $0 \in S$, and (2) if $n \in S$, then $2n+2 \in S$. Give a simple description of the integers in S.*

Problem 15.3. *Give a recursive definition of the set*

$$\{3^n - 3 | n \text{ a positive integer}\} = \{0,6,24,78,240,726,2184,\ldots\}.$$

Problem 15.4. *A set, S, of strings over the alphabet $\Sigma = \{a, b, c\}$ is defined recursively by (1) $a \in S$ and (2) if $x \in S$ then $bxc \in S$. List all the strings in S of length seven or less.*

Problem 15.5. *A set, S, of positive integers is defined recursively by the rule: (1) $1 \in S$, and (2) If $n \in S$, then $2n - 1 \in S$. List all the elements in the set S.*

Problem 15.6. *Give a recursive definition of the set of positive integers that end with the digit 1.*

Problem 15.7. *Give a recursive definition of the set of strings over the alphabet $\Sigma = \{a, b, c\}$ of the form $aaa \cdots abccc \cdots c$. More carefully: zero or more a's followed by a single b followed by the same number of c's as a's.*

Problem 15.8. *Describe the strings in the set S of strings over the alphabet $\Sigma = \{a, b, c\}$ defined recursively by (1) $a \in S$ and (2) if $x \in S$ then $ax \in S$ and $xb \in S$ and $xc \in S$.*

Hint: Your description should be a sentence that provides an easy test to check if a given string is in the set or not. An example of such a description is: S consists of all strings of a's, b's, and c's, with more a's than b's. That isn't a correct description since abb is in S and doesn't have more a's than b's, and also baac isn't in S, but does have more a's than b's. So that attempted description is really terrible. One way to do this problem is to use the rules to build a bunch of strings in S until a suitable description becomes obvious. Alternatively, just thinking about the recursive rules might be sufficient for you to see a simple description of the strings in S.

Problem 15.9. *A set S of ordered pairs of integers is defined recursively by (1) $(1,1) \in S$, and (2) if $(m,n) \in S$, then $(m+2,n) \in S$, and $(m,n+2) \in S$, and $(m+1,n+1) \in S$. Give a simple description of the ordered pairs in S.*

Chapter 16

Mathematical Induction

As mentioned earlier, to show that a proposition of the form $\forall\, x\, P(x)$ is true, it is necessary to check that $P(c)$ is true for every possible choice of c in the domain of discourse. If that domain is not too big, it is feasible to check the truth of each $P(c)$ one by one. For instance, consider the proposition *For every page in these notes, the letter **e** appears at least once on the page.* To express the proposition in symbolic form we would let the domain of discourse be the set of pages in these notes, and we would let the predicate E be *has an occurrence of the letter **e***, so the proposition becomes $\forall\, p\, E(p)$. The truth value of this proposition can be determined by the tedious but feasible task of checking every page of the notes for an **e**. If a single page is found with no **e**'s, that page would constitute a counterexample to the proposition, and the proposition would be false. Otherwise it is true.

When the domain of discourse is a finite set, it is, in principle, always possible to check the truth of a proposition of the form $\forall\, x\, P(x)$ by checking the members of the domain of discourse one by one. But that option is no longer available if the domain of discourse is an infinite set since no matter how quickly the checks are made there is no practical way to complete the checks in a finite amount of time.

For example, consider the proposition *For every natural number n, $n^5 - n$ ends with a* 0. Here the domain of discourse is the set $\mathbb{N} = \{0, 1, 2, 3, \cdots\}$. The truth of the proposition could be established by checking:

$0^5 - 0 = 0$	$1^5 - 1 = 0$	$2^5 - 2 = 30$	$3^5 - 3 = 240$
$4^5 - 4 = 1020$	$5^5 - 5 = 3120$	$6^5 - 6 = 7770$	$7^5 - 7 = 16800$
$8^5 - 8 = 32760$	$9^5 - 9 = 59040$	$10^5 - 10 = 99990$	$11^5 - 11 = 161040$
.	.	.	.

(and so on forever.)

Checking these facts one by one is obviously a hopeless task, and, of course, just checking a few of them (or even a few billion of them) will never suffice to prove they are all true. And it is not sufficient to check a few and say that the facts are all clear. That's not a proof, it's only a suspicion. So verifying the truth of $\forall\, n\, (n^5 - n)\ \mathit{ends\ with\ a}\ 0$ for domain of discourse $\mathbb{N}$ seems tough.

In general, proving a universally quantified statement when the domain of discourse is an infinite set is a tough nut to crack. But, in the special case when the domain of discourse is the set $\mathbb{N} = \{0, 1, 2, 3, \cdots\}$, there is a technique called **mathematical induction** that comes to the rescue.

The method of proof by induction provides a way of checking that all the statements in the list are true without actually verifying them one at a time. The process is carried out in two steps. First (the **basis step**) we check that the first statement in the list is correct. Next (the **inductive step**), we show that if any statement in the list is known to be correct, then the one following must also be correct. Putting these two facts together, it ought to appear reasonable that all the statements in the list are correct. In a way, it's pretty amazing: we learn infinitely many statements are true just by checking two facts. It's like killing infinitely many birds with two stones.

So, suppose a list of statements, $p(0), p(1), p(2), \cdots, p(k), p(k+1) \cdots$ is presented and we want to show they are all true. The plan is to show two facts:

(1) $p(0)$ is true, and

(2) for any $n \in \mathbb{N}$, $p(n) \to p(n+1)$.

We then conclude all the statements in the list are true.

The **well ordering property** of the positive integers provides the justification for proof by induction. This property asserts that every non-empty subset of the natural numbers contains a smallest number. In fact, given any nonempty set of natural numbers, we can determine the smallest number in the set by the process of checking to see, in turn, if 0 is in the set, and, if the answer is *no*, checking for 1, then for 2, and so on. Since the set is nonempty, eventually the answer will be *yes, that number is in the set*, and in that way, the smallest natural number in the set will have been found. Now let's look at the proof that induction is a valid form of proof. The statement of the theorem is a little more general than described above. Instead of beginning with a statement $p(0)$, we allow the list to begin with a statement $p(k)$ for some integer k (almost always, $k = 0$ or $k = 1$ in practice). This does not have any effect of the concept of induction. In all cases, we have a list of statements, and we show the first statement is true, and then we show that if any statement is true, so is the next one. The particular name for the starting point of the list doesn't really matter. It only matters that there is a starting point.

Theorem 16.1 (Principle of Mathematical Induction)**.** *Suppose we have a list of statements* $p(k), p(k+1), p(k+2), \cdots, p(n), p(n+1) \cdots$.

(1) $p(k)$ *is true, and*

(2) $p(n) \to p(n+1)$ *for every* $n \geq k$,

then all the statements in the list are true.

Proof. The proof will be by contradiction.

Suppose that 1 and 2 are true, but that it is not the case that $p(n)$ is true for all $n \geq k$. Let $S = \{n | n \geq k$ and $p(n)$ is false$\}$, so that $S \neq \emptyset$. Since S is a non-empty set of integers $\geq k$ it has a least element, say t. So t is the smallest positive integer for which $p(n)$ is false. In the ever colorful jargon of mathematics, t is usually called the *minimal criminal*.

Since $p(k)$ is true, $k \notin S$. Therefore $t > k$. So $t - 1 \geq k$. Since t is the smallest integer $\geq k$ for which p is false, it must be that $p(t-1)$ is true. Now, by part 2, we also know $p(t-1) \to p(t)$ is true. So it must be that $p(t)$ is true, and that is a contradiction. □

Many people find proofs by induction a little bit black-magical at first, but just keep the goals in mind (namely check [1] the first statement in the list is true, and [2] that if any statement in the list is true, so is the one that follows it) and the process won't seem so confusing.

A handy way of viewing mathematical induction is to compare proving the sequence $p(k) \wedge p(k+1) \wedge p(k+2) \wedge ... \wedge p(m) \wedge ...$ to knocking down a set of dominos set on edge and numbered consecutively $k, k+1, ...$ If we want to knock all of the dominos down, which are numbered k and greater, then we must knock the kth domino down, and ensure that the spacing of the dominos is such that every domino will knock down its successor. If either the spacing is off ($\exists m \geq k$ with $p(m)$ not implying $p(m+1)$), or if we fail to knock down the kth domino (we do not demonstrate that $p(k)$ is true), then there may be dominoes left standing.

When checking the inductive step, $p(n) \rightarrow p(n+1)$, the statement $p(n)$, is called the **inductive hypothesis**.

To discover how to prove the inductive step most people start by explicitly listing several of the first instances of the inductive hypothesis $p(n)$. Then, look for how to make, in a general way, an argument from one, or more, instances to the next instance of the hypothesis. Once an argument is discovered that allows us to advance from the truth of previous one, or more, instances, that argument, in general form, becomes the pattern for the proof on the inductive hypothesis. Let's examine an example.

Example 16.2. *Let's prove that, for each positive integer* n, *the sum of the first* n *positive integers is* $n(n+1)/2$. *Here is the list of statements we want to verify.*

$$p(1): \quad 1 = \frac{1(1+1)}{2} \qquad \textit{To get } p(2) \textit{ add } 2 \textit{ to both sides.}$$

$$p(2): \quad 1+2 = \frac{2(2+1)}{2} \qquad \textit{To get } p(3) \textit{ add } 3 \textit{ to both sides.}$$

$$p(3): \quad 1+2+3 = \frac{3(3+1)}{2} \qquad \textit{You will need to simplify each step.}$$

$$\vdots$$

$$p(n): \quad 1+2+\cdots+n = \frac{n(n+1)}{2}$$

$$p(n+1): \quad 1+2+\cdots+(n+1) = \frac{(n+1)(n+2)}{2}$$

$$\vdots$$

Once you figure out the general form of the argument that takes us from one instance of $p(l)$ *to the next, you have the form of the inductive argument.*

Proof. Basis: Let's check the first statement in the list, $p(1)$: 1 = 1(1 + 1)/2, is correct. The left-hand side is 1, and the right-hand side is 1(1 + 1)/2 = 2/2 = 1, so the two sides are equal as claimed.

Inductive Step: Suppose $p(k)$ is true for some integer $k \geq 1$. To be as precise as possible we should suppose $1 + 2 + \cdots + k = \frac{k(k+1)}{2}$ is true for some integer $k \geq 1$.

(We need to show $p(k+1)$ is true. In other words, we need to verify $1 + 2 + \cdots + (k+1) = \frac{(k+1)(k+2)}{2}$. But we don't write this down since it would be **assuming the conclusion**.)

Here are the computations which provide justification of the inductive step:

$$\begin{aligned} 1 + 2 + \cdots + (k+1) &= 1 + 2 + \cdots + k + (k+1) \\ &= \frac{k(k+1)}{2} + (k+1) \text{ using the inductive hypothesis} \\ &= \frac{k(k+1)}{2} + \frac{2(k+1)}{2} \\ &= \frac{k(k+1) + 2(k+1)}{2} \\ &= \frac{(k+1)(k+2)}{2} \end{aligned}$$

as we needed to show. So we conclude all the statements in the list are true. □

Notice that in the previous proof we used the following strategy to prove equality: start on one side of the equation, $p(n+1)$ in this case, and work until we obtain the other side. We did this through a series of algebraic manipulations and using the induction hypothesis along the way. This will be our general strategy when writing induction proofs.

The next example reproves the useful formula for the sum of the terms in a geometric sequence. Recall that to form a geometric sequence, fix a real number $r \neq 1$, and list the integer powers of r starting with $r^0 = 1$: $1, r, r^2, r^3, ..., r^n, ...$. The formula given in the next example shows the result of adding $1 + r + r^2 + \cdots + r^n$. You may be familiar with the extension from calculus which allows us to sum $a + a \cdot r + a \cdot r^2 + \cdots + a \cdot r^n$. For a finite sum the previous sum is simply $a(1 + r + r^2 + \cdots + r^n)$. So the formula derived is sufficient for our purposes.

Example 16.3. *For all* $n \geq 0$, *we have* $1 + r + r^2 + \cdots + r^n = \sum_{k=0}^{n} r^k = \frac{r^{n+1}-1}{r-1}$, $(if\ r \neq 1)$.

Proof. We assume $r \neq 1$.

Basis: When $n = 0$ we have $\sum_{k=0}^{0} r^k = r^0 = 1$. We also have $\frac{r^{n+1}-1}{r-1} = \frac{r-1}{r-1} = 1$.

Inductive Step: Now suppose that $\sum_{k=0}^{m} r^k = \frac{r^{m+1}-1}{r-1}$ is true for some $m \geq 0$. Then, we see that

$$\sum_{k=0}^{m+1} r^k = \left(\sum_{k=0}^{m} r^k\right) + r^{m+1} \quad \text{by the recursive definition of a sum.}$$

$$= \frac{r^{m+1}-1}{r-1} + r^{m+1} \quad \text{by induction hypothesis,}$$

$$= \frac{r^{m+1}-1}{r-1} + \frac{r^{m+2}-r^{m+1}}{r-1}$$

$$= \frac{r^{m+1}-1+r^{m+2}-r^{m+1}}{r-1}$$

$$= \frac{r^{m+2}-1}{r-1}.$$

□

Example 16.4. *For every integer* $n \geq 2$, $2^n > n + 1$.

Proof. Basis: When $n = 2$, the inequality to check is $2^2 > 2 + 1$, and that is correct.

Inductive Step: Now suppose that $2^n > n + 1$ for some integer $n \geq 2$. Then $2^{n+1} = 2 \cdot 2^n > 2(n+1) = 2n + 2 > n + 2$, as we needed to show. □

Example 16.5. *Of historical interest is the fact that one can show that using only* 5¢ *stamps and* 9¢ *stamps, any postage amount* 32¢ *or greater can be formed.*

To re-phrase: Any integer $n \geq 32$ *is a linear combination of* 5 *and* 9 *with natural number coefficients. That is: If* n *is an integer and* $n \geq 32$, *then* $n = 5k + 9l$ *for some* $k, l \in \mathbb{N}$.

Proof. Basis: $32 = 1 \cdot 5 + 3 \cdot 9$ so the base case is true.

Inductive Step: Now suppose we can write $n = 5k + 9l$ for some integer $n \geq 32$, where $k, l \in \mathbb{N}$. We need to show we can write $(n+1)$ as a natural number linear combination of 5 and 9. Since $n \geq 32$ we must use either have (1) $k \geq 7$, or (2) $l \geq 1$. If not, then $n \leq 5 \cdot 6 + 0 \cdot 9 = 30 \rightarrow\leftarrow$.

case 1: If $k \geq 7$ we can write $n + 1 = (k-7) \cdot 5 + (l+4) \cdot 9$.

case 2: If $l \geq 1$ we can write $n + 1 = (k+2) \cdot 5 + (l-1) \cdot 9$.

So, in either case, if we can write n as a linear combination of 5 and 9 with natural number coefficients, then we can also write $n + 1$ in such a fashion. □

Example 16.6. *Let's now look at an example of an induction proof with a geometric flavor. Suppose we have a* 4×5 *chess board:*

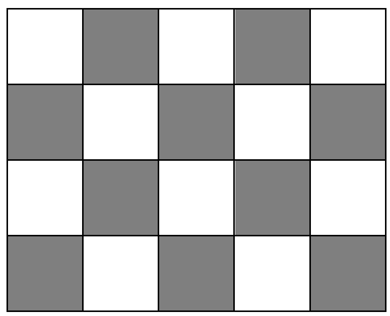


Figure 16.1: 4×5 chessboard

and a supply of 1×2 *dominoes:* 

Each domino covers exactly two squares on the board. A ***perfect cover*** *of the board consists of a placement of dominoes on the board so that each domino covers two squares on the board (dominoes can be either vertically or horizontally orientated), no dominoes overlap, no dominoes extend beyond the edge of the board, and all the squares on the board are covered by a domino. It's easy to see that the* 4×5 *board above has a perfect cover. More generally, it is not hard to prove:*

Theorem 16.7. *An* $m \times n$ *board has a perfect cover with* 1×2 *dominoes if and only if at least one of* m *and* n *is even.*

There is a second version of mathematical induction. Anything that can be proved with this second version can be proved with the method described above, and vice versa, but this second version is often easier to use. The change occurs in the induction assumption made in the inductive step of the proof. The inductive step of the method described above ($p(n) \to p(n+1)$ for all $n \geq k$) is replaced with $[p(k) \wedge p(k+1) \wedge \cdots \wedge p(n)] \to p(n+1)$ for all $n > k$. The effect is that we now have a lot more hypotheses to help us derive $p(n+1)$. In more detail, the second form of mathematical induction is described in the following theorem.

Theorem 16.8 (Second Principle of Mathematical Induction)**.**
For integers k and n, if

(1) $p(k)$ is true, and

(2) $[p(k) \wedge p(k+1) \wedge \ldots \wedge p(n)] \to p(n+1)$ for an arbitrary $n \geq k$,

then $p(n)$ is true for all $n \geq k$.

This principle is shown to be valid in the same way the first form of induction was justified. The utility lies in dealing with cases where we want to use inductive reasoning, but cannot deduce the $(n+1)$st case from the nth case directly. Let's do a few examples of proofs using this second form of induction. One more comment before doing the examples. In many induction proofs, it is convenient to check several initial cases in the basis step to avoid having to include special cases in the inductive step. The examples below illustrate this idea.

Example 16.9. *Show that any integer $n \geq 32$ can be written in the form $n = 5 \cdot k + 9 \cdot l$ for some $k, l \in \mathbb{N}$.*

Proof.
Basis: We can certainly write

$$32 = (1)5 + (3)9$$
$$33 = (3)5 + (2)9$$
$$34 = (5)5 + (1)9$$
$$35 = (7)5 + (0)9$$
$$36 = (0)5 + (4)9$$

Inductive Step: Suppose for some integer m with $m \geq 36$ we can write $j = 5k + 9l$, where $k, l \in \mathbb{N}$ for all integers $32 \leq j \leq m$. Then since $32 \leq m-4$, by inductive hypothesis we can write $m-4 = 5k+9l$ for some natural numbers k, l. Thus $m+1 = 5(k+1) + 9l$, where $k+1, l \in \mathbb{N}$. □

In that example, the basis step was a little messier than our first solution to the problem, but to make up for that, the inductive step required much less cleverness.

Example 16.10. *Induction can be used to verify a guessed closed from formula for a recursively defined sequence. Consider the sequence defined recursively by the initial conditions* $a_0 = 2$, $a_1 = 5$ *and the recursive rule, for* $n \geq 2$, $a_n = 5a_{n-1} - 6a_{n-2}$. *The first few terms of this sequence are* $2, 5, 13, 35, 97, \cdots$. *A little experimentation leads to the guess* $a_n = 2^n + 3^n$. *Let's verify that guess using induction. For the basis of the induction we check our guess gives the correct value of* a_n *for* $n = 0$ *and* $n = 1$. *That's easy. For the inductive step, let's suppose our guess is correct up to* n *where* $n \geq 2$. *Then, we have*

$$\begin{aligned} a_{n+1} &= 5a_n - 6a_{n-1} \\ &= 5(2^n + 3^n) - 6(2^{n-1} + 3^{n-1}) \\ &= (5 \cdot 2 - 6)2^{n-1} - (6 - 5 \cdot 3)3^{n-1} \\ &= 4 \cdot 2^{n-1} - (-9) \cdot 3^{n-1} \\ &= 2^{n+1} + 3^{n+1} \text{ as we needed to show.} \end{aligned}$$

Example 16.11. *In the game of* ***Nim****, two players are presented with a pile of matches. The players take turns removing one, two, or three matches at a time. The player forced to take the last match is the loser. For example, if the pile initially contains* 8 *matches, then first player can, with correct play, be sure to win. Here's how: player 1: take* 3 *matches leaving* 5*; player 2's options will leave* 4, 3, *or* 2 *matches, and so player 1 can reduce the pile to* 1 *match on her turn, thus winning the game. Notice that if player 1 takes only* 1 *or* 2 *matches on her first turn, she is bound to lose to good play since player 2 can then reduce the pile to* 5 *matches.*

Let's prove that if the number of matches in the pile is 1 *more than a multiple of* 4*, the second player can force a win; otherwise, the first player can force a win.*

Proof. For the basis, we note that obviously the second player wins if there is 1 match in the pile, and for 2, 3, or 4 matches the first player wins by taking 1, 2, or 3 matches in each case, leaving 1 match.

For the inductive step, suppose the statement we are to prove is correct for the number of matches anywhere from 1 up to k for some $k \geq 4$. Now consider a pile of $k + 1$ matches.

case I: If $k + 1$ is 1 more than a multiple of 4, then when player 1 takes her matches, the pile will not contain 1 more than a multiple of 4 matches, and so the next player can force a win by the inductive assumption. So player 2 can force a win.

case 2: If $k + 1$ is not 1 more than a multiple of 4, then player 1 can select matches to make it 1 more than a multiple of 4, and so the next player is bound to lose (with best play) by the inductive assumption. So player 1 can force a win. □

So, to win at Nim, when it is your turn, make sure you leave 1 more than a multiple of 4 matches in the pile (which is easy to do unless your opponent knows the secret as well, in which case you can just count the number of matches in the pile to see who will win, and skip playing the game altogether!).

Exercises

Exercise 16.1. *Use induction to prove that for every integer* $n \geq 1$, *is*

$$1 \cdot 3 + 2 \cdot 4 + 3 \cdot 5 + \cdots + n(n+2) = \frac{n(n+1)(2n+7)}{6}.$$

Exercise 16.2. *Prove that for every integer* $n \geq 1$,

$$1 \cdot 2^1 + 2 \cdot 2^2 + 3 \cdot 2^3 + \ldots + n \cdot 2^n = (n-1)\, 2^{n+1} + 2.$$

Exercise 16.3. *The Fibonacci sequence is defined recursively by* $f_0 = 0$, $f_1 = 1$, *and, for* $n \geq 2$, $f_n = f_{n-1} + f_{n-2}$. *Use induction to prove that for all* $n \geq 0$, $f_0 + f_1 + f_2 + \cdots + f_n = f_{n+2} - 1$.

Exercise 16.4. *Prove by induction: For every integer* $n > 4$, *we have* $2^n > n^2$.

Exercise 16.5. *Prove by induction: For every integer* $n \geq 0$, $11^n - 6$ *is divisible by* 5.

Exercise 16.3. *A pizza is cut into pieces (maybe some pretty oddly shaped) by making some integer* $n \geq 0$ *number of straight-line cuts. Prove: The maximum number of pieces is* $(n^2 + n + 2)/2$.

Exercise 16.7. *A sequence is defined recursively by* $a_0 = 0$, *and, for* $n \geq 1$, $a_n = 5a_{n-1} + 1$. *Use induction to prove the closed form formula for* a_n *is*

$$a_n = \frac{5^n - 1}{4}.$$

Exercise 16.8. *A sequence is defined recursively by* $a_0 = 1, a_1 = 4$, *and for* $n \geq 2$, $a_n = 5a_{n-1} - 6a_{n-2}$. *Use induction to prove that the closed form formula for* a_n *is* $a_n = 2 \cdot 3^n - 2^n, n \geq 0$.

Problems

Problem 16.1. *Prove for every integer* $n \geq 1$,

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

Problem 16.2. *Prove by induction: For* $n \geq 2$,

$$\left(1 - \frac{1}{4}\right)\left(1 - \frac{1}{9}\right) \cdot \ldots \cdot \left(1 - \frac{1}{n^2}\right) = \frac{n+1}{2n}.$$

Problem 16.3. *Show that any integer* $n \geq 8$ *can be written as a linear combination of the integers* 3 *and* 5 *using nonnegative integers as coefficients. That is if* $n \geq 8$, *there exist nonnegative integers* k_n, l_n *so that* $n = 3 \cdot k_n + 5 \cdot l_n$. *Do this twice, using both styles of induction.*

Problem 16.4. *Prove by induction: For every integer* $n \geq 1$,

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

Problem 16.5. *Prove by induction: For every integer* $n \geq 1$, *the number* $n^5 - n$ *is divisible by* 5.

Problem 16.6. *Prove by induction: For the Fibonacci sequence and for all* $n \geq 2$,

$$f_0^2 + f_1^2 + f_3^2 + \cdots + f_n^2 = f_n f_{n+1}.$$

Problem 16.7. *Prove by induction: For the Fibonacci sequence and for all* $n \geq 2$,

$$f_{n-1} f_{n+1} = f_n^2 + (-1)^n.$$

Problem 16.8. *Here is a proof that for* $n \geq 0, 1 + 2 + 2^2 + \cdots + 2^n = 2^{n+1}$.

Proof. Suppose $1 + 2 + 2^2 + \cdots + 2^n = 2^{n+1}$ for some $n \geq 0$. Then

$$\begin{aligned} 1 + 2 + 2^2 + \cdots + 2^n + 2^{n+1} &= 2^{n+1} + 2^{n+1} \qquad \text{using the inductive hypothesis} \\ &= 2 \cdot 2^{n+1} = 2^{n+2} = 2^{(n+1)+1}. \end{aligned}$$

as we needed to show. □

Now, obviously there is something wrong with this proof since, for example, $1 + 2 + 2^2 = 7$, *but we have* $2^{2+1} = 2^3 = 8$. *Where does the proof go bad?*

Chapter 17

Algorithms

An **algorithm** is a recipe to solve a problem. For example, here is an algorithm that solves the problem of finding the distance traveled by a car given the time it has traveled, t, and its average speed, s: multiply t and s.

Over time, the requirements of what exactly constitutes an algorithm have matured. A really precise definition would be filled with all sorts of technical jargon, but the ideas are commonsensible enough that an informal description will suffice for our purposes. So, suppose we have in mind a certain class of problems (such as determine the distance traveled given time traveled and average speed). The properties of an algorithm to solve examples of that class of problems are:

(1) **Input**: The algorithm is provided with data.

(2) **Output**: The algorithm produces a solution.

(3) **Definiteness**: The instructions that make up the algorithm are precisely described. They are not open to interpretation.

(4) **Finiteness**: The output is produced in a finite number of steps.

(5) **Generality**: The algorithm produces correct output for any set of input values.

The algorithm for finding distance traveled given time traveled and average speed obviously meets all five requirements of an algorithm. Notice that, in this example, we have assumed the user of the algorithm understands what it means to multiply two numbers. If we cannot make that assumption, then we would need to add a number of additional steps to the algorithm to solve the problem of multiplying two numbers together. Of course, that would make the algorithm significantly longer. When describing algorithms, we'll assume the user knows the usual algorithms for solving common problems such as addition, subtraction, multiplication, and division of numbers, and knows how to determine if one number is larger than another, and so on.

Just as important as an example of what an algorithm is, is an example of what is not an algorithm. For example, we might describe the method by which most people used to look up a number in a phone book. One would open the book and look to see if the listing you're looking for is on that page or not. If it is you find the number using the fact that the listings are alphabetized and you're done. If the number you're looking for is not on the page, you use the fact that the listings are alphabetized to either flip back several pages, or forward several pages. This page is checked to see if the listing is on it. If it is not we repeat the process. One problem in this case is that this description is not definite. The phrase *flip back several pages* is too vague, it violates the definiteness requirement. Another problem is that someone could flip back and forth between two pages and never find the number, and so violate the finiteness requirement. So this method is not an algorithm.

Continuing the example from above of looking up an item indexed by a sorted list, one algorithm for completing this is to look at the first entry in the list. If it's the item you're looking for you're done. Else move to the next entry. It's either the item you're looking for or you move to the next entry. This is an example of a **linear search algorithm**. It's not too bad for finding an early entry in the list, but awful for later entries in the list.

Another algorithm to complete that task of looking up an item indexed by a sorted list is the **binary search algorithm**. We first consider the middle entry. If it's the item we're looking for, we're done. Else we know the item is in the first half of the list, or the last half of the list since the entries are ordered. We then pick the middle entry of the appropriate half, and repeat the halving process on that half, until eventually the item is located. There are a few details to fix up to make this a genuine algorithm. For example, what is the middle entry if there are an even number of items listed? Also, what happens if the item we are looking for isn't in the list? But it is clear with a little effort we can add a few lines to the instructions to make this process into an algorithm.

It is traditional to present algorithms in a *pseudocode* form similar to a program for a computer. For instance, the linear search name lookup algorithm given above could be written in pseudocode form, as shown in Algorithm 1 on page 121.

Algorithm 1 Linear search (for loop)

Input: *(name,phonelist)*
Output: phonelist(namespot)
Output: *phonelist(namespot)* = phone number of *name* in *phonelist*

```
for namespot ∈ {1, 2, . . . , length(phonelist)} do
    if list(namespot) is name then
        output phonelist(namespot)
        stop
    end if
end for
output name not found
stop
```

Example 17.1. *Here is an algorithm for determining* $\lfloor m/n \rfloor$ *for positive integers* m, n.

Algorithm 2 Calculate $\lfloor m/n \rfloor$

Input: positive integers m and n
Output: integer value of $\lfloor m/n \rfloor$

```
1: k ← 0                      ▷ k will eventually hold our answer
2: while m ≥ 0 do
3:     m ← m − n              ▷ We're doing division by repeated subtraction
4:     k ← k + 1              ▷ k counts the number of subtractions
5: end while
6: output k − 1               ▷ We counted one too many subtractions!(How?)
```

Here are the sequence of steps this algorithm would carry out with input $m = 23$ *and* $n = 7$*:*
[initial status. $m = 23$*,* $n = 7$*,* $k =$ *(undefined)]*

instr 1: Set k to be 0 [*status:* $m = 23$, $n = 7$, $k = 0$]

instr 2: is $m \geq 0$*?* Yes, $(23 \geq 0)$ is true. Do next instruction (i.e. instr 3).

instr 3: m reset to be $m - n = 23 - 7 = 16$. [*status:* $m = 16$, $n = 7$, $k = 0$]

instr 4: k reset to be $k + 1 = 0 + 1 = 1$. [*status:* $m = 16$, $n = 7$, $k = 1$]

instr 5: jump back to the matching **while** (i.e. instr 2).

instr 2: is $m \geq 0$*?* Yes, $(16 \geq 0)$ is true. Do next instruction.

instr 3: m reset to be $m - n = 16 - 7 = 9$. [*status:* $m = 9$, $n = 7$, $k = 1$]

instr 4: k reset to be $k + 1 = 1 + 1 = 2$. [*status:* $m = 9$, $n = 7$, $k = 2$]

instr 5: jump back to the matching **while**.

instr 2: is $m \geq 0$? Yes, $(9 \geq 0)$ is true. Do next instruction.

instr 3: m reset to be $m - n = 9 - 7 = 2$. [*status:* $m = 2, n = 7, k = 2$]

instr 4: k reset to be $k + 1 = 2 + 1 = 3$. [*status:* $m = 2, n = 7, k = 3$]

instr 5: jump back to the matching **while**.

instr 2: is $m \geq 0$? Yes, $(2 \geq 0)$ is true. Do next instruction.

instr 3: m reset to be $m - n = 2 - 7 = -5$. [*status:* $m = -5, n = 7, k = 2$]

instr 4: k reset to be $k + 1 = 3 + 1 = 4$. [*status:* $m = -5, n = 7, k = 4$]

instr 5: jump back to the matching **while**.

instr 2: is $m \geq 0$? **No**, $(-5 \geq 0)$ is false. Jump to instr after **end while**.

instr 6: ***output*** *value of* $k - 1$ *(i.e. 3).* [*status:* $m = -5, n = 7, k = 4$]

stop! *(This is what happens when there are no more instructions to execute.)*

Here is a second algorithm for the same problem.

Algorithm 3 Calculate $\lfloor m/n \rfloor$ (again)

Input: positive integers m and n
Output: integer value of $\lfloor m/n \rfloor$
  divide m by n to one place beyond the decimal, call the result r.
  output the digits of r preceding the decimal point.

So, again, given input $m = 23$ *and* $n = 7$ *we go through the steps.*

instr 1: $r = 3.2$

instr 2: Output 3

stop!

Problems

Problem 17.1. *Consider the following algorithm: The input will be two integers, $m \geq 0$, and $n \geq 1$.*

Input: positive integers $m \geq 0$ and $n \geq 1$
Output: (to be determined)
  $s \leftarrow 0$
  while $m \neq 0$ **do**
    $s \leftarrow n + s$
    $m \leftarrow m - 1$
  end while
  output s

Describe in words what this algorithm does. In other words, what problem does this algorithm solve?

Problem 17.2. *Consider the following algorithm: The input will be any integer n, greater than 1.*

Input: integer $n > 1$
Output: (to be determined)
  $t \leftarrow 0$
  while n is even **do**
    $t \leftarrow t + 1$
    $n \leftarrow n/2$
  end while
  output t

(a) *List the steps the algorithm follows for the input $n = 12$.*

(b) *Describe in words what this algorithm does. In other words, what problem does this algorithm solve?*

Problem 17.3. *Design an algorithm that takes any positive integer n and returns half of n if it is even and half of $n + 1$ if n is odd. (Such an algorithm is needed quite often in computer science.)*

Problem 17.4. *Consider the following algorithm. The input will be a function f together with its finite domain, $D = \{d_1, d_2, \ldots, d_n\}$.*

Input: function f with domain $D = \{d_1, d_2, \ldots, d_n\}$
Output: (to be determined)

```
i ← 1
while i < n do
    j ← i + 1
    while j < n do
        if f(d_j) = f(d_i) then
            output NO
            stop
        end if
        j ← j + 1
    end while
    i ← i + 1
end while
output YES
```

(a) *List the steps the algorithm follows for the input f: $\{a, b, c, d, e\} \to \{+, *, \&, \$, \#, @\}$ given by $f(a) = *$, $f(b) = \$$, $f(c) = +$, $f(d) = \$$, and $f(e) = @$.*

(b) *Describe in words what this algorithm does. In other words, what problem does this algorithm solve?*

Problem 17.5. *Design an algorithm that will convert the ordered triple (a, b, c) to the ordered triple (b, c, a). For example, if the input is $(7, \mathrm{X}, *)$, the output will be $(\mathrm{X}, *, 7)$.*

Problem 17.6. *Design an algorithm whose input is a finite list of positive integers and whose output is the sum of the even integers in the list. If there are no even integers in the list, the output should be 0.*

Problem 17.7. *A **palindrome** is a string of letters that reads the same in each direction. For example, refer and redder are palindromes of length five and six respectively. Design an algorithm that will take a string as input and output yes if the string is a palindrome, and no if it is not.*

Chapter 18

Algorithm Efficiency

There are many different algorithms for solving any particular class of problems. In the last chapter, we considered two algorithms for solving the problem of looking up a phone number given a person's name.

Algorithm L: Look at the first entry in the book. If it's the number you're looking for you're done. Else move to the next entry. It's either the number you're looking for or you move to the next entry, and so on. (The **linear search algorithm**)

Algorithm B: The second algorithm took advantage of the arrangement of a phone book in alphabetical order. We open the phone book to the middle entry. If it's the number we're looking for, we're done. Otherwise we know the number is listed in the first half of the book, or the last half of the book. We then pick the middle entry of the appropriate half, and repeat the process. After a number of repetitions, we will either be at the name we want, or learn the name isn't in the book. (The **binary search algorithm**)

The question arises, which algorithm is *better*? The question is pretty vague. Let's assume that *better* means *uses fewer steps*. Now if there are only one or two names in the phone book, it doesn't matter which algorithm we use, the look-up always takes one or two steps. But what if the phone book contains 10000 names? In this case, it is hard to say which algorithm is better: looking up Adam Aaronson will likely only take one step by the linear search algorithm, but binary search will take 14 steps or so. But for Zebulon Zyzniewski, the linear search will take 10000 steps, while the binary search will again take about 14 steps.

There are two lessons to be learned from those last examples:

(1) Small cases of the problem can be misleading when judging the quality of an algorithm, and

(2) It's unlikely that one algorithm will always be more efficient than another.

The common approach to compare the efficiency of two algorithms takes those two lessons into account by agreeing to the following protocol:

(1) only compare the algorithms when the size, n, of the problem it is applied to is huge. In the phone book example, don't worry about phone books of 100 names or even 10000 names. Worry instead about phone books with n names where n gets arbitrarily large.

(2) to compare two algorithms, first, for each algorithm, find the maximum number of steps ever needed when applied to a problem of size n. For a phone book of size n the linear search algorithm will require n steps in the worst possible case of the name not being in the book. On the other hand, the halving process of the binary search algorithm means that it will never take more than about $\log_2 n$ steps to locate a name (or discover the name is missing) in the phone book. This information is expressed compactly by saying the linear search algorithm has **worst case scenario** efficiency $w_L(n) = n$ while the binary search algorithm has worst case scenario efficiency $w_B(n) = \log_2 n$.

(3) we declare that algorithm #1 is **more efficient** than algorithm #2 provided, for all problems of huge sizes n, $w_1(n) < w_2(n)$, where w_1 and w_2 are the worst case scenario efficiencies for each algorithm.

Notice that for huge n, $w_B(n) < w_L(n)$. In fact, there is no real contest. For example, when $n = 1048576 = 2^{20}$, we get $w_B(n) = 20$ while $w_L(n) = 1048576$, and things only get better for $w_B(n)$ as n gets larger.

In summary, to compare two algorithms designed to solve the same class of problems we:

(1) Determine a number n that indicates the size of the a problem. For example, if the algorithm manipulates a list of numbers, n could be the length of the list. If the algorithm is designed to raise a number to a power, the size could be the power n.

(2) Decide what will be called a *step* when applying the algorithms. In the phone book example, we took a step to mean a comparison. When raising a number to a power, a step might consist of performing a multiplication. A step is usually taken to be the most time consuming action in the algorithm, and other actions are ignored. Also, when determining the function, w don't get hung up worrying about minuscule details. Don't spend time trying to determine if $w(n) = 2n + 7$ or $w(n) = 2n + 67$. For huge values of n, the +7 and +67 become unimportant. In such a case, $w(n) = 2n$ has all the interesting information. Don't sweat the small stuff.

(3) Determine the worst case scenario functions for the two algorithms, and compare them. The smaller of the two (assuming they are not essentially the same) is declared the more efficient algorithm.

Example 18.1. *Let's do a worst case scenario computation for the following algorithm designed to determine the largest number in a list of* n *numbers. It would be natural to use the number of items*

Algorithm 4 Maximum list value

Input: a list of n numbers $a_1, a_2, \ldots, a_n$
Output: maximum$(a_1, a_2, \ldots, a_n)$

1: max $\leftarrow a_1$
2: $k \leftarrow 2$
3: **while** $k \leq n$ **do**
4:     **if** max $< a_k$ **then**
5:         max $\leftarrow a_k$
6:         $k \leftarrow k + 1$
7:     **end if**
8: **end while**
9: **output** max

in the list, n*, to represent the size of a problem. And let's use the comparisons as steps. We are going to make two comparisons each for each of the items in the list in every case (every case is a worst case for this algorithm!). So we give this algorithm an efficiency* $w(n) = 2n$*. Notice that we actually only need comparisons for the last* $n - 1$ *items in the list, and the exact number of times the comparisons in instructions (3) and (4) are carried out might take a few minutes to figure out. But it's clear that both are carried out about* n *times, and since we are only interested in huge* n*'s, being off by a few (or a few billion) isn't really going to matter at all.*

Problems

Problem 18.1. *For the algorithm presented in problem 17.2 from the last chapter:*

(a) *Select a value to represent the size of an instance of the problem the algorithm is designed to solve.*

(b) *Decide what will constitute a step in the algorithm.*

(c) *Determine the worst case scenario function w(n).*

Problem 18.2. *Repeat problem 18.1 for the following algorithm:*

Input: Sets of reals: $\{x_1, x_2, \ldots, x_n\}$ and $\{y_1, y_2, \ldots, y_n\}$ of size n
Output: (to be determined)

```
S ← 0
i ← 1
while i ≤ n do
    S ← S + x_i · y_i
    i ← i + 1
end while
output S
```

Chapter 19

The Growth of Functions

Now that we have an idea of how to determine the efficiency of an algorithm by computing its worst case scenario function, *w(n)*, we need to be able to decide when one algorithm is better than another. For example, suppose we have two algorithms to solve a certain problem, the first with $w_1(n) = 10000n^2$, and the second with $w_2(n) = 2^n$ Which algorithm would be the better choice to implement based on these functions? To find out, let's assume that our computer can carry out one billion steps per second, and estimate how long each algorithm will take to solve a worst case problem for various values of *n*.

w(n)	$n = 10$	$n = 20$	$n = 50$	$n = 100$
$10000n^2$	.001 sec	.004 sec	.025 sec	.1 sec
2^n	.000001 sec	.001 sec	4.2 months	4×10^{11} centuries

Table 19.1: Problem size vs. CPU time used

So, it looks like the selection of the algorithm depends on the size of the problems we expect to run into. Up to size 20 or so, it doesn't look like the choice makes a lot of difference, but for larger values of *n*, the $10000n^2$ algorithm is the only practical choice.

It is worth noting that the values of the efficiency functions for small values of *n* can be deceiving. It is also worth noting that, from a practical point of view, simply designing an algorithm to solve a problem without analyzing its efficiency can be a pointless exercise.

There are a few types of efficiency functions that crop up often in the analysis of algorithms. In order of decreasing efficiency for large n they are $\log_2 n$, $\sqrt{n}$, n, n^2, n^3, 2^n, $n!$.

Assuming one billion steps per second, here is how these efficiency functions compare for various choices of n.

$w(n)$	$n = 10$	$n = 20$	$n = 50$	$n = 100$
$log_2 n$	.000000003 sec	.000000004 sec	.000000005 sec	.000000006 sec
$\sqrt{n}$	.000000003 sec	.000000004 sec	.000000006 sec	.000000008 sec
n	.00000001 sec	.00000002 sec	.00000004 sec	.00000006 sec
n^2	.0000001 sec	.0000004 sec	.0000016 sec	.0000036 sec
n^3	.000001 sec	.000008 sec	.000064 sec	.00022 sec
2^n	.000001 sec	.001 sec	18.3 minutes	36.5 years
$n!$	.0036 sec	77 years	2.6×10^{29} centuries	2.6×10^{29} centuries

Table 19.2: Common efficiency functions for small values of n

Even though the values in the first five rows of the table look reasonably close together, that is a false impression fostered by the small values of n. For example, when $n = 1000000$, those five entries would be as in table 19.3.

And, for even larger values of n, the $\sqrt{n}$ algorithm will require billions more years than the $\log_2 n$ algorithm.

There is a traditional method of estimating the efficiency of an algorithm. As in the examples above, one part of the plan is to ignore tiny contributions to the efficiency function. In other words, we won't write expressions such as $w(n) = n^2 + 3$, since the term 3 is insignificant for the large values of n we are interested in. As far as behavior for large values of n is concerned, the functions n^2

$w(n)$	$n = 1000000$
$log_2 n$	.00000002 sec
$\sqrt{n}$	.000001 sec
n	.001 sec
n^2	17 minutes
n^3	31.7 years

Table 19.3: Efficiency functions where $n = 1000000$

and $n^2 + 3$ are indistinguishable. A second part of the plan is to not distinguish between functions if one is always say 10 times the other. In other words, as far as analyzing efficiency, the functions n^2 and $10n^2$ are indistinguishable. And there is nothing special about 10 in those remarks. These ideas lead us to the idea of the order of growth with respect to n, $\mathcal{O}(\boldsymbol{g}(\boldsymbol{n}))$, in the next definition.

Definition 19.1. The function $\boldsymbol{w(n)}$ **is** $\mathcal{O}(\boldsymbol{g}(\boldsymbol{n}))$ provided there is a number $k > 0$ such that $w(n) \leq k{\cdot}g(n)$ for all n (or at least for all large values of n). The symbol $\mathcal{O}(g(n))$ is read in English as *big-oh of g(n)*.

As an example, $n^3 + 2n^2 + 10n + 4 \leq (1 + 2 + 10 + 4)\, n^3 = 17n^3$ is $O(n^3)$. So if we have an algorithm with efficiency function $w(n) = n^3 + 2n^2 + 10n + 4$, we can suppress all the unimportant details, and simply say the efficiency is $\mathcal{O}(n^3)$. In this example, it is also true that $w(n)$ is $\mathcal{O}(n^4)$, but that is less precise information. On the other hand, saying $w(n)$ is $\mathcal{O}(n^2)$ is certainly false. To indicate that we have the best big-oh estimate allowed by our analysis, we would say $w(n)$ is *at best* $\mathcal{O}(n^3)$.

$\mathcal{O}(g(n))$ actually represents the set of functions dominated by $g(n)$. So, it would be proper to write $w(n) = n^3 + 2n^2 + 10n + 4 \in O(n^3)$. Moreover, we could write $\mathcal{O}(n^2) \subset \mathcal{O}(n^3)$ since the functions dominated by n^2 are among those dominated by n^3. Loosely speaking, finding the $\mathcal{O}$ estimate for a function selects the most influential, or dominant, term (for large values of the variable) in the function, and suppresses any constant factor for that term.

In each of the following examples, we find a big-oh estimate for the given expression.

Example 19.2. *We have that* $n^4 - 3n^3 + 2n^2 - 6n + 14$ *is* $\mathcal{O}(n^4)$, *since for large* n *the first term dominates the others.*

Example 19.3. *For large* n, *we have the inequalities:*

$$\begin{aligned}&(n^3 \log_2 n + n^2 - 3)(\, n^2 + 2n + 8),\\ &\qquad \leq (1 + 1 + 3)\, n^3(\log_2 n)(1 + 2 + 8)\, n^2,\\ &\qquad \leq 55n^5 \log_2 n.\end{aligned}$$

Hence, $(n^3 \log_2 n + n^2 - 3)(\, n^2 + 2n + 8)$ *is* $\mathcal{O}(n^5 \log_2 n)$. *Alternatively, we have that* $n^3 \log_2 n$ *and* n^2 *dominate their respective factors. Thus, again, the product is* $\mathcal{O}(n^5 \log_2 n)$

The lesson learned is that dominant factors may be multiplied.

Example 19.4. *We see that* $n^5 + 3(2^n) - 14n^{22} + 13 \leq (1 + 3 + 14 + 13)2^n = 31(2^n)$. *Hence, the expression is* $\mathcal{O}(2^n)$. *Or, since* $3 \cdot 2^n$ *dominates all the other terms for large* n, *we see that the expression is* $\mathcal{O}(3 \cdot 2^n)$. *That is, it is of order* $\mathcal{O}(2^n)$ *since the constant factor is really irrelevant.*

Problems

Problem 19.1. *You have been hired for a certain job that can be completed in less than two months and offered two modes of payment. Method 1: You get $1,000,000,000 a day for as long as the job takes. Method 2: You get $1 the first day, $2 the second day, $4 the third day, $8 the fourth day, and so on, your payment doubling each day, for as long as the job lasts. Which method of payment do you choose?*

Problem 19.2. *Suppose an algorithm has efficiency function* $w(n) = n\log_2 n$. *Compute the worst-case time required for the algorithm to solve problems of sizes* n = 10, 20, 40, 60 *assuming the operations are carried out at the rate of one billion per second. Where does this function fit in the table on the second page of this chapter?*

Problem 19.3. *Repeat problem 19.2 for* $w(n) = n^n$.

Problem 19.4. *Explain why* $3n^3 + 400n^2 + 2\sqrt{n}$ *is at least* $\mathcal{O}(n^2)$.

Problem 19.5. *Explain why* $10n^2 + 4n + 2\sqrt{n}$ *is not* $\mathcal{O}(1000n)$.

Problem 19.6. *Find the best possible big-oh estimate for* $\sqrt{5n} + \log_2 10n + 1$.

Problem 19.7. *Find the best possible big-oh estimate for* $2n^2 + \frac{3}{n}$.

Problem 19.8. *Find the best possible big-oh estimate for* $\frac{2n^2+2n+1}{2n+1}$. *Hint: Use the idea of dominant terms first.*

Chapter 20

The Integers

Number theory is concerned with the integers and their properties. In this chapter the rules of the arithmetic of integers are reviewed. The surprising fact is that all the dozens of rules and tricks you know for working with integers (and for doing algebra, which is just arithmetic with symbols) are consequences of just a few basic facts.

The set of **integers**, $\{\cdots, -2, -1, 0, 1, 2, \cdots\}$, is denoted by the symbol $\mathbb{Z}$. The two familiar arithmetic operations for the integers, addition and multiplication, obey several basic rules. First, notice that addition and multiplication are **binary operations**. In other words, these two operations combine a pair of integers to produce a value. It is not possible to add (or multiply) three numbers at a time. We can figure out the sum of three numbers, but it takes two steps: we select two of the numbers, and add them up, and then add the third to the preliminary total. Never are more than two numbers added together at any time. A list of the seven fundamental facts about addition and multiplication of integers follows.

(1) The integers are **closed** with respect to addition and multiplication.

That means that when two integers are added or multiplied, the result is another integer. In symbols, we have

$$\forall a, b \in \mathbb{Z},\ ab \in \mathbb{Z} \text{ and } a + b \in \mathbb{Z}.$$

(2) Addition and multiplication of integers are **commutative** operations.

That means that the *order* in which the two numbers are combined has no effect on the final total. Symbolically, we have

$$\forall a, b \in \mathbb{Z},\ a + b = b + a \text{ and } ab = ba.$$

(3) Addition and multiplication of integers are **associative** operations. In other words, when we compute the sum (or product) of three integers, it does not matter whether we combine the first two and then add the third to the total, or add the first to the total of the last two. The final total will be the same in either case. Expressed in symbols, we have

$$\forall a, b, c \in \mathbb{Z},\ a(bc) = (ab)c \text{ and } a + (b + c) = (a + b) + c.$$

(4) There is an **additive identity** denoted by 0. It has the property that when it is added to any number the result is that number right back again. In symbols, we see that

$$0 + a = a = a + 0 \text{ for all } a \in \mathbb{Z}.$$

(5) Every integer has an **additive inverse**: $\forall n \in \mathbb{Z}, \exists m \in \mathbb{Z}$ so that $n + m = 0 = m + n$. As usual, m is denoted by $-n$. So, we write $n + (-n) = (-n) + n = 0$.

(6) 1 is a **multiplicative identity**. That is, we have $1a = a = a1$ for all $a \in \mathbb{Z}$.

And finally, there is a rule which establishes a connection between the operations of addition and multiplication.

(7) Multiplication **distributes** over addition. Again, we symbolically write

$$\forall a, b, c \in \mathbb{Z},\ a(b + c) = ab + ac.$$

The preceding facts tell all there is to know about arithmetic. Every other fact can be proved from these. For example, here is a proof of the cancellation law for addition using the facts listed above.

Theorem 20.1 (Integer cancellation law)**.** *For integers a, b, c, if $a + c = b + c$ then $a = b$.*

Proof. Suppose $a + c = b + c$. Add $-c$ to both sides of that equation (applying fact 5 above) to get $(a + c) + (-c) = (b + c) + (-c)$. Using the associative rule, that equation can be rewritten as $a + (c + (-c)) = b + (c + (-c))$, and that becomes $a + 0 = b + 0$. By property 4 above, that means $a = b$. □

Theorem 20.2. *For any integer a, $a0 = 0$.*

Proof. Here are the steps in the proof. You supply the justifications for the steps.

$$\begin{aligned} a0 &= a(0+0) \\ a0 &= a0 + a0 \\ a0 + (-(a0)) &= (a0 + a0) + (-(a0)) \\ a0 + (-(a0)) &= a0 + (a0 + (-(a0))) \\ 0 &= a0 + 0 \\ 0 &= a0 \end{aligned}$$

□

Your justification for each step should be stated as using one, or more, of the fundamental facts as applied to the specific circumstance in each line.

The integers also have an order relation, *a is less than or equal to b*: $a \le b$. This relation satisfies three fundamental order properties: $\le$ is a reflexive, antisymmetric, and transitive relation on $\mathbb{Z}$.

The notation $b \ge a$ means the same as $a \le b$. Also $a < b$ (and $b > a$) are shorthand ways to say $a \le b$ and $a \ne b$.

The **trichotomy law** holds: for $a \in \mathbb{Z}$ exactly one of $a > 0, a = 0$, or $a < 0$ is true.

The ordering of the integers is related to the arithmetic by several rules:

(1) If $a < b$, then $a + c < b + c$ for all $c \in \mathbb{Z}$.

(2) If $a < b$ and $c > 0$, then $ac < bc$.

(3) If $a < b$ and $c < 0$, then $bc < ac$.

And, finally, the rule that justifies proofs by induction:

The Well Ordering Principle for $\mathbb{Z}$: The set of positive integers is **well-ordered**: every nonempty subset of positive integers has a least element.

Exercises

Exercise 20.1. *Prove that if* $a > 0$ *and* $b > 0$*, then* $ab > 0$*.*

Exercise 20.2. *Prove that if* $ab = 0$*, then* $a = 0$ *or* $b = 0$*. Hint: Try an indirect proof with four cases. Case 1: Show that if* $a > 0$ *and* $b > 0$*, then* $ab \neq 0$*. Case 2: Show that if* $a > 0$ *and* $b < 0$*, then* $ab \neq 0$*. There are two more similar cases. (This fact is called the zero property.)*

Exercise 20.3. *Prove the cancellation law for multiplication: For integers* a, b, c*, with* $c \neq 0$*, if* $ac = bc$*, then* $a = b$*. (Hint: Use exercise 20.2)*

Problems

Problem 20.1. *Prove that if* $a > 0$ *and* $b < 0$*, then* $ab < 0$*.*

Problem 20.2. *Prove that if* n *is an integer, then* $n^2 \geq 0$*.*

Problem 20.3. *Prove that if* $m^2 = n^2$*, then* $m = n$ *or* $m = -n$*.*
(Hint from algebra: $a^2 - b^2 = (a + b)(a - b)$*.)*

Chapter 21

The divides Relation and Primes

Given integers a and b we say that a **divides** b and write $a|b$ provided there is an integer c with $b = ac$. So a divides b means a divides into b **evenly**. When a divides b we also say that a is a **factor** of b, or that a is a **divisor** of b, or that b is a **multiple** of a. For example $3|12$ since $12 = 3 \cdot 4$. Keep in mind that *divides* is a relation. When you see $a|b$ you should think *is that true or false*. Don't write things like $3|12 = 4$! If a does not divide b, write $a \nmid b$. For example, it is true that $3 \nmid 13$ since 3 does not divide into 13 **evenly**.

Here is a list of a few simple facts about the divisibility relation.

Theorem 21.1. *For $a, b, c \in \mathbb{Z}$ we have*

(1) $a|0$

(2) $\pm 1|a$

(3) If $a|b$, then $-a|b$

(4) If $a|b$ and $b|c$, then $a|c$. So $a|b$ is a transitive relation on $\mathbb{Z}$.

(5) $a| - a$

(6) If $a|b$ and $b \neq 0$, then $0 < |a| \leq |b|$

(7) If $a|1$, then $a = \pm 1$

(8) *If* $a|b$ *and* $b|a$, *then* $a = \pm b$

(9) *If* $a|b$ *and* $a|c$, *then* $a|(mb + nc)$ *for all* $m, n \in \mathbb{Z}$

(10) *If* $a|b$, *then* $a|bc$ *for all* $c \in \mathbb{Z}$

Here are the proofs of a few of these facts.

(1) *Proof. For any integer* a, $a0 = 0$, *so* $a|0$. □

(4) *Proof. Suppose* $a|b$ *and* $b|c$. *That means there are integers* s, t *so that* $as = b$ *and* $bt = c$. *Substituting* as *for* b *in the second equation gives* $(as)t = c$, *which is the same as* $a(st) = c$. *That shows* $a|c$. □

(9) *Proof. Suppose* $a|b$ *and* $a|c$. *That means there are integers* s, t *such that* $as = b$ *and* $at = c$. *Multiply the first equation by* m *and the second by* n *to get* $a(sm) = mb$ *and* $a(tn) = nc$. *Now add those two equations:* $a(sm) + a(tn) = mb + nc$. *Factoring out the* a *on the left shows* $a(sm + tn) = mb + nc$, *and so we see* $a|(mb + nc)$. □

The prime integers play a central role in number theory. A positive integer larger than 1 is said to be **prime** if its only positive divisors are 1 and itself.

The first few primes are $2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79$.

A positive integer larger than 1 which is not prime is **composite**. So a composite number n has a positive divisor a which is neither 1 nor n. By part (6) of the theorem above, $1 < a < n$.

So, to check if an integer n is a prime, we can trial divide it in turn by $2, 3, 4, 5, \cdots, n - 1$, and if we find one of these that divides n, we can stop, concluding that n is not a prime. On the other hand, if we find that none of those divide n, then we can conclude n is a prime. This algorithm for checking a number for primality can be made more efficient. For example, there is really no need to test to see if 4 divides n if we have already determined that 2 does not divide n. And the same reasoning shows that to test n for primality we need only check in to see if n is divisible by any of $2, 3, 5, 7, 11, 13$ and so on up to the largest prime less than n. For example, to test 15 for primality, we would trial divide by the six values $2, 3, 5, 7, 11, 13$. But even this improved algorithm can be made more efficient by the following theorem.

Theorem 21.2. *Every composite number n has a divisor a, with $2 \le a \le \sqrt{n}$.*

Proof. Suppose n is a composite integer. That means $n = ab$ where $1 < a, b < n$. Not both a and b are greater than $\sqrt{n}$, for if so $n = ab > \sqrt{n}\sqrt{n} = \sqrt{n^2} = n$, and that is a contradiction. □

So, if we haven't found a divisor of n by the time we reach $\sqrt{n}$, then n must be prime.

We can be a little more informative, as the next theorem shows.

Theorem 21.3. *Every integer $n > 1$ is divisible by a prime.*

Proof. Let $n > 1$ be given. The set, D, of all integers greater than 1 that divide n is nonempty since n itself is certainly in that set. Let m be the smallest integer in that set. Then m must be a prime since if k is an integer with $1 < k < m$ and $k|m$, then $k|n$, and so $k \in D$. That is a contradiction since m is the smallest element of D. Thus m is a prime divisor of n. □

Among the more important theorems in number theory is the following.

Theorem 21.4. *The set of prime integers is infinite.*

Proof. Suppose that there were only finitely many primes. List them all: $2, 3, 5, 7, \cdots, p$. Form the number $N = 1 + 2 \cdot 3 \cdot 5 \cdot 7 \cdots p$. According to the last theorem, there must be a prime that divides N, say q. Certainly q also divides $2 \cdot 3 \cdot 5 \cdot 7 \cdots p$ since that is the product of all the primes, so q is one of its factors. Hence q divides $N - 2 \cdot 3 \cdot 5 \cdot 7 \cdots p$. But that's crazy since $N - 2 \cdot 3 \cdot 5 \cdot 7 \cdots p = 1$. We have reached a contradiction, and so we can conclude there are infinitely many primes. □

Theorem 21.5 (The Division Algorithm for Integers)**.** *If $a, d \in \mathbb{Z}$, with $d > 0$, there exist unique integers q and r, with $a = qd + r$, and $0 \le r < d$.*

Proof. Let $S = \{a - nd | n \in \mathbb{Z}, \text{ and } a - nd \ge 0\}$. Then $S \ne \emptyset$, since $a - (-|a|)d \in S$ for sure. Thus, by the Well Ordering Principle, S has a least element, call it r. Say $r = a - qd$. Then we have $a = qd + r$, and $0 \le r$. If $r \ge d$, then $a = (q + 1)d + (r - d)$, with $0 \le r - d$ contradicting the minimality of r.

To prove uniqueness, suppose that $a = q_1 d + r_1 = q_2 d + r_2$, with $0 \leq + r_1, r_2 < d$. Then $d(q_1 - q_2) = r_2 - r_1$ which implies that $r_2 - r_1$ is a multiple of d. Since $0 \leq r_1, r_2 < d$, we have $-d < r_2 - r_1 < d$. Thus the only multiple of d which $r_2 - r_1$ can possibly be is $0d = 0$. So $r_2 - r_1 = 0$ which is the same thing as $r_1 = r_2$. Thus $d(q_1 - q_2) = 0 = d0$. Since $d \neq 0$ we can cancel d to get $q_1 - q_2 = 0$, whence $q_1 = q_2$. □

The quantities *q* and *r* in the division algorithm are called the **quotient** and **remainder** when *a* is divided by *d*.

Exercises

Exercise 21.1. *Determine the quotient and remainder when* 107653 *is divided by* 22869*.*

Exercise 21.2. *Determine if* 1297 *is a prime.*

Exercise 21.3. *Prove or give a counterexample: The divides relation is reflexive.*

Exercise 21.4. *Prove or give a counterexample: The divides relation is symmetric.*

Exercise 21.5. *Prove or give a counterexample: The divides relation is transitive.*

Exercise 21.6. *What is wrong with the expression* $4|12 = 3$*?*

Exercise 21.7. *Show that none of the* 1000 *consecutive integers* $1001! + 2$ *to* $1001! + 1001$ *are primes.*

Exercise 21.8. *Prove: For* $a, b \in \mathbb{Z}$*, if* $a|b$*, then* $-a|b$*.*

Problems

Problem 21.1. *For positive integers,* a *and* b*, if the quotient when* a *is divided by* b *is* q*, what are the possible quotients when* $a + 1$ *is divided by* b*?*

Problem 21.2. *For positive integers,* a *and* b*, if the quotient when* a *is divided by* b *is* q*, what are the possible quotients when* $2a$ *is divided by* b*?*

Problem 21.3. *Prove or give a counterexample: If* p *is a prime, then* $2p + 1$ *is a prime.*

Problem 21.4. *Determine all the integers that* 0 *divides. (Hint: Think about the definition of the divides relation. The correct answer is probably not what you expect.)*

Problem 21.5. *Determine if* 3599 *is a prime. (Hint: This is easy since* $3599 = 3600 - 1$*.)*

Problem 21.6. *Determine if* 5129 *is a prime.*

Problem 21.7. *Prove property 10 of Theorem 21.1: For integers* a, b, c*, if* $a|b$*, then* $a|bc$*.*

Chapter 22

GCD's and the Euclidean Algorithm

The **greatest common divisor** of a and b, not both 0, is the largest integer which divides both a and b. For example, the greatest common divisor of 21 and 35 is 7. We write $\gcd(a, b)$, as shorthand for the greatest common divisor of a and b. So $\gcd(35,21) = 7$.

There are several ways to find the gcd of two integers, a and b (not both 0).

First, we could simply list all the positive divisors of a and b and pick the largest number that appears in both lists. Notice that 1 will appear in both lists. For the example above the positive divisors of 35 are 1, 5, 7, and 35. For 21 the positive divisors are 1, 3, 7, and 21. The largest number appearing in both lists is 7, so $\gcd(35,21) = 7$.

Another way to say the same thing: If we let D_a denote the set of positive divisors of a, then $\gcd(a, b)$ = the largest number in $D_a \cap D_b$.

The reason $\gcd(0, 0)$ is not defined is that every positive integer divides 0, and so there is no largest integer that divides 0. From now on, when we use the symbol $\gcd(a, b)$, we will tacitly assume a and b are not both 0. The integers a and b can be negative. For example if $a = -34$ and $b = 14$, then the set of positive divisors of -34 is $\{1, 2, 17, 34\}$ and the set of positive divisors of 14 is $\{1, 2, 7, 14\}$. The set of positive common divisors of 14 and -34 is the set $\{1, 2, 17, 34\} \cap \{1, 2, 7, 14\} = \{1, 2\}$.

The largest number in this set is $2 = \gcd(-34,14)$.

Obviously then $\gcd(a,b) = \gcd(-a,b)$ since a and $-a$ have the same set of positive divisors. So when computing the $\gcd(a,b)$ we may as well replace a and b by their absolute values if one or both happen to be negative.

Here are a few easy facts about gcd's:

(1) If $a \neq 0$, then $\gcd(a,a) = |a|$.

(2) $\gcd(a,1) = 1$.

(3) $\gcd(a,b) = \gcd(b,a)$. (The order a and b are given is not important, but it is traditional to list them with $a \geq b$.)

(4) If $a \neq 0$ and $a|b$, then $\gcd(a,b) = |a|$.

(5) If $a \neq 0$, $\gcd(a,0) = |a|$.

If $\gcd(a,b) = 1$, we say that a and b are **relatively prime**. When a and b are relatively prime, they have no common prime divisor. For example 12 and 35 are relatively prime.

It's pretty clear that computing $\gcd(a,b)$ by listing all the positive divisors of a and all the positive divisors of b, and selecting the largest integers that appears in both lists is not very efficient. There is a better way of computing $\gcd(a,b)$.

Theorem 22.1. *If a and b are integers (not both 0) and $a = sb + t$ for integers s and t, then $\gcd(a,b) = \gcd(b,t)$.*

Proof. To prove the theorem, we will show that the list of positive integers that divide both a and b is identical to the list of positive integers that divide both b and $t = a - sb$. So, suppose $d|a$ and $d|b$. Then $d|(a - sb)$ so $d|t$. Hence d divides both b and t. On the other hand, suppose $d|b$ and $d|t$. Then $d|(sb + t)$, so that $d|a$. Hence d divides both a and b. It follows that $\gcd(a,b) = \gcd(b,t)$. □

Euclid is given the credit for discovering this fact, and its use for computing gcd's is called the **Euclidean algorithm** in his honor. The idea is to use the theorem repeatedly until a pair of numbers is reached for which the gcd is obvious. Here is an example of the Euclidean algorithm in action.

Example 22.2. *Since* $14 = 1\cdot 10+4$, $\gcd(14,10) = \gcd(10,4)$. *In turn* $10 = 2\cdot 4+2$ *so* $\gcd(10,4) = \gcd(4,2)$. *Since* $4 = 2\cdot 2$, $\gcd(4,2) = \gcd(2,0) = 2$. *So* $\gcd(14,10) = 2$.

The same example, presented a little more compactly, and without explicitly writing out the divisions, looks like

$$\gcd(14,10) = \gcd(10,4) = \gcd(4,2) = \gcd(2,0) = 2$$

At each step, the second number is replaced by the remainder when the first number is divided by the second, and the second moves into the first spot. The process is repeated until the second number is a 0 *(which must happen eventually since the second number never will be negative, and it goes down by at least* 1 *with each repetition of the process). The* gcd *is then the number in the first spot when the second spot is* 0 *in the last step of the algorithm.*

Now, a more exciting example.

Example 22.3. *Find the greatest common divisor of* 540 *and* 252. *We may present the computations compactly, without writing out the divisions. We have*

$$\gcd(540,252) = \gcd(252,36) = \gcd(36,0) = 36.$$

Using the Euclidean algorithm to find gcd's is extremely efficient. Using a calculator with a ten-digit display, you can find the gcd of two ten-digit integers in a matter of a few minutes at most using the Euclidean algorithm. On the other hand, doing the same problem by first finding the positive divisors of the two ten-digit integers would be a tedious project lasting several days. Some modern cryptographic systems rely on the computation of the gcd's of integers of hundreds of digits. Finding the positive divisors of such large integers, even with a computer, is, at present, a hopeless task. But a computer implementation of the Euclidean algorithm will produce the gcd of integers of hundreds of digits in the blink of an eye.

The Euclidean algorithm can also be written out as a sequence of divisions:

$$a = q_1 \cdot b + r_1, 0 < r_1 < b$$

$$b = q_2 \cdot r_1 + r_2, 0 < r_2 < r_1$$

$$r_1 = q_3 \cdot r_2 + r_3, 0 < r_3 < r_2$$

$$\vdots$$

$$r_k = q_{k+2} r_{k+1} + r_{k+2}, 0 < r_{k+2} < r_{k+1}$$

$$\vdots$$

$$r_{n-2} = q_n \cdot r_{n-1} + r_n, 0 < r_n < r_{n-1}$$

$$r_{n-1} = q_{n+1} \cdot r_n + 0$$

The sequence of integer remainders $b > r_1 > \cdots > r_k > \cdots \geq 0$ must eventually reach 0. Let's say $r_n \neq 0$, but $r_{n+1} = 0$, so that $r_{n-1} = q_{n+1} r_n$. That is, in the sequence of remainders, r_n is the last non-zero term. Then, just as in the examples above we see that the gcd of a and b is the last nonzero remainder:

$$\gcd(a, b) = \gcd(b, r_1) = \gcd(r_1, r_2) = \cdots = \gcd(r_{n-1}, r_n) = \gcd(r_n, r_{n+1}) = \gcd(r_n, 0) = r_n.$$

Let's find gcd(317,118) using this version of the Euclidean algorithm. Here are the steps:

$$317 = 2 \cdot 118 + 81$$

$$118 = 1 \cdot 81 + 37$$

$$81 = 2 \cdot 37 + 7$$

$$37 = 5 \cdot 7 + 2$$

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

$$2 = 2 \cdot 1 + 0$$

Since the last non-zero remainder is 1, we conclude that gcd(317,118) = 1. So, in the terminology introduced above, we would say that 317 and 118 are relatively prime.

Exercises

Exercise 22.1. *Use the Euclidean algorithm to compute* $\gcd(a, b)$ *in each case.*
a) $a = 233$, $b = 89$ *b)* $a = 1001$, $b = 13$ *c)* $a = 2457$, $b = 1458$ *d)* $a = 567$, $b = 349$

Exercise 22.2. *Compute* $\gcd(987654321, 123456789)$.

Exercise 22.3. *Write a step-by-step algorithm that implements the Euclidean algorithm for finding* gcd*'s.*

Exercise 22.4. *If* n *is a positive integer, what is* $\gcd(n, 2n)$*?*

Problems

Problem 22.1. *Use the Euclidean algorithm to compute* $\gcd(a, b)$ *in each case.*

a) $a = 216$, $b = 111$ *b)* $a = 1001$, $b = 11$ *c)* $a = 663$, $b = 5168$ *d)* $a = 1357$, $b = 2468$

Problem 22.2. *Compute* $\gcd(733103, 91637)$.

Problem 22.3. *If* p *is a prime, and* n *is any integer, what are the possible values of* $\gcd(p, n)$*?*

Problem 22.4. *Prove or give a counterexample: If* p *and* q *are distinct primes, then* $\gcd(2p, 2q) = 2$.

Problem 22.5. *If* p *is a prime, and* m *is a positive integer, determine* $\gcd(p, p^m)$.

Problem 22.6. *If* p *is a prime, and* $m \leq n$ *are positive integers, determine* $\gcd(p^n, p^m)$.

Problem 22.7. *Show that if* n *is a positive integer, then* $\gcd(n, n + 1) = 1$.

Chapter 23

GCD's Reprised

The greatest common divisor of two integers a and b, not both zero, is defined to be the largest integer $\gcd(a, b)$ that divides them both. But there is another way to describe the greatest common divisor. First, a little vocabulary: recall that a **linear combination** of a and b is any expression of the form $as + bt$ where s, t are integers. For example, $4 \cdot 5 + 10 \cdot 2 = 40$ is a linear combination of 4 and 10. Here are some more linear combinations of 4 and 10:

$$4 \cdot 1 + 10 \cdot 1 = 14, \quad 4 \cdot 0 + 10 \cdot 0 = 0, \text{ and,} \quad 4 \cdot (-11) + 10 \cdot 1 = -34.$$

If we make a list of all possible linear combinations of 4 and 10, an unexpected pattern appears: $\cdots, -6, -4, -2, 0, 2, 4, 6, \cdots$. Since 4 and 10 are both even, we are sure to see only even integers in the list of linear combinations, but the surprise is that *every* even number is in the list. Now here's the connection with gcd's: The gcd of 4 and 10 is 2, and the list of all linear combinations is exactly all multiples of 2. Let's prove that was no accident.

Theorem 23.1. *Let a, b be two integers (not both zero). Then the smallest positive number in the list of the linear combinations of a and b is $\gcd(a, b)$. In other words, the $\gcd(a, b)$ is the smallest positive integer that can be written as a linear combination of a and b.*

Proof. Let $L = \{as + bt \mid s, t \text{ are integers and } as + bt > 0\}$. Since a, b are not both 0, we see this set is nonempty. As a nonempty set of positive integers, it must have a least element, say m. Since $m \in L$, m is a linear combination of a and b. Say $m = as_0 + bt_0$. We need to show $m = \gcd(a, b) = d$.

As noted above, since $d|a$ and $d|b$, it must be that $d|(as_0 + bt_0)$, so $d|m$. That implies $d \le m$.

We complete the proof by showing m is a common divisor of a and b. The plan is to divide a by m and show the remainder must be 0. So write $a = qm + r$ with $0 \le r < m$. Solving for r we get $0 \le r = a - qm = a - q(as_0 + bt_0) = a(1 - qs_0) + b(-qt_0) < m$. That shows r is a linear combination of a and b that is less than m. Since m is the smallest positive linear combination of a and b, the only option for r is $r = 0$. Thus $a = qm$, and so $m|a$. In the same way, $m|b$. Since m is a common divisor or a and b, it follows that $m \le d$. Since the reverse inequality is also true, we conclude $m = d$. □

And now we are ready for the punch-line.

Theorem 23.2. *Let a, b be two integers (not both zero). Then the list of all the linear combinations of a and b consists of all the multiples of* $\gcd(a, b)$.

Proof. Since $\gcd(a, b) = d$ certainly divides any linear combination of a and b, only multiples of d stand a chance to be in the list. Now we need to show that if n is a multiple of the d then n will appear in the list for sure. According to the last theorem, we can find integers s_0, t_0 so that $d = as_0 + bt_0$. Now since n is a multiple of d, we can write $n = de$. Multiplying both sides of $d = as_0 + bt_0$ by e gives $a(s_0e) + b(t_0e) = de = n$, and that shows n does appear in the list of linear combinations of a and b. □

So, without doing any computations, we can be sure that the set of all linear combinations of 15 and 6 will be all multiples of 3.

In practice, finding integers s and t so that $as + bt = d = \gcd(a, b)$ is carried out by using the Euclidean algorithm applied to a and b and then *back-solving*.

Example 23.3. *Let $a = 35$ and $b = 55$. Then the Euclidean algorithm gives*

$$55 = 35 \cdot 1 + 20$$
$$35 = 20 \cdot 1 + 15$$
$$20 = 15 \cdot 1 + 5$$
$$15 = 5 \cdot 3 + 0$$

The penultimate equation allows us to write $5 = 1 \cdot 20 + (-1) \cdot 15$ *as a linear combination of* 20 *and* 15. *We then use the equation* $35 = 20 \cdot 1 + 15$, *to write* $15 = 1 \cdot 35 + (-1) \cdot 20$. *We* *can substitute this into the previous expression for* 5 *as a linear combination of* 20 *and* 15 *to get* $5 = 1 \cdot 20 + (-1) \cdot 15 = 1 \cdot 20 + (-1) \cdot (1 \cdot 35 + (-1) \cdot 20)$*. Which can be simplified by collecting* 35*'s and* 20*'s to write* $5 = 2 \cdot 20 + (-1) \cdot 35$*. Now we can use the top equation to write* $20 = 1 \cdot 55 + (-1) \cdot 35$ *and substitute this into the expression giving* 5 *as a linear combination of* 35 *and* 20*. We get* $5 = 2 \cdot (1 \cdot 55 + (-1) \cdot 35) + (-1) \cdot 35$*. This simplifies to* $5 = 2 \cdot 55 + (-3) \cdot 35$*.*

This sort of computation gets a little tedious, keeping track of equations and coefficients. Moreover, the back-substitution method isn't very pleasant from a programming perspective since all the equations in the Euclidean algorithm need to be saved before solving for the coefficients in a linear combination for the $\gcd(a, b)$. A streamlined version called the continued fraction method uses forward-substitution to allow us to compute $\gcd(a, b)$ as a linear combination.

For the example above we first build a four-row table with $4 + 2 = 6$ columns. We begin by entering a and b in the top row left hand spaces. We put $*$'s below them in the second row (these entries are not used). Then we fill out the last two rows of the first two columns as shown.

55	35				
*	*				
0	1				
1	0				

We complete the first row using the remainders from the Euclidean Algorithm.

$$55 = 35 \cdot 1 + \mathbf{20}$$
$$35 = 20 \cdot 1 + \mathbf{15}$$
$$20 = 15 \cdot 1 + \mathbf{5}$$
$$15 = 5 \cdot 3 + \mathbf{0}$$

55	35	**20**	**15**	**5**	**0**
*	*				
0	1				
1	0				

We complete the second row using the quotients from the Euclidean Algorithm.

$$55 = 35 \cdot \mathbf{1} + 20$$
$$35 = 20 \cdot \mathbf{1} + 15$$
$$20 = 15 \cdot \mathbf{1} + 5$$
$$15 = 5 \cdot \mathbf{3} + 0$$

55	35	20	15	5	0
*	*	**1**	**1**	**1**	**3**
0	1				
1	0				

Finally, the remaining entries in each of the bottom two rows are filled in one after the other from left to right by taking the quotient entry in a column times the row entry one column back and then adding the row entry two columns back. For example, to compute the c in the following table use $c = q \cdot j + k$.

*	*			q	
0	1	k	j	c	
1	0				

Using that rule to complete the table for 55 and 35 we obtain the result

55	35	20	15	5	0
*	*	1	1	1	3
0	1	1	2	3	11
1	0	1	1	2	7

We don't actually need the bottom two entries in the last column. But observe that $7 \cdot 55 = 11 \cdot 35$. This will always be true! So we have a nice way to check that we have correctly filled out the table.

Anyway we now find integers s and t so that $\gcd(a, b) = s \cdot a + t \cdot b$. Obviously one of s and t will be positive and the other negative - but which one? The answer is in counting the number of back-substitutions we would have made. Start with a + at the one in the first column of the bottom row, and step along alternating + and − in each column. This will give the correct sign for s which is the second to last entry in the bottom row. The sign of t is the opposite and the absolute value of t is the entry in the 3rd row and second to last column.

In our running example we have $s = +2$ and $t = -3$ and can check that

$$5 = \gcd(55,35) = 55 \cdot 2 + 35 \cdot (-3).$$

In general to use the continued fraction table method to write $\gcd(a, b) = s \cdot a + t \cdot b$ we build a table with 4 rows and $n + 2$ columns, where the Euclidean Algorithm applied to a and b requires n steps and $r_n = 0$. In practice, as noticed above, we can omit the last column.

Exercises

Exercise 23.1. *Determine* $\gcd(13477,7667)$ *and write it as a linear combination of* 13447 *and* 7667. *Try both the method of back-substitution and the Extended Euclidean Algorithm to determine a suitable linear combination.*

Exercise 23.2. *What can you conclude about* $\gcd(a,b)$ *if there are integers* s,t *with* $as+bt=1$?

Exercise 23.3. *What can you conclude about* $\gcd(a,b)$ *if there are integers* s,t *with* $as+bt=19$?

Exercise 23.4. *What can you conclude about* $\gcd(a,b)$ *if there are integers* s,t *with* $as+bt=18$?

Problems

Problem 23.1. *Determine* $\gcd(41559,39417)$ *and write it as a linear combination of* 41559 *and* 39417. *Try both the method of back-substitution and the Extended Euclidean Algorithm to determine a suitable linear combination.*

Problem 23.2. *What can you conclude about* $\gcd(a,b)$ *if there are integers* s,t *with* $as+bt=12$?

Problem 23.3. *What is the smallest positive integer that can be written as a linear combination of* 2191 *and* 1351?

Problem 23.4. Definition: *The least common multiple of the positive integers* a *and* b *is the smallest positive integer that is divisible by both* a *and* b.

Example: the least common multiple of 24 *and* 18 *is* 72. *Write that as* $\text{lcm}(24,18)=72$.

You might recall the notion of least common multiple from the time you learned how to add fractions. The idea was that to add two fractions, $\frac{a}{b}$ *and* $\frac{c}{d}$, *first write the two as equivalent fractions with the same denominator. For example, to add* $\frac{2}{3}$ *and* $\frac{5}{4}$, *write them as* $\frac{8}{12}$ *and* $\frac{15}{12}$, *then add to get* $\frac{23}{12}$. *The least common denominator when adding* $\frac{a}{b}$ *and* $\frac{c}{d}$ *is the least common multiple of the denominators,* $\text{lcm}(b,d)$.

Determine $\text{lcm}(22,33)$. *(The answer is not* 726.*)*

Problem 23.5. *The result when* $\gcd(a,b)$ *and* $\text{lcm}(a,b)$ *are multiplied is always a simple combination of* a *and* b. *Example:* $\gcd(6,4)\cdot\text{lcm}(6,4)=2\cdot 12=24$. *Try a few more examples and see if you can guess the value of* $\gcd(a,b)\cdot\text{lcm}(a,b)$.

Chapter 24

The Fundamental Theorem of Arithmetic

The Fundamental Theorem of Arithmetic states the familiar fact that every positive integer greater than 1 can be written in exactly one way as a product of primes. For example, the prime factorization of 60 is $2^2 \cdot 3 \cdot 5$, and the prime factorization of 625 is 5^4. The factorization of 60 can be written is several different ways: 60 = 2 · 2 · 3 · 5 = 5 · 2 · 3 · 2, and so on. The order in which the factors are written does not matter. The factorization of 60 into primes will always have two 2's, one 3, and one 5. One more example: The factorization if 17 consists of the single factor 17. In the *standard form* of the factorization of an integer greater than 1, the primes are written in order of size, and exponents are used for primes that are repeated in the factorization. So, for example, the standard factorization of 60 is $60 = 2^2 \cdot 3 \cdot 5$.

Before proving the Fundamental Theorem of Arithmetic, we will need to assemble a few facts.

Theorem 24.1 (Euclid's Lemma)**.** *If $n|ab$ and n and a are relatively prime, then $n|b$.*

Proof. Suppose $n|ab$ and that $\gcd(n, a) = 1$. We can find integers s, t such that $ns + at = 1$. Multiply both sides of that equation by b to get $nsb + abt = b$. Since n divides both terms on the left side of that equation, it divides their sum, which is b. □

One consequence of this theorem is that if a prime divides a product of some integers, then it must divide one of the factors. That is so since if a prime does not divide an integer, then it is relatively prime to that integer. That is useful enough to state as a theorem.

Theorem 24.2. *If p is a prime, and $p|a_1a_2 \dots a_n$, then $p|a_j$ for some $j = 1, 2, \dots, n$.*

Theorem 24.3 (Fundamental Theorem of Arithmetic)**.** *If $n > 1$ is an integer, then there exist prime numbers $p_1 \leq p_2 \leq \dots \leq p_r$ such that $n = p_1p_2 \dots p_r$ and there is only one such prime factorization of n.*

Proof. There are two things to prove: (1) every $n > 1$ can be written in at least one way as a product of primes (in increasing order) and (2) there cannot be two different such expressions equal to n.

We will prove these by induction. For the basis, we see that 2 can be written as a product of primes (namely 2 = 2) and, since 2 is the smallest prime, this is the only way to write 2 as a product of primes.

For the inductive step, suppose every integer from 2 to k can be written uniquely as a product of primes. Now consider the number $k + 1$. We consider two cases:

(1) If $k + 1$ is a prime, then $k + 1$ is already an expression for $k + 1$ as a product of primes. There cannot be another expression for $k + 1$ as a product of primes, for if $k + 1 = pm$ with p a prime, then $p|k + 1$ and p and $k + 1$ both prime tells us $p = k + 1$ and $m = 1$.

(2) If $k + 1$ is not a prime, then we can write $k + 1 = ab$ with $2 \leq a, b \leq k$. By the inductive hypothesis, each of a and b can be written as products of primes, say $a = p_1p_2 \dots p_s$ and $b = q_1q_2 \dots q_t$. That means $k + 1 = p_1p_2 \dots p_r q_1 q_2 \dots q_t$, and we can rearrange the primes in increasing order. To complete the proof, we need to show $k + 1$ cannot be written in more than one way as a product of an increasing list of primes. So suppose $k + 1$ has two different such expressions: $k + 1 = u_1u_2 \dots u_l = v_1v_2 \dots v_m$. Since $u_1|v_1v_2 \dots v_m, u_1$ must divide some one of the v_i's and since u_1 and that v_i are both primes, they must be equal. As the v's are listed in increasing order, we can conclude $u_1 \geq v_1$. The same reasoning shows $v_1 \geq u_1$. Thus $u_1 = v_1$. Now cancel u_1, v_1 from each side of $u_1u_2 \dots u_l = v_1v_2 \dots v_m$ to get $u_2 \dots u_l = v_2 \dots v_m$. Since $k + 1$ was not a prime, both sides of this equation are greater than 1. Both sides are also less than $k + 1$. Since we started with two different factorizations, and canceled the same thing from both sides, we now have two different factorizations of a number between 2 and k. That contradicts the ind-

uctive assumption. We conclude that the prime factorization of $k + 1$ is unique. Thus, our induction proof is complete. □

We can apply the Fundamental Theorem of Arithmetic to the problem of counting the number of positive divisors of an integer greater than 1. For example, consider the integer $12 = 2^2 3$. It follows from the Fundamental Theorem that the positive divisors of 12 must look like $2^a 3^b$ where $a = 0, 1, 2$, $b = 0, 1$. So there are six positive divisors of 12:

$$2^0 3^0 = 1, 2^1 3^0 = 2, 2^2 3^0 = 4, 2^0 3^1 = 3, 2^1 3^1 = 6 \text{ and } 2^2 3^1 = 12.$$

Exercises

Exercise 24.1. *Determine the prime factorization of* 345678.

Exercise 24.2. *Determine the prime factorization of* 1016.

Exercise 24.3. *List all the positive divisors of* 1016.

Exercise 24.4. *How many positive divisors does* 345678 *have?*

Problems

Problem 24.1. *Determine the prime factorization of* 13579.

Problem 24.2. *List all the positive divisors of* 13579.

Problem 24.3. *Prove that if* n *is an even integer bigger than* 2*, then* $2^n - 1$ *is not a prime. Examples:* $2^4 - 1 = 15 = (3)(5)$, $2^{10} - 1 = 1023 = (3)(11)(31)$, *and Hint: Recall the factorization from college algebra* $s^2 - t^2 = (s+t)(s-t)$.

Problem 24.4. *Prime factorizations can be used to find greatest common divisors and least common multiples. The method is very inefficient compared to the Euclidean Algorithm since there is no known fast method of finding prime factorizations. Suppose* a *and* b *are two positive integers. Factor them each as product of primes. Say*

$$a = p_1^{e_1} p_2^{e_2} p_3^{e_3} \cdots p_n^{e_n} \text{ and } b = p_1^{f_1} p_2^{f_2} p_3^{f_3} \cdots p_n^{f_n}.$$

Note that the same list of primes is used for both factorizations, so we will need to allow exponents to be 0 *or more. For example, for* $a = 12$ *and* $b = 15$ *we will write* $a = 12 = 2^2 \cdot 3^1 \cdot 5^0$ *and* $15 = 2^0 \cdot 3^1 \cdot 5^1$.

Prove: $\gcd(a,b) = p_1^{\min(e_1,f_1)} p_2^{\min(e_2,f_2)} \cdots p_n^{\min(e_n,f_n)}$.

Example: $\gcd(12{,}15) = 2^{\min(2,0)} 3^{\min(1,1)} 5^{\min(0,1)} = 2^0 3^1 5^0 = 3$.

Prove: $\operatorname{lcm}(a,b) = p_1^{\max(e_1,f_1)} p_2^{\max(e_2,f_2)} \cdots p_n^{\max(e_n,f_n)}$.

Example: $\operatorname{lcm}(12{,}15) = 2^{\max(2,0)} 3^{\max(1,1)} 5^{\max(0,1)} = 2^2 3^1 5^1 = 60$.

Problem 24.5. *Using the formulas for* gcd *and* lcm *in the previous two problems, prove the result you guessed in the last chapter for the product of* $\gcd(a,b)\operatorname{lcm}(a,b)$. *(Hint: The fact you need is* $\min(s,t) + \max(s,t) = s + t$, *which is not hard to prove.)*

Chapter 25

Linear Diophantine Equations

Consider the following problem:

Al buys some books at \$25 each, and some magazines at \$3 each. If he spent a total of \$88, how many books and how many magazines did Al buy? At first glance, it does not seem we are given enough information to solve this problem. Letting x be the number of books Al bought, and y the number of magazines, then the equation we need to solve is $25x + 3y = 88$. Thinking back to college algebra days, we recognize $25x + 3y = 88$ as the equation of a straight line in the plane, and any point along the line will give a solution to the equation. For example, $x = 0$ and $y = 88/3$ is one solution. But, in the context of this problem, that solution makes no sense because Al cannot buy a fraction of a magazine. We need a solution in which x and y are both integers. In fact, we need even a little more care than that. The solution $x = -2$ and $y = 46$ is also unacceptable since Al cannot buy a negative number of books. So, we really need solutions in which x and y are both nonnegative integers. The problem can be solved by brute force: If $x = 0$, y is not an integer. If $x = 1$, then $y = 21$, so that is one possibility. If $x = 2, y$ is not an integer. If $x = 3, y$ is not an integer. And, if x is 4 or more, then y would have to be negative. So, it turns out there is only one possible solution: Al bought one book, and 21 magazines.

The above question is an example of a Diophantine problem. Pronounce Diophantine as *dee-uh-FAWN-teen* or *dee-uh-FAWN-tine*, or, the more common variations, *die-eh-FAN-teen* or *die-eh-FAN-tine*.

http://www.merriam-webster.com/audio.php?file=diopha01&word=Diophantineequation

For a modern pronunciation of Diophantus's name (Διοφαντοζ) see http://www.pronouncenames.com/Diophantus. In general, problems in which we are interested in finding solutions in which the variables are to be integers are called **Diophantine** problems.

In this chapter we will learn how to easily find the solutions to all linear Diophantine equations: $ax + by = c$ where a, b, c are given integers. To show some of the subtleties of such problems, here are two more examples:

(1) Al buys some books at \$24 each, and some magazines at \$3 each. If he spent a total of \$875, how many books and how many magazines did Al buy? For this question we need to solve the Diophantine equation $24x + 3y = 875$. In this case there are no possible solutions. For any integers x and y, the left-hand side will be a multiple of 3 and so cannot be equal to 875 which is not a multiple of 3.

(2) Al buys some books at \$26 each, and some magazines at \$3 each. If he spent a total of \$157, how many books and how many magazines did Al buy? Setting up the equation as before, we need to solve the Diophantine equation $26x + 3y = 157$. A little trial and error, testing $x = 0, 1, 2, 3$, and so on shows there are two possible answers this time: $(x, y) \in \{(2, 35), (5, 9)\}$.

Determining all the solutions to $ax + by = c$ is closely connected with the idea of gcd's. One connection is theorem 23.2. Here is how solutions of $ax + by = c$ are related.

Theorem 25.1. *$ax + by = c$ has a solution in the integers if and only if $\gcd(a, b)$ divides c.*

So, for example, $9x + 6y = 211$ has no solutions (in the integers) while $9x + 6y = 213$ does have solutions. To find a solution to the last equation, apply the Extended Euclidean Algorithm method to write the gcd(9,6) as a linear combination of 9 and 6 (actually, this one is easy to do by sight): $9 \cdot 1 + 6 \cdot (-1) = 3$, then multiply both sides by $\frac{213}{\gcd(9,6)} = \frac{213}{3} = 71$ to get $(71)9 + (-71)6 = 213$. That shows $x = 71$, $y = -71$ is a solution to $9x + 6y = 213$.

But that is only one possible solution. When a linear Diophantine equation has one solution it will have infinitely many. In the example above, another solution will be $x = 49$ and $y = -38$. Checking shows that $(49)9 + (-38)6 = 213$.

There is a simple recipe for all solutions, once one particular solution has been found.

Theorem 25.2. *Let* $d = \gcd(a, b)$. *Suppose* $x = s$ *and* $y = t$ *is one solution to* $ax + by = c$. *Then all solutions are given by*

$$x = s + k\left(\frac{b}{d}\right) \text{ and } y = t - k\left(\frac{a}{d}\right) \text{ where } k \text{ is any integer.}$$

Proof. It is easy to check that all the displayed x, y pairs are solutions simply by plugging in:

$$a\left(s + k\left(\frac{b}{d}\right)\right) + b\left(t - k\left(\frac{a}{d}\right)\right) = as + \frac{abk}{d} + bt - \frac{abk}{d} = as + bt = c.$$

Checking that the displayed formulas for x and y give all possible solutions is trickier. Let's assume $a \neq 0$. Suppose $x = u$ and $y = v$ is a solution. That means $au + bv = c = as + bt$. It follows that $a(u - s) = b(t - v)$. Divide both sides of that equation by d to get

$$\frac{a}{d}(u - s) = \frac{b}{d}(t - v).$$

That equation shows that $\frac{a}{d} \mid \frac{b}{d}(t - v)$. Since a/d and b/d are relatively prime, we conclude that a/d divides $(t - v)$. Let's say $k\left(\frac{a}{d}\right) = t - v$. Rearrange that equation to get

$$v = t - k\left(\frac{a}{d}\right).$$

Next, replacing $t - v$ in the equation $\frac{a}{d}(u - s) = \frac{b}{d}(t - v)$ with $k\left(\frac{a}{d}\right)$ gives

$$\frac{a}{d}(u - s) = \frac{b}{d}(t - v) = \frac{b}{d}\left(k\frac{a}{d}\right).$$

Since $a/d \neq 0$, we can cancel that factor. So we have

$$u - s = k\left(\frac{b}{d}\right) \text{ so that } u = s + k\left(\frac{b}{d}\right).$$

That proves the solution x = u, y = v is given by the displayed formulas. □

Example 25.3. *Determine all the solutions to* $221x + 91y = 39$.

Using the Extended Euclidean Algorithm method, we learn that $\gcd(221{,}91) = 13$ *and since* $13|39$, *the equation will have infinitely many solutions. The Extended Euclidean Algorithm table provides a linear combination of* 221 *and* 91 *equal to* 13*:* $221(-2) + 91(5) = 13$. *Multiply both sides by* 3 *and we get* $221(-6) + 91(15) = 39$. *So one particular solution to* $221x + 91y = 39$ *is* $x = -6$, $y = 15$. *According the the theorem above, all solutions are given by*

$$x = -6 + k\frac{91}{13} = -6 + 7k \text{ and } y = 15 - k\frac{221}{14} = 15 - 17k,$$

where k *is any integer.*

Example 25.4. *Armand buys some books for* \$25 *each and some cd's for* \$12 *each. If he spent a total of* \$331, *how many books and how many cd's did he buy?*

Let x = *the number of books, and* y = *the number of cd's. We need to solve* $25x + 12y = 331$. *The* gcd *of* 25 *and* 12 *is* 1, *and there is an obvious linear combination of* 25 *and* 12 *which equals* 1*:* $25(1) + 12(-2) = 1$. *Multiplying both sides by* 331 *gives* $25(331) + 12(-662) = 331$. *So one particular solution to* $25x + 12y = 331$ *is* $x = 331$ *and* $y = -662$. *Of course, that won't do for an answer to the given problem since we want* $x, y \geq 0$. *To find the suitable choices for* x *and* y, *let's look at all the possible solutions to* $25x + 12y = 331$. *We have that*

$$x = 331 + 12k \text{ and } y = -662 - 25k.$$

We want x *and* y *to be at least* 0, *and so we need*

$$331 + 12k \geq 0 \text{ and } -662 - 2kk \geq 0.$$

Which means that

$$k \geq -\frac{331}{12} \text{ and } k \leq -\frac{662}{25}, \text{or } -\frac{331}{12} \leq k \leq -\frac{662}{25}.$$

The only option for k *is* $k = -27$, *and so we see Armand bought* $x = 331 + 12(-27) = 7$ *books and* $y = -662 - 25(-27) = 13$ *cd's.*

Exercises

Exercise 25.1. *Find all integer solutions to* $21x + 48y = 8$.

Exercise 25.2. *Find all integer solutions to* $21x + 48y = 9$.

Exercise 25.3. *Find all integer solutions to* $33x + 12y = 7$.

Exercise 25.4. *Find all integer solutions to* $33x + 12y = 6$.

Exercise 25.5. *Sal sold some ceramic vases for* \$59 *each, and a number of bowls for* \$37 *each. If she took in a total of* \$4270, *how many of each item did she sell?*

Problems

Problem 25.1. *Find all integer solutions to* $14x + 77y = 69$.

Problem 25.2. *Find all integer solutions to* $14x + 77y = 70$.

Problem 25.3. *Beth stocked her video store with a number of video game machines at \$79 each, and a number of video games at \$41 each. If she spent a total of \$6358, how many of each item did she purchase?*

Problem 25.4. *If you all you have are dimes and quarters, in how many ways can you pay a \$7 bill?*
(For example, one way would be 10 *dimes and* 24 *quarters.)*

Problem 25.5. *How many integer solutions are there to the equation* $11x + 7y = 137$ *if the value of* x *has to be at least* -15 *and not more than* 20.

Problem 25.6. *Determine all integer solutions to* $5x - 7y = 99$. *(Watch that minus sign!)*

Chapter 26

Modular Arithmetic

Karl Friedrich Gauss made the important discovery of modular arithmetic. Modular arithmetic is also called *clock arithmetic,* and we are actually used to doing modular arithmetic *all the time* (pun intended). For example, consider the question *If it is* 7 *o'clock now, what time will it be in* 8 *hours?*. Of course the answer is 3 o'clock, and we found the answer by adding $7 + 8 = 15$, and then subtracting 12 to get $15 - 12 = 3$. Actually, we are so accustomed to that sort of calculation, we probably just immediately blurt out the answer without stopping to think how we figured it out. But trying a less familiar version of the same sort of problem makes it plain exactly what we needed to do to answer such questions: *If it is* 7 *o'clock now, what time will it be in* 811 *hours?* To find out, we add $7 + 811 = 818$, then divide that by 12, getting $818 = (68)(12) + 2$, and so we conclude it will be 2 o'clock. The general rule is: to find the time h hours after t o'clock, add $h + t$, divide by 12 and take the remainder.

There is nothing special about the number 12 in the above discussion. We can imagine a clock with any integer number of hours (greater than 1) on the clock. For example, consider a clock with 5 hours. What time will it be 61 hours after 2 o'clock. Since $61 + 2 = 63 = (12)(5) + 3$, the answer is 3 o'clock.

In the general case, if we have a clock with m hours, then the time h hours after t o'clock will be the remainder when $t + h$ is divided by m. So, the reason it is 2 o'clock 811 hours after 7 o'clock is that

$$811 + 7 \equiv 2 \pmod{12}.$$

This can all be expressed in more mathematical sounding language. The key is obviously the notion of remainder. That leads to the following definition:

Definition 26.1. Given an integer $m > 1$, we say that two integers a and b are **congruent modulo** m, and write $a \equiv b \pmod{m}$, in case a and b leave the same remainder when divided by m.

Theorem 26.2. *Congruence modulo m defines an equivalence relation on $\mathbb{Z}$.*

Proof. The relation is clearly reflexive since every number leaves the same remainder as itself when divided by m. Next, if a and b leave the same remainder when divided by m, so do b and a, so the relation is symmetric. Finally, if a and b leave the same remainder, and b and c leave the same remainder, then a and c leave the same remainder, and so the relation is transitive. □

There is an alternative way to think of congruence modulo m.

Theorem 26.3. $a \equiv b \pmod{m}$ *if and only if* $m|(a-b)$.

Proof. Suppose $a \equiv b \pmod{m}$. That means a and b leave the same remainder, say r when divided by m. So we can write $a = jm + r$ and $b = km + r$. Subtracting the second equation from the first gives $a - b = (jm + r) - (km + r) = jm - km = (j - k)m$, and that shows $m|(a - b)$.

For the converse, suppose $m|(a - b)$. Divide a, b by m to get quotients and remainders: $a = jm + r$ and $b = km + s$, where $0 \le r, s < m$. We need to show that $r = s$. Subtracting the second equation from the first gives $a - b = m(j - k) + (r - s)$. Since m divides $a - b$ and m divides $m(j - k)$, we can conclude m divides $(a - b) - m(j - k) = r - s$. Now since $0 \le r, s < m$, the quantity $r - s$ must be one of the numbers $m-1, m-2, \cdots, 2, 1, 0, -1, -2, \cdots -(m-1)$. The only number in that list that m divides is 0, and so $r - s = 0$. That is, $r = s$, as we wanted to show. □

The equivalence class of an integer a with respect to congruence modulo m will be denoted by $[a]$, or $[a]_m$ in case we are employing more than one number m as a *modulus*. In other words, $[a]$ is the set of all integers that leave the same remainder as a when divided by m. Or, another way to say the same thing, $[a]$ comprises all integers b such that $b - a$ is a multiple of m. That means $b - a = km$, or $b = a + km$.

That last version is often the easiest way to think about the integers that appear in $[a]$: start with a and add and subtract any number of m's. For example, the equivalence class of 7 modulo 11 would be

$$[7] = \{\cdots, -15, -4, 7, 18, 29, 40, \cdots\}.$$

We know that the distinct equivalence classes partition $\mathbb{Z}$. Since dividing an integer by m leaves one of $0, 1, 2, ..., m-1$ as a remainder, we can conclude that there are exactly m equivalence classes modulo m. In particular, $[0], [1], [2], [3], ..., [m-1]$ is a list of all the different equivalence classes modulo m. It is traditional when working with modular arithmetic to drop the [ ] symbols denoting the equivalence classes, and simply write the representatives. So we would say, modulo m, there are m numbers: $0, 1, 2, 3, \ldots, m-1$. But keep in mind that each of those numbers really represents a set, and we can replace any number in that list with another equivalent to it modulo m. For example, we can replace the 0 by m. The list $1, 2, 3, \ldots, m-1, m$ still consists of all the distinct values modulo m.

One reason the relation of congruence modulo m useful is that addition and multiplication of numbers modulo m acts in many ways just like arithmetic with ordinary integers.

Theorem 26.4. *If $a \equiv c \pmod{m}$, and $b \equiv d \pmod{m}$, then $a + b \equiv c + d \pmod{m}$ and $ab \equiv cd \pmod{m}$.*

Proof. Suppose $a \equiv c \pmod{m}$ and $b \equiv d \pmod{m}$. Then there exist integers k and l with $a = c + km$ and $b = d + lm$. So $a + b = c + km + d + lm = (c + d) + (k + l)m$. This can be rewritten as $(a + b) - (c + d) = (k + l)m$, where $k + l \in \mathbb{Z}$. So $a + b \equiv c + d \pmod{m}$. The other part is done similarly. □

Example 26.5. *What is the remainder when $1103 + 112$ is divided by 11? We can answer this problem in two different ways. We could add 1103 and 112, and then divide by 11. Or, we could determine the remainders when each of 1103 and 112 is divided by 11, then add those remainders before dividing by 11. The last theorem promises us the two answers will be the same. In fact $1103 + 112 = 1215 = (110)(11) + 5$ so that $1103 + 112 \equiv 5 \pmod{11}$. On the other hand $1103 = (100)(11) + 3$ and $112 = (10)(11) + 2$, so that $1103 + 112 \equiv 3 + 2 \equiv 5 \pmod{11}$.*

Example 26.6. *A little more impressive is the same sort of problem with operation of multiplication: what is the remainder when $(1103)(112)$ is divided by 11? The calculation looks like $(1103)(112) \equiv (3)(2) \equiv 6 \pmod{11}$.*

Example 26.7. *For a really awe inspiring example, let's find the remainder when 1103^{112} is divided by 11. In other words, we want to find $x = 0, 1, 2, ..., 10$ so that $1103^{112} \equiv x \pmod{11}$.*

Now 1103^{112} *is a pretty big number (in fact, since* $\log 1103^{112} = 112 \log 1103 = 340.7\cdots$*, the number has* 341 *digits). In order to solve this problem, let's start by thinking small: Let's compute* 1103^n*, for* $n = 1, 2, 3, \ldots$.

$$1103^1 \equiv 1103 \equiv 3 \pmod{11}$$
$$1103^2 \equiv 3^2 \equiv 9 \pmod{11}$$
$$1103^3 \equiv 1103(1103^2) \equiv 3(9) \equiv 27 \equiv 5 \pmod{11}$$
$$1103^4 \equiv (1103)(1103^3) \equiv 3(5) \equiv 15 \equiv 4 \pmod{11}$$
$$1103^5 \equiv (1103)(1103^4) \equiv 3(4) \equiv 12 \equiv 1 \pmod{11}$$

Now that last equation is very interesting. It says that whenever we see 1103^5 *we may just as well write* 1 *if we are working modulo* 11*. And now we see there is an easy way to determine* 1103^{112} *modulo* 11*:*

$$1103^{112} \equiv 1103^{5(22)+2} \equiv (1103^5)^{22}(1103^2) \equiv 1^{22}(9) \equiv 9 \pmod{11}$$

The sort of computation in example 26.7 appears to be just a curiosity, but in fact the last sort of example forms the basis of one version of public key cryptography. Computations of exactly that type (but with much larger integers) are made whenever you log into a secure Internet site. It's reasonable to say that e-commerce owes its existence to the last theorem.

While modular arithmetic in many ways behaves like ordinary arithmetic, there are some differences to watch for. One important difference is the familiar *rule of cancellation*: in ordinary arithmetic, if $ab = ac$ and $a \neq 0$, then $b = c$. This rule fails in modular arithmetic. For example, $3 \not\equiv 0 \pmod 6$ and $(3)(5) \equiv (3)(7) \pmod 6$, but $5 \not\equiv 7 \pmod 6$.

Solving congruence equations is a popular sport. Just as with regular arithmetic with integers, if we want to solve $a + x \equiv b \pmod m$, we can simply set $x \equiv b - a \pmod m$. So, for example, solving $55 + x \equiv 11 \pmod 6$ we would get $x \equiv 11 - 55 \equiv -44 \equiv 4 \pmod 6$.

Equations involving multiplication, such as $ax \equiv b \pmod m$, are much more interesting. If the modulus m is small, equations of this sort can be solved by trial-and-error: simply try all possible choices for x. For example, testing $x = 0, 1, 2, 3, 4, 5, 6$ in the equation $4x \equiv 5 \pmod 7$, we see $x \equiv 3 \pmod 7$ is the only solution. The equation $4x \equiv 5 \pmod 8$ has no solutions at all. And the equation $2x \equiv 4 \pmod 6$ has $x \equiv 2, 5 \pmod 6$ for solutions.

Trial-and-error is not a suitable approach for large values of m. There is a method that will produce all solutions to $ax \equiv b \pmod{m}$. It turns out that such equations are really just linear Diophantine equations in disguise, and that is the key to the proof of the following theorem.

Theorem 26.8. *The congruence* $ax \equiv b \pmod{m}$ *can be solved for* x *if and only if* $d = \gcd(a, m)$ *divides* b.

Proof. Solving $ax \equiv b \pmod{m}$ is the same as finding x so that $m|(ax - b)$ and that's the same as finding x and y so that $ax - b = my$. Rewriting that last equation in the form $ax + (-m)y = b$, we can see solving $ax \equiv b \pmod{m}$ is the same as solving the linear Diophantine equation $ax + (-m)y = b$. We know that equation has a solution if and only if $\gcd(a, m)\,|b$, so that proves the theorem. □

This is why $4x \equiv 5 \pmod{7}$ has a solution: $\gcd(4,7) = 1$ and $1|5$. And, why $4x \equiv 5 \pmod{8}$ has no solutions: $\gcd(4,8) = 4$, because $4 \nmid 5$. The theorem also shows that $2x \equiv 4 \pmod{6}$ has a solution since $\gcd(2,6) = 2$ and $2|4$. But why does this last equation have two solutions? The answer to that is also provided by the results concerning linear Diophantine equations.

Let $\gcd(a, m) = d$. The solutions to $ax \equiv b \pmod{m}$ are the same as the solutions for x to $ax + (-m)y = b$. Supposing that last equation has a solution with $x = s$, then we know all possible choices of x are given by $x = s + k\left(\frac{m}{d}\right)$. So if $x = s$ is one solution to $ax \equiv b \pmod{m}$, then all solutions are given by $x = s + k\left(\frac{m}{d}\right)$, where k is any integer. In other words, all solutions are given by $x \equiv s \pmod{\frac{m}{4}}$ and so there are d solutions modulo m.

Example 26.9. *Let's find all the solutions to* $2x \equiv 4 \pmod{6}$. *Since* $x = 2$ *is obviously one solution, we see all solutions are given by* $x = 2 + k\frac{6}{2} = 2 + 3k$, *where* k *is any integer. When* $k = 0, 1$ *we get* $x = 2, 5$, *and other values of* k *repeat these two modulo* 6. *Looking at the solutions written as* $x = 2 + 3k$, *we see that another way to express the solutions is* $x \equiv 2 \pmod{3}$.

Example 26.10. *Find all solutions to* $42x \equiv 35 \pmod{91}$.

First we see that $\gcd(91,42) = 7$ *and, since* $7|35$, *the equation will have a solution. In fact, since* $\gcd(91,42) = 7$, *there are going to be seven solutions modulo* 91. *All we need is to find one particular solution, then the others will all be easy to determine. Again using the continued fraction method (or just playing with* 42 *and* 91 *a little bit) we discover* $(42)(-2) + (91)(1) = 7 = \gcd(91,42)$. *Multiplying by* 5 *gives* $(42)(-10) + (91)(5) = 35$. *Thus* $x = -10$ *is one solution to* $42x \equiv 35 \pmod{91}$. *It follows that all solutions are given by* $x \equiv -10 \pmod{\frac{91}{\gcd(91,42)}}$ *That's the same as* $x \equiv -10 \pmod{13}$, *or, even more neatly,* $x \equiv 3 \pmod{13}$. *In other words, the solutions are* $3, 16, 29, 42, 55, 68, 81$ *modulo* 91.

Exercises

Exercise 26.1.

(a) *On a military (24-hour) clock, what time is it* 3122 *hours after* 16 *hundred hours?*

(b) *What day of the week is it* 3122 *days after a Monday?*

(c) *What month is it* 3122 *months after November?*

Exercise 26.2. *List the integers in* $[7]_{11}$.

Exercise 26.3. *In a listing of the five equivalence classes modulo* 5*, four of the values are* 1211*,* 218*,* −100*, and* −3333*. What are the possible choices for the fifth value?*

Exercise 26.4. *Determine* n *between* 0 *and* 24 *such that* $2311 + 3912 \equiv n \pmod{25}$.

Exercise 26.5. *Determine* n *between* 0 *and* 24 *such that* $(2311)(3912) \equiv n \pmod{25}$.

Exercise 26.6. *Determine* n *between* 0 *and 8 such that* $1111^{2222} \equiv n \pmod{9}$.

Exercise 26.7. *Solve:* $4x \equiv 3 \pmod{7}$.

Exercise 26.8. *Solve* $11x \equiv 8 \pmod{57}$.

Exercise 26.9. *Solve:* $14x \equiv 3 \pmod{231}$.

Exercise 26.10. *Solve* $8x \equiv 16 \pmod{28}$.

Exercise 26.11. *Solve:* $91x \equiv 189 \pmod{231}$.

Exercise 26.12. *Let* $d = \gcd(a, m)$ *and let* s *be a solution to* $ax \equiv b \pmod{m}$.

(a) *Show that if* $ax \equiv b \pmod{m}$*, then there is an integer* r *such that* $x = s + r\left(\frac{m}{d}\right)$.

(b) *If* $0 \leq r_1 < r_2 < d$*, then the numbers* $x_1 = s + r_1$ *and* $x_2 = s + r_2$ *are not congruent modulo* m.

Problems

Problem 26.1. *Suppose we have a* 52 *card deck with the cards in order ace,* 2, 3, . . . , *queen,king for clubs, then diamonds, then hearts, then spades from top to bottom. A step consists of taking the top card and moving it to the bottom of the deck. We start with the ace of clubs as the top card. After two steps, the top card is the* 3 *of clubs. What is the top card after* 735 *steps?*

Problem 26.2. *The marks on a combination lock are numbered* 0 *to* 39*. If the lock is at mark* 19*, and the dial is turned one mark clockwise, it will be at mark* 18*. If the lock is at mark* 19 *and turned* 137 *marks clockwise at what mark will it be?*

Problem 26.3. *List the integers in* $[11]_7$.

Problem 26.4. *Arrange the numbers* −39, −27, −8, 11, 37*, 68, 91 so they are in the order 0, 1, 2, 3, 4, 5, 6 modulo 7.*

Problem 26.5. *Determine* n *between* 0 *and* 16 *such that* $311 + 891 \equiv n \pmod{17}$.

Problem 26.6. *Determine* n *between* 0 *and* 16 *such that* $(405)(777) \equiv n \pmod{17}$.

Problem 26.7. *Determine* n *between* 0 *and* 16 *such that* $710^{447} \equiv n \pmod{17}$.

Problem 26.8. *Solve:* $3x \equiv 5 \pmod{8}$.

Problem 26.9. *Solve:* $13x \equiv 12 \pmod{68}$.

Problem 26.10. *Solve:* $15x \equiv 12 \pmod{27}$.

Problem 26.11. *Solve:* $12x \equiv 9 \pmod{88}$.

Problem 26.12. *Solve:* $33x \equiv 183 \pmod{753}$.

Problem 26.13. *There is exactly one* n *between* 0 *and* 55 *such that* $n \equiv 6 \pmod{7}$ *and* $n \equiv 1 \pmod{8}$. *Determine that* n.

Problem 26.14. *There is exactly one* n *between* 0 *and* 19548 *such that* $n \equiv 22 \pmod{173}$ *and* $n \equiv 80 \pmod{113}$. *Determine that* n.

Chapter 27

Integers in Other Bases

The usual way of writing integers is in terms of groups of ones (units), and groups of tens, and groups of tens of tens (hundreds), and so on. Thus 237 stands for 7 units plus 3 tens and 2 hundreds, or $2(10^2) + 3(10) + 7$. This is the familiar decimal notation for numbers (deci = ten). But there is really nothing special about the number ten here, and it could be replaced by any integer bigger than one. That is, we could use say 7 the way 10 was used above to describe a number. Thus we would specify how many units, how many 7's and 7^2's and 7^3's, and so on are needed to make up the number. When a number is expressed in this fashion with b in place of the 10, the result is called the **base-*b*** expansion (or **radix-*b*** expansion) of the integer.

For example, the decimal integer 132, is made up of two 7^2's, four 7's and finally six units. Thus we express the base ten number 132 as 246 in base 7, or as 246_7, the little 7 indicating the base. For small numbers, with a couple of minutes practice, conversion from base 10 (decimal) to other bases, and back again can be carried out mentally. For larger numbers, mental arithmetic will prove a little awkward. Luckily there is a handy algorithm to do the conversion automatically.

For base 10 integers, we use the decimal digits: $0, 1, 2, 3, 4, 5, 6, 7, 8, 9$. In general, for base b, the digits will be 0, 1, 2, 3 ... , $b - 1$. So, for example, a base 7 numbers use digits 0, 1, 2, 3, 4, 5, 6.

Conversion from the base b expansion of a number to its decimal version is a snap: For example, the meaning of 2302_5 is

$$2302_5 = 2 \cdot 5^3 + 3 \cdot 5^2 + 0 \cdot 5 + 2 = 2(125) + 3(25) + 0(5) + 2 = 327.$$

That sort of computation is so easy because we have been practicing base 10 arithmetic for so many years. If we were as good at arithmetic in some base b, then conversion from base 10 to base b would be just as simple. But, lacking that comfort with base b arithmetic, we need to describe the conversion algorithm from decimal to base b a little more formally. Here's the idea.

Suppose we have a decimal number n that we want to convert to some base b. Let's say the base b expansion is $d_k d_{k-1} \ldots d_2 d_1 d_0$, with the base b digits between 0 and $b-1$. That means

$$n = d_k \cdot b^k + d_{k-1} \cdot b^{k-1} + \cdots + d_2 \cdot b^2 + d_1 \cdot b + d_0$$

Now, if we divide n by b, we can see the equation above tells us

$$n = (d_k \cdot b^{k-1} + d_{k-1} \cdot b^{k-2} + \cdots + d_2 \cdot b + d_1)b + d_0.$$

So the quotient is $q = d_k \cdot b^{k-1} + d_{k-1} \cdot b^{k-2} + \cdots + d_2 \cdot b + d_1$, and the remainder is the base b digit d_0 of n. So we have found the units digit in the base b expansion of n. If we repeat that process on the quotient q, the result is

$$q = (d_k \cdot b^{k-2} + d_{k-1} \cdot b^{k-3} + \cdots + d_2)b + d_1$$

so the next base b digit, d_1 appears as the remainder. Continuing in this fashion, the base b expansion is produced one digit at a time.

Briefly, to convert a positive decimal integer n to its base b representation, divide n by b, to find the quotient and the remainder. That remainder will be needed units digit. Then divide the quotient by b again, to get a new quotient and a new remainder. That remainder gives the next base b digit. Then divide the new quotient by b again, and so on. In this way producing the base b digits one after the other.

Example 27.1. *To convert* 14567 *from decimal to base* 5*, the steps are:*

$$\begin{aligned}
14567 &= 2913 \cdot 5 + 2\\
2913 &= 582 \cdot 5 + 3\\
582 &= 116 \cdot 5 + 2\\
116 &= 23 \cdot 5 + 1\\
23 &= 4 \cdot 5 + 3\\
4 &= 0 \cdot 5 + 4.
\end{aligned}$$

So, we see that $14567 = 431232_5$.

We can convert between two integer bases, neither of which is 10.

Example 27.2. *Convert* $n = 3355_7$ *to base* 5*.*

The least confusing way to do such a problem would be to convert n *from base* 7 *to base* 10*, and then convert the base* 10 *expression for* n *to base* 5*. This method allows us to do all our work in base* 10 *where we are comfortable. The computations start with:*

$$n = 3 \cdot 7^3 + 3 \cdot 7^2 + 5 \cdot 7 + 5 = 1216.$$

Then, we calculate:

$$\begin{aligned}
1216 &= 243 \cdot 5 + 1\\
243 &= 48 \cdot 5 + 3\\
48 &= 9 \cdot 5 + 3\\
9 &= 1 \cdot 5 + 4\\
1 &= 0 \cdot 5 + 1.
\end{aligned}$$

So, we have $3355_7 = 14331_5$.

An alternative method, not for the faint of heart, is convert directly from base 7 to base 5 skipping the middle man, base 10. In this method, we simply divide *n* by 5, take the remainder, getting the units digit, then divide the quotient by 5 to get the next digit, and so on, just as described above. The rub is that the arithmetic must all be done in base 7, and we don't know the base 7 times table very well. For example, in base 7, $3 \cdot 5 = 21$ is correct since three 5's add up to two 7's plus one more.

The computation would now look like (all the $_7$'s indicating base 7 are suppressed for readability):

$$\begin{aligned} 3355 &= 465 \cdot 5 + 1 \qquad \text{(yes, that's really correct!)}, \\ 465 &= 66 \cdot 5 + 3 \\ 66 &= 12 \cdot 5 + 3 \\ 12 &= 1 \cdot 5 + 4 \\ 1 &= 0 \cdot 5 + 1. \end{aligned}$$

Hence, once again, we have $3355_7 = 14331_5$.

Particularly important in computer science applications of discrete mathematics are the bases 2 (called binary), 8 (called octal) and 16 (called hexadecimal, or simply hex). Thus the decimal number 75 would be 1001011_2 (binary), 113_8 (octal) and $4B_{16}$ in hex. Note that for hex numbers, symbols will be needed to represent hex digits for 10, 11, 12, 13, 14 and 15. The letters A, B, C, D, E and F are traditionally used for these digits.

Exercises

Exercise 27.1. *Convert to decimal:* 21_3, 321_4, 4321_5, *and* FED_{16}.

Exercise 27.2. *Convert the decimal integer* 11714 *to bases* 2, 6, *and* 16. *Remember to use* $A, B, \cdots, F$ *to represent base* 16 *digits from* 10 *to* 15, *if needed.*

Exercise 27.3. *Complete the following* base 7 *multiplication table.*

×	1	2	3	4	5	6
1	1	2	3	4	5	6
2	2	4				
3			12			
4				22		
5						
6						

Exercise 27.4. *Make* base 6 *addition and multiplication tables similar to the base* 7 *multiplication table of exercise* 27.3.

Exercise 27.5. ***(For those with a sweet tooth for punishment!)*** *Use the Euclidean algorithm to compute* $\gcd(5122_7, 1312_7)$ *without converting the numbers to base* 10.

Problems

Problem 27.1. *Convert to decimal:* 12_3, 123_4, 1234_5, *and* DDD_{16}.

Problem 27.2. *Convert the decimal integer* 3177 *to bases* 2, 6, *and* 16. *Use* $A, B, \cdots, F$ *to represent base* 16 *digits from* 10 *to* 15 *as needed.*

Problem 27.3. *Make* base 8 *addition and multiplication tables similar to the base* 7 *multiplication table of exercise* 27.3.

Problem 27.4. *Using the table in problem* 27.3, *add* $126_8 + 457_8$.

Problem 27.5. *Using the table in problem* 27.3, *multiply* $(126_8)(457_8)$.

Chapter 28

The Two Fundamental Counting Principles

The next few chapters will deal with the topic of **combinatorics:** the art of counting. By counting we mean determining the number of different ways of arranging objects in certain patterns or the number of ways of carrying out a sequence of tasks. For example, suppose we want to count the number of ways of making a bit string of length two. Such a problem is small enough that the possible arrangements can be counted by *brute force*. In other words, we can simply make a list of all the possibilities: $00, 01, 10, 11$. So the answer is four. If the problem were to determine the number of bit strings of length fifty, the brute force method loses a lot of its appeal. For problems where brute force counting is not a reasonable alternative, there are a few principles we can apply to aid in the counting. In fact, there are just two basic principles on which all counting ultimately rests.

Throughout this chapter, all sets mentioned will be finite sets, and if A is a set, $|A|$ will denote the number of elements in A.

The **sum rule** says that if the sets A and B are disjoint, then

$$|A \cup B| = |A| + |B|.$$

Example 28.1. *For example, if* $A = \{a, b, c\}$ *and* $B = \{j, k, l, m, n\}$, *then* $|A| = 3, |B| = 5$, *and, sure enough,*

$$|A \cup B| = |\{a, b, c, j, k, l, m, n\}| = 8 = 3 + 5.$$

Care must be used when applying the sum principle that the sets are disjoint. If $A = \{a, b, c\}$ and $B = \{b, c, d\}$, then $|A \cup B| = 4$, and not 6.

Example 28.2. *As another example of the sum principle, if we have a collection of* 3 *dogs and* 5 *cats, then we can select one of the animals in* 8 *ways.*

The sum principle is often expressed in different language: If we can do task 1 in m ways and task 2 in n ways, and the tasks are *independent* (meaning that both tasks cannot be done at the same time), then there are $m + n$ ways to do one of the two tasks. The independence of the tasks is the analog of the disjointness of the sets in the set version of the sum rule.

A serious type of error is trying to use the sum rule for tasks that are not independent. For instance, suppose we want to know *in how many different ways we can select either a deuce or a six from an ordinary deck of* 52 *cards.* We could let the first task be the process of selecting a deuce from the deck. That task can be done in 4 ways since there are 4 deuces in the deck. For the second task, we will take the operation of selecting a six from the deck. Again, there are 4 ways to accomplish that task. Now these tasks are independent since we cannot simultaneously pick a deuce and a six from the deck. So, according to the sum rule, there are 4 + 4 = 8 ways of selecting one card from a deck, and having that card be either a deuce or a six.

Now consider the similar sounding question: *In how many ways can we select either a deuce or a diamond from a deck of* 52 *cards?* We could let the first task again be the operation of selecting a deuce from the deck, with 4 ways to carry out that task. And we could let the second task be the operation of selecting a diamond from the deck, with 13 ways to accomplish that. But in this case, the answer to the question is not 4 + 13 = 17, since these tasks are not independent. It is possible to select a card that is both a deuce and a diamond. So the sum rule cannot be used. What is the correct answer? Well, there are 13 diamonds, and there are 3 deuces besides the two of diamonds, and so there are actually 16 cards in the deck that are either a deuce or a diamond. That means there are 16 ways to select a card from a deck and have it turn out to be either a deuce or a diamond.

The sum rule can be extended to the case of more than two sets (or more than two tasks): If $A_1, A_2, \dots, A_n$ is a collection of *pairwise disjoint* sets, then $|A_1 \cup A_2 \cup \dots \cup A_n| = |A_1| +$

$|A_2| + |A_3| + \cdots + |A_n|$. Or, in terms of tasks: If task 1 can be done in k_1 ways, and task 2 in k_2, and task 3 in k_3 ways, and so on, until task n can be done in k_n ways, and if the tasks are all independent, then we can do one task in $k_1 + k_2 + k_3 + \cdots + k_n$ ways.

Notice that the tasks must be pairwise independent.

Example 28.3. *For example, if we own three cars, two bikes, a motorcycle, four pairs of roller skates, and two scooters, then we can select one of these modes of transportation in* $3+2+1+4+2 = 12$ *ways.*

The sum rule is related to the logical connective *or*. That is reasonable since the sum rule counts the number of elements in the set $A \cup B = \{x \mid x \in A \text{ or } x \in B\}$. In terms of tasks, the sum rule counts the number of ways to do either task 1 or task 2. Generally speaking, when the word *or* occurs in a counting problem, the sum rule is the tool to use.

The logical connective *and* is related to the second fundamental counting principle: the **product rule**. The product rule says:

$$|A \times B| = |A| \cdot |B|.$$

An explanation of this is that $A \times B$ consists of all ordered pairs (a, b) where $a \in A$ and $b \in B$. There are $|A|$ choices for a and then $|B|$ choices for b.

In terms of tasks, the product rule says that if task 1 can be done in m ways and task 2 can be done in n ways after task 1 has been done, then there are mn ways to do both tasks, the first then the second. Here the relation with the logical connective *and* is also obvious. We need to do task 1 and task 2. Generally speaking, the appearance of *and* in a counting problem suggests the product rule will come into play.

As with the sum rule, the product rule can be used for situations with more than two sets or more than two tasks. In terms of sets, the product rule reads $|A_1 \times A_2 \times \cdots \times A_n| = |A_1| \cdot |A_2| \cdots |A_n|$. In terms of tasks, it reads, if task 1 can be done in k_1 ways, and for each of those ways, task 2 can be done in k_2 ways, and for each of those ways, task 3 can be done in k_3 ways, and so on, until for each of those ways, task n can be done in k_n ways, then we can do task 1 followed by task 2 followed by task 3, etc, followed by task n in $k_1 \cdot k_2 \cdot \ldots \cdot k_n$ ways. That sounds worse than it really is.

Example 28.4. *How many bit strings are there of length five?*

Solution. *We can think of task* 1 *as filling in the first (right hand) position, task* 2 *as filling in the second position, and so on. We can argue that we have two ways to do task* 1, *and then two ways* *to do task* 2*, and then two ways to do task* 3*, and then two ways to do task* 4*, and then two ways to do task* 5*. So, by the product rule, there are* $2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 = 2^5 = 32$ *ways to do all five tasks, and so there are* 32 *bit strings of length five.*

The same reasoning shows that, in general, there are 2^n *bit strings of length* n*, when* $n \geq 0$*.*

Example 28.5. *Suppose we are buying a car with five choices for the exterior color and three choices for the interior color. Then there is a total of* $3 \cdot 5 = 15$ *possible color combinations that we can choose from. The first task is to select an exterior color, and there are* 5 *ways to do that. The second task is to select an interior color, and there are* 3 *ways to do that. So the product rule says there are* 15 *ways total to do both tasks. Notice that there is no requirement of independence of tasks when using the product rule. However, also notice that the number of ways of doing the second task* ***must*** *be the same no matter what choice is made for doing the first task.*

Example 28.6. *For another, slightly more complicated, example of the product rule in action, suppose we wanted to make a two-digit number using the digits* $1, 2, 3, 4, 5, 6, 7, 8,$ *and* 9*. How many different such two-digit numbers could we form? Let's make the first task filling in the left digit, and the second task filling in the right digit. There are* 9 *ways to do the first task. And, no matter how we do the first task, there are* 9 *ways to do the second task as well. So, by the product rule, there are* $9 \cdot 9 = 81$ *possible such two-digit numbers.*

Example 28.7. *Now, let's change the problem in example 28.6 a little bit. Suppose we wanted two-digit numbers made up of those same nine digits, but we do not want to use a digit more than once in any of the numbers. In other words,* 37 *and* 91 *are OK, but we do not want to count* 44 *as a possibility. We can still make the first task filling in the left digit, and the second task filling in the right digit. And, as before, there are* 9 *ways to do the first task. But now, once the first task has been done, there are only* 8 *ways to do the second task, since the digit used in the first task is no longer available for doing the second task. For instance, if the digit* 3 *was selected in the first task, then for the second task, we will have to choose from the eight digits* $1, 2, 4, 5, 6, 7, 8,$ *and* 9*. So, according to the product rule, there are* $9 \cdot 8 = 72$ *ways of building such a number.*

No matter in what way the first task was done, there are always 8 *ways to to the second task in sequence. What if you chose to pick the second digit first?*

Example 28.8. *Just for fun, here is another way to see the answer in example 28.7 is* 72*. We saw above that there are* 81 *ways to make a two-digit number when we allow repeated digits. But there are* 9 *two digit numbers that do have repeated digits (namely* $11, 22, \cdots, 99$*). That means there must be* $81 - 9 = 72$ *two-digit numbers without repeated digits.*

The trick we used in example 28.8 looks like a new counting principle, but it is really the sum rule being applied in a tricky way. Here's the idea. Call the set of all the two-digit numbers (not using 0) T, call the set with no repeated digits N, and call the set with repeated digits R. By the sum rule, $|T| = |N| + |R|$, so $|N| = |T| - |R|$. This is a very common trick.

Generally, suppose we are interested in counting some arrangements, let's call them the *Good* arrangements. But it is not easy for some reason to count the *Good* arrangements directly. So, instead, we count the *Total* number of arrangements, and subtract the number of *Bad* arrangements:

$$\text{Good} = \text{Total} - \text{Bad}.$$

Let's have another example of this trick.

Example 28.9. *By a word of length five, we will mean any string of five letters from the* 26 *letter alphabet. How many words contain at least one vowel. The vowels are:* a,e,i,o,u.

By the product rule, there is a total of 26^5 *possible words of length five. The bad words are made up of only the* 21 *non-vowels. So, by the sum rule, the number of good words is* $26^5 - 21^5$.

As in example 28.9, most interesting counting problems involve a combination of both the sum and product rules.

Example 28.10. *Suppose we wanted to count the number of different possible bit strings of length five that start with either three* 0*'s or with two* 1*'s. Recall that a bit string is a list of* 0*'s and* 1*'s, and the length of the bit string is the total number of* 0*'s and* 1*'s in the list. So, here are some bit strings that satisfy the stated conditions:* 00001, 11111, 11011, *and* 00010. *On the other hand, the bit strings* 00110 *and* 10101 *do not meet the required condition.*

To do this problem, let's first count the number of good bit strings that start with three 0*'s. In this case, we can think of the construction of such a bit string as doing five tasks, one after the other, filling in the leftmost bit, then the next one, then the third, the next, and finally the last bit. There is only one way to do the first three tasks, since we need to fill in* 0*'s in the first three positions. But there are two ways to do the last two tasks, and so, according to the product rule there are* $1 \cdot 1 \cdot 1 \cdot 2 \cdot 2 = 4$ *bit strings of length five starting with three* 0*'s. Using the same reasoning, there are* $1 \cdot 1 \cdot 2 \cdot 2 \cdot 2 = 8$ *bit strings of length five starting with two* 1*'s. Now, a bit string cannot both start with three* 0*'s and also with two* 1*'s, (in other words, starting with three* 0*'s and starting with two* 1*'s are independent). And so, according to the sum rule, there will be a total of* $4 + 8 = 12$ *bit strings of length five starting with either three* 0*'s or two* 1*'s.*

Example 28.11. *How many words of six letters (repeats OK) contain exactly one vowel?*

Solution. *Let's break the construction of a good word down into a number of tasks.*

Task 1: Select a spot for the vowel: 6 *choices.*

Task 2: Select a vowel for that spot: 5 *choices.*

Task 3: Fill first empty spot with a non-vowel: 21 *choices*

Task 4: Fill next empty spot with a non-vowel: 21 *choices*

Task 5: Fill next empty spot with a non-vowel: 21 *choices*

Task 6: Fill next empty spot with a non-vowel: 21 *choices*

Task 7: Fill last empty spot with a non-vowel: 21 *choices*

By the product rule, the number of good words is $6 \cdot 5 \cdot 21^5$*.*

Example 28.12. *Count the number of strings on license plates which either consist of three capital English letters, followed by three digits, or consist of two digits followed by four capital English letters.*

Solution. *Let* A *be the set of strings which consist of three capital English letters followed by three digits, and* B *be the set of strings which consist of two digits followed by four capital English letters. By the product rule* $|A| = 26^3 \cdot 10^3$ *since there are 26 capital English letters and 10 digits. Also by the product rule* $|B| = 10^2 \cdot 26^4$*. Since* $A \cap B = \emptyset$*, by the sum rule the answer is* $26^3 \cdot 10^3 + 10^2 \cdot 26^4$*.*

In the previous examples we might continue on with the arithmetic. For instance, in the last, example 28.12, using the distributive law on our answer to factor out common terms we see $|A \cup B| = 10^2 \cdot 26^3(10 + 26)$ is an equivalent answer. This, in turn, simplifies to $|A \cup B| = 10^2 \cdot 26^3 \cdot 36$, and that gives

$$|A \cup B| = 100 \cdot 17576 \cdot 36 = 63{,}273{,}600.$$

Of all of these answers the most valuable is probably $26^3 \cdot 10^3 + 10^2 \cdot 26^4$, since **the form of the answer is indicative of the manner of solution.** We can readily observe that the sum rule was applied to two disjoint subcases. For each subcase the product rule was applied to compute the intermediate answer. As a general rule, answers to counting problems should be left in this uncomputed form.

Exercises

Exercise 28.1. *To meet the science requirement a student must take one of the following courses: a choice of* 5 *biology courses,* 4 *physics courses, or* 6 *chemistry courses. In how many ways can the one course be selected?*

Exercise 28.2. *Using the data of problem 1, a student has decided to take one biology, one physics, and one chemistry course. How many different such selections are possible?*

Exercise 28.3. *A serial code is formed in one of three ways: (1) two letters followed by two digits, or (2) three letters followed by one digit, or (3) four letters. How many different codes are there? (Unless otherwise indicated, letters will means upper case letters chosen from the usual* 26*-letter alphabet and digits are selected from* $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$*.)*

Exercise 28.4. *How many words of length six are there if letters may be repeated? (Examples: BBBXBB, ABATBC are OK).*

Exercise 28.5. *How many words of length six are there if letters may not be repeated? (Examples: BBBBXB, ABATJC are bad but ABXHYR is OK).*

Exercise 28.6. *A true/false test contains 25 questions.*

(a) How many ways can a student complete the test if every question must be answered?

(b) How many ways can a student complete the test if questions can be left unanswered?

Exercise 28.7. *How many binary strings of length less than or equal to nine are there?*

Exercise 28.8. *How many eight-letter words contain at least one* A*?*

Exercise 28.9. *How many seven-letter words contain at most one* A*?*

Exercise 28.10. *How many nine-letter words contain at least two* A*'s?*

Problems

Problem 28.1. *My piggy bank contains* 20 *pennies,* 4 *nickels,* 7 *dimes, and* 2 *quarters. In how many ways can I select one coin?*

Problem 28.2. *My piggy bank contains* 20 *pennies,* 4 *nickels,* 7 *dimes, and* 2 *quarters. In how many ways can I select four coins, one of each value?*

Problem 28.3. *A multiple choice test contains 10 questions. There are four possible answers for each question.*

(a) *How many ways can a student complete the test if every question must be answered?*

(b) *How many ways can a student complete the test if questions can be left unanswered?*

Problem 28.4. *Computer ID's are length seven strings made up of any combination of seven different letters and digits. How many different ID's are there?*

Problem 28.5. *Computer ID's are length seven strings made up of any combination of seven letters and digits, with repeats allowed. How many different ID's are there?*

Problem 28.6. *A code word is either a sequence of three letters followed by two digits or two letters followed by three digits. (Unless otherwise indicated, letters will means upper case letters chosen from the usual* 26*-letter alphabet and digits are selected from* $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$*.) How many different code words are possible?*

Problem 28.7. *Code words consist of five letters followed by five digits. How many code words contain at least one* X*?*

Problem 28.8. *Code words consist of five letters followed by five digits. How many code words contain exactly one* X*?*

Problem 28.9. *Code words consist of five letters followed by five digits. How many code words contain exactly two* X*'s?*

Problem 28.10. *How many bit strings of length ten begin and end with* 1*'s?*

Problem 28.11. *How many bit strings of length at least two but no more than ten begin and end with* 1*'s?*

Chapter 29

Permutations and Combinations

By a **permutation** of a set of objects we mean a listing of the objects of the set in a specific order. For example, there are six possible permutations of the set $A = \{a, b, c\}$. They are

$$abc, \quad acb, \quad bac, \quad bca, \quad cab, \quad cba.$$

The product rule explains why there are six permutation of A: there are 3 choices for the first letter, once that choice has been made there are 2 choices for the second letter, and finally that leaves 1 choice for the last letter. So the total number of permutations is $3 \cdot 2 \cdot 1 = 6$.

A set with n elements is called an n**-set**. We have just shown that a 3-set has 6 permutations. The same reasoning shows that an n-set has $n\cdot(n-1)\cdot(n-2)\cdots 2\cdot 1 = n!$ permutations. So the total number of different ways to arrange a deck of cards is 52!, a number with 68 digits:

80658175170943878571660636856403766975289505440883277824000000000000.

Instead of forming a permutation of all the elements of an n-set, we might consider the problem of first selecting some of the elements of the set, say r of them, and then forming a permutation of just those r elements. In that case we say we have formed an r**-permutation** of the n-set. All the possible 2-permutations of the 4-set $A = \{a, b, c, d\}$ are

$$ab, \quad ac, \quad ad, \quad ba, \quad bc, \quad bd, \quad ca, \quad cb, \quad cd, \quad da, \quad db, \quad dc$$

Hence, there are twelve 2-permutations of a 4-set.

Notation. *In general, $P(n,r)$ denotes the number of different r-permutations of an n-set.*

So for example The number of 2-permutations of a 4-set is $P(4,2) = 12$.

The product rule provides a simple formula for $P(n,r)$. There are n choices for the first element, and once that choice has been made, there are $n-1$ choices for the second element, then $n-2$ for the third, and so on, until finally, there are $n-(r-1) = n-r+1$ choices for the rth element. So $P(n,r) = n(n-1)\cdot \ldots \cdot (n-r+1)$. That expression can be written more neatly as follows:

$$P(n,r) = n(n-1)\cdot \ldots \cdot (n=r+1)$$
$$= \frac{n(n-1)\cdot \ldots \cdot (n-r+1)(n-r)(n-r-1)\ldots \cdot 2\cdot 1}{(n-r)(n-r-1)\ldots \cdot 2\cdot 1} = \frac{n!}{(n-r)!}$$

Example 29.1. *As an example, the number of ways of selecting a president, vice-president, secretary, and treasurer from a group of 20 people is $P(20,4)$ (assuming no person can hold more than one office). If you want the actual numerical value, it is $\frac{20!}{(20-4)!} = 20\cdot 19\cdot 18\cdot 17 = 116280$, but the best way to write the answer in most cases would be $P(20,4) = \frac{20!}{16!}$, and skip the numerical computations.*

Example 29.2. *How many one-to-one functions are there from a 5-set to a 7-set?*

While this question doesn't sound on the surface like a problem of permutations, it really is. Suppose the 5-set is $A = \{1,2,3,4,5\}$ and the 7-set is $B = \{a,b,c,d,e,f,g\}$. One example of a one-to-one function from A to B would be $f(1) = a, f(2) = c, f(3) = g, f(4) = b, f(5) = d$. But, if we agree to think of the elements of A listed in their natural order, we could specify that function more briefly as $acgbd$. In other words, each one-to-one function specifies a 5-permutation of B, and, conversely, each 5-permutation of B specifies a one-to-one function. So the number of one-to-one functions from a 5-set to a 7-set is equal to the number of 5-permutations of a 7-set, and that is $P(7,5) = 2520$.

The same reasoning shows there are $P(n,r)$ one-to-one functions from an r-set to an n-set.

Example 29.3. *Here are a few easily seen values of $P(n,r)$:*

(1) $P(n,n) = n!$

(2) $P(n,1) = n$

(3) $P(n,0) = 1$

(4) $P(n,r) = 0$ *if* $r > n$

When forming permutations, the order in which the elements are listed is important. But there are many cases when we are interested only in which elements are selected and we do not care about the order. For example, when playing poker, a hand consists of five cards dealt from a standard 52-card deck. The order in which the cards arrive in a hand does not matter, only the final selection of the five cards is important. When order is not important, the selection is called a combination rather than a permutation. More carefully, an ***r*-combination** from an *n*-set is an *r*-subset of the *n*-set. In other words an *r*-combination of an *n*-set is an unordered selection of *r* distinct elements from the n-set.

Example 29.4. *The* 2*-combinations of the* 5*-set* $\{a, b, c, d, e\}$ *are*

$$\{a, b\},\quad \{a, c\},\quad \{a, d\},\quad \{a, e\},\quad \{b, c\},$$
$$\{b, d\},\quad \{b, e\},\quad \{c, d\},\quad \{c, e\},\quad \{d, e\}.$$

Notation. *The number of r-combinations from an n-set is denoted by* $C(n,r)$ *or, sometimes* $\binom{n}{r}$.

Example 29.5. *Here are a few easily seen values of* $C(n,r)$*:*

(1) $C(n,n) = 1$

(2) $C(n,1) = n$

(3) $C(n,0) = 1$ *the only set with* 0 *elements is the empty set.*

(4) $C(n,r) = 0$ *if* $r > n$ *since a subset cannot contain more objects than its superset.*

There is a compact formula for $C(n,r)$ which can be derived using the product rule in a sort of back-handed way. An r-permutation of an n-set can be built using a sequence of two tasks. First, select r elements of the n-set. There are $C(n,r)$ ways to do that task. Next, arrange those r elements in some specific order. There are $r!$ ways to do that task. So, according to the product rule, the number of r-permutations of an n-set will be $C(n,r)r!$. However, we know that the number of r-permutations of an n-set is $P(n,r)$. So we may conclude that $P(n,r) = C(n,r)r!$, or, rearranging that, we see

$$C(n,r) = \binom{n}{r} = \frac{P(n,r)}{r!} = \frac{n!}{r!\,(n-r)!}.$$

Example 29.6. *Suppose we have a club with* 20 *members. If we want to select a committee of* 4 *members, then there are*

$$C(20,4) = \frac{20!}{4!\,16!} = \frac{20 \cdot 19 \cdot 18 \cdot 17}{4 \cdot 3 \cdot 2 \cdot 1} = 4845.$$

ways to do this since the order of people on the committee doesn't matter.

Compare this answer with example 29.1 where we counted the number of possible selections for president, vice-president, secretary, and treasurer from the group of 20. The difference between the two cases is that the earlier example is a question about permutations (order matters), whereas this example is a question about combinations (order does not matter).

Exercises

Exercise 29.1. *In how many ways can the* 26 *volumes (labeled* A *through* Z*) of the Encyclopedia of Pseudo-Science be placed on a shelf?*

Exercise 29.2. *In how many ways can those same* 26 *volumes be placed on a shelf if superstitions demand the volumes labeled with vowels must be adjacent? In how many ways can they be placed on the shelf obeying the conflicting superstition that volumes labeled with vowels cannot touch each other?*

Exercise 29.3. *For those same* 26 *volumes, how many ways can they be placed in a two shelf bookcase if volumes* A-M *go on the top shelf and* N-Z *go on the bottom shelf?*

Exercise 29.4. *In how many ways can seven men and four women sit in a row if the men must sit together?*

Exercise 29.5. 20 *players are to be divided into two* 10*-man teams. In how many ways can that be done?*

Exercise 29.6. *A lottery ticket consists of five different integers selected from* 1 *to* 99*. How many different lottery tickets are possible? How many tickets would you need to buy to have a one-in-a-million chance of winning by matching all five randomly selected numbers?*

Exercise 29.7. *A committee of size six is selected from a group of nine deans and thirteen professors.*

(a) *How many different committees are possible?*

(b) *How many committees are possible if there must be exactly two deans on the committee?*

(c) *How many committees are possible if professors must outnumber deans on the committee?*

Problems

Problem 29.1. *In how many ways can the ten digits be written in a row?*

Problem 29.2. *In how many ways can the ten digits be written in a row if the odd digits have to be adjacent?*

Problem 29.3. *In how many ways can the ten digits be written in a row if the even and odd digits have to alternate?*

Problem 29.4. *How many bit strings of length ten have exactly four* 0*'s?*

Problem 29.5. *How many bit strings of length ten have at most four* 0*'s?*

Problem 29.6. *How many length twenty strings of* a*'s,* b*'s, and* c*'s have ten* a*'s, six* b*'s, and four* c*'s?*

Problem 29.7. *How many bit strings of length ten have more* 0*'s than* 1*'s?*

Problem 29.8. *In how many ways can a subset of two numbers from* 1 *to* 100 *(inclusive) be selected if the selected numbers cannot be consecutive?*

Chapter 30

The Binomial Theorem and Pascal's Triangle

The quantity $C(n,k)$ is also written $\binom{n}{k}$ and called a *binomial coefficient*. It gives the number of k-subsets of an n-set, or, equivalently, it gives the number of ways of selecting k distinct items from n items.

Facts involving the binomial coefficients can be proved algebraically, using the formula $\binom{n}{k} = \frac{n!}{k!(n-k)!}$. But often the same facts can be proved much more neatly by recognizing $\binom{n}{k}$ gives the number of k-subsets of an n-set. This second sort of proof is called a combinatorial proof. Here is an example of each type of proof.

Theorem 30.1 (Pascal's Identity). *Let n and k be non-negative integers, then*

$$\binom{n+1}{k} = \binom{n}{k-1} + \binom{n}{k}.$$

Proof. (a combinatorial proof) Let S be a set with $n + 1$ elements. Select one particular element $a \in S$. There are two ways to produce a subset of S of size k. We can include a in the subset, and toss in $k - 1$ of the remaining elements of S. There are $\binom{n}{k-1}$ ways to do that. Or, we can avoid a and choose all k elements from the other n elements of S. There are $\binom{n}{k}$ ways to do that. So, according to the sum rule, there is a total of $\binom{n}{k-1} + \binom{n}{k}$ subsets of size k of S. But we know there are $\binom{n+1}{k}$ subsets of size k of S. So it must be that $\binom{n+1}{k} = \binom{n}{k-1} + \binom{n}{k}$. □

Proof. (an algebraic proof)

$$\begin{aligned}
\binom{n}{k-1} + \binom{n}{k} &= \frac{n!}{(k-1)!\,\big(n-(k-1)\big)!} + \frac{n!}{k!\,(n-k)!} \\
&= \frac{n!}{(k-1)!\,(n-k+1)!} + \frac{n!}{k!\,(n-k)!} \\
&= \frac{n!}{(k-1)!\,(n-k)!}\left(\frac{1}{n-k+1} + \frac{1}{k}\right) \\
&= \frac{n!}{(k-1)!\,(n-k)!}\left(\frac{k}{k(n-k+1)} + \frac{n-k+1}{k(n-k+1)}\right) \\
&= \frac{n!}{(k-1)!\,(n-k)!}\frac{k+(n-k)+1}{k(n-k+1)} \\
&= \frac{n!}{(k-1)!\,(n-k)!}\frac{n+1}{k(n-k+1)} \\
&= \frac{n!\,(n+1)}{(k-1)!\,k(n-k)!\,(n-k+1)} \\
&= \frac{(n+1)!}{k!\,(n+1-k)!} \\
&= \binom{n+1}{k}
\end{aligned}$$

□

The idea of a combinatorial proof is to ask a counting problem that can be answered in two different ways, and then conclude the two answers must be equal. In the proof above, we asked how many k-subsets there are of an $(n+1)$-set. We provided an argument to show two answers were correct: $\binom{n+1}{k}$ and $\binom{n}{k-1} + \binom{n}{k}$, and so we could conclude the two answers must be equal:

$$\binom{n+1}{k} = \binom{n}{k-1} + \binom{n}{k}.$$

As in the combinatorial proof of Pascal's Identity 30.1, such arguments can be much less work, far less tedious, and much more illuminating, than algebraic proofs. Unfortunately, they can also be much more difficult to discover since it is necessary to dream up a good counting problem that will have as answers the two expressions we are trying to show are equal, and there is no algorithm for coming up with such a suitable counting problem.

Example 30.2. *Give a combinatorial proof of* $\binom{n}{k} = \binom{n}{n-k}$.

Solution. *To provide a combinatorial proof, we ask how many ways are there to grab k elements of an* n*-set? One answer is of course* $\binom{n}{k}$.
But here is a second way to view the problem. We can select k *elements by deciding on* $n-k$ *elements* ***not*** *to pick. Since there are* $\binom{n}{n-k}$ *ways to select the* $n-k$ *not to pick, there must be* $\binom{n}{n-k}$ *ways to select* k *elements of an* n*-set. Since the two answers must be equal we conclude that* $\binom{n}{k} = \binom{n}{n-k}$.

□

Example 30.3. *Give a combinatorial proof of* ***Vandermonde's Identity****:*

Solution. *Consider the set* $\{a_1, a_2, \dots, a_n, b_1, b_2, \dots, b_m\}$ *of* $n+m$ *elements. We ask, how may* k*-subsets does the set have?.*

One answer of course is $\binom{n+m}{k}$.

But here's another way to answer the question:

we can select 0 *of the* a*'s and* k *of the* b*'s. There are* $\binom{n}{0}\binom{m}{k}$ *ways to do that.*

Or we select 1 *of the* a*'s and* $k-1$ *of the* b*'s. There are* $\binom{n}{1}\binom{m}{k-1}$ *ways to do that.*

Or we select 2 *of the* a*'s and* $k-2$ *of the* b*'s. There are* $\binom{n}{2}\binom{m}{k-2}$ *ways to do that.*

And so on, until we reach the option of selecting k *of the* a*'s and* 0 *of the* b*'s.*

There are $\binom{n}{k}\binom{m}{0}$ *ways to do that.*

By the sum rule, it follows that another way to count the number of k*-subsets is*

$$\binom{n}{0}\binom{m}{k} + \binom{n}{1}\binom{m}{k-1} + \binom{n}{2}\binom{m}{k-2} + \cdots + \binom{n}{k}\binom{m}{0}.$$

□

Row 0: $\binom{0}{0}$

Row 1: $\binom{1}{0}$ $\binom{1}{1}$

Row 2: $\binom{2}{0}$ $\binom{2}{1}$ $\binom{2}{2}$

Row 3: $\binom{3}{0}$ $\binom{3}{1}$ $\binom{3}{2}$ $\binom{3}{3}$

Row 4: $\binom{4}{0}$ $\binom{4}{1}$ $\binom{4}{2}$ $\binom{4}{3}$ $\binom{4}{4}$

Row 5: $\binom{5}{0}$ $\binom{5}{1}$ $\binom{5}{2}$ $\binom{5}{3}$ $\binom{5}{4}$ $\binom{5}{5}$

Figure 30.1: Pascal's Triangle

The binomial coefficients are so named because they appear when a binomial $x + y$ is raised to an integer power $k \geq 0$. To appreciate the connection, let's look at a table of the binomial coefficients $\binom{n}{k}$. The table is arranged in rows starting with row $n = 0$, and within each row, the entries are arranged from left to right for $k = 0, 1, 2, \dots, n$. The result, called Pascal's Triangle, is shown in figure 30.1.

Filling in the numerical values for the binomial coefficients gives the table shown in figure 30.2.

Note that we number the rows starting with 0. We already know quite a bit about the entries in Pascal's Triangle.

Since $\binom{n}{0} = \binom{n}{n} = 1$ for all n, each row begins and ends with a 1.

From the symmetry formula $\binom{n}{n} = \binom{n}{n-k}$, the rows read the same in both directions.

From Pascal's Identity, each entry in a row is the sum of the two entries diagonally above it. This last fact makes is easy to add new rows to Pascal's Triangle

The numbers in the first, second, and third rows of Pascal's triangle probably seem familiar. In fact, we see that

Row 0:	1
Row 1:	1 1
Row 2:	1 2 1
Row 3:	1 3 3 1
Row 4:	1 4 6 4 1
Row 5:	1 5 10 10 5 1

Figure 30.2: Pascal's Triangle (numeric)

$$(x+y)^0 = 1$$

$$(x+y)^1 = x + y = 1 \cdot x + 1 \cdot y$$

$$(x+y)^2 = x^2 + 2xy + y^2 = 1 \cdot x^2 + 2 \cdot xy + 1 \cdot y^2$$

$$(x+y)^3 = x^3 + 3x^2y + 3xy^2 + y^3 = 1 \cdot x^3 + 3 \cdot x^2y + 3 \cdot xy^2 + 1 \cdot y^3$$

The coefficients in these binomial expansions are exactly the entries in the corresponding rows of Pascal's Triangle. This even works for the 0^{th} row: $(x+y)^0 = 1$.

The fact that the coefficients in the expansion of the binomial $(x+y)^n$ (where $n \geq 0$ is an integer) can be read off from the n^{th} row of Pascal's Triangle is called the Binomial Theorem. We will give two proofs of this theorem, one by induction, and the other a combinatorial proof.

Theorem 30.4 (The Binomial Theorem)**.** *When n is a non-negative integer and $x, y \in \mathbb{R}$*

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

Proof: We proceed by induction on n. When $n = 0$ the result is clear. So, suppose that for some $m \geq 0$ we have

$$(x+y)^m = \sum_{k=0}^{m} \binom{m}{k} x^k y^{m-k} \text{ , for any } x, y \in \mathbb{R}.$$

Then $(x+y)^{m+1} = (x+y)^m(x+y)$, by recursive definition of integral exponents.

$$= \left(\sum_{k=0}^{m} \binom{m}{k} x^k y^{m-k}\right)(x+y), \text{ by inductive hypothesis.}$$

$$= \left[\sum_{k=0}^{m} \binom{m}{k} x^{k+1} y^{m-k}\right] + \left[\sum_{k=0}^{m} \binom{m}{k} x^k y^{m+1-k}\right], \text{ by distribution}$$

$$= \binom{m}{m} x^{m+1} + \left[\sum_{k=0}^{m-1} \binom{m}{k} x^{k+1} y^{m-k}\right] + \left[\sum_{k=1}^{m} \binom{m}{k} x^k y^{m+1-k}\right] + \binom{m}{0} y^{m+1}$$

$$= \binom{m}{m} x^{m+1} + \left[\sum_{l=1}^{m} \binom{m}{l-1} x^l y^{m-(l-1)}\right] + \left[\sum_{k=1}^{m} \binom{m}{k} x^k y^{m+1-k}\right] + \binom{m}{0} y^{m+1}$$

$$= \binom{m}{m} x^{m+1} + \left[\sum_{l=1}^{m} \binom{m}{l-1} x^l y^{m+1-l}\right] + \left[\sum_{k=1}^{m} \binom{m}{k} x^k y^{m+1-k}\right] + \binom{m}{0} y^{m+1}$$

$$= \binom{m}{m} x^{m+1} + \left[\sum_{k=1}^{m} \left\{\binom{m}{k-1} + \binom{m}{k}\right\} x^k y^{m+1-k}\right] + \binom{m}{0} y^{m+1}$$

$$= \binom{m+1}{m+1} x^{m+1} + \left[\sum_{k=1}^{m} \binom{m+1}{k} x^k y^{m+1-k}\right] + \binom{m+1}{0} y^{m+1}$$

$$= \sum_{k=0}^{m+1} \binom{m+1}{k} x^k y^{m+1-k}.$$

□

Now, let's look at a combinatorial proof of the Binomial Theorem.

Proof. When the binomial $(x+y)^n = (x+y)(x+y)(x+y) \cdots (x+y)$ is expanded, the terms are produced by selecting either the x or the y from each of the n factors $x+y$ appearing on the left-hand side of the equation. The number of ways of selecting exactly k x's from the n

r�

available is $\binom{n}{k}$ and so that will be the coefficient of the term $x^k y^{n-k}$ in the expansion. □

Example 30.5. *The coefficient of* x^7y^3 *in the expansion of* $(x+y)^{10}$ *is*

$$\binom{10}{7} = \frac{10!}{7!\,3!} = \frac{10\cdot 9\cdot 8}{3\cdot 2\cdot 1} = 120.$$

Example 30.6. *The coefficient of* x^7y^3 *in the expansion of* $(2x-3y)^{10}$ *is*

$$\binom{10}{7}2^7(-3)^3 = \frac{10!}{7!\,3!}2^7(-3)^3 = -\frac{10\cdot 9\cdot 8}{3\cdot 2\cdot 1}2^7 3^3 = -120\cdot 128\cdot 27 = -414720.$$

From the binomial theorem we can derive facts such as

Theorem 30.7. *A finite set with* n *elements has* 2^n *subsets.*

Proof. By the sum rule the number of subsets of an *n*-set is

$$\sum_{k=0}^{n}\binom{n}{k} = \sum_{k=0}^{n}\binom{n}{k}1^k 1^{n-k}.$$

By the Binomial Theorem $\sum_{k=0}^{n}\binom{n}{k}1^k1^{n-k} = (1+1)^n = 2^n$.

□

Theorem 30.8. *If* $n \geq 1$, *then*

$$\binom{n}{0} - \binom{n}{1} + \binom{n}{2} - \binom{n}{3} + \cdots + (-1)^n\binom{n}{n} = 0.$$

Proof. By the Binomial Theorem,

$$\binom{n}{0} - \binom{n}{1} + \binom{n}{2} - \binom{n}{3} + \cdots + (-1)^n\binom{n}{n} = \left(1+(-1)\right)^n = 0^n = 0.$$

□

Exercises

Exercise 30.1. *Determine the sixth row of Pascal's Triangle.*

Exercise 30.2. *Determine the coefficient of* x^3y^7 *in the expansion of* $(3x - 2y)^{10}$.

Exercise 30.3. *Give an algebraic proof that*

$$\binom{2n}{2} = 2\binom{n}{2} + n^2.$$

Exercise 30.4. *Give an algebraic proof that*

$$\binom{r}{s}\binom{s}{t} = \binom{r}{t}\binom{r-t}{s-t}.$$

Exercise 30.5. *Give a combinatorial proof that*

$$\binom{r}{s}\binom{s}{t} = \binom{r}{t}\binom{r-t}{s-t}.$$

Exercise 30.6. *Using the same reasoning as in the combinatorial proof of the Binomial Theorem, determine the coefficient of* $x^4y^5z^6$ *in the expansion of* $(x + y + z)^{15}$.

Exercise 30.7. *Show that if* p *is a prime and* $0 < k < p$, *then* p *divides* $\binom{p}{k}$

Hint: When $\binom{p}{k}$ *is written out, how many times does* p *occur as a factor of the numerator and how many times as a factor of the denominator?*

Problems

Problem 30.1. *Determine the seventh row of Pascal's Triangle.*

Problem 30.2. *Determine the coefficient of x^5y^2 in the expansion of $(2x-5y)^7$.*

Problem 30.3. *Give a combinatorial proof that $\binom{2n}{2} = 2\binom{n}{2} + n^2$.*

Hint: How many ways are there to select a 2-subset of $\{a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_n\}$?

Problem 30.4. *Give an algebraic proof that*

$$\binom{n}{2}\binom{n-2}{m-2} = \binom{n}{m}\binom{m}{2}.$$

Problem 30.5. *Give a combinatorial proof that*

$$\binom{n}{2}\binom{n-2}{m-2} = \binom{n}{m}\binom{m}{2}.$$

Problem 30.6. *Give an algebraic proof that*

$$\binom{n}{m} = n\binom{n-1}{m-1}.$$

Problem 30.7. *Give a combinatorial proof that*

$$\binom{n}{m} = n\binom{n-1}{m-1}.$$

Problem 30.8. *Give a combinatorial proof that*

$$\binom{3m}{3} = \binom{m}{3} + 2m\binom{m}{2} + m\binom{2m}{2} + \binom{2m}{3}.$$

Problem 30.9. *Determine the coefficient of $x^3y^2z^3$ in the expansion of $(x+2y-3z)^8$.*

Chapter 31

Inclusion-Exclusion Counting

The sum rule says that **if A and B are disjoint sets, then** $|A \cup B| = |A| + |B|$. If the sets are not disjoint, then this formula over counts the number of elements in the union of A and B. For example, if $A = \{a, b, c\}$ and $B = \{c, d, e\}$, then

$$|A \cup B| = |\{a, b, c\} \cup \{c, d, e\}| = |\{a, b, c, d, e\}| = 5.$$

So, we see that $|A \cup B| \neq 3 + 3 = |A| + |B|$.

The correct way to count the number of elements in $|A \cup B|$ when A and B might not be disjoint is via the **inclusion-exclusion** formula. To derive this formula, notice that $A \cup B = (A - B) \cup B$, and that the sets $A - B$ and B are disjoint. So we can apply the sum rule to conclude

$$|A \cup B| = |(A - B) \cup B| = |A - B| + |B|.$$

Next, notice that $A = (A - B) \cup (A \cap B)$, and the two sets on the right are disjoint. So, using the sum rule, we get

$$|A| = |(A - B) \cup (A \cap B)| = |A - B| + |A \cap B|,$$

which we can rearrange as

$$|A - B| = |A| - |A \cap B|.$$

So, replacing $|A - B|$ by $|A| - |A \cap B|$ in the formula $|A \cup B| = |(A - B) \cup B| = |A - B| + |B|$, we

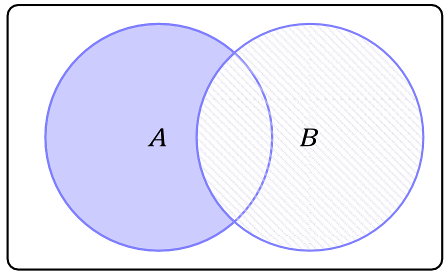


Figure 31.1: $A \cup B = (A - B) \cup B$

end up with the **inclusion-exclusion** formula:

$$|A \cup B| = |A| + |B| - |A \cap B|.$$

In words, to count the number of items in the union of two sets, include one for everything in the first set, and include one for everything in the second set, then exclude one for each element in the overlap of the two sets (since those elements will have been counted twice).

Example 31.1. *How many students are there in a discrete math class if 15 students are computer science majors, 7 are math majors, and 3 are double majors in math and computer science?*

Solution. *Let C denote the subset of computer science majors in the class, and M denote the math majors. Then $|C| = 15$, $|M| = 7$ and $|C \cap M| = 3$. So by the principle of inclusion-exclusion there are $|C| + |M| - |C \cap M| = 15 + 7 - 3 = 19$ students in the class.* □

Example 31.2. *How many integers between 1 and 1000 are divisible by either 7 or 11?*

Solution. *Let S denote the set of integers between 1 and 1000 divisible by 7, and E denote the set of integers between 1 and 1000 divisible by 11. We need to count the number of integers in $S \cup E$. By the principle of inclusion-exclusion, we have*

$$|S \cup E| = |S| + |E| - |S \cap E| = \left\lfloor \frac{1000}{7} \right\rfloor + \left\lfloor \frac{1000}{11} \right\rfloor - \left\lfloor \frac{1000}{77} \right\rfloor$$

$$= 142 + 90 - 12 = 220.$$

□

The inclusion-exclusion principle can be extended to the problem of counting the number of elements in the union of three sets. The trick is the think of the union of three sets as the union of two sets.

It goes as follows:

$$
\begin{aligned}
|A \cup B \cup C| &= |(A \cup B) \cup C| \\
&= |A \cup B| + |C| - |(A \cup B) \cap C| \\
&= |A| + |B| + |C| - |A \cap B| - |(A \cup B) \cap C| \\
&= |A| + |B| + |C| - |A \cap B| - |(A \cap C) \cup (B \cap C)| \\
&= |A| + |B| + |C| - |A \cap B| - |(A \cap C) \cup (B \cap C)| \\
&= |A| + |B| + |C| - |A \cap B| - (|A \cap C| + |B \cap C| - |(A \cap C) \cap (B \cap C)|) \\
&= |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|.
\end{aligned}
$$

This might more appropriately be named the inclusion-exclusion-inclusion formula, but nobody calls it that. In words, the formula says that to count the number of elements in the union of three sets, first, include everything in each set, then exclude everything in the overlap of each pair of sets, and finally, re-include everything in the overlap of all three sets.

Example 31.3. *How many integers between* 1 *and* 1000 *are divisible by at least one of* 7, 9, *and* 11?

Solution. *Let* S *denote the set of integers between* 1 *and* 1000 *divisible by* 7, *let* N *denote the set of integers between* 1 *and* 1000 *divisible by* 9, *and* E *denote the set of integers between* 1 *and* 1000 *divisible by* 11. *We need to count the number of integers in* $S \cup N \cup E$. *By the principle of inclusion-exclusion,*

$$
\begin{aligned}
|S \cup N \cup E| &= |S| + |N| + |E| - |S \cap N| - |S \cap E| - |N \cap E| + |S \cap N \cap E| \\
&= \left\lfloor \frac{1000}{7} \right\rfloor + \left\lfloor \frac{1000}{9} \right\rfloor + \left\lfloor \frac{1000}{11} \right\rfloor - \left\lfloor \frac{1000}{63} \right\rfloor - \left\lfloor \frac{1000}{77} \right\rfloor - \left\lfloor \frac{1000}{99} \right\rfloor + \left\lfloor \frac{1000}{693} \right\rfloor \\
&= 142 + 111 + 90 - 15 - 12 - 10 + 1 = 307.
\end{aligned}
$$

There are similar inclusion-exclusion formulas for the union of four, five, six, $\cdots$ sets. The formulas can be proved by induction with the inductive step using the trick we used above to go from two sets to three. However, there is a much neater way to prove the formula based on the Binomial Theorem.

Theorem 31.4. *Given finite sets* $A_1, A_2, \ldots, A_n$

$$
\left|\bigcup_{k=1}^{n} A_k\right| = \sum_{k=1}^{n} |A_k| - \sum_{1 \le j < k \le n} |A_j \cap A_k| + \cdots + (-1)^{n-1} \left|\bigcap_{k=1}^{n} A_k\right|.
$$

Proof. Suppose $x \in \bigcup_{k=1}^{n} A_k$. We need to show that x is counted exactly once by the right-hand side of the promised formula. Say $x \in A_i$ for exactly p of the sets A_i, where $1 \leq p \leq n$.

The key to the proof is being able to count the number of intersections in each summation on the right-hand side of the offered formula that contain *x* since we will account for *x* once for each such term. The number of such terms in the first sum is $\binom{p}{1}$, the number in the second sum is $\binom{p}{2}$, and , in general the number of terms in the jth sum will be $\binom{p}{j}$ provided $j \leq p$. If $j > p$ this is still true since x will not be in any of the intersections of the j sets and $\binom{p}{j} = 0$.

So, the total number of times x is accounted for on the right-hand side is

$$\binom{p}{1} - \binom{p}{2} + \binom{p}{3} - \cdots + (-1)^{p-1}\binom{p}{p}$$
$$= 1 - \left(\binom{p}{0} - \binom{p}{1} + \binom{p}{2} - \binom{p}{3} + \cdots + (-1)^p\binom{p}{p}\right)$$
$$= 1 - \left(1 + (-1)\right)^p = 1 - 0 = 1.$$

Just as we hoped. □

Example 31.5. *How many students are in a calculus class if* 14 *are math majors,* 22 *are computer science majors,* 15 *are engineering majors, and* 13 *are chemistry majors, if* 5 *students are double majoring in math and computer science,* 3 *students are double majoring in chemistry and engineering,* 10 *are double majoring in computer science and engineering,* 4 *are double majoring in chemistry and computer science, none are double majoring in math and engineering and none are double majoring in math and chemistry, and no student has more than two majors?*

Solution. *Let A_1 denote the math majors, A_2 denote the computer science majors, A_3 denote the engineering majors, and A_4 the chemistry majors. Then the information given is*

$$|A_1| = 14, \quad |A_2| = 22, \quad |A_3| = 15, \quad |A_4| = 13,$$
$$|A_1 \cap A_2| = 5, \quad |A_1 \cap A_3| = 0, \quad |A_1 \cap A_4| = 0,$$
$$|A_{22} \cap A_3| = 10, \quad |A_2 \cap A_4| = 4, \quad |A_3 \cap A_4| = 3,$$
$$|A_1 \cap A_2 \cap A_3| = 0 = |A_1 \cap A_2 \cap A_4| = |A_1 \cap A_3 \cap A_4| = |A_2 \cap A_3 \cap A_4|,$$

and

$$|A_1 \cap A_2 \cap A_3 \cap A_4| = 0.$$

So, by inclusion-exclusion, the number of students in the class is

$$14 + 22 + 15 + 13 - 5 - 10 - 4 - 3 = 42.$$

Example 31.6. *How many ternary strings (using* 0*'s,* 1*'s and* 2*'s) of length* 8 *either start with a* 1*, end with two* 0*'s or have* 4*th and* 5*th positions* 12*, respectively?*

Solution. *Let* A_1 *denote the set of ternary strings of length* 8 *which start with a* 1*,* A_2 *denote the set of ternary strings of length* 8 *which end with two* 0*'s, and* A_3 *denote the set of ternary strings of length* 8 *which have* 4*th and* 5*th positions* 12*. By inclusion-exclusion, the answer is*

$$3^7 + 3^6 + 3^6 - 3^5 - 3^5 - 3^4 + 3^3.$$

The inclusion-exclusion formula is often used along with the *Good=Total-Bad* trick.

Example 31.7. *How many integers between* 1 *and* 1000 *are divisible by none of* 7, 9, *and* 11*?*

Solution. *There are* 1000 *numbers between* 1 *and* 1000 *(assuming* 1 *and* 1000 *are included). As counted before, there are* 307 *of those that are divisible by at least one of* 7, 9, *and* 11*. That means there are* 1000 – 307 = 693 *that are divisible by none of* 7, 9, *or* 11.

Exercises

Exercise 31.1. *At a certain college no student is allowed more than two majors. How many students are in the college if there are* 70 *math majors,* 160 *chemistry majors,* 230 *biology majors,* 56 *geology majors,* 24 *physics majors,* 35 *anthropology majors,* 12 *double math-physics majors,* 10 *double math-chemistry majors,* 4 *double biology-math majors,* 53 *double biology-chemistry majors,* 5 *double biology-anthropology majors, and no other double majors?*

Exercise 31.2. *How many bit strings of length 15 start with the string* 1111*, end with the string* 1000 *or have* 4^{th} *through* 7^{th} *bits* 1010*?*

Exercise 31.3. *How many positive integers between* 1000 *and* 9999 *inclusive are divisible by any of* 4, 10 *or* 25 *(careful!)?*

Exercise 31.4. *How many permutations of the digits* 1, 2, 3, 4, 5*, have at least one digit in its own spot? In other words, a* 1 *in the first spot, or a* 2 *in the second, etc. For example,* 35241 *is OK since it has a* 4 *in the fourth spot, and* 14235 *is OK, since it has a* 1 *in the first spot (and also a* 5 *in the fifth spot). But* 31452 *is no good. Hint: Let* A_1 *be the set of permutations that have* 1 *in the first spot, let* A_2 *be the set of permutations that* 2 *in the second spot, and so on.*

Exercise 31.5. *How many permutations of the digits* 1, 2, 3, 4, 5 *have no digit in its own spot?*

Problems

Problem 31.1. *The membership of a language club consists of seven people who speak only English, eight speak only French, five speak only Spanish, seven speak only English and Spanish, two speak only French and Spanish, there are none who speak only English and French, and there are four who speak all three languages. How many members are in the club?*

Problem 31.2. *How many integers between* 1 *and* 10000 *(inclusive) are divisible by at least one of* 9, 10, *or* 11*?*

Problem 31.3. *Suppose* p, q *are two different primes. How many integers between* 1 *and the product* pq *are relatively prime to* pq *(or, same thing, how many are divisible by neither of* p *and* q*)? (The correct answer will factor neatly.)*

Problem 31.4. *Suppose* p, q, r *are three different primes. How many integers between* 1 *and the product* pqr *are relatively prime to* pqr *(or, same thing, how many are divisible by none of* p, q *and* r*)? (The correct answer will factor neatly.)*

Problem 31.5. *Suppose* p, q, r, s *are four different primes. How many integers between* 1 *and the product* $pqrs$ *are relatively prime to* $pqrs$ *(or, same thing, how many are divisible by none of* p, q, r *and* s*)? (The correct answer will factor neatly.)*

Problem 31.6. *Based on the results of the previous three problems, can you guess the neat formula for five, six, seven, and so on, different primes?*

Problem 31.7. *Of the words of length ten using the alphabet* $\Sigma = \{a, b, c\}$, *how many either begin* abc *or end* cba *or have* $cccccc$ *as the middle six letters?*

Problem 31.8. *There are* $6!$ *permutations of the numbers* $1, 2, 3, 4, 5, 6$. *In some of these there is a run of three (or more) consecutive numbers that increase (left to right) such as* 514632 *which has the increasing run* 146. *Others do not have any increasing runs of length three such as (a cheap example)* 654321 *and (not quite as cheap)* 615243. *How many of the* $6!$ *permutations contain no increasing runs of length three (or more)? (Hint: runs of length three can start with the first, second, third, or fourth spot in the permutation.)*

Chapter 32

The Pigeonhole Principle

The pigeonhole principle, like the sum and product rules, is another one of those absolutely obvious counting facts. The statement is simple: If $n + 1$ objects are divided into n piles (some piles can be empty), then at least one pile must have two or more objects in it. Or, more colorfully, if $n + 1$ pigeons land in n pigeonholes, then at least one pigeonhole has two or more pigeons. What could be more obvious? The pigeonhole principle is used to show that no matter how a certain task is carried out, some specific result must always happen.

As a simple example, suppose we have a drawer containing ten identical black socks and ten identical white socks. How many socks do we need to select to be sure we have a matching pair? The answer is three. Think of the pigeonholes as the colors black and white, and as each sock is selected put it in the pigeonhole of its color. After we have placed the third sock, one of the two pigeonholes must have at least two socks in it, and we will have a matching pair. Of course, we may have been lucky and had a pair after picking the second sock, but the pigeonhole principle *guarantees* that with the third sock we will have a pair.

As another example, suppose license plates are made consisting of four digits followed by two letters. Are there enough license plates for a state with seven million cars? No, since there are only $10^4 \cdot 26^2 = 6760000$ possible license plates, and so, by the pigeonhole principle, at least two of the seven million plates assigned would have to be the same.

A slightly fancier version of the pigeonhole principle says that if N objects are distributed in k

piles, then there must be a least one pile with $\lceil N/k \rceil$ objects in it.

That formula looks impressive, but actually is easy to understand. For example, if there are 52 people in a room, we can be absolutely certain that there are at least eight born on the same day of the week. Think of it this way: with 49 people, it would be possible to have seven born on each of the seven days of the week. But when the 50^{th} one is reached, it must boost one day up to an eighth person. That is really about all there is to it. The general proof of the fancy pigeonhole principle uses this same sort of reasoning. It is a proof by contradiction, and goes as follows:

Theorem 32.1 (Pigeonhole Principle)**.** *If N objects are distributed in k piles, then there must be at least one pile with* $\lceil N/k \rceil$ *objects in it.*

Proof. Suppose we have N objects distributed in k piles, and suppose that every pile has fewer than $\lceil N/k \rceil$ objects in it. That means that each of the piles each contain $\lceil N/k \rceil - 1$ or fewer objects. We will use the fact the $\lceil N/k \rceil < \left(\frac{N}{k}\right) + 1$ to complete the proof. The total number of objects will be at most $k\left(\left\lceil\frac{N}{k}\right\rceil + 1\right) < k\left(\left(\frac{N}{k} + 1\right) - 1\right) = N$. That is a contradiction since we know there is a total of N objects in the k piles.

□

Even though the pigeonhole principle sounds very simple, clever applications of it can produce totally unexpected results.

Example 32.2. *Five misanthropes move to a perfectly square deserted island that measures two kilometers on a side. Of course, being misanthropes, they want to live as far from each other as possible. Show that, no matter where they build on the island, some two will be no more than* $\sqrt{2}$ *kilometers of each other.*

Solution. *Divide the island into four one kilometer by one kilometer squares by drawing lines joining the midpoints of opposite sides. Since there are five people and four squares, the pigeonhole principle guarantees there will be two people living in one of those four squares. But people in one of those squares cannot be further apart than the length of the diagonal of the square which is, according to Pythagoras,* $\sqrt{2}$*.* □

Example 32.3. *For any positive integer* n*, there is a positive multiple of* n *made up of a number of* 1*'s followed by a number of* 0*'s. For example, for* $n = 1084$*, we see* $1084 \cdot 1025 = 1111100$*.*

Solution. *Consider the* $n+1$ *integers* $1, 11, 111, \cdots, 11\cdots1$*, where the last one consists of* 1 *repeated* $n+1$ *times. Some two of these must be the same modulo* n*, and so* n *will divide the difference of some two of them. But the difference of two of those numbers is of the required type.* □

Example 32.4. *Bill has 20 days to prepare his tiddlywinks title defense. He has decided to practice at least one hour every day. But, to avoid burn-out, he will not practice more than a total of 30 hours. Show there is a sequence of consecutive days during which he practices exactly 9 hours.*

Solution. *For $j = 1, 2, \cdots 20$, let t_j = the total number of hours Bill practices up to and including day j. Since he practices at least one hour every day, and the total number of hours is no more than 30, we see*

$$0 < t_1 < t_2 < \cdots < t_{20} \leq 30.$$

Adding 9 to each term we get

$$9 < t_1 + 9 < t_2 + 9 < \cdots < t_{20} + 9 \leq 39.$$

So we have 40 integers $t_1, t_2, \ldots, t_{20}, t_1 + 9, \cdots t_{20} + 9$, all between 1 and 39. By the pigeonhole principle, some two must be equal, and the only way that can happen is for $t_i = t_j + 9$ for some i and j. It follows that $t_i - t_j = 9$, and since the difference $t_i - t_j$ is the the total number of hours Bill practiced from day $j + 1$ to day i, that shows there is a sequence of consecutive days during which he practiced exactly 9 hours. □

Exercises

Exercise 32.1. *Show that in any group of eight people, at least two were born on the same day of the week.*

Exercise 32.2. *Show that in any group of* 100 *people, at least* 15 *were born on the same day of the week.*

Exercise 32.3. *How many cards must be selected from a deck to be sure that at least six of the selected cards have the same suit?*

Exercise 32.4. *Show that in any set of* n *integers, where* $n \geq 2$*, there must be a pair with a difference that is a multiple of* $n - 1$.

Exercise 32.5. *Al has* 75 *days to master discrete mathematics. He decides to study at least one hour every day, but no more than a total of* 125 *hours. Show there must be a sequence of consecutive days during which he studies exactly* 24 *hours.*

Exercise 32.6. *Show that in any set of* 217 *integers, there must be a pair with a difference that is a multiple of* 216.

Problems

Problem 32.1. *Show that in a town with population* 18,000*, there must be at least two people with the same three initials.*

Problem 32.2. *What is the smallest town population that will guarantee there will be at least two people with the same three initials?*

Problem 32.3. *What is the smallest town population that will guarantee there will be at least five people with the same three initials?*

Problem 32.4. *How many cards have to be selected from a* 52 *card deck to be sure there will be two cards of the same suit?*

Problem 32.5. *How many cards have to be selected from a* 52 *card deck to be sure there will be two cards of the same rank?*

Problem 32.6. *Five misanthropes buy a six mile by eight mile rectangular plot in the arctic. Show that no matter where they build their houses, there will be at least two people that are no more than five miles apart. (You can assume the ice sheet they buy is perfectly flat.)*

Problem 32.7. *In any list of* n *integers, there will be a chunk of consecutive entries from the list that add up to a multiple of* n. *For example: in the list* $-8, 4, 22, -11, 7$, *we have* $4 + 22 - 11 = 15$ *is a multiple of* 5.

Problem 32.8. *Suppose* $a_1, a_2, \ldots, a_{99}$ *is a permutation of* $1, 2, \ldots, 99$. *Show that the product*

$$(a_1 + 1)(a_2 + 2)(a_3 + 3)\ldots(a_{99} + 99)$$

is even.

Problem 32.9. *In a rematch, Bill has* 30 *days to train for a new defense of his tiddlywinks title. He plans to practice at least one hour every day, but no more than* 45 *hours total. Show there is a sequence of consecutive days during which he practices exactly* 14 *hours.*

Chapter 33

Tougher Counting Problems

All of the counting exercises you've been asked to complete so far have not been realistic. In general it won't be true that a counting problem fits neatly into a section. So we need to work on the bigger picture.

When we start any counting exercise it is true that there is an underlying exercise at the basic level that we want to consider first. So instead of answering the question immediately we might first want to decide on what type of exercise we have. So far we have seen three types which are distinguishable by the answers to two questions.

(1) In forming the objects we want to count, is repetition allowed?

(2) In forming the objects we want to count, does the order of selection matter?

The three scenarios we have seen so far are described in table 33.1.

There are two problems to address. First of all, table 33.1 is incomplete. What about, for example, counting objects where repetition is allowed, but order doesn't matter. Second of all, there are connections among the types which make some solutions appear misleading. But as a general rule of thumb, if we correctly identify the type of problem we are working on, then all we have to do is use the principles of addition, multiplication, inclusion/exclusion, or exclusion to decompose our problem into subproblems. The solutions to the subproblems often have the same form as the

Order	Repetition	Type	Form
Y	Y	r-strings	n^r
Y	N	r-permutations	$P(n,r)$
N	N	r-combinations	$\binom{n}{r}$

Table 33.1 Basic Counting Problems

underlying problem. The principles we employed direct us on how the sub-solutions should be recombined to give the final answer.

Example 33.1. *As an example of the second problem, if we ask how many binary strings of length* 10 *contain exactly three* $1'$*s, then the underlying problem is an* r*-string problem. But in this case the answer is* $\binom{10}{3}$*. Of course this is really* $\binom{10}{3}1^3 1^7$ *from the binomial theorem. In this case the part of the answer that looks like* n^r *is suppressed since it's trivial. To see the difference we might ask how many ternary strings of length* 10 *contain exactly three* $1'$*s. Now the answer is* $\binom{10}{3}1^3 2^7$*, since we choose the three positions for the* $1'$*s, and then fill in each of the* 7 *remaining positions with a* 0 *or a* 2*.*

To begin to address the first problem we introduce the basic donut shop problem: If you get to the donut shop before the cops get there, you will find that they have a nice variety of donuts. You might want to order several dozen. They will put your order in a box. You don't particularly care what order the donuts are put into the box. You do usually want more than one of several types. The number of ways for you to complete your order is therefore a counting problem where order doesn't matter, and repetition is allowed.

In order to answer the question of how many ways you can complete your order, we first recast the problem mathematically. From among n types of objects we want to select r objects. If x_i denotes the number of objects of the ith type selected, we have $0 \le x_i$, (since we cannot choose a negative number of chocolate donuts), also $x_i \in \mathbb{Z}$, (since we cannot select fractional parts of donuts). So, the different ways to order are in one-to-one correspondence with the solutions to

$$x_1 + x_2 + \ldots + x_n = r, \text{ with } x_i \ge 0,\ x_i \in \mathbb{Z}, \text{ for } i = 1, 2, \ldots, n.$$

Next, in order to compute the number of solutions in non-negative integers to $x_1 + \ldots + x_n = r$, we model each solution as a string (possibly empty) of x_1 $1'$s followed by a +, then a string of x_2 $1'$s followed by a +, ... then a string of x_{n-1} $1'$s followed by a +, then a string of x_n $1'$s. So, for example, if $x_1 = 2, x_2 = 0, x_3 = 1, x_4 = 3$ is a solution to $x_1 + x_2 + x_3 + x_4 = 6$ the string we get is 11 + +1 + 111.

Finally, we see that the total number of solutions in non-negative integers to $x_1 + \ldots + x_n = r$, is the number of binary strings of length $r + n - 1$ with exactly r 1's and $(n - 1)$ +'s. From the remark above, the number of ways to select r donuts from n different types is

$$\binom{n+r-1}{r}.$$

The basic donut shop problem is not very realistic in two ways. First it is common that some of your order will be determined by other people. You might for example canvas the people in your office before you go to see if there is anything you can pick up for them. So whereas you want to order r donuts, you might have been asked to pick up a certain number of various types.

Now suppose that we know that we want to select r donuts from among n types so that at least $a_i (a_i \geq 0)$ donuts of type i are selected. In terms of our equation we have $x_1 + x_2 + \cdots + x_n = r$, where $a_i \leq x_i$, and $x_i \in \mathbb{Z}$. Set $y_i = x_i - a_i$ for $i = 1,2,\ldots,n$, and $a = a_1 + a_2 + \cdots + a_n$, so that $0 \leq y_i, y_i \in \mathbb{Z}$ and

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} x_i - a_i = \left[\sum_{i=1}^{n} x_i\right] - \left[\sum_{i=1}^{n} a_i\right] = r - a.$$

So, the number of ways to complete our order is $C(n + (r - a) - 1, r - a)$.

Still, we qualified the donut shop problem by supposing that we arrived before the cops did.

If we arrive at the donut shop after canvassing our friends, we want to select r donuts from among n types. The problem is that if the cops have been there, there are probably only a few donuts left of each type. This may place an upper limit on how often we can select a particular type. So now we wish to count solutions to

$$x_1 + x_2 + \cdots + x_n = r \text{ with } a_i \leq x_i \leq b_i, x_i \in \mathbb{Z}.$$

We proceed by replacing r with $s = r - a$, where a is the sum of lower bounds. We also replace b_i with $c_i = b_i - a_i$ for $i = 1,2,\ldots,n$. So, we want to find the number of solutions with $0 \leq y_i \leq c_i$, $y_i \in \mathbb{Z}$ and $y_1 + y_2 + \cdots + y_n = s$.

There are several ways to proceed. First, we choose inclusion/exclusion. Let $\mathcal{U}$ be the set of all solutions in nonnegative integers to $y_1 + y_2 + \cdots + y_n = s$. Next let A_i denote those solutions in nonnegative integers to $y_1 + y_2 + \cdots + y_n = s$ where $c_i < y_i$. Then we want to compute $|\overline{A_1} \cap \overline{A_2} \cap \ldots \cap \overline{A_n}|$. This we can do by general inclusion/exclusion, and the ideas from the more realistic donut shop problem.

Example 33.2. *Let us count the number of solutions to*

$$x_1 + x_2 + x_3 + x_4 = 34$$

where $0 \le x_1 \le 4$, $0 \le x_2 \le 5$, $0 \le x_3 \le 8$, *and* $0 \le x_4 \le 40$.

As discussed above we have $c_1 = 4, c_2 = 5, c_3 = 8$, *and* $c_4 = 40$. *Hence we see that*

$$|\mathcal{U}| = \binom{34+4-1}{34}.$$

Now A_i *will denote the solutions in nonnegative integers to*

$$x_1 + x_2 + x_3 + x_4 = 34, with\ x_i > c_i, i = 1,2,3,4.$$

Next realize that $A_4 = \emptyset$, *so* $\overline{A_4} = \mathcal{U}$ *and* $\overline{A_1} \cap \overline{A_2} \cap \overline{A_3} \cap \overline{A_4} = \overline{A_1} \cap \overline{A_2} \cap \overline{A_3}$.

Now to compute $|A_1|$, *we must first rephrase* $x_1 > 4$ *as a non-strict inequality, i.e.* $5 \le x_1$. *So it follows that*

$$|A_1| = \binom{29+4-1}{29}.$$

Similarly,

$$|A_2| = \binom{28+4-1}{28}\ and\ |A_3| = \binom{25+4-1}{25}.$$

Next, we observe that $A_1 \cap A_2$ *represents all solutions in nonnegative integers to*

$$x_1 + x_2 + x_3 + x_4 = 34\ with\ 5 \le x_1\ and\ 6 \le x\_2.$$

So we have that

$$|A_1 \cap A_2| = \binom{23+4-1}{23}.$$

Also, we find that

$$|A_1 \cap A_3| = \binom{20+4-1}{20} \text{ and } |A_2 \cap A_3| = \binom{19+4-1}{19}.$$

Finally, we see that

$$|A_1 \cap A_2 \cap A_3| = \binom{14+4-1}{14}.$$

Hence, the final answer is

$$|\overline{A_1} \cap \overline{A_2} \cap \overline{A_3} \cap \overline{A_4}| = \binom{34+4-1}{34} - \binom{29+4-1}{29} - \binom{28+4-1}{28} - \binom{25+4-1}{25} + \binom{23+4-1}{23} + \binom{20+4-1}{20} + \binom{19+4-1}{19} - \binom{14+4-1}{14}.$$

We leave the answer in this form for clarity. The numerical value is not illuminating.

We can now solve general counting exercises where order is unimportant and repetition is restricted somewhere between no repetition, and full repetition.

To complete the picture we should be able to also solve counting exercises where order is important and repetition is partial. This is somewhat easier. It suffices to consider the sub-cases in example 33.3.

Example 33.3. *Let us take as initial problem the number of quaternary strings of length* 15*. There are* 4^{15} *of these.*
Now, if we ask how many contain exactly two 0*'s, the answer is*

$$\binom{15}{2} 3^{13}.$$

If we ask how many contain exactly two 0*'s and four* 1*'s, the answer is*

$$\binom{15}{2}\binom{13}{4} 2^9.$$

And, if we ask how many contain exactly two 0*'s, four* 1*'s and five* 2*'s, the answer is*

$$\binom{15}{2}\binom{13}{4}\binom{9}{5}\binom{4}{4} = \frac{15!}{2!\,4!\,5!\,4!}.$$

Order	Repetition	Form
Yes	Yes	n^r
Yes	No	$P(n,r)$
No	Yes	$C(r,r+n-1)$
No	No	$C(n,r)$
Yes	some	$C(r;k_1,k_2,\ldots,k_n)$
No	some	$C(r,r+n-1)$, w/ I/E

Table 33.2: Six counting problems

So, in fact many types of counting are related by what we call the multinomial theorem.

Theorem 33.4. *When r is a nonnegative integer and $x_1, x_2, \ldots, x_n \in \mathbb{R}$, then*

$$(x_1 + x_2 + \cdots + x_n)^r = \sum_{\substack{e_1+e_2+\cdots+e_n=r \\ 0\le e_i, e_i\in\mathbb{Z}}} \binom{r}{e_1, e_2, \ldots, e_n} x_1^{e_1} x_2^{e_2} \ldots x_n^{e_n}$$

where $\binom{r}{e_1,e_2,\ldots,e_r} = \frac{r!}{e_1!e_2!\ldots e_n!}$.

To recap, when we have a counting exercise, we should first ask whether order is important and then ask whether repetition is allowed. This will get us into the right ballpark as far as the form of the solution. We must use basic counting principles to decompose the exercise into sub-problems. Solve the sub-problems and put the pieces back together. Solutions to sub-problems usually take the same form as the underlying problem, though they may be related to it via the multinomial theorem. Table 33.2 synopsizes our six fundamental cases.

Exercises

Exercise 33.1. *How many quaternary strings of length n are there (a quaternary string uses 0's, 1's, 2's, and 3's)?*

Exercise 33.2. *How many quaternary strings of length less than or equal to 7 are there?*

Exercise 33.3. *How many solutions in integers are there to $x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 = 54$, where $3 \leq x_1$, $4 \leq x_2$, $5 \leq x_3$, and $6 \leq x_4, x_5, x_6, x_7$?*

Exercise 33.4. *How many ternary strings of length n start* 0101 *and end* 212*?*

Exercise 33.5. *A doughnut shop has 8 kinds of doughnuts: chocolate, glazed, sugar, cherry, strawberry, vanilla, caramel, and jalapeño. How many ways are there to order three dozen doughnuts, if at least 4 are jalapeño, at least 6 are cherry, and at least 8 are strawberry, but there are no restrictions on the other varieties?*

Exercise 33.6. *How many strings of twelve lowercase English letters are there*

(a) which start and end with the letter x, if letters may be repeated?

(b) which contain the letter x exactly once, if letters can be repeated?

(c) which contain each of the letters x and y both exactly once, if letters can be repeated?

(d) which contain at least one letter from the first half of the alphabet (a through m), where letters may be repeated?

Exercise 33.7. *How many bit strings of length 19 either begin* 0101*, or have 4th, 5th and 6th digits* 101*, or end* 1010*?*

Exercise 33.8. *How many pentary strings (i.e. strings using the digits 0,1,2,3,4) of length 15 consist of three 0's, four 1's, three 2's, four 3's and one 4?*

Exercise 33.9. *Seven lecturers and fourteen professors are on the faculty of a math department.*

(a) How many ways are there to form a committee with seven members which contains more lecturers than professors?

(b) How many ways are there to form a committee with seven members where the professors outnumber the lecturers on the committee by at least a two-to-one margin?

(c) *How many ways are there to form a committee consisting of at least five lecturers?*

Exercise 33.10. *In how many ways can twenty people form a line at a ticket window if Hans and wife Brunhilda are having a spat, and refuse to stand in consecutive places in the line?*

Exercise 33.11. *Prove that a set with $n \geq 1$ elements has the same number of subsets with an even number of elements, as subsets with an odd number of elements.*

Problems

Problem 33.1. *A doughnut shop has 8 kinds of doughnuts: chocolate, glazed, sugar, cherry, strawberry, vanilla, caramel, and jalapeño. How many ways are there to order three dozen doughnuts, if at most 4 are jalapeño, at most 6 are cherry, and at most 8 are strawberry, but there are no restrictions on the other varieties?*

Problem 33.2. *How many strings of twelve lowercase English letters are there*

(a) *which start with the letter x, if letters may be repeated?*

(b) *which contain the letter x at least once, if letters can be repeated?*

(c) *which contain each of the letters x and y at least once, if letters can be repeated?*

(d) *which contain at least one vowel, where letters may not be repeated?*

Problem 33.3. *How many ternary strings of length 9 have*

(a) *exactly four 1's?*

(b) *at least three 0's?*

(c) *at most three 1's?*

(d) *exactly four 0's, three 1's and two 2's?*

Problem 33.4. *How many ways are there to seat six people at a circular table where two seatings are considered equivalent if one can be obtained from the other by rotating the table?*

Problem 33.5. *A donut shop sells six types of donut. You buy a scratch-off ticket that promises you will win anywhere from one to two dozen donuts. How many different prizes are possible?*

Problem 33.6. *How many ternary strings of length* n *contain no two adjacent identical symbols? Examples:* ($n = 8$) 13123231 *is good, but* 13112321 *is bad.*

Problem 33.7. *How many ternary strings of length* n *contain at least two adjacent identical symbols? Examples:* ($n = 9$) 131212321 *is bad, but* 131123321 *is good.*

Chapter 34

Counting Using Recurrence Relations

It is not always convenient to use the methods of earlier chapters to solve counting problems. Another technique for finding the solution to a counting problem is **recursive counting**. The method will be illustrated with several examples.

Example 34.1. *Recall that a bit string is a list of* 0*'s and* 1*'s, and the length of a bit string is the total number of* 0*'s and* 1*'s in the string. For example,* 10111 *is a bit string of length five, and* 000100 *is a bit string of length six. The problem of counting the number of bit strings of length* n *is duck soup. There are two choices for each bit, and so, applying the product rule, there are* 2^n *such strings. However, consider the problem of counting the number of bit strings of length* n *with no adjacent* 0*'s.*

Let's use a_n *to denote the number of bit strings of length* n *with no adjacent* 0*'s. Here are a few sample cases for small values of* n.

n=0: *Just one good bit string of length zero, and that is* λ*, the empty bit string. So* $a_0 = 1$.

n=1: *There are two good bit strings of length one. Namely* 0 *and* 1*. So* $a_1 = 2$.

n=2: *There are three good bit strings of length two. Namely,* 01*,* 10 *and* 11*. (Of course,* 00 *is a bad bit string.) That means* $a_2 = 3$.

n=3: *Things start to get confusing now. But here is the list of good bit strings of length three:* 010, 011, 101, 110, *and* 111. *So* $a_3 = 5$.

n=4: *A little scratch work produces the good bit strings* 0101, 0111, 1011, 1101, 1111, 0110, 1010, *and* 1110, *for a total of eight. That means* $a_4 = 8$.

We can do a few more, but it is hard to see a formula for a_n *like the* 2^n *formula that gives the total number of all bit strings of length* n*. Even though a formula for* a_n *is difficult to spot, there is a pattern to the list of values for* a_n *which looks like the Fibonacci sequence pattern. In fact, the list so far looks like* 1, 2, 3, 5, 8, *and if a few more are worked out by brute force, it turns out the list continues* 13, 21, 34. *So, it certainly seems that the solution to the counting problem can be expressed recursively as* $a_0 = 1$, $a_1 = 2$, *and for* $n \geq 2$, $a_n = a_{n-1} + a_{n-2}$. *If this guess is really correct, then we can quickly compute the number of good bit strings of length* n. *We just calculate* a_0, a_1, a_2, *etc., until we reach the* a_n *we are interested in.*

Such a recursive solution to counting problems is certainly less satisfactory than a simple formula, but some counting problems are so messy that a simple formula might not be possible, and the recursive solution is better than nothing in such a case.

There is one problem with the recursive solution offered in example 34.1. We said that *it seems that* the solution to the counting problem can be expressed recursively as $a_0 = 1$, $a_1 = 2$, and for $n \geq 2$, $a_n = a_{n-1} + a_{n-2}$. That *it seems that* is not an acceptable justification of the formula. After all, we are basing that guess on just eight or ten values of the infinite sequence a_n, and it is certainly possible that those values happen to follow the pattern we've guessed simply by accident. Maybe the true pattern is much more complicated, and we have been tricked by the small number of cases we have considered. It is necessary to show that the guessed pattern is correct by supplying a logical argument.

Our argument would begin by checking the initial conditions we offered. In other words, we would verify by hand that $a_0 = 1$ and $a_1 = 2$. This serves as a basis for the verification of the recursive formula. Now what we want to do is assume that we have already calculated all the values a_0, a_1, $\cdots$, a_k for some $k \geq 1$, and **show** that a_{k+1} must equal $a_k + a_{k-1}$. It is very important to understand that *we do not want to compute the value of* a_{k+1}. *We only want to prove that* $a_{k+1} = a_k + a_{k-1}$. The major error made doing these types of problems is attempting to compute the specific value of a_{k+1}. **Don't fall for that trap!** After all, if it were possible to actually compute the specific value of a_{k+1}, then we could find a formula for a_n in general, and we wouldn't have to be seeking a recursive relation at all.

Here is how the argument would go in the bit string example. Suppose we have lists of the good bit strings of lengths 0, 1, … , k. Here is how to make a list of all the good bit strings of length $k + 1$. First, take any good bit string of length k and add a 1 on the right hand end. The result must be a good bit string of length $k + 1$ (since we added a 1 to the end, and the original bit string didn't have two consecutive 0's, the new bit string cannot have two consecutive 0's either). In that way we form some good bit strings of length $k + 1$. In fact, we have built exactly a_k good bit strings of length $k + 1$. But wait, there's more! (as they say in those simple-minded TV ads). Another way to build a good bit string of length $k + 1$ is to take a good bit string of length $k - 1$ and add 10 to the right end. Clearly these will also be good bit strings of length $k + 1$. And these all end with a 0, so they are all new ones, and not ones we built in the previous step. How many are there of this type? One for each of the good bit strings of length $k - 1$, or a total of a_{k-1}. Thus, so far we have built $a_{k^k} + a_{k-1}$ good bit strings of length $k + 1$. Now we will show that in fact we have a complete list of all good bit string of length $k + 1$, and that will complete the proof that $a_{k+1} = a_k + a_{k-1}$. But before driving that last nail into the coffin, let's look at the steps outlined above for the case $k + 1 = 4$.

The previous paragraph essentially provides an algorithm for building good bit strings of length $k + 1$ from good bits strings of lengths k and $k - 1$. The algorithm instructs us to add 1 to the right end of all the good bit strings of length k and 10 to the right of all the good bit strings of length $k - 1$. Applying the algorithm for the case $k + 1 = 4$, gives the following list, where the added bits are put in parentheses to make them stand out. 010(1), 011(1), 101(1), 110(1), 111(1), 01(10), 10(10), and 11(10).

There remains one detail to iron out. It is clear that the algorithm will produce good bit strings of length $k + 1$. But, does it produce *every* good bit string of length $k + 1$? If it does not, then the recursive relation we are offering for the solution to the counting problem will eventually begin to produce answers that are too small, and we will undercount the number of good bit strings. To see that we do count all good bit strings of length $k + 1$, consider any particular good bit string of length $k + 1$, call it s for short, and look at the right most bit of s. There are two possibilities for that bit. It could be a 1. If that is so, then when the 1 is removed the remaining bit string is a good string of length k (it can't have two adjacent 0's since s doesn't have two adjacent 0's). That means the bit string s is produced by adding a 1 to the right end of a good bit string of length k, and so s is produced by the first step in the algorithm. The other option for s is that the right most bit is a 0. But then the second bit in from the right must be a 1, since s is a good bit string, so it doesn't have adjacent 0's. So the last two bits on the right of s are 10. If those two bits are removed, there remains a good bit string of length $k - 1$. Thus s is produced by adding 10 to the right end of a good bit string of length $k - 1$, and so s is produced by the second case in the algorithm.

In a nutshell, we have shown our algorithm produces $a_k + a_{k-1}$ good bit strings of length $k + 1$, and that the algorithm does not miss any good bit strings of length $k + 1$. Thus we have proved that $a_{k+1} = a_k + a_{k-1}$ for all $k \geq 2$.

Example 34.1 was explained in excruciating detail. Normally, the verifications will be much more briefly presented. It takes a while to get used to recursive counting, but once the light goes on, the beauty and simplicity of the method will become apparent.

Example 34.2. *This example is a little silly since it is very easy to write down a formula to solve the counting problem. But the point of the example is not find the solution to the problem but rather to exhibit recursive counting in action. The problem is to compute the total number of individual squares on an $n \times n$ checkerboard. If we let the total number of squares be denoted by s_n, then obviously $s_n = n^2$. For example, an ordinary checkerboard is an 8×8 board, and it has a total of $s_8 = 8^2 = 64$ individual squares. But let's count the number of squares recursively. Clearly $s_0 = 0$. Now suppose we have computed the values of $s_0, s_1, \ldots, s_k$, for some $k \geq 0$. We will show how to compute s_{k+1} from those known values. To determine s_{k+1}, draw a $(k+1) \times (k+1)$ checkerboard. (You should make a little sketch of such a board for say $k+1=5$ so you can follow the process described next.) From that $(k+1) \times (k+1)$ board, slice off the right hand column of squares, and the bottom row of squares. What is left over will be a $k \times k$ checkerboard, so it will have s_k individual squares. That means that*

$$s_{k+1} = s_k + \text{ the number of squares sliced off}$$

Now ignore the lower right hand corner square for a moment. There are k other squares in the right hand column that was sliced off. Likewise, ignoring the corner square, there are k other squares in the bottom row that was sliced off. Hence the total number of squares sliced off was $k+k+1$, the 1 accounting for the corner square. Thus

$$s_{k+1} = s_k + k + k + 1 = s_k + 2k + 1$$

So a recursive solution to the problem of counting s_n = number of individual squares on an $n \times n$ checkerboard is

$$s_0 = 0, \text{ and}$$
$$s_{k+1} = s_k + 2k + 1, \text{ for } k \geq 0.$$

Using the recursive relation, we get $s_0 = 0$, $s_1 = s_0 + 2(0) + 1 = 0 + 0 + 1 = 1$, $s_2 = s_1 + 2(1) + 1 = 1 + 2 + 1 = 4$, $s_3 = s_2 + 2(2) + 1 = 4 + 4 + 1 = 9$, *and so on, giving what we recognize as the correct answers.*

Example 34.3. *Suppose we have available an unlimited number of pennies and nickels to deposit in a vending machine (a really old vending machine it seems, since it even accepts pennies). Let* d_n *be the number of different ways of depositing a total of* n *cents in the machine. Just to make sure we understand the problem, let's compute* d_n *for a few small values of* n*. Clearly* $d_0 = 1$ *since there is only one way to deposit no money in the machine (namely don't put any money in the machine!).* $d_1 = 1$ *(put in one penny),* $d_2 = 1$ *(put in two pennies),* $d_3 = 1$ *(put in three pennies),* $d_4 = 1$ *(put in four pennies). Now things start to get exciting!* $d_5 = 2$ *(put in five pennies or put in one nickel). And even more thrilling is* $d_6 = 3$ *(the three options are (1) six pennies, (2) one penny followed by a nickel, and (3) one nickel followed by a penny). That last count indicates a fact that may not have been clear: the order on which pennies and nickels are deposited is considered important. With a little more trial and error with pencil and paper, further values are found to be* $d_7 = 4$, $d_8 = 5$, $d_9 = 6$, $d_{10} = 8$, $d_{11} = 11$, *and* $d_{12} = 15$*. It is hard to see a formula for these values. But it is relatively easy to write down a recursive relation that produces this sequence of values. Think of it this way, suppose we wanted to put* n *cents in the machine, where* $n \geq 5$*. We can make the first coin either a penny or a nickel. If we make the first coin a penny, then we will need to add* $n - 1$ *more cents, which can be done in* d_{n-1} *ways. On the other hand, if we make the first coin a nickel, we will need to deposit* $n - 5$ *more cents, and that can be done in* d_{n-5} *ways. By the sum rule of counting, we conclude that the number of ways of depositing* n *cents is* $d_{n-1} + d_{n-5}$*. In other words,* $d_n = d_{n-1} + d_{n-5}$ *for* $n \geq 5$*.*

Since our recursive relation for d_n *does not kick in until* n *reaches* 5*, we will need to include* d_0, d_1, d_2, d_3*, and* d_4 *as initial terms. So the recursive solution to this counting problem is*

$$d_0 = 1 \qquad d_1 = 1 \qquad d_2 = 1 \qquad d_3 = 1 \qquad d_4 = 1$$
$$\text{for } n \geq 5, \qquad d_n = d_{n-1} + d_{n-5}.$$

Example 34.4 (The Tower of Hanoi). *The classic example of recursive counting concerns the story of the Tower of Hanoi. A group of monks wished a magical tower to be constructed from 1000 stone rings. The rings were to be of 1000 different sizes. The size and composition of the rings was to be designed so that any ring could support the entire weight of all of the rings smaller than itself, but each ring would be crushed beneath the weight of any larger ring.*

The monks hired the lowest bidder to construct the tower in a clearing in the dense jungle nearby. Upon completion of construction the engineers brought the monks to see their work. The monks *admired the exquisite workmanship but informed the engineers that the tower was not in the proper clearing.*

In the jungle there were only three permanent clearings. The monks had labelled them A, B and C. The engineers had labelled them in reverse order. The monks instructed the engineers to move the tower from clearing A to clearing C!

Because of the massive size of the rings, the engineers could only move one per day. No ring could be left anywhere in the jungle except one of A, B, or C. Finally each clearing was only large enough so that rings could be stored there by stacking them one on top of another.

The monks then asked the engineers how long it would take for them to fix the problem.

Before they all flipped a gasket, the most mathematically talented engineer came upon the following solution.

Let H_n denote the minimum number of days required to move an n *ring tower from A to C under the constraints given. Then $H_1 = 1$, and in general an n ring tower can be moved from A to C by first moving the top $(n-1)$ rings from A to B leaving the bottom ring at A, then moving the bottom ring from A to C, and then moving the top $(n-1)$ rings from clearing B to clearing C. That shows $H_n \leq 2 \cdot H_{n-1} + 1$, for $n \geq 2$, and a little more thought shows the algorithm just described cannot be improved upon. Thus $H_n = 2 \cdot H_{n-1} + 1$.*

Using the initial condition $H_1 = 1$ together with the recursive relation $H_n = 2 \cdot H_{n-1} + 1$, we can generate terms of the sequence:

$$1, 3, 7, 15, 31, 63, 127, 255, 511, \cdots ,$$

and it looks like $H_n = 2^n - 1$ for $n \geq 1$, which can be verified by an easy induction.

So, the problem would be fixed in $2^{1000} - 1$ days, or approximately 2.93564×10^{296} centuries. Now, that is job security!

Here are a few general rules for solving counting problems recursively:

(1) do a few small cases by brute force,

(2) think recursively: how can a larger case be solved if the solutions to smaller cases are known, and

(3) check the numbers produced by the recursive solution to make sure they agree with the values obtained by brute force.

Exercises

Exercise 34.1. *On day zero, a piggy bank contains* $0. *Each day, one more penny is added to the bank than the day before. So, on day* 1*, one penny is added, on day* 2*, two pennies are added. Write a recursive formula for the total number of pennies in the bank each day,* $n = 0, 1, 2, \ldots$.

Exercise 34.2. *Al climbs stairs by taking either one or two steps at a time. For example, he can climb a flight of three steps in three different ways: (1) one step, one step, one step or (2) two step, one step, or (3) one step, two step. Determine a recursive formula for the number of different ways Al can climb a flight of* n *steps.*

Exercise 34.3. *Find a recurrence relation for the number of bit strings of length* n *that contain an even number of* 0*'s.*

Exercise 34.4. *Find a recurrence relation for the number of bit strings of length* n *that contain two consecutive* 0*'s.*

Exercise 34.5. *Find a recurrence relation for the number of bit strings of length* n *that contain the string* 01.

Exercise 34.6. *Find a recurrence relation for the number of ternary strings of length* n *that contain two consecutive* 0*'s.*

Exercise 34.7. *Find a recurrence relation for the number of subsets of* $\{1, 2, 3, \ldots, n\}$ *that do not contain any consecutive integers. Examples for* $n = 9$*: the subset* $\{1, 3, 8\}$ *is good, but the subset* $\{2, 5, 6, 9\}$ *is bad since it contains the consecutive integers* $5, 6$.

Exercise 34.8. *Suppose in the original Tower of Hanoi problem there are four clearings* A, B, C, D. *Find a recursive relation for* J_n*, the minimum number of moves needed to transfer the tower from clearing* A *to clearing* D.

Problems

Problem 34.1. *Suppose on December* 31, 2000*, a deposit of* \$100 *is made in a savings account that pays* 10% *annual interest (Ah, those were the days!). So one year after the initial deposit, on December* 31, 2001*, the account will be credited with* \$10*, and have a value of* \$110*. On December* 31, 2002 *that account will be credited with an additional* \$11*, and have value* \$121*. Find a recursive relation that gives the value of the account* n *years after the initial deposit.*

Problem 34.2. *Sal climbs stairs by taking either one, two, or three steps at a time. Determine a recursive formula for the number of different ways Sal can climb a flight of* n *steps. In how many ways can Sal climb a flight of* 10 *steps?*

Problem 34.3. *Passwords for a certain computer system are strings of uppercase letters. A valid password must contain an even number of* X*'s. Determine a recurrence relation for the number of valid passwords of length* n*.*

Problem 34.4. *A (cheap) vending machine accepts pennies, nickels, and dimes. Let* d_n *be the number of ways of depositing* n *cents in the machine, where the order in which the coins are deposited matters. Determine a recurrence relation for* d_n*. Give the initial conditions.*

Problem 34.5. *Suppose the Tower of Hanoi rules are changed so that stones may only be transferred to an adjacent clearing in one move. Let* I_n *be the minimum number of moves required to transfer tower from clearing* A *to clearing* C*?*

(a) By brute force, determine I_1, I_2*, and* I_3*.*

(b) Find a recursive relation for I_n*.*

(c) Guess a formula for I_n*.*

Problem 34.6. *Find a recurrence relation for the number of binary strings of length* n *which do not contain the substring* 010*.*

Problem 34.7. *Find a recurrence relation for the number of ternary strings of length* n *that contain three consecutive zeroes.*

Problem 34.8. *Find a recurrence relation for the number of quaternary strings of length* n *which contain two consecutive* 1*'s.*

Problem 34.9. *Let* n *be a positive integer. Find a recurrence relation that counts the number of increasing sequences of distinct integers that start with* 1 *and end with* n*. Example: For* $n = 4$*, there are* 4 *such sequences. They are* 1, 4*;* 1, 2, 4*;* 1, 3, 4*; and* 1, 2, 3, 4*.*

Chapter 35

Solutions to Recurrence Relations

In chapter 34, it was pointed out that recursively defined sequences suffer from one major drawback: In order to compute a particular term in the sequence, it is necessary to first compute all the terms of the sequence leading up to the one that is wanted. Imagine the chore to calculate the 250^{th} Fibonacci number, f_{250}! For problems of computation, there is nothing like having a formula like $a_n = n^2$, into which it is merely necessary to plug the number of interest.

It may be possible to find a formula for a sequence that is defined recursively. When that can be done, you have the best of both the formula and recursive worlds. If we find a formula for the terms of a recursively defined sequence, we say we have **solved** the recursion.

Example 35.1. *Here is an example: The sequence* $\{a_n\}$ *is defined recursively by the initial condition* $a_0 = 2$, *and the recursive formula* $a_n = 2a_{n-1} - 1$ *for* $n \geq 1$. *If the first few terms of this sequence are written out, the results are*

$$2, 3, 5, 9, 17, 33, 65, 129, \cdots ,$$

and it shouldn't be too long before the pattern becomes clear. In fact, it looks like $a_n = 2^n + 1$ *is the formula for* a_n. *You do have to recognize the slightly hidden powers of* 2*:* 1, 2, 4, 8, 16, 32, 64, ...

To prove that guess is correct, induction would be the best way to go. Here are the details. Just to make everything clear, here is what we are going to show: If $a_0 = 2$, and $a_n = 2a_{n-1} + 1$ for $n \geq 1$, then $a_n = 2^n + 1$ for all $n \geq 0$. The basis for the inductive proof is the case $n = 0$. The correct value for a_0 is 2, and the guessed formula has value 2 when $n = 0$, so that checks out. Now for the inductive step: suppose that the formula for a_k is correct for a particular $k \geq 0$. That is, assume $a_k = 2^k + 1$ for some $k \geq 0$. Let's show that the formula must also be correct for a_{k+1}. That is, we want to show $a_{k+1} = 2^{k+1} + 1$. Well, we know that $a_{k+1} = 2a_k - 1$, and hence $a_{k+1} = 2(2^k + 1) - 1 = 2^{k+1} + 2 - 1 = 2^{k+1} + 1$, just as was to be proved. It can now be concluded that the formula we guessed is correct for all $n \geq 0$.

In example 35.1, it was possible to guess the correct formula for a_n after looking at a few terms. In most cases the formula will be so complicated that that sort of guessing will be out of the question.

There is a method that will nearly automatically solve any recurrence of the form $a_0 = a$ and for, $n \geq 1$, $a_n = ba_{n-1} + c$ (where a, b, c are constants). The method is called **unfolding**.

Example 35.2. *As an example, let's solve $a_0 = 2$ and, for $n \geq 1$, $a_n = 5 + 2a_{n-1}$. The plan is to write down the recurrence relation, and then substitute for a_{n-1}, then for a_{n-2}, and so on, until we reach a_0. It looks like this*

$$\begin{aligned} a_n &= 5 + 2a_{n-1} \\ &= 5 + 2(5 + 2a_{n-2}) = 5 + 5(2) + 2^2 a_{n-2} \\ &= 5 + 5(2) + 2^2(5 + 2a_{n-3}) = 5 + 5(2) + 5(2^2) + 2^3 a_{n-3}. \end{aligned}$$

If this substitution is continued, eventually we reach an expression we can compute in closed form:

$$\begin{aligned} a_n &= 5 + 5(2) + 5(2^2) + 5(2^3) + \cdots + 5(2^{n-1}) + 2^n a_0 \\ &= 5(1 + 2 + 2^2 + \cdots + 2^{n-1}) + 2^n(2) \\ &= 5\frac{2^n - 1}{2 - 1} + 2(2^n) \\ &= 5(2^n - 1) + 2(2^n) \\ &= 7(2^n) - 5. \end{aligned}$$

In the next to last step we use the formula for adding the terms of a geometric sequence.

Exercises

Exercise 35.1. *Guess the solution to* $a_0 = 2$, *and* $a_1 = 4$, *and, for* $n \geq 2$, $a_n = 4a_{n-1} - 3a_{n-2}$ *and prove your guess is correct by induction.*

Exercise 35.2. *Solve by unfolding:* $a_0 = 2$, *and, for* $n \geq 1$, $a_n = 5a_{n-1}$.

Exercise 35.3. *Solve by unfolding:* $a_0 = 2$, *and, for* $n \geq 1$, $a_n = 5a_{n-1} + 3$.
Hint: This one will involve applying the geometric sum formula.

Problems

Problem 35.1. *Guess the solution to* $a_0 = 1$, *and* $a_1 = 5$, *and, for* $n \geq 2$, $a_n = a_{n-1} + 2a_{n-2}$ *and prove your guess is correct by induction.*

Problem 35.2. *Solve by unfolding:* $a_0 = 2$, *and, for* $n \geq 1$, $a_n = 7a_{n-1}$..

Problem 35.3. *Solve by unfolding:* $a_0 = 2$, *and, for* $n \geq 1$, $a_n = 7a_{n-1} + 3$.

Chapter 36

The Method of Characteristic Roots

There is no method that will solve all recurrence relations. However, for one particular type, there is a standard technique. The type is called a **linear recurrence relation with constant coefficients**. In such a recurrence relation, the recurrence formula has the form

$$a_n = c_1 a_{n-1} + c_2 a_{n-2} + \cdots + c_k a_{n-k} + f(n)$$

where $c_1, \ldots, c_k$ are constants with $c_k \neq 0$, and $f(n)$ is any function of n.

The **degree** of the recurrence is k, the number of terms we need to go back in the sequence to compute each new term. If $f(n) = 0$, then the recurrence relation is called homogeneous. Otherwise it is called **non-homogeneous**.

In chapter 35, we noted that some simple non-homogeneous linear recurrence relations with constant coefficients can be solved by unfolding. This method is not powerful enough for more general problems. In this chapter we introduce a basic method that, in principle at least, can be used to solve any homogeneous linear recurrence relation with constant coefficients.

We begin by considering the degree 2 case. That is, we have a recurrence relation of the form $a_n = c_1 a_{n-1} + c_2 a_{n-2}$, for $n \geq 2$, where c_1, c_2 are real constants. We must also have two initial conditions a_0 and a_1. That is, we are given a_0 and a_1 and the formula $a_n = c_1 a_{n-1} + c_2 a_{n-2}$, for $n \geq 2$. Notice that $c_2 \neq 0$ or else we have a linear recurrence relation with constant coefficients and degree 1. What we seek is a **closed form formula** for a_n, which is a function of n alone, and which is therefore independent of the previous terms of the sequence.

Here's the technique in a specific example: The problem we will solve is to find a formula for the terms of the sequence

$$a_0 = 4 \text{ and } a_1 = 8, \text{ with}$$
$$a_n = 4a_{n-1} + 12a_{n-2}, \text{for } n \geq 2.$$

The first thing to do is to ignore the initial conditions, and concentrate on the recurrence relation. And the way to solve the recurrence relation is to guess the solution. Well, actually, it is to guess the **form** of the solution - an *educated* guess! For such a recurrence you should guess that the solution looks like $a_n = r^n$, for some constant r. In other words, guess the solution is simply the powers of some fixed number. The good news is that this guess will always be correct! You will always find some solutions of this form. When this guess is plugged into the recurrence relation and the equation is simplified, the result is an equation that can be solved for r. That equation is called the **characteristic equation** for the recurrence. In our example, when $a_n = r^n$ for each n, the result is $r^n = 4r^{n-1} + 12r^{n-2}$, and canceling r^{n-2} from each term, and rearranging the equation, we get $r^2 - 4r - 12 = 0$. That's the characteristic equation. The left side can be factored, and the equation then looks like $(r-6)(r+2) = 0$, and we see the solutions for r are $r = 6$ and $r = -2$. And, sure enough, if you check it out, you will see that $a_n = 6^n$ and $a_n = (-2)^n$ both satisfy the given recurrence relation. In other words, we find that

$$6^n = 4 \cdot 6^{n-1} + 12 \cdot 6^{n-2}, \text{and } (-2)^n = 4 \cdot (-2)^{n-1} + 12 \cdot (-2)^{n-2}, \text{for all } n \geq 2.$$

Using the characteristic equation, we have a method of finding some solutions to a recurrence relation. This method will not find all possible solutions however. *BUT...* if we find **all** the solutions to the characteristic equation, then they can be combined in a certain way to produce all possible solutions to the recurrence relation. The fact to remember is that if $r = a, b$ are the two solutions to the characteristic equation (for a recurrence of order two), then every possible solution to the linear homogeneous recurrence relation must look like $\alpha a^n + \beta b^n$ for some constants α, β.

In the example we have been working on, every possible solution looks like

$$a_n = \alpha(6)^n + \beta(-2)^n.$$

This expression is called the **general solution** of the recurrence relation.

Once we have figured out the general solution to the recurrence relation, it is time to think about the initial conditions. In our case, the initial conditions are $a_0 = 4$ and $a_1 = 8$. The idea is to select the constants α and β of the general solution $a_n = \alpha(6)^n + \beta(-2)^n$ so it will produce the correct two initial values. For $n = 0$ we see we need $4 = a_0 = \alpha(6)^0 + \beta(-2)^0 = \alpha + \beta$, and for $n = 1$, we need $8 = a_1 = \alpha(6)^1 + \beta(-2)^1 = 6\alpha - 2\beta$. Now, we solve the following pair of equations for α and β:

$$\alpha + \beta = 4,$$
$$6\alpha - 2\beta = 8.$$

Performing a bit of algebra, we learn that $\alpha = 2$ and $\beta = 2$. Thus the solution to the recurrence is

$$a_n = 2(6)^n + 2(-2)^n.$$

The steps in solving a recurrence problem are:

(1) Determine the characteristic equation.

(2) Find the solutions to the characteristic equation.

(3) Write down the general solution to the recurrence relation.

(4) Select the constants in the general solution to produce the correct initial conditions.

One catch with the method of characteristic equation occurs when the equation has repeated roots. Suppose, for example, that when the characteristic equation is factored the result is $(r-2)(r-2)(r-3)(r+5) = 0$. The characteristic roots are 2, 2, 3 and −5. Here 2 is a repeated root. If we follow the instructions given above, then the general solution we would write down is

$$a_n = \alpha 2^n + \beta 2^n + \gamma 3^n + \delta(-5)^n. \tag{36.1}$$

However, this expression will **not** include all possible solutions to the recurrence relation. Happily, the problem is not too hard to repair: each time a root of the characteristic equation is repeated, multiply it by an additional factor of n in the general solution, and then proceed with step 4 as described earlier.

For our example, we modify one of the 2^n terms in equation 36.1. The correct general solution looks like

$$a_n = \alpha 2^n + \beta n 2^n + \gamma 3^n + \delta(-5)^n.$$

Notice the extra factor of n in the second term. If $(r - 2)$ had been a four fold factor of the characteristic equation, in other words, if 2 had been a characteristic root four times, then the part of the general solution involving the 2's would look like

$$\alpha 2^n + \beta n \cdot 2^n + \gamma n^2 \cdot 2^n + \delta n^3 \cdot 2^n$$

Each new occurrence of a 2 is multiplied by one more factor of n.

Let's describe the method of **characteristic equation** a little more formally. First, the characteristic equation is denoted by $\chi(x) = 0$. Notice that the degree of $\chi(x)$ coincides with the degree of the recurrence relation. Notice also that the non-leading coefficients of $\chi(x)$ are simply the negatives of the coefficients of the recurrence relation. In general, the characteristic equation of $a_n = c_1 a_{n-1} + \ldots + c_k a_{n-k}$ is

$$\chi(x) = x^k - c_1 x^{k-1} - \ldots - c_{k-1} x - c_k = 0.$$

A number r (possibly complex) is a **characteristic root** if $\chi(r) = 0$. From basic algebra we know that r is a root of a polynomial if and only if $(x - r)$ is a factor of the polynomial. When $\chi(x)$ is a degree 2 polynomial, by the quadratic formula, either $\chi(x) = (x - r_1)(x - r_2)$, where $r_1 \neq r_2$, or $\chi(x) = (x - r)^2$, for some r. So there are two theorems about degree 2 linear recurrence relations with constant coefficients.

Theorem 36.1. *Let c_1 and $c_2 \neq 0$ be real numbers. Suppose that the polynomial $x^2 - c_1 x - c_2$ has two distinct roots r_1 and r_2. Then a sequence $a\colon \mathbb{N} \to \mathbb{R}$ satisfies the recursive formula*

$$a_n = c_1 a_{n-1} + c_2 a_{n-2}, for\ n \geq 2$$

if and only if $a_m = \alpha r_1^m + \beta r_2^m$, for all $m \geq 0$, for some constants α and β.

Proof: If $a_m = \alpha r_1^m + \beta r_2^m$ for all $m \geq 0$, where α and β are some constants, then since $r_i^2 - c_1 r_i - c_2 = 0$, for $i = 1,2$, we have $r_i^2 = c_1 r_i + c_2$, for $i = 1,2$.

Thus, for $n \geq 2$

$$\begin{aligned} c_1 a_{n-1} + c_2 a_{n-2} &= c_1(\alpha r_1^{n-1} + \beta r_2^{n-1}) + c_2(\alpha r_1^{n-2} + \beta r_2^{n-2}) \\ &= \alpha r_1^{n-2}(c_1 r_1 + c_2) + \beta r_2^{n-2}(c_1 r_2 + c_2) \text{ distributing and combining} \\ &= \alpha r_1^{n-2} \cdot r_1^2 + \beta r_2^{n-2} \cdot r_2^2, \text{by the remark above.} \\ &= \alpha r_1^n + \beta r_2^n = a_n \end{aligned}$$

Conversely, if the terms of a satisfy the recursive formula and has initial terms a_0 and a_1, then one checks that the sequence $a_m = \alpha r_1^m + \beta r_2^m$ for $m \geq 0$ with

$$\alpha = \frac{a_1 - a_0 r_2}{r_1 - r_2} \text{ and } \beta = \frac{a_0 r_1 - a_1}{r_1 - r_2}$$

also satisfies the recursive formula and has the same initial conditions. The equations for α and β come from solving the system of linear equations

$$\begin{aligned} a_0 &= \alpha(r_1)^0 + \beta(r_2)^0 = \alpha + \beta \\ a_1 &= \alpha(r_1)^1 + \beta(r_2)^1 = \alpha r_1 + \beta r_2. \end{aligned}$$

This system is solved using techniques from a prerequisite course. □

Example 36.2. *Solve the recurrence relation* $a_0 = 2, a_1 = 3$ *and* $a_n = a_{n-2}$, *for* $n \geq 2$.

Solution. *The recurrence relation is a linear homogeneous recurrence relation of degree 2 with constant coefficients* $c_1 = 0$ *and* $c_2 = 1$. *The characteristic polynomial is*

$$\chi(x) = x^2 - 0 \cdot x - 1 = x^2 - 1.$$

The characteristic polynomial has two distinct roots since

$$x^2 - 1 = (x - 1)(x + 1).$$

Let's say $r_1 = 1$ *and* $r_2 = -1$. *Then, we find the system of equations:*

$$\begin{aligned} 2 = a_0 &= \alpha 1^0 + \beta(-1)^0 = \alpha + \beta \\ 3 = a_1 &= \alpha 1^1 + \beta(-1)^1 = \alpha + \beta(-1) = \alpha - \beta. \end{aligned}$$

Adding the two equations eliminates β *and gives* $5 = 2\alpha$, *so* $\alpha = 5/2$. *Substituting this into the first equation,* $2 = 5/2 + \beta$, *we see that* $\beta = -1/2$. *Thus, our solution is*

$$a_n = \frac{5}{2} \cdot 1^n + \frac{-1}{2}(-1)^n = \frac{5}{2} - \frac{1}{2}(-1)^n.$$

Example 36.3. *Solve the recurrence relation* $a_1 = 3$, $a_2 = 5$, *and* $a_n = 5a_{n-1} - 6a_{n-2}$ *for* $n \geq 3$.

Solution. *Here the characteristic polynomial is*

$$\chi(x) = x^2 - 5x + 6 = (x - 2)(x - 3),$$

with roots $r_1 = 2$ *and* $r_2 = 3$. *Now, we suppose that*

$$a_m = \alpha 2^m + \beta 3^m, \text{ for all } m \geq 1.$$

The initial conditions give rise to the system of equations

$$3 = a_1 = \alpha 2^1 + \beta 3^1 = 2\alpha + 3\beta$$
$$5 = a_2 = \alpha 2^2 + \beta 3^2 = 4\alpha + 9\beta.$$

If we multiply the top equation through by 2, *we obtain*

$$6 = 4\alpha + 6\beta$$
$$5 = 4\alpha + 9\beta.$$

Subtracting the second equation from the first eliminates α *and yields* $1 = -3\beta$. *So, we have found that* $\beta = -1/3$. *Substitution into the first equation yields* $3 = 2\alpha + 3 \cdot (-1/3)$, *so* $\alpha = 2$. *Thus*

$$a_m = 2 \cdot 2^m - \frac{1}{3}3^m = 2^{m+1} - 3^{m-1} \text{ for all } m \geq 1.$$

The other case we mentioned had a characteristic polynomial of degree two with one repeated root. Since the proof is similar we simply state the theorem.

Theorem 36.4. *Let* c_1 *and* c_2 *be real numbers with* $c_2 \neq 0$ *and suppose that the polynomial* $x^2 - c_1x - c_2$ *has a root* r *with multiplicity 2, so that* $x^2 - c_1x - c_2 = (x - r)^2$. *Then a sequence* $a\colon \mathbb{N} \to \mathbb{R}$ *satisfies the recursive formula*

$$a_n = c_1a_{n-1} + c_2a_{n-2}, \text{for } n \geq 2$$

if and only if $a_m = \alpha r^m + \beta m r^m$, *for all* $m \geq 0$, *for some constants* α *and* β.

Example 36.5. *Solve the recurrence relation* $a_0 = -1, a_1 = 4$ *and* $a_n = 4a_{n-1} - 4a_{n-2}$, *for* $n \geq 2$.

Solution. *In this case we have* $\chi(x) = x^2 - 4x + 4 = (x - 2)^2$. *So, we may suppose that*

$$a_m = (\alpha + \beta m)2^m, \text{ for all } m \in \mathbb{N}.$$

The initial conditions give rise to the system of equations

$$-1 = a_0 = (\alpha + \beta \cdot 0)2^0 = (\alpha) \cdot 1 = \alpha$$
$$4 = a_1 = (\alpha + \beta \cdot 1)2^1 = 2(\alpha + \beta) \cdot 2.$$

Substituting $\alpha = -1$ *into the second equation gives* $4 = 2(\beta - 1)$, *so* $2 = \beta - 1$ *and* $\beta = 3$. *Therefore* $a_m = (3m - 1)2^m$, *for all* $m \in \mathbb{N}$.

Finally, we state, without proof, the general method of characteristic roots.

Theorem 36.6. *Let* $c_1, c_2, \dots, c_k \in \mathbb{R}$ *with* $c_k \neq 0$. *Suppose that*

$$\chi(x) = x^k - c_1 x^{k-1} - c_2 x^{k-2} - \cdots - c_{k-1} x - c_k = (x - r_1)^{j_1}(x - r_2)^{j_2} \cdot \dots \cdot (x - r_s)^{j_s}$$

where $r_1, r_2, \dots, r_s$ *are distinct roots of* $\chi(x)$, *and* $j_1, j_2, \dots, j_s$ *are positive integers so that*

$$j_1 + j_2 + j_3 + \cdots + j_s = k.$$

Then a sequence $a: \mathbb{N} \to \mathbb{R}$ *satisfies the recursive formula*

$$a_n = c_1 a_{n-1} + c_2 a_{n-2} + \cdots + c_k a_{n-k}, for\ n \geq k$$

if and only if $a_m = p_1(m) r_1^m + p_2(m) r_2^m + \cdots + p_s(m) r_s^m$ *for* $m \geq 0$, *where*

$$p_i(m) = \alpha_{0,i} + \alpha_{1,i} m + \alpha_{2,i} m^2 + \cdots + \alpha_{j_i - 1, i} m^{j_i - 1}, for\ 1 \leq i \leq s$$

and the $\alpha_{j,i}$*'s are constants*.

There is a problem with the general case. It is true that given the recurrence relation we can simply write down the characteristic polynomial. However it can be quite a challenge to factor it as required by the theorem. Even if we succeed in factoring it we are faced with the tedious task of setting up and solving a system of k linear equations in k unknowns (the $\alpha_{l,i}$'s). While in theory such a system can be solved using the methods of elimination or substitution covered in a college algebra course, in practice, the amount of labor involved can become overwhelming. For this reason, computer algebra systems are often used in practice to help solve systems of equations, or even the original recurrence relation.

Exercises

Exercise 36.1. *For each of the following sequences find a recurrence relation satisfied by the sequence. Include a sufficient number of initial conditions to completely specify the sequence.*

(a) $a_n = 2n + 3, n \geq 0$

(b) $a_n = 3 \cdot 2^n, n \geq 1$

(c) $a_n = n^2, n \geq 1$

(d) $a_n = n + (-1)^n, n \geq 0$

Solve each of the following recurrence relations:

Exercise 36.2. $a_0 = 3, a_1 = 6,$ *and* $a_n = a_{n-1} + 6a_{n-2}$, *for* $n \geq 2$.

Exercise 36.3. $a_0 = 4, a_1 = 7,$ *and* $a_n = 5a_{n-1} - 6a_{n-2}$, *for* $n \geq 2$.

Exercise 36.4. $a_2 = 5, a_3 = 13,$ *and* $a_n = 7a_{n-1} - 10a_{n-2}$, *for* $n \geq 4$.

Exercise 36.5. $a_1 = 3, a_2 = 5,$ *and* $a_n = 4a_{n-1} - 4a_{n-2}$, *for* $n \geq 3$.

Exercise 36.6. $a_0 = 1, a_1 = 6,$ *and* $a_n = 6a_{n-1} - 9a_{n-2}$, *for* $n \geq 2$.

Exercise 36.7. $a_1 = 2, a_2 = 8,$ *and* $a_n = a_{n-2}$, *for* $n \geq 3$.

Exercise 36.8. $a_0 = 2, a_1 = 5, a_2 = 15,$ *and* $a_n = 6a_{n-1} - 11a_{n-2} + 6a_{n-3}$, *for* $n \geq 3$.

Exercise 36.9. *Find a closed form formula for the terms of the Fibonacci sequence:* $f_0 = 0, f_1 = 1,$ *and for* $n \geq 2$, $f_n = f_{n-1} + f_{n-2}$.

Problems

Solve each of the recurrence relations using the method of characteristic roots:

Problem 36.1. $a_0 = 1,$ *and* $a_n = 2a_{n-1},$ *for* $n \geq 1.$

Problem 36.2. $a_0 = 2, a_1 = 5$ *and* $a_n = a_{n-1} + 6a_{n-2},$ *for* $n \geq 2.$

Problem 36.3. $a_0 = 3, a_1 = 7,$ *and* $a_n = 6a_{n-1} - 5a_{n-2},$ *for* $n \geq 2.$

Problem 36.4. $a_2 = 5, a_3 = 13,$ *and* $a_n = 3a_{n-1} + 10a_{n-2},$ *for* $n \geq 4.$

Problem 36.5. $a_1 = 3, a_2 = 5,$ *and* $a_n = 8a_{n-1} - 16a_{n-2},$ *for* $n \geq 3.$

Problem 36.6. $a_1 = 2, a_2 = 8,$ *and* $a_n = 4a_{n-2},$ *for* $n \geq 3.$

Problem 36.7. $a_0 = 0, a_1 = 1, a_2 = 2,$ *and* $a_n = -a_{n-1} + 4a_{n-2} + 4a_{n-3},$ *for* $n \geq 3.$

Problem 36.8. $a_0 = 0,$ $a_1 = 1,$ *and for* $n \geq 2,$ $a_n = 2a_{n-1} + a_{n-2}.$

Chapter 37

Solving Non-homogeneous Recurrences

When a linear recurrence relation with constant coefficients for a sequence $\{s_n\}$ looks like

$$s_n = c_1 s_{n-1} + c_2 s_{n-2} + \cdots + c_k s_{n-k} + f(n),$$

where $f(n)$ is some (nonzero) function of n, then the recurrence relation is said to be **non-homogeneous**. For example, $s_n = 2s_{n-1} + n^2 + 1$ is a non-homogeneous recurrence. Here $f(n) = n^2 + 1$. The methods used in the last chapter are not adequate to deal with non-homogeneous problems. But it wasn't all a waste since those methods do provide one step in the solution of non- homogeneous problems.

The steps used to solve non-homogeneous linear recurrence relations with constant coefficients are:

Step (1): Replace the $f(n)$ by 0 to create a homogeneous recurrence relation,

$$s_n = c_1 s_{n-1} + c_2 s_{n-2} + \cdots + c_k s_{n-k}.$$

Now solve this and write down the general solution. We learned to do this in chapter 36. For

$f(n)$	Particular Solution Guess
c (a constant)	A (constant)
n	$An + B$
n^2	$An^2 + Bn + C$
n^3	$An^3 + Bn^2 + Cn + D$
2^n	$A2^n$
r^n (r constant)	Ar^n

Table 37.1: Particular solution patterns

example, in the case of no repeated roots, the general solution will look something like:

$$s_n = a_1 r_1^n + a_2 r_2^n + \cdots + a_k r_k^n$$

where the constants $a_1, a_2, \ldots, a_k$ are to be determined.

Step (2): Next, find one **particular solution** to the original non-homogeneous recursion. In other words, one specific sequence that obeys the recursive formula (ignoring the initial conditions). A method for finding a particular solution that works in many cases is to guess! Actually, it is to make an educated guess. Reasonable guesses depend on the form of $f(n)$. There is an algorithm that will produce the correct guess, but it is so complicated it isn't worth learning for the few simple examples we will be doing. Instead, rely on the following guidelines to guess the form of a particular solution.

Roughly, the plan is the guess a particular solution that is the most general function of the same type as $f(n)$. Specifically, table 37.1 shows reasonable guesses.

These guesses can be *mixed-and-matched*. For example, if

$$f(n) = 3n^2 + 5^n,$$

then a reasonable candidate particular solution would be

$$An^2 + Bn + C + D5^n.$$

Once a guess has been made for the form of a particular solution, that guess is plugged into the recurrence relation, and the coefficients $A, B, \cdots$ are determined. In this way a specific particular solution will be found.

It will sometimes happen that when the equations are set up to determine the coefficients of the particular solution, an inconsistent system will appear. In such a case, as with repeated characteristic roots, the trick is (more-or-less) to multiply the guess for the particular solution by n, and try again.

Step (3): Once a particular solution has been found, add the particular solution of step (2) to the general solution of the homogeneous recurrence found in step (1). If we denote a particular solution by $h(n)$, then the total general solution looks like

$$s_n = a_1 r_1^n + a_2 r_2^n + \cdots + a_k r_k^n + h(n).$$

Step (4): Invoke the initial conditions to determine the values of the coefficients $a_1, a_2, \ldots, a_k$ just as we did for the homogeneous problems in chapter 36.

The major oversight made solving a non-homogeneous recurrence relation is trying to determine the coefficients $a_1, a_2, \ldots, a_k$ before the particular solution is added to the general solution. This mistake will usually lead to inconsistent information about the coefficients, and no solution to the recurrence will be found.

Example 37.1. *Let's solve the Tower of Hanoi recurrence using this method.*

The recurrence is $H_0 = 0$, and, for $n \geq 1$, $H_n = 2H_{n-1} + 1$. We know the closed form formula for H_n is $2^n - 1$ already, but let's work it out using the method outlined above.

Step (1): Find the general solution of related homogeneous recursion (indicated by the superscript (h)):

$H_n^{(h)} = 2H_{n-1}^{(h)}$. That will be $H_n^{(h)} = A \cdot 2^n$.

Step (2): Guess the particular solution (indicated by superscript (p)): $H_n^{(p)} = B$, a constant.

Plugging that guess into the recurrence gives $B = 2B + 1$, and so we see $B = -1$.

Step (3): Hence, the general solution to the Tower of Hanoi recurrence is

$$H_n = H_n^{(h)} + H_n^{(p)} = A \cdot 2^2 - 1.$$

Step (4): Now, use the initial condition to determine A: When $n = 0$, we want $0 = A2^0 - 1$ which means $A = 1$. Thus, we find the expected result:

$$H_n = 2^n - 1, \text{ for } n \geq 0.$$

Example 37.2. *Here is a more complicated example worked out in detail to exhibit the method.*

Let's solve the recurrence

$$s_1 = 2, \quad s_2 = 5 \quad \textit{and,}$$
$$s_n = s_{n-1} + 6s_{n-2} + 3n - 1, \ \textit{for } n \geq 3.$$

Step (1): Find the general solution of $s_n = s_{n-1} + 6s_{n-2}$. *After finding the characteristic equation, and the characteristic roots, the general solution turns out to be* $s_n = a_1 3^n + a_2(-2)^n$.

Step (2): To find a particular solution let's guess that there is a solution $h(n)$ *that looks like* $h(n) = an + b$, *where* a *and* b *are to be determined. To find values of* a *and* b *that work, we substitute this guess for a solution into the original recurrence relation. In this case, the result of plugging in the guess* ($s_n = h(n) = an + b$) *gives us:*

$$an + b = a(n-1) + b + 6(a(n-2) + b) + 3n - 1.$$

which can be rearranged to

$$(6a + 3)n + (-13a + 6b - 1) = 0.$$

If this equation is to be correct for all n, *then, in particular, it must be correct when* $n = 0$ *and when* $n = 1$, *and that tells us that*

$$-13a + 6b - 1 = 0 \quad \textit{and,}$$
$$6a + 3 - 13a + 6b - 1 = 0.$$

Solving this pair of equations we find $a = -1/2$ *and* $b = -11/12$. *And sure enough, if you plug this alleged solution into the original recurrence, you will see it checks.*

Step (3): Write down the general solution to the original non-homogeneous problem by adding the particular solution of step (2) to the general solution from step (1) getting:

$$s_n = a_1 3^n + a_2(-2)^n - \frac{1}{2}n - \frac{11}{12}.$$

Step (4): Now a_1, a_2 *can be calculated: For* $n = 1$, *the first initial condition gives*

$$s_1 = 2 = a_1 3^1 + a_2(-2)^1 - \frac{1}{2} \cdot 1 - \frac{11}{12}.$$

And for $n = 2,$ *we get*

$$s_2 = 5 = a_1 3^2 + a_2(-2)^2 - \frac{1}{2} \cdot 2 - \frac{11}{12}.$$

Solving these two equations for a_1 *and* a_2*, we find that* $a_1 = 11/12$ *and* $a_2 = -1/3$*. So the solution to the recurrence is*

$$s_n = \frac{11}{12} 3^n - \frac{1}{3}(-2)^n - \frac{1}{2} n - \frac{11}{12}.$$

Exercises

Use the general solutions for the related homogeneous problems of chapter 36 to help solve the following non-homogeneous recurrence relations with initial conditions.

Exercise 37.1. $a_0 = 3, a_1 = 6$ *and* $a_n = a_{n-1} + 6a_{n-2} + 1,$ *for* $n \geq 2.$

Exercise 37.2. $a_2 = 5, a_3 = 13$ *and* $a_n = 7a_{n-1} - 10a_{n-2} + n,$ *for* $n \geq 4.$

Exercise 37.3. $a_1 = 3, a_2 = 5$ *and* $a_n = 4a_{n-1} - 4a_{n-2} + 2^n,$ *for* $n \geq 3.$

Exercise 37.4. $a_0 = 1, a_1 = 6$ *and* $a_n = 6a_{n-1} - 9a_{n-2} + n,$ *for* $n \geq 2.$

Exercise 37.5. $a_0 = 2, a_1 = 5, a_2 = 15,$ *and* $a_n = 6a_{n-1} - 11a_{n-2} + 6a_{n-3} + 2n + 1,$ *for* $n \geq 3.$

Problems

Solve the following non-homogeneous recurrence relations with initial conditions.

Problem 37.1. $a_0 = 1,$ *and* $a_n = 2a_{n-1} + 1,$ *for* $n \geq 1.$

Problem 37.2. $a_0 = 2, a_1 = 5$ *and* $a_n = a_{n-1} + 6a_{n-2} + 2,$ *for* $n \geq 2.$

Problem 37.3. $a_0 = 3, a_1 = 7,$ *and* $a_n = 6a_{n-1} - 5a_{n-2} + n,$ *for* $n \geq 2.$

Problem 37.4. $a_2 = 5, a_3 = 13,$ *and* $a_n = 3a_{n-1} + 10a_{n-2} + n + 2,$ *for* $n \geq 4.$

Problem 37.5. $a_1 = 3, a_2 = 5,$ *and* $a_n = 8a_{n-1} - 16a_{n-2} + n^2,$ *for* $n \geq 3.$

Problem 37.6. $a_1 = 2, a_2 = 8,$ *and* $a_n = 4a_{n-2} + 2^n,$ *for* $n \geq 3.$

Problem 37.7. $a_0 = 0, a_1 = 1, a_2 = 2,$ *and* $a_n = -a_{n-1} + 4a_{n-2} + 4a_{n-3} + 2n,$ *for* $n \geq 3.$

Problem 37.8. $a_0 = 0,$ $a_1 = 1,$ *and for* $n \geq 2,$ $a_n = 2a_{n-1} + a_{n-2} + 1.$

Chapter 38

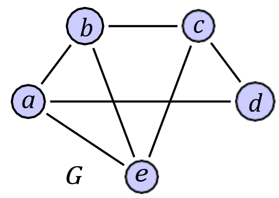
Graphs

In chapter 8 we represented a relation with a graph. In this chapter we discuss a more general notion of a graph.

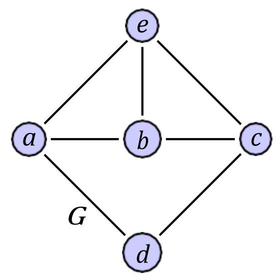
There is a lot of new vocabulary to absorb concerning graphs! For this chapter, a **graph** will consist of a number of points (called **vertices**) (singular: **vertex**) together with lines (called **edges**) joining some (possibly none, possibly all) pairs of vertices. Unlike the graphs of earlier chapters, we will not allow an edge from a vertex back to itself (so no loops allowed), we will not allow multiple edges between vertices, and the edges will not be directed (there will be no edges with arrowheads on one or both ends). All of our graphs will have a finite vertex sets, and consequently a finite number of edges. Graphs are typically denoted by an uppercase letter such as G or H.

If you would like a formal definition: a **graph**, G consists of a set of vertices V and a set E of edges, where an edge $t \in E$ is written as an unordered pair of vertices $\{u, v\}$, (in other words, a set consisting of two different vertices). We say that the edge $t = \{u, v\}$ has **endpoints** u and v, and that the edge t is **incident** to both u and v. The vertices u and v are **adjacent** when there is an edge with endpoints u and v; otherwise they are not adjacent. Such a formal definition is necessary, but a more helpful way to think of a graph is as a diagram.

Here is an example of a graph G with vertex set $\{a, b, c, d, e\}$ illustrating these concepts.



The placement of the vertices in a diagram representing a graph is (within reason!) not important. Here is another diagram of that same graph G.



In this diagram, we again have vertex set a, b, c, d, e, and edges $\{a, b\}$, $\{b, c\}$, $\{c, d\}$, $\{a, d\}$, $\{a, e\}$, $\{b, e\}$, $\{c, e\}$, and that is all that matters. It is a good idea to draw a diagram that is easy to understand! In particular, while any curve can be used to represent an edge between two vertices, whenever it is reasonable, edges are normally drawn as straight lines. The vertices b and e are adjacent and the vertices b and d are not adjacent. The vertices a and c are not adjacent since there is no edge $\{a, c\}$. If we use s to denote the edge joining b to c, then s has endpoints b and c, and s is incident to b and c.

Applying the *a-picture-is-worth-a-thousand-words* principle, for the small graphs we will be working with, a graph diagram is generally the easiest way to represent a graph.

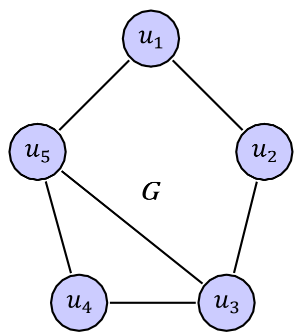
There are two standard ways to represent a graph in computer memory, both involving matrices (in other words, tables of numbers). The matrices are of a special type called $0, 1$-matrices since the table entries will all be either 0 or 1.

<u>**Adjacency matrix**</u>: If there are n vertices in the graph G, the adjacency matrix is an n by n square table of numbers. The rows and columns of the table are labeled with the symbols used to name the vertices. The names are used in the same order for the rows and columns, so there are $n!$ possible labelings. Often there will be some *natural* choice of the order of the labels, such as

alphabetic or numeric order. The entries in the table are determined as follows: the matrix entry with row label x and column label y is 1 if x and y are adjacent, and 0 otherwise.

Incidence matrix: Suppose the graph G has n vertices and m edges. The table will have n rows, labeled with the names of the vertices, and m columns labeled with the edges. Which of the $n!m!$ possible orderings of these labelings has to be specified in some way. The entry in the row labeled with vertex u and column labeled with edge e is 1 if e is incident with u, and 0 otherwise. Since every edge is incident to exactly two vertices, every column of the incidence matrix will have exactly two 1's.

Example 38.1. *Let G have vertex set $\{u_1, u_2, u_3, u_4, u_5\}$ and edges $\{u_1, u_2\}, \{u_2, u_3\}, \{u_3, u_4\}, \{u_4, u_5\}, \{u_5, u_1\}, \{u_5, u_3\}$. A graphical representation of G is*



Here are the adjacency matrix A_G, and the incidence matrix M_G of G using the vertices and edges in the orders given above:

$$A_G = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \end{bmatrix}, M_G = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}$$

Unlike most areas of mathematics, it is possible to point to a specific person as the creator of graph theory and a specific problem that led to its creation. The problem is called the *Seven Bridges of Königsberg* problem. Leonhard Euler used a graph theoretic approach to a exhibit a solution in 1736. The citizens of Königsberg would attempt to cross every bridge in town without retracing steps and finishing where they started. Euler demonstrated that this could not be done.

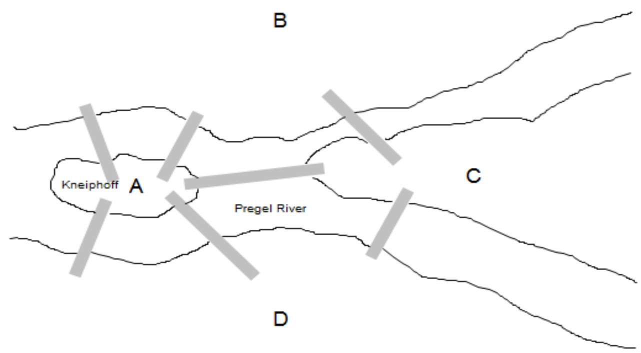


Figure 38.1 The seven bridges of Konigsberg

The notion of a graph discussed is a little more general than the graphs we will be working with in the chapter. To model the bridge problem as a graph, Euler allowed multiple edges between vertices. In modern terminology, graphs with multiple edges are called **multigraphs**.

While we are on the topic of extensions of the definition of a graph, let's also mention the case of graphs with *loops*. Here we allow an edge to connect a vertex to itself, forming a loop. Multigraphs with loops allowed are called **pseudographs**. Another generalization of the basic concept of a graph is **hypergraph**: in a hypergraph, a single edge is allowed to connect not just two, but any number of vertices.

Finally, for all these various types of graphs, we can consider the **directed** versions in which the edges are given arrowheads on one or both ends to indicate the permitted direction of travel along that edge.

For a vertex v in a graph we denote the number of edges incident to v as the **degree** of v, written as $\deg(v)$. For example, consider the graph of example 38.1

Vertices u_1, u_2, u_4 each have degree 2, while $\deg(u_3)$ and $\deg(u_5)$ are each 3. The list of the degrees of the vertices of a graph is called the **degree sequence** of the graph. The degrees are traditionally listed in increasing order. So the degree sequence of the graph G above is $2, 2, 2, 3, 3$.

The following theorem is usually referred to as the *First Theorem of Graph Theory*, it's also called the *Hand-Shaking Theorem*.

Theorem 38.2. *The sum of the degrees of the vertices of a graph equals twice the number of edges. In particular, the sum of the degrees is even.*

Proof. Notice that, when adding the degrees for the vertices, each edge will contribute two to the total, once for each end. So the sum of the degrees is twice the number of edges. □

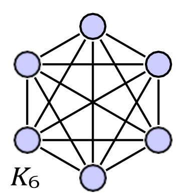
For example, in the graph G above, there are 6 edges, and the sum of the degrees of the vertices is $2+2+2+3+3 = 12 = 2(6)$.

Corollary 38.3. *A graph must have an even number of vertices of odd degree.*

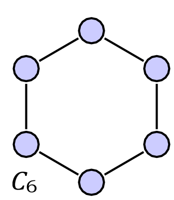
Proof. Split the vertices into two groups: the vertices with even degree and the vertices with odd degree. The sum of all the degrees is even, and the sum of all the even degrees is also even. That implies that the sum of all the odd degrees must also be even. Since an odd number of odd integers adds up to an odd integer, it must be that there is an even number of odd degrees. □

It is convenient to have names for some particular types of graphs that occur frequently.

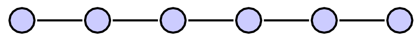
For $n \geq 1$, K_n denotes the graph with n vertices where every pair of vertices is adjacent. K_n is the **complete graph** on n vertices. So K_n is the largest possible graph with n vertices in the sense that it has the maximum possible number of edges.



For $n \geq 3$, C_n denotes the graph with n vertices, $v_1, \ldots, v_n$, where each vertex in that list is adjacent to the vertex that follows it and v_n is adjacent to v_1. The graph C_n is called the n**-cycle**. The graph C_3 is called a **triangle**.

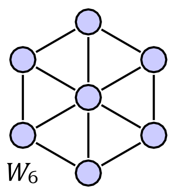


For $n \geq 2$, L_n denotes the n**-link**. An n-link is a row of n vertices with each vertex adjacent to the following vertex. Alternatively, for $n \geq 3$, an n-link is produced by erasing one edge from an n-cycle.



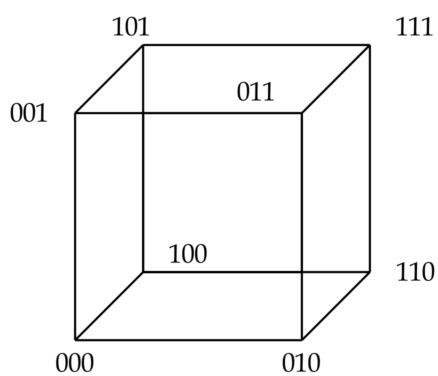
L_6

For $n \geq 3$, W_n denotes the n**-wheel**. To form W_n add one vertex to C_n and make it adjacent to every other vertex. Notice that the n-wheel has $n + 1$ vertices.



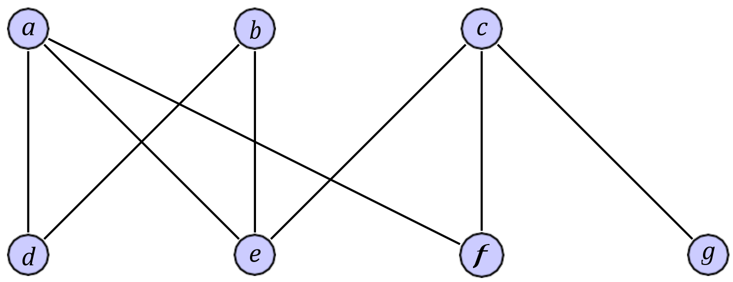
W_6

For $n \geq 1$, the n**-cube**, Q_n, is the graph whose vertices are labeled with the 2^n bit strings of length n. The unusual choice of names for the vertices is made so it will be easy to describe the edges in the graph: two vertices are adjacent provided their labels differ in exactly one bit. Except for $n = 1, 2, 3$ it is not easy to draw a convincing diagram of Q_n. The graph Q_3 can be drawn so it looks like what you would probably draw if you wanted a picture of a 3-dimensional cube. In the graph below, there is a vertex placed at each of the eight corners of the 3-cube labeled with the name of the vertex.

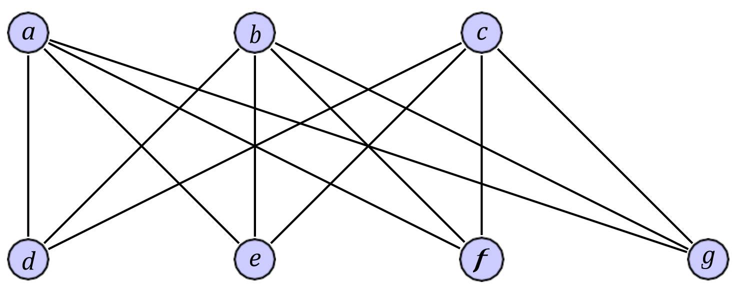


A graph is **bipartite** if it is possible to split the vertices into two subsets, let's call them T and B for top and bottom, so that all the edges go from a vertex in one of the subsets to a vertex in the other subset.

For example, the graph below is a bipartite graph with $T = \{a, b, c\}$ and $B = \{d, e, f, g\}$.

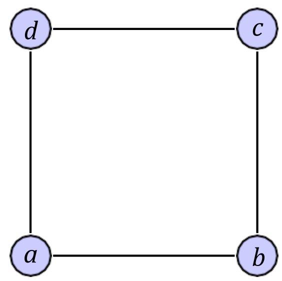


If T has m vertices and B has n vertices, and every vertex in T is adjacent to every vertex in B, the graph is called the **complete bipartite graph**, and it is denoted by $K_{m,n}$. Here is the graph $K_{3,4}$

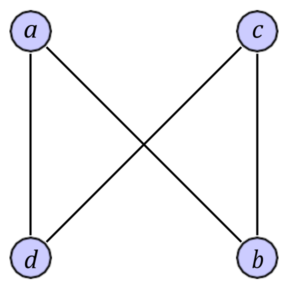


$K_{3,4}$

It is not always obvious if a graph is bipartite or not when looking at a diagram. For example the square

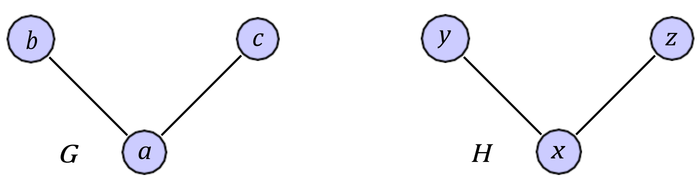


is bipartite since the graph can be redrawn as



so we can see the graph is actually $K_{2,2}$ in disguise.

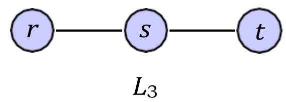
The graphs G and H are obviously really the same except for the labels used for the vertices.



This idea of *sameness* (the official phrase is the graphs G and H are **isomorphic**) for graphs is defined as follows: Two graphs G and H are isomorphic provided we can relabel the vertices of one of the graphs using the labels of the other graph in such a way that the two graphs will have exactly the same edges. As you can probably guess, the notion of isomorphic graphs is an equivalence relation on the collection of all graphs.

In the example above, if the vertices of H are relabeled as $a \to x$ (meaning replace x with a), and $b \to y$, $c \to z$, then the graph H will have edges $\{a, b\}$ and $\{a, c\}$ just like the graph G. So we have proved G and H are isomorphic graphs. The set of replacement rules, $a \to x$, $b \to y$, $c \to z$, is called an **isomorphism**.

The graph G is also isomorphic to the 3-link L_3:



In this case, an isomorphism is $a \to s$, $b \to r$, $c \to t$.

On the other hand, G is certainly not isomorphic to the 4-cycle, C_4 since that graph does not even have the same number of vertices as G. Also G is not isomorphic to the 3-cycle C_3. In this case, the two graphs do have the same number of vertices, but not the same number of edges. For two graphs have a chance of being isomorphic, the two graphs must have the same number of vertices and the same number of edges. But **warning**: even if two graphs have the same number of vertices and the same number of edges, they need not be isomorphic. For example L_4 and $K_{1,3}$ are both graphs with 4 vertices and 3 edges, but they are not isomorphic. This is so since L_4 does not have a vertex of degree 3, but $K_{1,3}$ does.

Extending that idea: to have a chance of being isomorphic, two graphs will have to have the same degree sequences since they will end up with the same edges after relabeling. But even having the same degree sequences is not enough to conclude two graphs are isomorphic as the graphs in 38.1. We can see those two graphs are not isomorphic since G has three vertices that form a triangle, but there are no triangles in H.

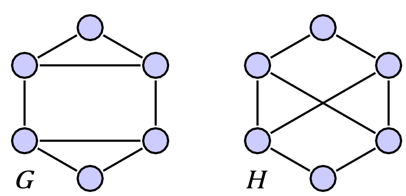


Figure 38.1: Nonisomorphic graphs with the same degree sequences.

For graphs with a few vertices and a few edges, a little trial and error is typically enough to determine if the graphs are isomorphic. For more complicated graphs, it can be very difficult to

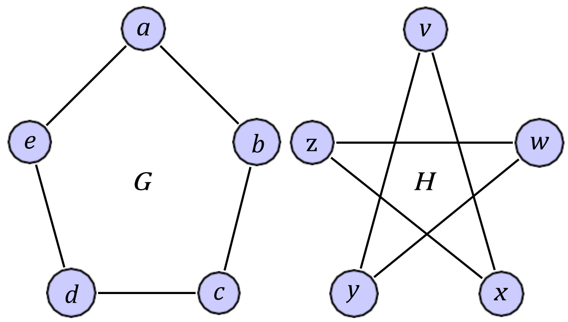


Figure 38.2: Isomorphic graphs

determine if they are isomorphic or not. One of the big goals in theoretical computer science is the design of efficient algorithms to determine if two graphs are isomorphic.

Example 38.4. *Let G be a 5-cycle on a, b, c, d, e drawn as a regular pentagon with vertices arranged clockwise, in order, at the corners. Let H have vertex set v, w, x, y, z and graphical presentation as a pentagram (five-pointed star), where the vertices of the graph are the ends of the points of the star, and are arranged clockwise, (see figure 38.2).*
An isomorphism is $a \to v,\ b \to x,\ c \to z,\ d \to w,\ e \to y$.

Example 38.5. *The two graphs in figure 38.3 are isomorphic as shown by using the relabeling*

$$u_1 \to v_1, u_2 \to v_2, u_3 \to v_3, u_4 \to v_4, u_5 \to v_9,$$

$$u_6 \to v_{10}, u_7 \to v_5, u_8 \to v_7, u_9 \to v_8, u_{10} \to v_6.$$

The graph G is the traditional presentation of the **Petersen Graph**. It could be described as the graph whose vertex set is labeled with all the two element subsets of a five element set, with an edge joining two vertices if their labels have exactly one element in common.

The origins of graph theory had to do with bridges, and possible routes crossing the bridges. In this section we will consider that sort of question in graphs in general. We will think of walking along edges, from one vertex in the list to the next, and visiting vertices. Remember that we do not allow multiple edges or loops in our graphs.

We begin with a collection of definitions. Warning: These terms are used differently in different texts. If you look at another graph theory text, be sure to see how the terms are used there.

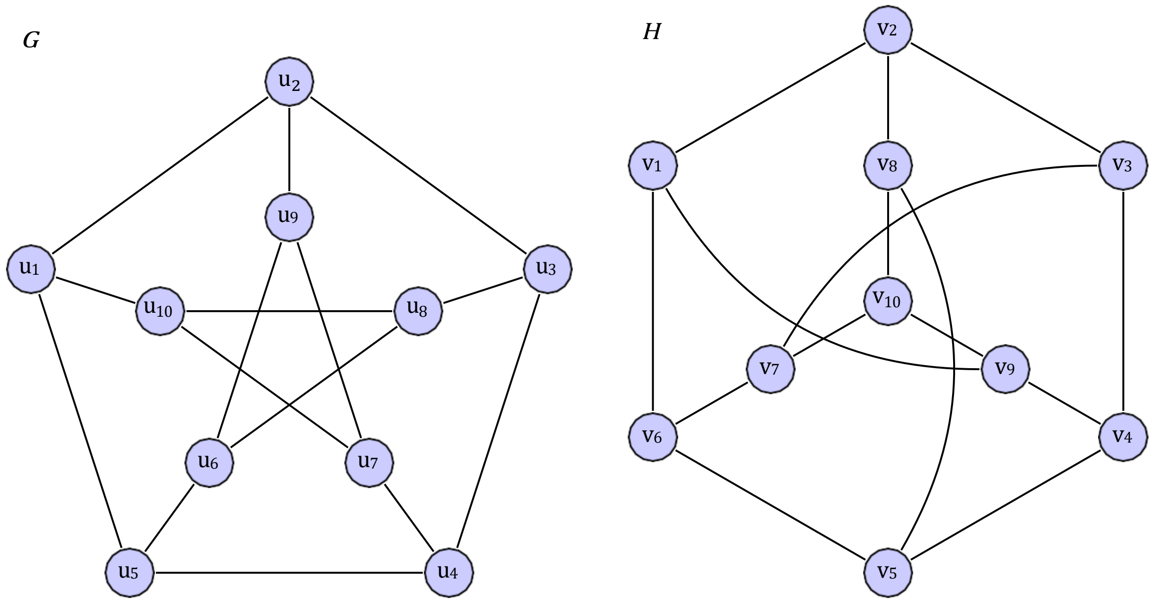


Figure 38.3: More Isomorphic graphs

A **path** of **length** n in a graph is a sequence of $n+1$ vertices $v_0, v_1, \ldots, v_n$, where each vertex in the list is adjacent to the following vertex. Repeated vertices and repeated edges in a path are allowed. The vertices v_0 and v_n are the **endpoints** of the path. Think of starting at v_0, walking along the edges, and ending up at v_n. The length n of the walk is the number of edges transversed in the path. A path of length three or more for which the endpoints are the same (so $v_0 = v_n$) is called a **circuit**. A **simple** path (or circuit) is one that does not repeat any edges. A single vertex v will be considered to be a path (but not a circuit!) of length 0.

Here is an example illustrating these definitions.

Example 38.6. *In the graph shown in figure 38.4, a, b, e, c, f, c is a path of length 5. That is an example of an a, c-path, meaning it starts at vertex a and ends at vertex c. That path is not simple since the edge c, f is repeated. Note that direction does not matter. The vertex sequence a, b, c, f, e is an a, e-path. Here are two simple circuits in that graph: a, b, e, d, a and a, b, c, e, d, a. Notice that the circuit a, e, b, c, f, e, d, a is also simple even though it repeats the vertex e. It does not repeat any edges.*

A graph is **connected** if there is a path between any two vertices. In plain English, a connected graph consists of a single piece. The individual connected pieces of a graph are called its **connected components**. The length of the shortest path between two vertices in a connected component of

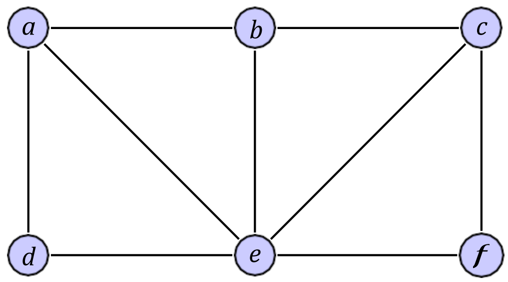


Figure 38.4: paths and circuits

a graph is called the **distance** between the vertices. In figure 38.4, the distance between a and f is 2.

Theorem 38.7. *In a connected graph there is a simple path between any two vertices. In other words, if there is a way to get from one vertex to another vertex along edges, then there is a way to get between those two vertices without repeating any edges.*

Proof. Problem 38.7. The idea is simple: in a path with a repeated edge, just eliminate the *side trip* made between the two occurrences of that edge from the path. Do that until all the repeated edges are eliminated. For example, in the graph shown in figure 38.4, The a, c-path a, e, b, e, c can be reduced to the path a, e, c, eliminating the side trip to b. □

A vertex in a graph is a **cut vertex**, if removal of the vertex and edges incident to it results in a graph with more connected components. Similarly a **bridge** is an edge whose removal (keeping the vertices it is incident to) yields a graph with more connected components.

We close this section with a discussion of two special types of paths.

An **Eulerian path** in a graph is a simple path which transverses every edge of the graph. In other words, an Eulerian path in a graph is a path that transverses every edge of the graph exactly once. An interesting property of a graph with an Eulerian path is that it can be drawn completely without lifting pencil from paper and without retracing any edges.

An **Eulerian circuit** is a simple circuit in a graph that transverses every edge of the graph. So an Eulerian circuit is a path of length three or more that transverses every edge of the graph and ends up at its initial vertex. A graph is called **Eulerian** if it has an Eulerian circuit.

Example 38.8. *The graph C_5 is an Eulerian graph. In fact, the graph itself is an Eulerian circuit.*

Example 38.9. *The graph K_5 is an Eulerian graph.*

Example 38.10. *The graph L_n is itself an Eulerian path, but does not have an Eulerian circuit.*

Example 38.11. *The graph K_4 is not an Eulerian graph. Try it!.*

A **Hamiltonian path** in a graph is a simple path that visits every vertex in the graph exactly once. A **Hamiltonian circuit** in a graph is a simple circuit that, except for the last vertex of the circuit, visits every vertex in the graph exactly once. A graph is **Hamiltonian** if it has a Hamiltonian circuit.

Example 38.12. *K_n is Hamiltonian for $n \geq 3$.*

Example 38.13. *W_n has a Hamiltonian circuit for $n \geq 3$.*

Example 38.14. *L_n has no Hamiltonian circuit for $n \geq 2$*

A few easy observations: if G is a graph with either an Eulerian circuit or Hamiltonian circuit, then

(1) G is connected.

(2) every vertex has degree at least 2.

(3) G has no bridges.

If G has a Hamiltonian circuit, then G has no cut vertices.

Leonhard Euler gave a simple way to determine exactly when a graph is Eulerian. On the other hand, despite considerable effort, no one has been able to devise a test to distinguish between Hamiltonian and non-Hamiltonian graphs that is much better than a brute force trial-and-error search for a Hamiltonian circuit.

Theorem 38.15. *A connected graph is Eulerian if and only if every vertex has even degree.*

Proof. Let G be an Eulerian graph, and suppose that v is a vertex in G with odd degree, say $2m + 1$. Let i denote the number of times an Eulerian circuit passes through v. Since every edge is used exactly once in the circuit, and each time v is visited two different edges are used, we have $2i = 2m + 1$, which is impossible. $\rightarrow\leftarrow$. So G cannot have any vertices of odd degree.

Conversely, let G be a connected graph where every vertex has even degree. Select a vertex u and build a simple path starting at u as long as possible: each time we visit a vertex we select an unused edge leaving that vertex to extend the simple path. For any vertex $v \neq u$ we visit, its even degree guarantees there will be an unused edge out, since each time v is visited used two edges incident to v and one more edge to arrive at v, for a total of an odd number of edges incident to v, and the vertex has even degree, so there must be at least one unused edge leading out of v. Since the process of extending the simple path must eventually come to an end, that shows the end must be at u when the simple path cannot be extended, and so we have constructed an Eulerian circuit.

If this simple path contains every edge we are done. Otherwise when these edges are removed from G we obtain a set of connected components $H_1, \ldots, H_m$ which are subgraphs of G and which each satisfy that all vertices have even degree. Since their sizes are smaller, we may inductively construct an Eulerian circuit for each H_i. Since each G is connected, each H_i contains a vertex of the initial circuit, say v_i. If we call the Eulerian circuit of H_i, C_i, then $v_0, \ldots, v_i, C_i, v_i, \ldots, v_n, v_0$ is a circuit in G. Since the H_i are disjoint, we may insert each Eulerian *partial* circuit thus obtaining an Eulerian circuit for G. ■

As a corollary we have

Theorem 38.16. *A connected graph has an Eulerian path, but not an Eulerian circuit, if and only if it has exactly two vertices of odd degree.*

The following theorem is an example of a sufficient (but not necessary) condition for a graph to have a Hamiltonian circuit.

Theorem 38.17. *Let* G *be a connected graph with* $n \geq 3$ *vertices. If* $\deg(v) \geq n/2$ *for every vertex* v*, then* G *is Hamiltonian.*

Proof. Suppose that the theorem is false. Let G be a connected graph with $\deg(v) \geq n/2$ for every vertex v. Moreover suppose that of all counterexamples on n vertices, G is a graph with the largest possible number of edges.

G is not complete, since K_n has a Hamiltonian circuit, for $n \geq 3$. Therefore G has two nonadjacent vertices v_1 and v_n. By maximality the graph G_1 formed by adding the edge $\{v_1, v_n\}$ to G has a Hamiltonian circuit. Moreover this circuit uses the edge $\{v_1, v_n\}$, since otherwise G has a Hamiltonian circuit. So we may suppose that the Hamiltonian circuit in G_1 is of the form $v_1, v_2, \ldots, v_n, v_1$. Thus $v_1, \ldots, v_n$ is a Hamiltonian path in G.

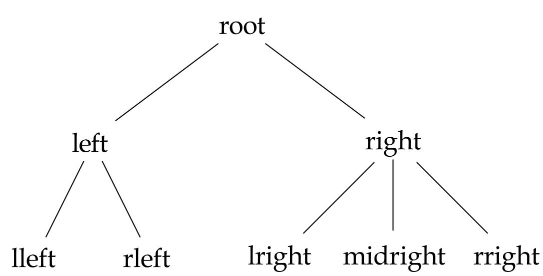
Let $k = \deg(v_1)$. If v_{i+1} is adjacent to v_1, then v_i cannot be adjacent to v_n since otherwise $v_1, ..., v_i, v_n, v_{n-1}, ..., v_{i+1}, v_1$ is a Hamiltonian circuit in G. Therefore, we have the contradiction

$$\deg(v_n) \leq (n-1) - k \leq n - 1 - n/2 = n/2 - 1. \rightarrow\leftarrow$$

□

WARNING: Do not read too much into this theorem. The condition is not a necessary condition. The 5-cycle, C_5, is obviously Hamiltonian, but the vertices all have degree 2 which is less than 5/2.

Trees form an important class of graphs. A **tree** is a connected graph with no circuits. Trees are traditionally drawn *upside down*, with the tree growing down rather than up, starting at a root vertex.



Theorem 38.18. *A graph G is a tree if and only if there is a unique path between any two vertices.*

Proof. Suppose that G is a tree, and let u and v be two vertices of G. Since G is connected, there is a path of the form $u = v_0, v_1, ..., v_n = v$. If there is a different path from u to v, say $u = w_0, w_1, ..., w_m = v$ let i be the smallest subscript so that $w_i = v_i$, but $v_{i+1} \neq w_{i+1}$. Also let j be the next smallest subscript where $w_j = v_j$. By construction $v_i, v_{i+1}, ..., v_j, w_{j-1}, w_{j-2}, ..., w_i$ is a circuit in G $\rightarrow\leftarrow$.

Conversely, if G is a graph where there is a unique path between any pair of vertices, then by definition G is connected. If G contained a circuit, C, then any two vertices of C would be joined by two distinct paths.$\rightarrow\leftarrow$ Therefore G contains no circuits, and is a tree. □

A consequence of theorem 38.18 is that given any vertex r in a tree, we can draw the tree with r at the top, as the root vertex, and the other vertices in levels below. The neighbors of r that appear at the first level below r are called r's **children**. The children of r's children are put in the second level below r, and are r's **grandchildren**. In general the ith level consists of those vertices in the tree which are at distance i from r. The result is called a **rooted tree**. The **height** of a rooted tree is the maximum level number.

Naturally, besides child and parent, many genealogical terms apply to rooted trees, and are suggestive of the structure. For example if a rooted tree has root r, and $v \neq r$, the **ancestors** of v are all vertices on the path from r to v, including r, but excluding v. The **descendants** of a vertex, w consist of all vertices which have w as one of their ancestors. The **subtree rooted at** w is the rooted tree consisting of w, its descendants, and all the required edges. A vertex with no children is a **leaf**, and a vertex with at least one child is called an **internal vertex**.

To distinguish rooted trees by breadth, we use the term m**-ary** to mean that any internal vertex has at most m children. An m-ary tree is **full** if every internal vertex has exactly m children. When $m = 2$, we use the term **binary**.

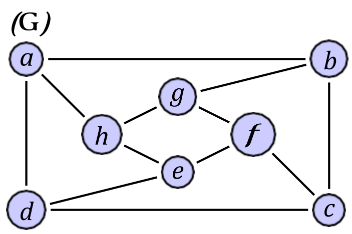
Theorem 38.19. *A tree on n vertices has $n-1$ edges.*

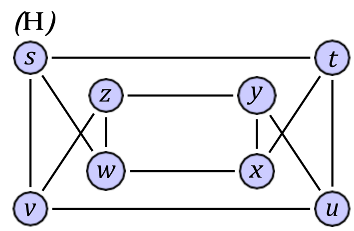
Proof. (*by induction on n.*)
Basis: Let $n = 1$, this is the trivial tree with 0 edges. So true the theorem is true for $n = 1$.

Inductive Step: Suppose that for some $n \geq 1$ every tree with n vertices has $n - 1$ edges. Now suppose T is a tree with $n + 1$ vertices. Let v be a leaf of T. If we erase v and the edge leading to it, we are left with a tree with n vertices. By the inductive hypothesis, this new tree will have $n-1$ edges. Since it has one less edge than the original tree, we conclude T has n edges. □

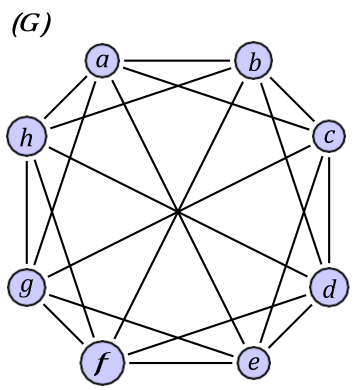
Exercises

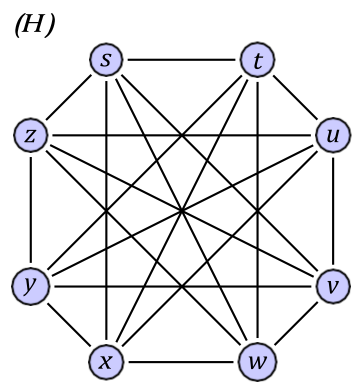
Exercise 38.1. *Find a graph isomorphism $\phi : G \to H$. Verify the adjacency preserving property by showing the adjacency matrices satisfy $A_G = A_H^{\phi}$*





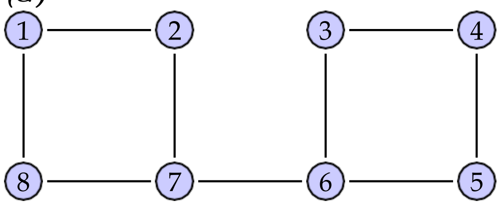
Exercise 38.2. *Prove that G and H are not isomorphic.*





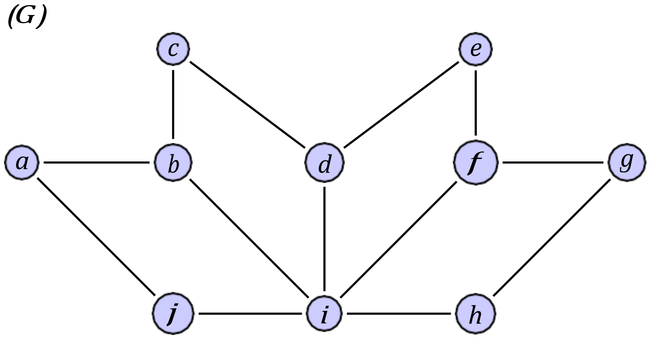
Exercise 38.3. *Redraw the graph* G *as a bipartite graph.*

(G)



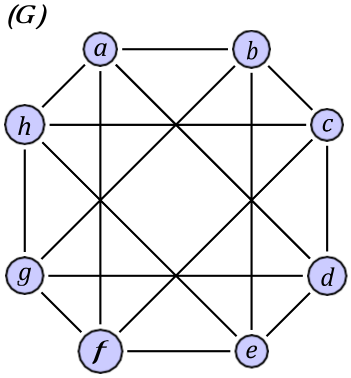
Exercise 38.4. *Explain why the graph* G *is not Eulerian, but is Hamiltonian.*

(G)

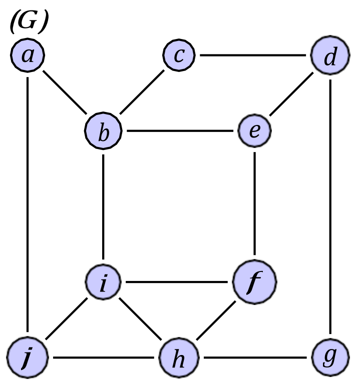


Exercise 38.5. *Find an Eulerian circuit for the graph* G *as a list of vertices.*

(G)



Exercise 38.6. *Prove that the graph G has no Hamiltonian circuit.*



Problems

Problem 38.1.

(a) *How many edges are there in K_n, the complete graph with n vertices?*

(b) *How many edges are there in C_n, the n-cycle with n vertices?*

(c) *How many edges are there in L_n, the n-link with n vertices?*

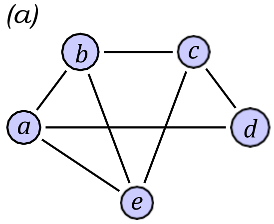
(d) *How many edges are there in W_n, the n-wheel with $n+1$ vertices?*

(e) *How many edges are there in Q_n, the n-cube with 2^n vertices?*

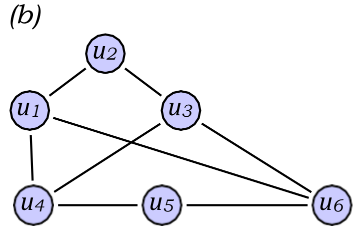
(f) *How many edges are there in $K_{m,n}$, the complete bipartite graph with m top and n bottom vertices?*

Problem 38.2. *Determine whether each graph is bipartite. If it is, redraw it as a bipartite graph.*

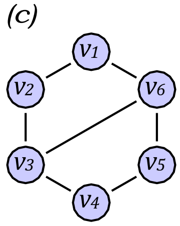
(a)



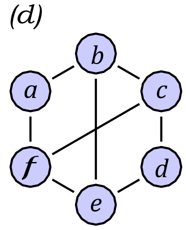
(b)



(c)



(d)

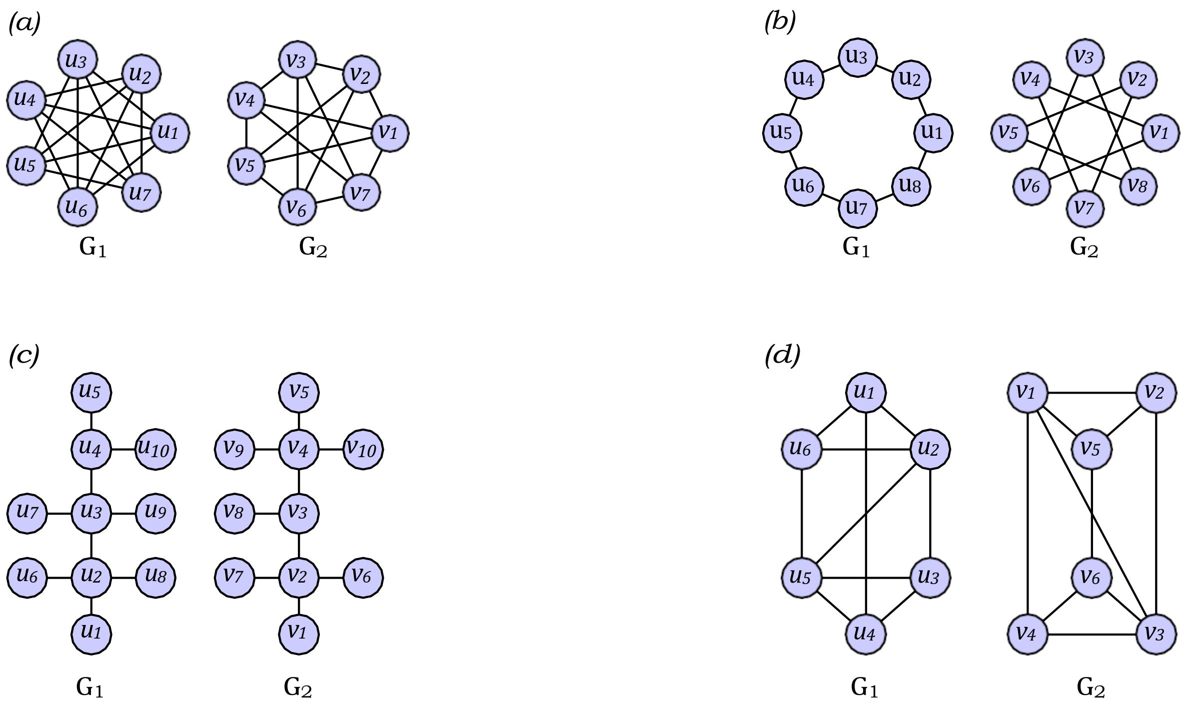


Problem 38.3. *(a) For which values of n is C_n bipartite?*

(b) For which values of n is Q_n bipartite?

Problem 38.4. *Draw the Petersen graph with vertices labeled with the ten different subsets of the five element set $\{a,b,c,d,e\}$ as suggested in example 38.5.*

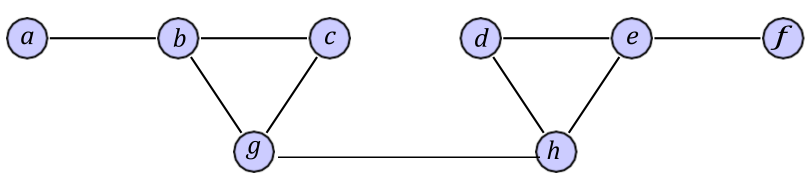
Problem 38.5. *For each pair of graphs either prove that* G_1 *and* G_2 *are not isomorphic, or else show they are isomorphic by exhibiting a graph isomorphism.*



Problem 38.6. *For the graph below*

(1) *Determine all the bridges.*

(2) *Determine all the cut vertices.*



Problem 38.7. *Prove theorem 38.7:* If G is a connected graph, then there is a simple path between any two different vertices.

Problem 38.8. *For each candidate degree sequence below, either draw a graph with that degree sequence or explain why that list cannot be the degree sequence of a graph.*

(1) $4, 4, 4, 4, 4$

(2) $6, 4, 4, 4, 4$

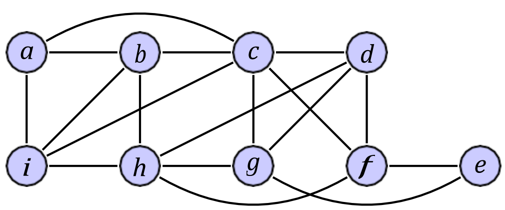
(3) $3, 2, 1, 1, 1$

(4) $3, 3, 2, 2, 1$

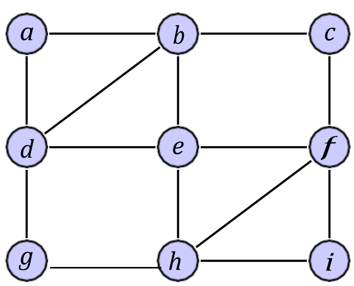
Problem 38.9. *A tree is called* **star-like** *if there is exactly one vertex with degree greater than* 2*. How many different (that it, nonisomorphic) star-like trees are there with six vertices? (Note: If you draw the graph with the vertex of of degree greater than* 2 *having the arms of the tree radiating out from it like spokes on a wheel, the name star-like will make sense.)*

Problem 38.10. *For each graph below (i) find an Eulerian circuit, or prove that none exists, and (ii) find a Hamiltonian circuit or prove that none exists.*

(a)



(b)



(c) The Petersen Graph. (See figure 38.3.)

(d) The 3*-cube* Q_3.

Problem 38.11. *Answer the following questions about the rooted tree shown below.*

(a) *Which vertex is the root?*

(b) *Which vertices are internal?*

(c) *Which vertices are leaves?*

(d) *Which vertices are children of* b*?*

(e) *Which vertices are grandchildren of* b*?*

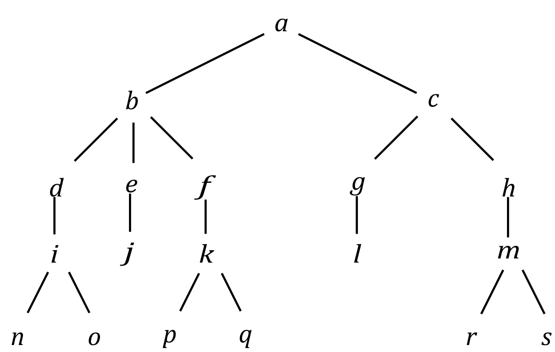
(f) *Which vertex is the parent of* m*?*

(g) *Which vertices are siblings of* q*?*

(h) *Which vertices are ancestors of* p*?*

(i) *Which vertices are descendants of* d*?*

(j) *What level is* i *at?*



Problem 38.12. *A forest is a graph consisting of one or more (separate) trees. If the total number of vertices in a forest is* f*, and the number of trees in the forest is* t*, what is the total number of edges in the forest?*

Appendix A

Answers to Exercises

Chapter 1

1.1.

a) yes

b) no

c) no

d) yes

1.2.

a)

p	q	p	$\oplus$	$\neg$	q
T	T	T	T	F	T
T	F	T	F	T	F
F	T	F	F	F	T
F	F	F	T	T	F

b)

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

c)

p	q	q	$\wedge$	$\neg$	p
T	T	T	F	F	T
T	F	F	F	F	T
F	T	T	T	T	F
F	F	F	F	T	F

d)

p	q	$\neg$	q	$\vee$	p
T	T	F	T	T	T
T	F	T	F	T	T
F	T	F	T	F	F
F	F	T	F	T	F

e)

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

1.3. a) (1101 0111 $\oplus$ 1110 0010) $\wedge$ 1100 1000 = (0011 0101) $\wedge$ 1100 1000 = 0000 0000

b) (1111 1010 $\wedge$ 0111 0010) $\vee$ (0101 0001) = (0111 0010) $\vee$ (0101 0001) = 0111 0011

c) (1001 0010 $\vee$ 0101 1101) $\wedge$ (0110 0010 $\vee$ 0111 0101) = (1101 1111) $\wedge$ (0111 0111) = 0101 0111

1.4.

a) $s \wedge \neg f$ b) $f \wedge \neg s$ c) $\neg s \to \neg f$

1.5.

a) Jordan did not play and the Wizards won.

b) If Jordan played, then the Wizards lost.

c) The Wizards won or Jordan played.

d) Jordan didn't play when the Wizards won. OR If the Wizards won, then Jordan did not play.

1.6.

a) If Sam plays chess with the white pieces, she wins.

b) $(c \wedge \neg w) \to b$

Chapter 2

2.1.

	p	q	r	(p	∨	q)	∨	r	p	∨	(q	∨	r)
	T	T	T	T	T	T	T	T	T	T	T	T	T
	T	T	F	T	T	T	T	F	T	T	T	T	F
	T	F	T	T	T	F	T	T	T	T	F	T	T
a)	T	F	F	T	T	F	T	F	T	T	F	F	F
	F	T	T	F	T	T	T	T	F	T	T	T	T
	F	T	F	F	T	T	T	F	F	T	T	T	F
	F	F	T	F	F	F	T	T	F	T	F	T	T
	F	F	F	F	F	F	F	F	F	F	F	F	F

	p	q	¬	p	∧	(p	∨	q)	¬	(q	→	p)
	T	T	F	T	F	T	T	T	F	T	T	T
b)	T	F	F	T	F	T	T	F	F	F	T	T
	F	T	T	F	T	F	T	T	T	T	F	F
	F	F	T	F	F	F	F	F	F	F	T	F

	p	q	r	p	∨	(q	∧	r)	(p	∨	q)	∧	(p	∨	r)
	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
	T	T	F	T	T	T	F	F	T	T	T	T	T	T	F
	T	F	T	T	T	F	F	T	T	T	F	T	T	T	T
c)	T	F	F	T	T	F	F	F	T	T	F	T	T	T	F
	F	T	T	F	T	T	T	T	F	T	T	T	F	T	T
	F	T	F	F	F	T	F	F	F	T	T	F	F	F	F
	F	F	T	F	F	F	F	T	F	F	F	F	F	T	T
	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F

2.2.

a)

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

Consider $(p, q, r) = (\mathrm{F}, \mathrm{T}, \mathrm{T})$, $(p, q, r) = (\mathrm{F}, \mathrm{T}, \mathrm{F})$, $(p, q, r) = (\mathrm{F}, \mathrm{F}, \mathrm{T})$, or $(p, q, r) = (\mathrm{F}, \mathrm{F}, \mathrm{F})$.

b)

p	q	p	$\to$	q	q	$\to$	p
T	T	T	T	T	T	T	T
T	F	T	F	F	F	T	T
F	T	F	T	T	T	F	F
F	F	F	T	F	F	T	F

Consider either (p, q) = (T, F) xor (p, q) = (F, T).

c)

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

Consider either $(p, q) = (\mathrm{T}, \mathrm{F})$ or $(p, q) = (\mathrm{F}, \mathrm{T})$.

2.3.

a)

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

b)

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

2.4. a) If I mow the lawn, then it is Saturday.

b) If it is not Saturday, then I do not mow the lawn.

c) If I do not mow the lawn, then it is not Saturday.

2.5. An implication is logically equivalent to its contrapositive. So, (If it is Saturday, then I mow the lawn) is logically equivalent to (If I do not mow the lawn, then it is not Saturday). The inverse and converse of an implication are logically equivalent (but not logically equivalent to the implication) so (If it is not Saturday, then I do not mow the lawn) is logically equivalent to (If I mow the lawn, then it is Saturday).

2.6.

a) $(p, q, r) = (\text{T}, \text{F}, \text{F}),\ (\text{F}, \text{T}, \text{F})$

b) $(p, q, r) = (\text{F}, \text{T}, \text{T}),\ (\text{F}, \text{F}, \text{T})$

2.7. *Proof.*

$(p \wedge \neg r) \to \neg q$	$\equiv \neg(p \wedge \neg r) \vee \neg q$	Disjunctive Form
	$\equiv (\neg p \vee \neg\neg r) \vee \neg q$	De Morgan's Law
	$\equiv (\neg p \vee r) \vee \neg q$	Double Negation Law
	$\equiv \neg p \vee (r \vee \neg q)$	Associative Law
	$\equiv \neg p \vee (\neg q \vee r)$	Commutative Law
	$\equiv \neg p \vee (q \to r)$	Disjunctive Form
	$\equiv p \to (q \to r)$	Disjunctive Form

□

Chapter 3

3.1.

a) T   b) F

c) F (e.g., $(0 \le 10) \to (2 \cdot 0 \ge 4)$ is false.)

d) T (e.g., $\neg(2 \cdot 0 \ge 4)$ is true.)

3.2.

a) Everyone in class has read *War and Peace*.

b) Someone in class has not read *The Great Gatsby*.

c) Someone in class has read every novel.

d) For each novel there is at least one person in the class who has read that novel.

3.3.

a) $\forall y\ F(\mathrm{I}, y)$

b) $\neg\exists y\, F(\text{George}, y)] \equiv \forall y\, \neg F(\text{George}, y)$

c) $\neg\exists x\, F(x, x) \equiv \forall x\, \neg F(x, x)$

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

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

f) $\exists y\, \exists z\, ((y \neq z) \wedge F(\text{Ralph}, y) \wedge F(\text{Ralph}, z))$

3.4.

a) Someone in class has not read *War and Peace*.

b) Everyone in class has read *The Great Gatsby*.

c) No student in class has read every novel.

d) There is a novel that no one in class has read.

3.5.

a) $\exists y\, \neg F(I, y)$

b) $\exists y\, F(\text{George}, y)$

c) $\exists x\, F(x, x)$

d) $\forall x\, \exists y\, \neg F(x, y)$

e) $\forall y\, \exists x\, \neg F(x, y)$

f) This one is a little complicated.
$\forall y \forall z((y = z) \vee \neg F(\text{Ralph}, y) \vee \neg F(\text{Ralph}, z))$.
In plain English name any two (not necessarily different) people. Either they are the same person, or else Ralph cannot fool at least one of the two. A logically equivalent, and maybe easier to fathom version converting the disjunctive form above to an implication:
$\forall y \forall z((F(\text{Ralph}, y) \wedge F(\text{Ralph}, z)) \rightarrow (y = z))$.
In plain English: If Ralph can fool both y and z, then y and z are actually the same person. Or, in more natural sounding English, Ralph can fool at most one person.

3.6. $\forall x\, \forall y\, ((E(x) \wedge O(y)) \rightarrow E(x \cdot y))$,
where $E(x)$: *x is an even integer* and $O(y)$: *y is an odd integer*.

3.7. "There is an x so P(x) holds, and, for any x and y, if both $P(x)$ and $P(y)$ hold, then x and y are equal."
Or, more succinctly: "There is exactly one x for which $P(x)$ is true."

Chapter 4

4.1.

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

The statement is a tautology, hence a valid rule of inference.

4.2. Proof: Make the assignments Porsche(x): "x owns a Porsche", Speeder(x): "x is a speeder", Sedan(x): "x owns a sedan",
and BuysPrem(x):"x buys premium fuel". The argument has the symbolic form:

$\forall x\,(\text{Porsche}(x) \to \text{Speeder}(x))$
$\neg\exists x\,(\text{Sedan}(x) \wedge \text{BuysPrem}(x))$
$\underline{\forall x\,(\neg\text{BuyPrem}(x) \to \neg\text{Speeder}(x))}$
$\therefore \forall x\,(\text{Porsche}(x) \to \neg\text{Sedan}(x))$

Proof:

1) $\forall x\,(\text{Porsche}(x) \to \text{Speeder}(x))$	Hypothesis
2) $\text{Porsche}(c) \to \text{Speeder}(c)$	Universal Instantiation (1)
3) $\forall x\,(\neg\text{BuyPrem}(x) \to \neg\text{Sedan}(x))$	Hypothesis
4) $\neg\text{BuyPrem}(c) \to \neg\text{Speeder}(c)$	Universal Instantiation (3)
5) $\text{Speeder}(c) \to \text{BuyPrem}(c)$	Contrapositive (4)
6) $\text{Porsche}(c) \to \text{BuyPrem}(c)$	Hypothetical syllogism (2) and (5)
7) $\neg\exists x(\text{Sedan}(x) \wedge \text{BuysPrem}(x))$	Hypothesis
8) $\forall x\,\neg(\text{Sedan}(x) \wedge \text{BuysPrem}(x))$	De Morgan's Law (Existential negation) (7)
9) $\neg(\text{Sedan}(c) \wedge \text{BuysPrem}(c))$	Universal Instantiation
10) $\neg\text{Sedan}(c) \vee \neg\text{BuysPrem}(c)$	De Morgan's Law (9)
11) $\neg\text{BuysPrem}(c) \vee \neg\text{Sedan}(c)$	Commutative law (10)
12) $\text{BuysPrem}(c) \to \neg\text{Sedan}(c)$	Disjunctive form (11)
13) $\text{Porsche}(c) \to \neg\text{Sedan}(c)$	Hypothetical syllogism (6) and (12)
14) $\forall x\,(\text{Porsche}(x) \to \neg\text{Sedan}(x))$	Universal generalization (13)

4.3.

1)	$\neg w$	Hypothesis
2)	$u \vee w$	Hypothesis
3)	u	Disjunctive syllogism
4)	$u \to \neg p$	Hypothesis
5)	$\neg p$	Modus Ponens (3) and (4)
6)	$\neg p \to (r \wedge \neg s)$	Hypothesis
7)	$r \wedge \neg s$	Modus Ponens (5) and (6)
8)	$\neg s$	Simplification
9)	$t \to s$	Hypothesis
10)	$\neg t$	Modus Tollens (8) and (9)
11)	$\neg t \vee w$	Addition

Chapter 5

5.1.

a) $\{2, 3, 4\}$ b) $\{-5, 5\}$ c) $\{4, 5, 6\}$

5.2.

a) $\{5x \mid x \in \mathbb{Z} \text{ where } -1 \leq x \leq 3\}$
b) $\{x \mid x \in \mathbb{N} \text{ where } x \leq 4\}$ or $\{x \mid x \in \mathbb{Z} \text{ where } -1 < x < 5\}$
c) $\{x \in \mathbb{R} \mid \pi \leq x < 4\}$ or $\{x \mid x \in \mathbb{R} \text{ where } \pi \leq x < 4\}$

5.3. Ex. **5.1**: a) 3 b) 2 c) 3 Ex. **5.2**: a) 5 b) 5 c) infinite set

5.4. This proposition is true. We may write this statement using symbols as $\forall x\,((x \in \emptyset) \to$ ("x has three toes")). The hypothesis, $(x \in \emptyset)$, of the implication is false, which makes the implication true.

5.5. $\mathcal{P}\{1, \{2\}\} = \{\emptyset, \{1\}, \{\{2\}\}, \{1, \{2\}\}\}$

5.6. True. Order and repetitions do not matter when the elements of a set are listed.

5.7. False. For example, 2 is in the set of even integers, but not in the set of integers that are multiples of four. In fact, just the reverse is true since if an integer is a multiple of four, then it is certainly even. So the set of integers that are multiples of four is a subset of the set of even integers.

Chapter 6

6.1.

a) $A \cap B = \{2, 4, 6, 7, 8\}$

b) $A \cup B = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$

c) $A - B = \{3, 5\}$

d) $B - A = \{1, 9\}$

6.2.

a) $A = (A - B) \cup (A \cap B) = \{1, 2, 5, 6, 7, 8, 9\}$

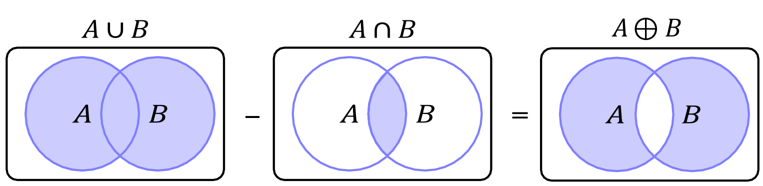
b) $B = (B - A) \cup (A \cap B) = \{3, 4, 5, 6, 9, 10\}$

6.3.

A	B	A	$\oplus$	B	$(A$	$\cup$	$B)$	$-$	$(A$	$\cap$	$B)$
1	1	1	0	1	1	1	1	0	1	1	1
1	0	1	1	0	1	1	0	1	1	0	0
0	1	0	1	1	0	1	1	1	0	0	1
0	0	0	0	0	0	0	0	0	0	0	0

The columns in red are identical, and that shows the set equality is correct.

6.4.



6.5. Let $\mathcal{U}$ denote the universal set.

Proof:

$A \cup (A \cap B)$	$= (A \cap \mathcal{U}) \cup (A \cap B)$	$\cap$-identity
	$= A \cap (\mathcal{U} \cup B)$	Distributive law
	$= A \cap \mathcal{U}$	$\cup$-domination
	$= A$	$\cap$-identity

6.6.

a) $A \times B = \{(1,a),(1,b),(1,c),(2,a),(2,b),(2,c),(3,a),(3,b),(3,c),(4,a),(4,b),(4,c)\}$

b) $B \times A = \{(a,1),(a,2),(a,3),(a,4),(b,1),(b,2),(b,3),(b,4),(c,1),(c,2),(c,3),(c,4)\}$

c) $C \times B \times D = \{(\alpha,a,7),(\alpha,a,8),(\alpha,a,9),(\alpha,b,7),(\alpha,b,8),(\alpha,b,9),(\alpha,c,7),(\alpha,c,8),(\alpha,c,9),$
$(\beta,a,7),(\beta,a,8),(\beta,a,9),(\beta,b,7),(\beta,b,8),(\beta,b,9),(\beta,c,7),(\beta,c,8),(\beta,c,9)\}$

6.7. $A = \emptyset$, $B = \emptyset$, or $A = B$

6.8. $\chi(\{1,2,4,8\}) = 0110100010$

6.9. The elements in B are $(1,1),(2,1),(2,2),(3,1),(3,2),(3,3)$.

Chapter 7

7.1. Proof Specification: $(\forall m,n \in \mathbb{Z})[((m \text{ is even}) \wedge (n \text{ is even})) \rightarrow (m+n \text{ is even})]$
Proof: (List form)

- ► Let $m,n \in \mathbb{Z}$ be given.
- ► Suppose m and n are even.
- ► $\Rightarrow$ $m = 2i$ and $n = 2j$, for some $i,j \in \mathbb{Z}$, by the definition of *even*.
- ► $\Rightarrow$ $m + n = 2i + 2j = 2(i + j)$

- $\Rightarrow$ $m + n = 2\ell$, where $\ell = i + j$.
- $\Rightarrow$ $m + n$ is even, by the definition of *even*.
- $\therefore$ if m and n are even, then $m + n$ is even. □

Proof. (Prose form) Suppose m and n are even integers. Then, by the definition of *even*, $m = 2i$ and $n = 2j$, for some $i, j \in \mathbb{Z}$. Thus, we have

$$m + n = 2i + 2j = 2(i + j) = 2\ell, \text{ where } \ell = i + j.$$

Hence, $m + n$ is even, by the definition of *even*. □

7.2. $(\forall n \in \mathbb{Z})[(n^2 \text{ is odd}) \to (n \text{ is odd})]$
$\equiv (\forall n \in \mathbb{Z})[\neg(n \text{ is odd}) \to \neg(n^2 \text{ is odd})]$
$\equiv (\forall n \in \mathbb{Z})[(n \text{ is even}) \to (n^2 \text{ is even})]$

- **Proof:** (List form)
- Let n $\in \mathbb{Z}$ be given.
- Suppose n is even.
- $\Rightarrow$ $n = 2j$, for some $j \in \mathbb{Z}$, by the definition of *even*.
- $\Rightarrow$ $n^2 = (2j)^2$
- $==\Rightarrow$ $n^2 = 2(2j^2) = 2\ell$, where $\ell = 2j^2$
- $==\Rightarrow$ n^2 is even, by the definition of *even*.
- $\therefore$ if n is even, then n^2 is even. □

Proof. (Prose form) Suppose n is an even integer. Then, by the definition of *even*, $n = 2j$, for some $j \in \mathbb{Z}$. Thus, we have

$$n^2 = (2j)^2 = 2(2j^2) = 2\ell, \text{ where } \ell = 2j^2.$$

Hence, n^2 is even, by the definition of *even*. □

7.3.

$$(\forall x, y \in \mathbb{R})[((x \text{ is rational}) \land (y \text{ is irrational})) \to (x + y \text{ is irrational})] \equiv$$
$$(\forall x, y \in \mathbb{R})[((x \text{ is rational}) \land (y \text{ is irrational}) \land \neg(x + y \text{ is irrational})) \to \mathbb{F}] \equiv$$
$$(\forall x, y \in \mathbb{R})[((x \text{ is rational}) \land (y \text{ is irrational}) \land (x + y \text{ is rational})) \to \mathbb{F}]$$

- **Proof:** (List form)
- Let $x, y \in \mathbb{R}$ be given
- Suppose x is rational and y is irrational
- Suppose for the sake of argument that $x + y$ is rational
- $\Rightarrow y = (x + y) - x$ is rational because the difference of rational numbers is rational
- $\Rightarrow$ y is rational and y is irrational $\rightarrow\leftarrow$
- This is impossible
- $\therefore$ $x + y$ must have been irrational. □

Proof. (Prose form) Suppose x is rational and y is irrational. Suppose, for the sake of argument, that $x + y$ is rational. Then, $y = (x + y) - x$ would be rational, since the difference of two rational numbers is rational. But, y was given to be irrational. This is impossible. Therefore, $x + y$ must have been irrational. □

7.4.

$$(\forall n \in \mathbb{Z})[(5n - 1 \text{ is odd}) \rightarrow (n \text{ is even})] \equiv$$
$$(\forall n \in \mathbb{Z})[((5n - 1 \text{ is odd}) \wedge \neg(n \text{ is even})) \rightarrow \mathbb{F}] \equiv$$
$$(\forall n \in \mathbb{Z})[((5n - 1 \text{ is odd}) \wedge (n \text{ is odd})) \rightarrow \mathbb{F}]$$

- **Proof:** (List form)
- Let $n \in \mathbb{Z}$ be given
- Suppose $5n - 1$ is odd.
- Suppose for the sake of argument that n is odd
- $\Rightarrow$ $n = 2j + 1$, for some $j \in \mathbb{Z}$, by the definition of *odd*
- $\Rightarrow$ $5n - 1 = 5(2j + 1) - 1 = 10j + 4 = 2(5j + 2)$
- $\Rightarrow$ $5n - 1 = 2\ell$, where $\ell = 5j + 2$
- $\Rightarrow$ $5n - 1$ is even, by the definition of *even*
- But, $5n - 1$ was given to be odd $\rightarrow\leftarrow$
- This is impossible
- $\therefore$, n must have been even. □

Proof. (Prose form) Suppose that $5n - 1$ is odd for some $n \in \mathbb{Z}$. And, suppose, for the sake of argument, that n is odd. That is, $n = 2j + 1$, for some $j \in \mathbb{Z}$. Then, we have

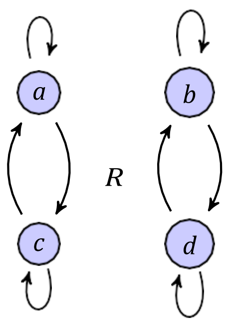
$$5n - 1 = 5(2j + 1) - 1 = 10j + 4 = 2(5j + 2).$$

That is, $5n - 1$ is even, since $5n - 1 = 2\ell$, for $\ell = 5j + 2$. But, $5n - 1$ was given to be odd. This is impossible. Therefore, n must have been even. $\square$

7.5. The positive integer 77 ends in a 7, but is not prime since $77 = 7 \cdot 11$.

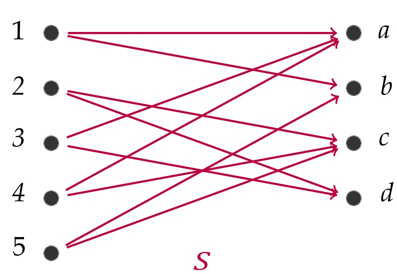
Chapter 8

8.1.



$$M_R = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}.$$

8.2. $S = \{(1, a), (1, b), (2, c), (2, d), (3, a), (3, d), (4, a), (4, c), (5, b), (5, c)\}$



8.3. $M_{R\circ S} = M_S \odot M_R = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}$

8.4. (a)

$R_1 \cup R_2$ = {(1, 2), (1, 3), (1, 5), (1, 6), (2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (3, 1), (3, 3), (3, 4), (3, 6), (4, 1), (4, 2), (4, 3), (4, 4), (4, 5), (5, 1), (5, 5), (5, 6), (6, 2), (6, 3), (6, 6)}

$R_1 \cap R_2$ = {(1, 2), (2, 1), (2, 2), (3, 3), (5, 5), (6, 6)}

$R_1 \oplus R_2$ = {(1, 3), (1, 5), (1, 6), (2, 3), (2, 4), (2, 5), (3, 1), (3, 4), (3, 6), (4, 1), (4, 2), (4, 3), (4, 4), (4, 5), (5, 1), (5, 6), (6, 2), (6, 3)}

$$b)\ M_{R_1 \cup R_2} = M_{R_1} \vee M_{R_2} = \begin{bmatrix} 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}$$

$$M_{R_1 \cap R_2} = M_{R_1} \wedge M_{R_2} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$M_{R_1 \oplus R_2} = M_{R_1} \oplus M_{R_2} = \begin{bmatrix} 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}$$

Chapter 9

9.1. $R = \{(1,1),(2,2),(3,3)\}$ (There are seven more correct answers since any subset of R would also be okay.)

9.2.

a) R is not reflexive since, for example $(1,1) \notin R$.
b) R is irreflexive since none of $(1,1),(2,2),(3,3),(4,4)$ are in R.
c) R is symmetric since the reverse of each ordered pair in R is also in R.
d) R is not antisymmetric since, for example, $(1,2)$ and $(2,1)$ are both in R, but $1 \neq 2$.
e) R is not transitive since, for example, $(1,2)$ and $(2,3)$ are both in R, but $(1,3)$ is not in R.

9.3.

a) reflexive, not irreflexive, symmetric, not antisymmetric, transitive
b) not reflexive, not irreflexive, not symmetric, antisymmetric, transitive
c) not reflexive, not irreflexive, symmetric, not antisymmetric, not transitive
d) reflexive, not irreflexive, symmetric, not antisymmetric, transitive

9.4. Recall that for a set A, the notation $|A|$ is the cardinal number of A (in other words, the number of elements in A).

a) C is reflexive since any set A has the same number of elements as itself!
b) C is not irreflexive since, for example, $|\{1\}| = |\{1\}|$, so $\{1\}\, C\, \{1\}$ is true.
c) C is not symmetric since, for example, $\{1\}\, C\, \{1,2\}$ is true, but $\{1,2\}\, C\, \{1\}$ is false.
d) C is not antisymmetric since, for example, $\{1\}\, C\, \{2\}$ and $\{2\}\, C\, \{1\}$ are both true, but $\{1\} \neq \{2\}$.
e) C is transitive since if $|A| \le |B|$ and $|B| \le |D|$, then $|A| \le |D|$ (by the transitive property of the *less than or equal to* relation for numbers).

9.5. For any universe of discourse $\mathcal{U}$, a relation on $\mathcal{U}$ is defined to be any subset of $\mathcal{U} \times \mathcal{U}$. The empty set is a subset of $\mathcal{U} \times \mathcal{U}$.

9.6. $\{1,2\}M\{1,2\}$ is false, so M is not reflexive. $\{1\}M\{1\}$ is true, so M is not irreflexive. If $|A \cap B| = 1$ then $|B \cap A| = 1$, (since $A \cap B = B \cap A$), so M is symmetric. $\{1\}M\{1,2\}$ and $\{1,2\}M\{1\}$ are true, but $\{1\} \neq \{1,2\}$ so M is not antisymmetric. $\{1\}M\{1,2\}$ and $\{1,2\}M\{2\}$ are true, but $\{1\}M\{2\}$ is false, so M is not transitive.

Chapter 10

10.1. The relation R is reflexive since $(0,0)$, $(1,1)$, $(2,2)$ are all in R. R is symmetric since the reverse of each ordered pair in R is also in R. Finally, R is transitive (there are a lot of cases to check for this condition. For example, $(1,1)$ and $(1,0)$ are both in R, so we need to check that $(1,0)$ is also in R! But it is of course. There is a total of nine such checks needed to verify that R is transitive, but they are all just as automatic as that one.) Alternatively, the transitive property follows from verifying $M_R \odot M_R \leq M_R$. So R is an equivalence relation on A. There are two equivalence classes: $[0] = [1] = \{0,1\}$ and $[2] = \{2\}$.

10.2. The relation R is not reflexive on A since $(3,3)$ is not in R. So R is not an equivalence relation on A. Notice that R is symmetric and transitive.

10.3. The relation R is not symmetric on A since $(1,0)$ is in R, but $(0,1)$ is not in R. So R is not an equivalence relation on A. Notice that R is reflexive on A and is transitive.

10.4. The relation R is not transitive on A since $(0,1)$ and $(1,2)$ are in R, but $(0,2)$ is not in R. So R is not an equivalence relation on A. Notice that R is reflexive on A and is symmetric.

10.5. True. The relation is symmetric since the reverse of each ordered pair in R is also in R (after all, the reverse of each ordered pair in R is just the ordered pair itself). To check the antisymmetric condition, we need to look at all cases where an ordered pair and its reverse are both in R, and make sure the two coordinates are equal in each such case. There are only two cases to check: $(1,1)$ and its reverse are both in R, and sure enough, $1 = 1$. The story is the same for $(2,2)$. So R passes the antisymmetry test. The moral of the story: It is possible for a relation to be both symmetric and antisymmetric.

10.6. The relation S is reflexive since, for any integer m, $m^2 = m^2$. S is symmetric since if $m^2 = n^2$, then $n^2 = m^2$. Finally, S is transitive since if $m^2 = n^2$ and $n^2 = t^2$, then $m^2 = t^2$. So S is an equivalence relation on $\mathbb{Z}$.

The equivalence class of an integer m the the set of all integers with the same square as m. For $m = 0$, the only element in the equivalence class would be 0 itself: $[0] = \{0\}$. For any integer, m, other than 0, we get $[\underline{m}] = \{m, -m\}$. For example, $[2] = \{2, -2\}$ since $2^2 = 4$ and $(-2)^2 = 4$.

10.7. The relation C is obviously reflexive and symmetric. But it is not transitive. For example, The lines $l_1\colon y = x$ crosses the $l_2\colon x + y = 2$ at the point $(1, 1)$ and the line $l_2\colon x + y = 2$ crosses the line $l_3\colon y = x + 2$ at the point $(0, 2)$. But the lines $l_1\colon y = x$ and $l_3\colon y = x + 2$ are distinct parallel lines, so they do not cross. In other words $l_1 C l_2$ and $l_2 C l_3$ are both true, but $l_1 C l_3$ is false. So C is not transitive.

10.8. Proof Specification: $(\forall a \in A)[aRa]$
Definitions: R is symmetric $\equiv (\forall a, b \in A)[aRb \to bRa]$
R is transitive $\equiv (\forall a, b, c \in A)[aRb \land bRc \to aRc]$
Supposition: $(\forall a \in A)(\exists b \in A)[aRb]$
Proof:

- Let $a \in A$ be given, by Universal Instantiation
- There is a $b \in A$ so that aRb, by supposition
- bRa, since R is symmetric
- Thus, we have aRb and bRa, by conjunction
- Thus, aRa, since $aRb \land bRa \to aRa$ by transitivity
- Therefore, R is reflexive
- Since R is reflexive, symmetric and transitive, it is an equivalence relation

10.9. Proof specification:

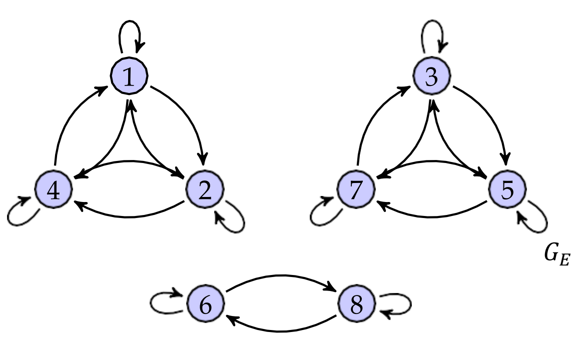
$$(\forall a, b \in A)\,[([a] \cap [b] = \emptyset) \lor ([a] = [b])] \equiv (\forall a, b \in A)\,[\neg([a] \cap [b] = \emptyset) \to ([a] = [b])]$$
$$\equiv (\forall a, b \in A)\,[([a] \cap [b] \neq \emptyset) \to ([a] = [b])] \equiv (\forall a, b \in A)\,[([a] \cap [b] \neq \emptyset) \to ([a] \subseteq [b] \land [b] \subseteq [a])]$$
$$\equiv (\forall a, b \in A)\,[([a] \cap [b] \neq \emptyset \to [a] \subseteq [b]) \land ([a] \cap [b] \neq \emptyset \to [b] \subseteq [a])]$$

Definition: $[a] \subseteq [b] \equiv (\forall x \in A)[x \in [a] \to x \in [b]]$

- **Proof:** (List form)
- Let $a, b \in A$ be given
- Suppose that $[a] \cap [b] \neq \emptyset$.
- $\Rightarrow$ there is a $c \in [a] \cap [b]$
- $\Rightarrow$ $c \in [a]$ and $c \in [b]$
- $\Rightarrow$ cEa and cEb, by definitions of $[a]$ and $[b]$
- $\Rightarrow$ (*) aEc and bEc, by symmetry
- We show that $[a] \subseteq [b]$
- Suppose $x \in [a]$ is given.
- $\Rightarrow$ xEa and aEc, by definition of $[a]$ and line $*$
- $\Rightarrow$ xEc, by transitivity
- $\Rightarrow$ xEc and cEb, by line $*$
- $\Rightarrow$ xEb, by transitivity
- $\Rightarrow$ $x \in [b]$, by definition of $[b]$
- Therefore, $[a] \subseteq [b]$.
- We show that $[b] \subseteq [a]$
- Suppose $y \in [b]$ is given.
- $\Rightarrow$ yEb and bEc, by definition of $[b]$ and line $*$
- $\Rightarrow$ yEc, by transitivity
- $\Rightarrow$ yEc and cEa, by $*$
- $\Rightarrow$ yEa, by transitivity
- $\Rightarrow$ $y \in [a]$, by definition of $[a]$
- Therefore, $[b] \subseteq [a]$.
- Thus, $[a] \cap [b] \neq \emptyset$ implies $[a] = [b]$.

10.10.

a)



b) Using the natural order [1, 2, 3, 4, 5, 6, 7, 8]:

$$M_E = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \end{bmatrix}$$

Using an order grouped by equivalence classes [1, 2, 4|3, 5, 7|6, 8]:

$$M_E = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

10.11. A matrix represents an equivalence relation if it defines a partition of $A = \{a, b, c, d, e, f, g, h\}$.

(a) $[a] = \{a, c, e, g\} = [c] = [e] = [g]$, $[b] = \{b, d, f\} = [d] = [f]$, and $[h] = \{h\}$ partition A.

(b) $[a] = \{a, b, g, h\} = [b] = [g] = [h]$, $[c] = \{c, d\} = [d]$, and $[e] = \{e, f\} = [f]$ partition A.

10.12. Here are the facts we know: (1) E is an equivalence relation on the set A, (2) $a \in [b]$ (a is in the equivalence class of b). Our job is to show that if $c \in [a]$, then $c \in [b]$. So, suppose $c \in [a]$. That means cEa is true according to the definition of equivalence class. We also know aEb is true, since a is in the equivalence class of b. Since cEa and aEb are true, the transitive condition tells us cEb is true, and that means $c \in [b]$, as we needed to show. □

Chapter 11

11.1. There are many different correct answers to these problems.

a) $f(x) = x + 1$

b) $f(x) = 1$

c) $f(x) = e^x$

d) $f(x) = x^3 - x$

11.2.

a) As a set of ordered pairs $f = \{(1, a), (2, b), (3, c), (4, d), (5, e)\}$

b) No such function can exist since $|B| > |A|$.

c) No such function can exist since $|A| < |B|$.

d) As a set of ordered pairs $g = \{(a, 1), (b, 2), (c, 3), (d, 4), (e, 5), (f, 5)\}$

11.3. The composition of two functions is a function, but the composition of two equivalence relations need not be an equivalence relation. However, the composition of an equivalence relation *with itself* will be an equivalence relation. Here is a proof.

Suppose E is an equivalence relation on a set A.

Since E is reflexive on A, for any $a \in A$, $(a, a) \in E$ is true, and since (a, a) and (a, a) are in E, the composition rule tells us $(a, a) \in E \circ E$! So $E \circ E$ is reflexive on A.

Next, suppose $(a, b) \in E \circ E$. That means there is an x in A such that (a, x) and (x, b) are in E. Since E is an equivalence relation, it is symmetric, so (b, x) and (x, a) are in E. The composition rule then tells us (b, a) is in $E \circ E$. That proves $E \circ E$ is symmetric.

Finally, suppose (a, b) and (b, c) are in $E \circ E$. That means there an x in A such that (a, x) and (x, b) are in E and there is a y in A such that (b, y) and (y, c) are in E. From (x, b) and (b, y) in E, we see (x, y) is in E. Then from (a, x) and (x, y) in E, we get (a, y) is in E. And finally, (a, y) and (y, c) in E tells us (a, c) is in $E \circ E$. So $E \circ E$ is transitive.

Putting the pieces together, we have proved $E \circ E$ is an equivalence relation on A.

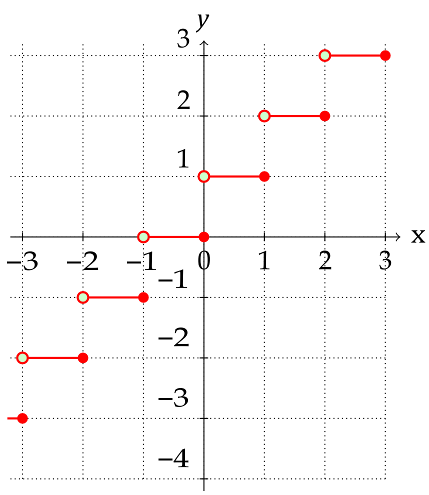
11.4. Suppose $g\colon A \to B$ and $f\colon B \to C$ are both one-to-one. To show $f \circ g\colon A \to C$ is one-to-one, suppose x and y are in A, and $(f \circ g)(x) = (f \circ g)(y)$. We want to show $x = y$.

Since $(f \circ g)(x) = (f \circ g)(y)$, the definition of the composition of functions tells us $f(g(x)) = f(g(y))$. Since f is one-to-one, we conclude $g(x) = g(y)$, and then, since g is one-to-one, we get $x = y$. □

Chapter 12

12.1. $\lceil x \rceil$ is the smallest integer greater than or equal to x.

12.2.



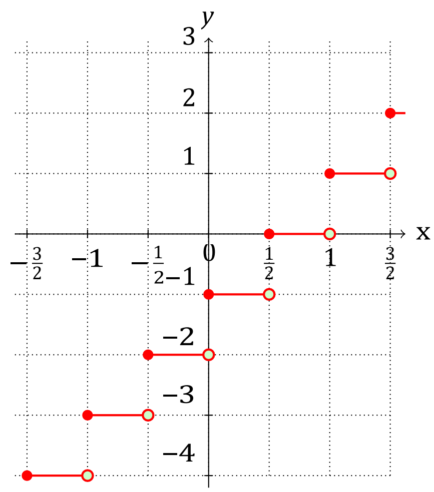
12.3.

We could fall back on the classic *plot-a-billion-points* method of graphing, but it is a better idea to think first. The graph ought to look a lot like the graph of $y = \lfloor x \rfloor$, but the jumps will move to new spots. In particular, there will be a jump up by 1 whenever x reaches a value where $2x - 1 = n$, for n equal an integer. In other words, there will be jumps when $x = (n+1)/2$ for integers n.

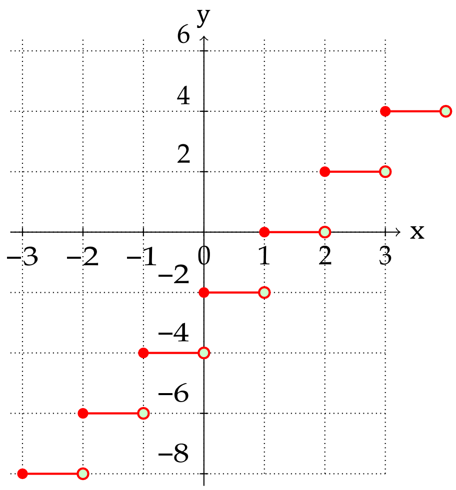
So the graph will show jumps at

$$\ldots, -2 = -\frac{4}{2}, -\frac{3}{2}, -1 = -\frac{2}{2}, -\frac{1}{2}, 0 = \frac{0}{2}, \frac{1}{2}, 1 = \frac{2}{2}, \frac{3}{2}, 2 = \frac{4}{2}, \ldots$$

Moreover, for the integer n, the jump at $x = (n+1)/2$ will be from $n - 1$ to n. That's enough to draw the graph: it looks exactly like the graph of $y = \lfloor x \rfloor$, except jumps come at multiples of $1/2$ instead of at the integers and care is needed with the heights of each step. That makes the graph easy to draw since we can *cheat* by dividing all the integer marks on the x-axis by 2 in the graph of $y = \lfloor x \rfloor$.



12.4. Thinking: the graph of $y = \lfloor x - 1 \rfloor$ will look just like the graph of $y = \lfloor x \rfloor$ shifted one unit to the right, and the factor of 2 in $y = 2\lfloor x - 1 \rfloor$ will move each horizontal segment of the graph to twice its original distance from the x-axis. Put those two pieces together:

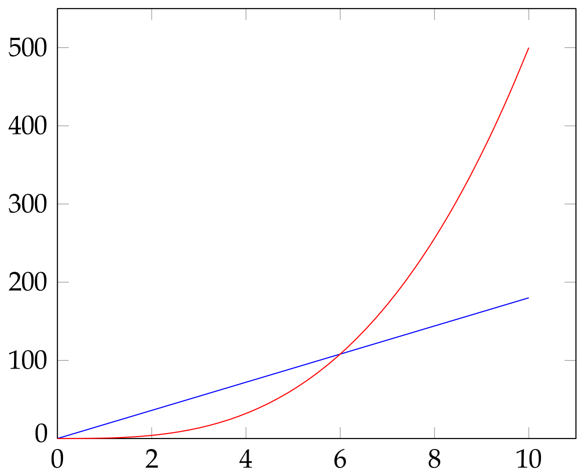


12.5. The functions $f(x) = 18x$ and $g(x) = x^3/2$ cross when $18x = \frac{x^3}{2}$. That is, when

$$\begin{aligned} 36x &= x^3 \\ x^3 - 36x &= 0 \\ x(x^2 - 36) &= 0 \\ x(x-6)(x+6) &= 0 \\ x &= -6, 0, 6 \end{aligned}$$

as seen in the graph.

So, for $x > 0$, g catches f at $x = 6$ for $x > 0$.



12.6. The exponential function $g(x) = 2^x$ ultimately beats the power function $f(x) = 4x^5$. To show that, lets find a number $x > 1$ so that $\ln(2^x) > \ln(4x^5)$. That inequality can be rewritten as $x \ln 2 > \ln 4 + 5 \ln x$. Since all we care about is showing there is an x for which that inequality is true, we can make some simplifying substitutions: if $x/2 > 2 + 5 \ln x$, that would guarantee $x \ln 2 > \ln 4 + 5 \ln x$ because $x \ln 2 > x/2$. So let's see if we can find x so that $x > 4 + 10 \ln x$, and since we can make x as large as we want, we can also guarantee $\ln x \geq 4$. Consequently we need only be sure we can pick x with $x > 11 \ln x$. Thinking about the graphs of $y = x$ and $y = 11 \ln x$, we can see there will certainly be such an x. If we know a bit of calculus, we can compute slopes of tangent lines to the two curves to show there is such an x. If we have a computer algebra system, we can find a specific value of such an x. In fact, it turns out that 41 is the smallest integer greater than 1 for which $x > 11 \ln x$.

12.7. Assuming the other buttons are in working order, we can use the fact $\log(2^{\sqrt{2}}) = \sqrt{2} \log 2$. Here are the steps:

- get square root of 2 (result: $1.41421 \cdots$)
- multiply by $\log 2$ (result: $.42572 \cdots$)
- take inverse log (10^x button for if you are using base 10 logs) (result $2.66514 \ldots = \log(2^{\sqrt{2}})$.

Chapter 13

13.1. A great source of information about numerical sequences is *The Online Encyclopedia of Integer Sequences* at http://oeis.org. That site finds over 200 sequences that either begin or are otherwise related to $1, 2, 4, 5, 7, 8$

One possible answer: It looks like the list of positive integers in order, skipping the multiples of 3. So the next few terms will be $10, 11, 13, 14, 16, 17$.

Another answer: The positive integers n (in increasing order) for which $2n+3$ is a prime. The next few terms would be $10, 13, 14, 17, 19, 20$.

13.2. $a_{100} = 2 + 6(99) = 596$ where the initial term is $a_1 = 2$.

13.3. Let a be the initial term and d be the common difference. We have the system $a + 9d = -4$ and $a + 15d = 47$. Subtracting the first from the second gives $6d = 51$, so $d = 51/6 = 17/2$, and then $a = -4 - 9d = -4 - 9(17/2) = -161/2$. That means the 11^{th} term is $-161/2 + 10(17/2) = 9/2$.

13.4. $g_5 = 6(2^4) = 6(16) = 96$ where the initial term is $g_1 = 6$.

13.5. If r is the common ratio, then $g_2 = -11 = g_1 r = 5r$, so $r = -11/5$. That means $g_5 = g_1 r^4 = 5\left(-\frac{11}{5}\right)^4 = \frac{14641}{125}$.

13.6. Suppose we have a sequence with initial term a that is arithmetic (with common difference d) as well as geometric (with common ration r). For the two terms following the initial term, we have $a + d = ar$ and $a + 2d = ar^2$. The first equation tells us that $d = a(r-1)$. Subtracting the first equation from the second gives $d = ar^2 - ar = ar(r-1)$. So $a(r-1) = ar(r-1)$, and that equation leaves only a few options for a and r. It could be that $a = 0$, and that means (1) $d = 0$ and the sequence is $0, 0, 0, 0 \ldots$, or (2) $r - 1 = 0$, so $r = 1$, and that means $d = 0$, and the sequence is $a, a, a, a, \ldots$, or (3) neither $a = 0$ nor $r = 1$ in which case $a(r-1) = ar(r-1)$ reduces to $r = 1$ which can't be in this case. Conclusion: the constant sequences $a, a, a, a, \ldots$ are the only sequences that are both arithmetic (initial term a and common difference 0) and geometric (initial term a and common ratio 1).

13.7. $\sum_{j=1}^{4}(j^2 + 1) = 2 + 5 + 10 + 17 = 34$

13.8. $\sum_{k=-2}^{4}(2k - 3) = -7 - 5 - 3 - 1 + 1 + 3 + 5 = -7$.

13.9. The initial terms is 2, the 100th term is 2 + (99)(6), so the total is $100\left(\frac{2+2+99*6}{2}\right) = 29900$.

13.10. 6 + 12 + 24 + 48 + 96 = 186

13.11. $1 - \frac{3}{2} + \frac{9}{4} - \frac{27}{8} + \frac{81}{16} = \frac{55}{16}$

13.12. $\sum_{k-1}^{n} \frac{1}{2k}$

Chapter 14

14.1. 3, 15, 255, 65535, 4294967295

14.2. 1, 1, 2, 3, 7, 22, 155

14.3. 1, 2, 3, 3, 4, 4, 4, 4, 5, 5

14.4. 1, 0, 3, 2, 5, 4, 7, 6, 9, 8. It looks like the terms alternately give one more and one less then the term index. That suggests $a_n = n + (-1)^n$.

14.5. This is the *Look-and-Say sequence* introduced by John Horton Conway. After the initial term equal to 1, each new term is produced by *reading* the previous term. Examples:

- for the second term, read the first term (1) as "*one* 1" (so 11)
- for the next term, read 11 as "*two* 1" (so 21)

- next, read 21 as "*one* 2 *and one* 1" (so 1211)
- next 1211 is "*one* 1, *one* 2, *and two* 1" (so 111221)

and so on. The term following 1113213211 is 31131211131221.

14.6. For $n = 1$, we define $1d$ to equal d. Now for the recursive part of the definition: For $n > 1$, we define $nd = (n-1)d + d$.

Chapter 15

15.1. S is the set of positive integers.

15.2. (1) $17 \in S$, and (2) If $n \in S$, then $n + 100 \in S$.

15.3. (1) $1, 2, 3 \in$ S, and (2) If $n \in S$, then $n + 4 \in S$.

15.4. Each string in S consists of zero or more a's followed on the right by an equal number of the two letter combination bc. Examples: *aabcbc* and *aaaabcbcbcbc*.

15.5. The strings in S consist of one or more c's preceded on the left by any combination of zero or more a's and b's. Examples: *cccc, abbbac, babbacc, aaacc*.

15.6. (1) $\lambda, a, b, c \in S$, and (2) If $x \in S$, then axa, bxb, and $cxc \in S$.

alternate solution: (1) $\lambda, a, b, c \in S$, and (2) If $x, y \in S$, the $xyx \in S$. In plain English, the recursive rule (2) says we can build longer palindromes by adding a palindrome to either end of a palindrome.

Chapter 16

16.1. basis: For n = 1, the left side is $1 \cdot 3 = 3$, and the right side is $\frac{1 \cdot 2 \cdot 9}{6} = 3$, so the equality is correct for this case.

inductive hypothesis: Suppose

$$1 \cdot 3 + 2 \cdot 4 + 3 \cdot 5 + \cdots + n(n+2) = \frac{n(n+1)(2n+7)}{6}$$

for some $n \geq 1$. Then

inductive step:

$$1 \cdot 3 + 2 \cdot 4 + 3 \cdot 5 + \cdots + n(n+2) + (n+1)(n+3)$$

$$= \frac{n(n+1)(2n+7)}{6} + (n+1)(n+3) \text{ by induction hypothesis.}$$

$$= \frac{n(n+1)(2n+7)}{6} + \frac{6(n+1)(n+3)}{6} \text{ (factor out } (n+1))$$

$$= \frac{(n+1)[n(2n+7) + 6(n+3)]}{6} \text{ (now it's just algebra)}$$

$$= \frac{(n+1)[2n^2 + 7n + 6n + 18]}{6}$$

$$= \frac{(n+1)[2n^2 + 13n + 18]}{6}$$

$$= \frac{(n+1)(n+2)(2n+9)}{6}$$

$$= \frac{(n+1)((n+1)+1)(2(n+1)+7)}{6}$$

as we needed to show. □

16.2. basis: For $n = 1$, $1 \cdot 2^1 = 2$ and $(1-1)2^{1+1} + 2 = 2$, so the equality is correct when $n = 1$.

inductive hypothesis: Suppose

$$1 \cdot 2^1 + 2 \cdot 2^2 + 3 \cdot 2^3 + \ldots + n \cdot 2^n = (n-1)\, 2^{n+1} + 2,$$

for some $n \geq 1$, Then

inductive step:

$$1 \cdot 2^1 + 2 \cdot 2^2 + 3 \cdot 2^3 + \ldots + n \cdot 2^n + (n+1)2^{n+1}$$
$$= (n-1)\,2^{n+1} + 2 + (n+1)\,2^{n+1} \quad \text{using the inductive hypothesis}$$
$$= ((n-1) + (n+1))\,2^{n+1} + 2 \quad \text{and the rest is just algebra}$$
$$= (2n)2^{n+1} + 2$$
$$= n2^{n+2} + 2$$
$$= ((n+1) - 1)\,2^{(n+1)+1} + 2,$$

as we needed to show. □

16.3. basis: $f_0 = 0 = 1 - 1 = f_2 - 1$, so the equation is correct for n = 1.
inductive hypothesis: Suppose

$$f_0 + f_1 + f_2 + \cdots + f_n = f_{n+2} - 1$$

for some $n \geq 0$. Then
inductive step:

$$f_0 + f_1 + f_2 + \cdots + f_n + f_{n+1}$$
$$= (f_{n+2} - 1) + f_{n+1} \quad \text{using the inductive hypothesis}$$
$$= (f_{n+1} + f_{n+2} - 1)$$
$$= f_{n+3} - 1 \quad \text{using the recursive definition of the Fibonacci sequence}$$
$$= f_{(n+1)+2} - 1,$$

as we needed to show. □

16.4. basis: For $n = 5$, the inequality is correct: $2^5 = 32 > 25 = 5^2$.
inductive hypothesis: Suppose $2^n > n^2$ for some $n > 4$. Then
inductive step:

$2^{n+1} = 2(2^n) > 2n^2$ using the inductive hypothesis

now $2n^2 = n^2 + n^2 = n^2 + (n)(n) > n^2 + 3n$ since for $n > 4$ it is true that $(n)(n) > 3n$

next $n^2 + 3n = n^2 + 2n + n > n^2 + 2n + 1$ since for $n > 4$ it is true that $n > 1$

and $n^2 + 2n + 1 = (n+1)^2$,

Putting the pieces together, we get $2^{n+1} > (n+1)^2$ as we needed to show.□

16.5. basis: For $n = 0$, $11^n - 6 = 1 - 6 = -5 = (5)(-1)$, so $11^0 - 6$ is divisible by 5.
inductive hypothesis: Suppose $11^n - 6$ is divisible by 5 for some $n \geq 0$. Then
inductive step:

$$11^{n+1} - 6 = 11(11^n) - 6 = 10(11^n) + (11^n - 6) = 2(5)(11^n) + (11^n - 6).$$

In that last expression, the term $2(5)(11^n)$ is certainly divisible by 5, and the expression $11^n - 6$ is divisible by 5 by the inductive hypothesis. That implies $2(5)(11^n) + (11^n - 6)$ is divisible by 5 as we needed to show.
Note: Here is another way to see that $11^n - 6$ is divisible by 5: for any integer $n \geq 0$, the number 11^n will have units digit 1, and so $11^n - 6$ will have units digit 5. Numbers with units digit 5 are divisible by 5. This proof does not answer the question posed however, since it is not a proof by induction.

16.6. basis: With 0 cuts, we end up with one piece of pizza, namely the whole thing. And, sure enough, for $n = 0$, $\frac{n^2+n+2}{2} = \frac{0^2+0+2}{2} = 1$.

inductive hypothesis: Suppose that for some $n \geq 0$, n straight lines cut produces a maximum number of $\frac{n^2+n+2}{2}$ pieces. Then
inductive step: Suppose we add one more cut. Notice that when the new cut crosses an old cut, it will slice one old piece into two new pieces. We can't get more than two pieces when one cut crosses another since two straight lines cannot cross each other more than once. So, to get the maximum number of new pieces, we should make the new cut not parallel to the n previous cuts (and so, with care, be sure to cross all the previous cuts, and not at a point where previous cuts cross each other). This will give the maximum number of new pieces equal to $n + 1$. Conclusion: the maximum number of pieces with $n + 1$ straight cuts is

$$\frac{n^2+n+2}{2} + (n+1) = \frac{n^2+n+2+2(n+1)}{2} = \frac{(n+1)^2+(n+1)+2}{2},$$

as we needed to show. □
The list of maximums begins 1, 2, 4, 7, 11, 16, 22, 29, 37, 46, 56, 67, 79, 92.

16.7. basis: For $n = 0$ we are given $a_0 = 0$ and we see $(5^0 - 1)/4 = 0$ so the base case is good.
inductive hypothesis: Suppose that $a_n = (5^n - 1)/4$ for some $n \geq 0$. Then

inductive step:

$$\begin{aligned}
\mathbf{a_{n+1}} &= \mathbf{5a_n + 1} && \text{using the recursive definition of the sequence}\\
&= \mathbf{5 \cdot \frac{5^n - 1}{4} + 1} && \text{by induction hypothesis}\\
&= \mathbf{\frac{5^{n+1} - 5}{4} + \frac{4}{4}} && \text{and the rest is algebra}\\
&= \mathbf{\frac{5^{n+1} - 5 + 4}{4}}\\
&= \mathbf{\frac{5^{n+1} - 1}{4}}
\end{aligned}$$

as we needed to show. □

16.8. basis: Since the recursive formula involves the two previous terms, we are going to have to check the closed form formula for the two terms a_0 and a_1. But they both work okay since $a_0 = 1$ and $2(3^0) - 2^0 = 2(1) - 1 = 1$ and $a_0 = 4$ and $2(3^1) - 2^1 = 2(3) - 2 = 4$.
inductive hypothesis: For $n \geq 2$, a_n depends on the two previous terms in the sequence, so it would be wise to use the second form of induction this time. So suppose that $a_k = 2 \cdot 3^k - 2^k$ for all k from 0 to some $n \geq 1$. Then
inductive step:

$$\begin{aligned}
a_{n+1} &= 5a_n - 6a_{n-1} && \text{using the recursive definition, okay since n + 1} \geq 2\\
&= 5(2 \cdot 3^n - 2^n) - 6(2 \cdot 3^{n-1} - 2^{n-1}) && \text{using the inductive hypothesis}\\
&= (5 \cdot 2 \cdot 3 - 6 \cdot 2)\, 3^{n-1} - (5 \cdot 2 - 6)\, 2^{n-1} && \text{and the rest is just algebra}\\
&= 18 \cdot 3^{n-1} - 4 \cdot 2^{n-1}\\
&= 2 \cdot 3^2 \cdot 3^{n-1} - 2^2 \cdot 2^{n-1}\\
&= 2 \cdot 3^{n+1} - 2^{n+1}.
\end{aligned}$$

as we needed to show. □

Chapter 20

20.1. Proof:
Suppose $a > 0$ and $b > 0$. Since $a > 0$, multiplying both sides of $b > 0$ by a gives $ab > a0$. We

know $a0 = 0$. It follows that $ab > 0$. □

20.2. Proof:
Suppose neither a nor b is 0. Consider four cases:

(1) $a > 0$ *and* $b > 0$: In this case, we proved $ab > 0$ in Exercise 1. In particular then, $ab \neq 0$ in this case.

(2) $a > 0$ *and* $b < 0$: Since $a > 0$, multiplying both sides of $b < 0$ by a gives $ab < a0$. We know $a0 = 0$. It follows that $ab < 0$. So $ab \neq 0$ in this case.

(3) $a < 0$ *and* $b > 0$: Since $a < 0$, multiplying both sides of $b > 0$ by a gives $ab < a0$. We know $a0 = 0$. It follows that $ab < 0$. So $ab \neq 0$ in this case.

(4) $a < 0$ *and* $b < 0$: Since $a < 0$, multiplying both sides of $b < 0$ by a gives $ab > a0$. We know $a0 = 0$. It follows that $ab > 0$. In particular then, $ab \neq 0$ in this case.

So, in any case, if neither a nor b is 0, then $ab \neq 0$. □

20.3. Proof:
Suppose $c \neq 0$ and $ac = bc$. We can rewrite $ac = bc$ as $ac - bc = 0$, and then use the distributive property to write that as $(a - b)c = 0$. Apply the result of Exercise 2 to conclude either $a - b = 0$ or $c = 0$. Since $c \neq 0$, it must be that $a - b = 0$. Adding b to each side of that equation shows $a = b$. □

Chapter 21

21.1. 107653 = (4)(22869) + 16177. So the quotient is 4, and the remainder is 16177.

21.2. The square root of 1297 is 36.01..., so if 1297 is not a prime, it must have a prime divisor no more than 36. Testing 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, and 31, (use a calculator if you want), we find none of those ten integers divides 1297. That means 1297 is a prime.

21.3. The divides relation is reflexive: For every integer a, $a|a$ is true since $a \cdot 1 = a$.

21.4. The divides relation is not symmetric. For example, $2|4$ is true, but $4|2$ is false.

21.5. The divides relation is transitive. Suppose $a|b$ and $b|c$ are both true. That means there are integers d and e such that $ad = b$ and $be = c$. Multiplying each side of the first of those two equations by e gives $ade = be$, so $a(de) = c$. Since de is an integer this shows $a|c$ is true.

□

21.6. The expression $4|12$ is a proposition, not a number. Note $4|12$ is true whereas $\frac{12}{4} = 3$ using correct notation.

21.7. The 1000 consecutive numbers are

$$1001! + 2, 1001! + 3, 1001! + 4, 1001! + 5, \ldots, 1001! + \mathrm{k}, \ldots, 1001! + 1001.$$

The first number in the list is not a prime since 2 is a factor of each term, and so 2 is a factor of $1001! + 2$. Likewise, 3 is a proper factor of the second number, 4 is a proper factor on the third number, and so on, until 1001 is a proper factor of the 1000^{th} number in the list. So none of the integers can be a prime. □

21.8. Proof:
Suppose $a|b$. That means $ac = b$ for some integer c. Then $(-a)(-c) = ac = b$, so $-a|b$. □

Chapter 22

22.1a.

$$\begin{aligned}
233 &= 2 \cdot 89 + 55 \\
89 &= 1 \cdot 55 + 34 \\
55 &= 1 \cdot 34 + 21 \\
34 &= 1 \cdot 21 + 13 \\
21 &= 1 \cdot 13 + 8 \\
13 &= 1 \cdot 8 + 5 \\
8 &= 1 \cdot 5 + 3 \\
5 &= 1 \cdot 3 + 2 \\
3 &= 1 \cdot 2 + 1 \\
2 &= 2 \cdot 1 + 0
\end{aligned}$$

$\gcd(233, 89) = 1$

22.1b.

$$1001 = 77 \cdot 13 + 0$$

$\gcd(1001, 13) = 13$

22.1c.

$$
\begin{aligned}
2457 &= 1 \cdot 1458 + 999 \\
1458 &= 1 \cdot 999 + 459 \\
999 &= 2 \cdot 459 + 81 \\
459 &= 5 \cdot 81 + 54 \\
81 &= 1 \cdot 54 + 27 \\
54 &= 2 \cdot 27 + 0
\end{aligned}
$$

$\gcd(2457, 1458) = 27$

22.1d.

$$
\begin{aligned}
567 &= 1 \cdot 349 + 218 \\
349 &= 1 \cdot 218 + 131 \\
218 &= 1 \cdot 131 + 87 \\
131 &= 1 \cdot 87 + 44 \\
87 &= 1 \cdot 44 + 43 \\
44 &= 1 \cdot 43 + 1 \\
43 &= 43 \cdot 1 + 0
\end{aligned}
$$

$\gcd(567, 349) = 1$

22.2.

$$
\begin{aligned}
987654321 &= 8 \cdot 123456789 + 9 \\
123456789 &= 13717421 \cdot 9 + 0
\end{aligned}
$$

$\gcd(987654321, 123456789) = 9$

22.3. There are many good answers to this problem.

Input: integers $m \geq 0$ and $n > 0$
Output: integer value of $\gcd(m, n)$

```
q ← m                          ▷ q represents a quotient of a division
r ← n                          ▷ r represents the remainder of a division
while r > 0 do
    g ← r                      ▷ Save current r. It will hold our gcd eventually
    q ← ⌊m/n⌋                  ▷ New quotient when m is divided by n
    q ← m − qn                 ▷ New remainder when m is divided by n
    m ← n                      ▷ Update m
    n ← r                      ▷ Update n
end while
output g                       ▷ The last nonzero remainder
```

22.4. Since n divides both n and $2n$, that means n is a common divisor of n and $2n$. On the other hand, no integer larger than n can divide n. So n is the largest common divisor of n and $2n$. So $\gcd(n, 2n) = n$.

Chapter 23

23.1. Part 1 (back-substitution method):

$$
\begin{aligned}
13447 &= 1 \cdot 7667 + 5780 \\
7667 &= 1 \cdot 5780 + 1887 \\
5780 &= 3 \cdot 1887 + 119 \\
1887 &= 15 \cdot 119 + 102 \\
119 &= 1 \cdot 102 + 17 \\
102 &= 6 \cdot 17 + 0
\end{aligned}
$$

$\gcd(13447, 7667) = 17$.

$$\begin{aligned}17 &= (1)(119) + (-1)(102)\\ &= (1)(119) + (-1)(1887 + (-15)(119)) = (-1)(1887) + (16)(119)\\ &= (-1)(1887) + (16)(5780 + (-3)(1887)) = (16)(5780) + (-49)(1887)\\ &= (16)(5780) + (-49)(7667 + (-1)(5780) = (-49)(7667) + (65)(5780)\\ &= (-49)(7667) + (65)(13447 + (-1)(7667)) = (65)(13447) + (-114)(7667)\end{aligned}$$

$(65)(13447) + (-114)(7667) = 17 = \gcd(13447, 7667)$

Part 2 (Extended Euclidean Algorithm Table):

13447	7667	5780	1887	119	102	17	0
		1	1	3	15	1	6
0	1	-1	2	-7	107	-114	791
1	0	1	-1	4	-61	65	-451

Thus, we may conclude that

$$\gcd(13447, 7667) = 17 = (65)(13447) + (-114)(7667).$$

You'll likely agree that the Extended Euclidean Algorithm table is a lot neater and much less prone to error.

23.2. Since 1 is a linear combination of a and b, we know that 1 is a multiple of $\gcd(a, b)$ from Theorem 23.2. Since gcd's are positive, it follows that $\gcd(a, b) = 1$.

23.3. Since 19 is a linear combination of a and b, we know that 19 is a multiple of $\gcd(a, b)$ from Theorem 23.2. Since gcd's are positive, the possible values for $\gcd(a, b)$ are 1 and 19.

23.4. Since 18 is a linear combination of a and b, we know that 18 is a multiple of $\gcd(a, b)$ from Theorem 23.2. Since gcd's are positive, the possible values for $\gcd(a, b)$ are 1, 2, 3, 6, 9, and 18.

Chapter 24

24.1. A calculator would be handy for this problem. The plan is to try divisions by 2, 3, 5, 7, 11, 13, and so on, until we have the complete factorization in to primes.

$$345678 = (2)(172839) = (2)(3)(57613) = (2)(3)(17)(3389)$$

Note that in testing for prime divisors of 3389, we needed only test 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53. With a calculator, the whole process took about two minutes.

24.2.

$$1016 = (2)(508) = (2)(2)(254) = (2)(2)(2)(127) = 2^3 \cdot 127$$

24.3. The positive divisors of 1016 will look like $2^a 127^b$ where $a = 0, 1, 2, 3$ and $b = 0, 1$. Since there are four choices for a and two choices for b, there will be a total of (4)(2) = 8 positive divisors of 1016. They are:

$2^0 127^0 = 1$	$2^1 127^0 = 2$
$2^2 127^0 = 4$	$2^3 127^0 = 8$
$2^0 127^1 = 127$	$2^1 127^1 = 254$
$2^2 127^1 = 508$	$2^3 127^1 = 1016$

24.4. The positive divisors of 345678 will look like $2^a 3^b 17^c 3389^d$ where a, b, c, d can each be either 0 or 1. So, there will be a total of (2)(2)(2)(2) = 16 positive divisors of 345678.

Chapter 25

25.1. Since gcd(21, 48) = 3 and 3 does not divide 8, there are no integer solutions to this equation.

25.2. Since gcd(21, 48) = 3 and 3 does divides 9, there are integer solutions to this equation. To find one solution, we could use the Extended Euclidean Algorithm, but in this case the numbers are small enough that we can find a solution by inspection (in other words, *we can guess an answer*). First, let's reduce the equation by dividing each side by $3 = \gcd(21, 48)$ to get the equation $7x + 16y = 3$. We can quickly see a solution: $x = 5, y = -2$ since $7(5) + 16(-2) = 35 - 32 = 3$. Multiplying both sides of $7(5) + 16(-2) = 3$ by 3 gives $21(5) + 48(-2) = 9$. Now that we have one solution to $21x + 48y = 9$ we can write down all solutions:

$$x = 5 + \frac{48}{\gcd(21,48)} k = 5 + 16k \quad \text{and} \quad y = -2 - \frac{21}{\gcd(21,48)} k = -2 - 7k$$

where k is any integer. Just to check our work, let's test the solution when $k = 10$ (so $x = 165$ and $y = -72$).

$$(21)(165) + (48)(-72) = 3465 - 3456 = 9$$

That looks good.

25.3. Since $\gcd(33, 12) = 3$ and 3 does not divide 7, there are no solutions.

25.4. We need one solution to get the ball rolling. We could use the Extended Euclidean Algorithm to write $3 = \gcd(33, 12)$ as a linear combination of 33 and 12, and we would probably have to do that if the numbers were larger. But with these small numbers we can do the work in our head: $3 = (33)(-1) + (12)(3)$.

Multiply that equation by 2 on each side to get $(33)(-2) + (12)(6) = 6$.

So now we have one solution to the given equation: $x = -2$ and $y = 6$.

Using the formulas that produce all solutions once one is known we get all solutions are given by

$$x = -2 + \frac{12}{3} k = -2 + 4k \quad \text{and} \quad y = 6 - \frac{33}{3} k = 6 - 11k$$

where k is any integer.

For example, when $k = 5$ we get the solution $x = 18$, $y = -49$.

25.5. First, let's find all solutions to $59x + 37y = 4270$ by applying the Extended Euclidean Algorithm.

59	37	22	15	7	1	0
		1	1	1	2	7
0	1	-1	2	-3	8	-59
1	0	1	-1	2	-5	37

The table shows $\gcd(59, 37) = 1 = 59(-5) + 37(8)$. Multiplying by 4270, we see one solution to $59x + 37y = 4270$ is given by $x = -5(4270)$ and $y = 8(4270)$. That means all solutions to that equations are given by

$$x = -5(4270) + 37k \quad \text{and} \quad y = 8(4270) - 59k \quad \text{where } k \text{ is any integer.}$$

We need to find values of k for which both x and y are 0 or more. In other words, we want to solve

$$-5(4270) + 37k \geq 0 \quad \text{and} \quad 8(4270) - 59k \geq 0.$$

That reduces to $\frac{5(4270)}{37} \leq \frac{8(4270)}{59}$ or $577.02 \ldots \leq k \leq 578.98 \ldots$. The only integer option is $k = 578$. We conclude the number of vases sold was $x = -5(4270) + 37(578) = 36$ and the number of bowls sold was $y = 8(4270) - 59(578) = 58$.

Chapter 26

26.1a. Taking 0 hours as midnight, the time 3122 hours after 16 hundred hours is $16 + 3122 \equiv 3138 \equiv 18 \pmod{24}$ hundred hours (or 6 p.m.).

26.1b. Taking Sunday as day 0 of a week, Monday will be 1. So, 3122 days after a Monday is $1 + 3122 \equiv 3123 \equiv 1 \pmod{7}$. So, it is a Monday.

26.1c. Taking January as month 1, November will be month 11. So, 3122 months later it will be $11 + 3122 \equiv 3133 \equiv 1 \pmod{12}$. So, it will be January.

26.2. The integers in $[7]_{11}$ are given by adding any number of 11's to 7. In other words, $7, 18, 29, 40, 51, \ldots$ and $-4, -15, -25, \ldots$. More compactly: $7 + 11k$, for all integers k.

26.3.

$$\begin{aligned} 1211 &\equiv 1 \pmod 5 \\ 218 &\equiv 3 \pmod 5 \\ -100 &\equiv 0 \pmod 5 \\ -3333 &\equiv 2 \pmod 5 \end{aligned}$$

The missing equivalence class is $[4]_5$. Any value in that equivalence class will do for the fifth value. So the possible answer is any number of the form $4 + 5k$, for an integer k. In particular, 4 would work (or -1, or 10004, or -6, and so on).

26.4. $2311 + 3912 \equiv 11 + 12 \equiv 23 \pmod{25}$

26.5. $(2311)(3912) \equiv (11)(12) \equiv 132 \equiv 7 \pmod{25}$

26.6. Since $1111 \equiv 4 \pmod 9$, the problem can be rewritten as $4^{2222} \equiv n \pmod 9$.

Let's check small powers of 4 modulo 9:

$$\begin{aligned} &4^1 \equiv 4 \pmod 9 \\ &4^2 \equiv 16 \equiv 7 \pmod 9 \\ &4^3 \equiv (4)(4^2) \equiv (4)(7) \equiv 28 \equiv 1 \pmod 9 \end{aligned}$$

Taking advantage of the fact that $4^3 \equiv 1 \pmod 9$, it follows that

$1111^{2222} \equiv 4^{2222} \equiv (4^3)^{740} \cdot 4^2 \equiv 1^{740} \cdot 16 \equiv 16 \equiv 7 \pmod 9$

26.7. Since $\gcd(4,7) = 1$ and 1 divides 3, there will be exactly one solution modulo 7 to $4x \equiv 3 \pmod 7$. Let's use trial-and-error to find that solution. Testing $x = 0, 1, 2, 3, 4, 5, 6$, we find $(4)(6) \equiv 24 \equiv 3 \pmod 7$, and so the solution is $x \equiv 6 \pmod 7$. Incidentally, it would be incorrect to say the solution is $x = 6$ since we are working modulo 7 and so the solution has to be given modulo 7.

26.8. Using the Extended Euclidean Algorithm, we get $\gcd(57, 11) = 1 = 57(-5) + 11(26)$. Since 1 divides 8, there will be exactly one solution to $11x \equiv 8 \pmod{57}$. To find that solution, multiply both sides of $57(-5)+11(26) = 1 = \gcd(57, 11)$ (cleverly) by 8 to get $57(-40)+11(208) = 8$. So, one solution to $11x \equiv 8 \pmod{57}$ is $x = 208$. That means all solutions are given by $x \equiv 208 \equiv 37 \pmod{57}$. Note that giving the solution as $x \equiv 208 \pmod{57}$ is correct, but people expect to see solutions to $x \equiv n \pmod m$ written with the value of n in the range 0 to $m - 1$.

26.9. Observe $\gcd(14, 231) = 7$ and 7 does not divide 3. So, $14x \equiv 3 \pmod{231}$ has no solutions.

26.10. To solve $8x \equiv 16 \pmod{28}$, we look for solutions to $8x + 28y = 16$, We can simplify that equation by dividing through by $4 = \gcd(8, 28)$ to get $2x+7y = 4$. Solving that equation is the same as solving $2x \equiv 4 \pmod 7$. (Short cut: when solving $ax \equiv b \pmod m$ where $\gcd(a, m) = d$ divides b, we can simplify the original equation by dividing a, b, m each by d to get $\frac{a}{d}x \equiv \frac{b}{d}\left(\frac{\text{mod } m}{d}\right)$.) The equation $2x \equiv 4 \pmod 7$ with have exactly one solution modulo 7 (but remember that $\gcd(8, 28) = 4$, so the original equation will have four solutions modulo 28). A little trial-and-error (or plain old common sense) shows $2x \equiv 4 \pmod 7$ has solution $x \equiv 2 \pmod 7$. It is acceptable to leave the answer in this form, but since the problem was given modulo 28, it is good manners to provide the solutions modulo 28. Since x has to be 2 modulo 7, the solutions modulo 28 will be the four values from 0 to 27 that are equal to 2 modulo 7. Final answer then: $x \equiv 2, 9, 16, 23 \pmod{28}$.

26.11. Since $\gcd(91, 231) = 7$, and $7|189$, we see the congruence will have seven solutions modulo 231. As in exercise 10, to simplify the work a bit we could cancel 7's in the given congruence, and rewrite the problem as $13x \equiv 27 \pmod{33}$. Solving that (using the Extended Euclidean Algorithm for example), we get $x \equiv 30 \pmod{33}$ for the solution. Since the problem was given modulo 231, we should express the solutions modulo 231 as well.
Solutions: $x \equiv 30, 63, 96, 129, 162, 195, 228 \pmod{231}$.

26.12a. Let $d = \gcd(a, m)$ and suppose s is a solution to $ax \equiv b \pmod m$ so that $as \equiv b \pmod m$. If x represents a solution to $ax \equiv b \pmod m$, then $ax \equiv as \pmod m$. Rearrange that as $ax - as \equiv$

0 (mod m), or $a(x-s) \equiv 0 \pmod{m}$. That means m divides $a(x-s)$, and so there is an integer k with $mk = a(x-s)$. Now d also divides a and m since $d = \gcd(a, m)$. So, we can divide both sides in that equation by d to get

$$\frac{m}{d}(k) = \frac{a}{d}(x-s).$$

So $\frac{m}{d}$ divides $\frac{a}{d}(x-s)$. But $\frac{m}{d}$ is relatively prime to $\frac{a}{d}$, and so $\frac{m}{d}$ must divide $x-s$. In other words, there is an integer r such that $x-s = r\left(\frac{m}{d}\right)$. Rearrange that as $x = s + r\left(\frac{m}{d}\right)$.

26.12b. Suppose $0 \le r_1 < r_2 < d$, and that d is a positive divisor m. We want to show the numbers $x_1 = s + r_1\left(\frac{m}{d}\right)$ and $x_2 = s + r_2\left(\frac{m}{d}\right)$ are not congruent modulo m. In other words we want to show that m does not divide

$$x_2 - x_1 = \left(s + r_2\left(\frac{m}{d}\right)\right) - \left(s + r_1\left(\frac{m}{d}\right)\right) = \frac{m}{d}(r_2 - r_1).$$

Cancel the common factor m, and multiply both sides by d to get $dk = r_2 - r_1$. That equation tells us $r_2 - r_1$ is a multiple of d. Since $0 \le r_1 < r_2 < d$, we know $0 < r_2 - r_1 < d$, and none of the integers in that range is a multiple of d. We have reached a contradiction, and so we can conclude the x_1 and x_2 are different modulo m. (Notice that this means that the numbers $s, s+\frac{m}{d}, s+2\frac{m}{d}, s+3\frac{m}{d}, \ldots, s+(d-1)\frac{m}{d}$ are d different values modulo m.)

Chapter 27

27.1. $21_3 = 7$

$321_4 = 57$

$4321_5 = 586$

$FED_{16} = 4077$

27.2. $11714 = 10110111000010_2$

$11714 = 130122_6$

$11714 = 2DC2_{16}$

27.3.

×	1	2	3	4	5	6
1	1	2	3	4	5	6
2	2	4	6	11	13	15
3	3	6	12	15	21	24
4	4	11	15	22	26	33
5	5	13	21	26	34	42
6	6	15	24	33	42	51

27.4.

+	1	2	3	4	5
1	2	3	4	5	10
2	3	4	5	10	11
3	4	5	10	11	12
4	5	10	11	12	13
5	10	11	12	13	14

×	1	2	3	4	5
1	1	2	3	4	5
2	2	4	10	12	14
3	3	10	13	20	23
4	4	12	20	24	32
5	5	14	23	32	41

27.5. Luckily we have a base 7 multiplication table above to make the work a little less onerous. For neatness, the subscript 7 will be omitted.

$$
\begin{aligned}
5122 &= (3)(1312) + 553 \\
1312 &= (1)(553) + 426 \\
553 &= (1)(426) + 124 \\
426 &= (3)(124) + 21 \\
124 &= (4)(21) + 10 \\
21 &= (2)(10) + 1 \\
10 &= (10)(1) + 0
\end{aligned}
$$

$\gcd(5122_7, 1312_7) = 1_7$.

Chapter 28

28.1. Number of options for one course = 5 + 4 + 6 = 15.

28.2. Number of options for a program of three courses, one from each area = $5 \cdot 4 \cdot 6 = 120$.

28.3. $26^2 \cdot 10^2 + 26^3 \cdot 10 + 26^4$

28.4. 26^6

28.5. $26 \cdot 25 \cdot 24 \cdot 23 \cdot 22 \cdot 21$

28.6a. 2^{25} (2 choices for each question, T or F.)

28.6b. 3^{25} (3 choices for each question, T or F or skip.)

28.7. $1 + 2 + 2^2 + \ldots + 2^9 = (2^{10} - 1)/(2 - 1) = 1023$.

If you don't want to include the empty string (of length 0) then the answer is 1022.

28.8. The total number of words of length 8 is 26^8. The number with no A's is 25^8. So the number with at least one A is $26^8 - 25^8$.

28.9. The number of seven letter words with no A's is 25^7. The number of seven letter words with exactly one A is $7 \cdot 25^6$. The 7 accounts for the number of options for placing the A, and the 25^6 accounts for the number of ways of filling in the remaining six spots. So, the number of seven letter words with at most one A is $25^7 + 7 \cdot 25^6$.

28.10. The number of nine letter words with at least two A's is the total number of nine letter words (26^9) minus the number with at most one A ($25^9 + 9 \cdot 25^8$). So, using the *good = total minus bad* rule the number of nine letter words with at least two A's is $26^9 - (25^9 + 9 \cdot 25^8)$.

Chapter 29

29.1. 26!

29.2. Part 1: (vowels together, assume the vowels are *a,e,i,o,u*) (Task 1) arrange the five vowels in some order: 5! ways to do that. (Task 2) arrange the 21 non-vowels in some order: 21! ways to do that. (Task 3): pick a spot in the row of 21 non-vowels to place the row of five vowels: 22 choices for that spot. Since we need to do all three tasks, the number of possible arrangements is $5! \cdot 21! \cdot 22$.

Part 2: (no adjacent vowels) (Task 1) arrange the 21 non-vowels in some order: 21! ways to do that. (Task 2) There are 22 gaps between those volumes, and we need to select five of them for the vowels: we can do that in $\binom{22}{5}$ ways. (Task 3): arrange the five vowels in some order to place in the five open spots: 5! ways to do that. Since we need to do all three tasks, the number of possible arrangements is $21! \binom{22}{5} \cdot 5!$.

29.3. There are 13! to arrange the books for each shelf. Since we need to arrange shelf 1 and shelf 2, there will be $(13!)(13!) = (13!)^2$ way to arrange the bookcase.

29.4. People are normally considered distinguishable. (Task 1) arrange the seven men in some order: 7! ways to do that. (Task 2) arrange the four women in some order: 4! ways to do that. (Task 3): pick a spot in the row of four women to place the row of seven men: 5 choices for that spot. Since we need to do all three tasks, the number of possible arrangements is $7! \cdot 4! \cdot 5$.

29.5. We need to select 10 of the 20 to form one of the teams (the remaining 10 will form the other team). That can be done in $\binom{20}{10}$ ways. If the teams have names (are distinguishable) we are done. Otherwise, this is a perfect double count and the final answer is $\frac{1}{2}\binom{20}{10}$.

29.6. The order of the numbers on the lottery ticker do not matter, so there are $\binom{99}{5}$ lottery tickets possible. In order to have at least a one-in-a-million chance of winning the lottery by matching all five numbers we need to have n tickets where $\frac{n}{\binom{99}{5}} \geq \frac{1}{1{,}000{,}000}$. In other words we need $n \geq \frac{\binom{99}{5}}{1{,}000{,}000} = \frac{71{,}523{,}144}{1{,}000{,}000} = 71.52 \ldots$. That means we need to buy 72 tickets.

29.7. Part a) We need to select six of the 9 + 13 = 22 people available. That can be done in $\binom{22}{6}$ ways.

Part b) (Task 1) Select two deans: $\binom{9}{2}$ ways.
(Task 2) Select four professors: $\binom{13}{4}$ ways.
We need to do task 1 and task 2, so there are $\binom{9}{2}\binom{13}{4}$ such committees.

Part c) The options are (1) six professors, (2) five professors and one dean, or (3) four professors and two deans. So the total number of acceptable committees is

$$\binom{13}{6}\binom{9}{0} + \binom{13}{5}\binom{9}{1} + \binom{13}{4}\binom{9}{2}.$$

Chapter 30

30.1. Adding one more row to the triangle as given in the text using Pascal's Identity, we get

Row 0:							1						
Row 1:						1		1					
Row 2:					1		2		1				
Row 3:				1		3		3		1			
Row 4:			1		4		6		4		1		
Row 5:		1		5		10		10		5		1	
Row 6:	1		6		15		20		15		6		1

30.2. $\binom{10}{3}3^3(-2)^7$

30.3. We have

$$\binom{2n}{2} = \frac{2n!}{2!\,(2n-2)!} = \frac{(2n)(2n-1)[(2n-2)!]}{2!\,(2n-2)!} = \frac{2n(2n-1)}{2} = n(2n-1) = 2n^2 - n,$$

and

$$2\binom{n}{2} + n^2 = 2\left(\frac{n!}{2!\,(n-2)!}\right) + n^2 = 2\left(\frac{n(n-1)[(n-2)!]}{2!\,(n-2)!}\right) + n^2$$
$$= 2\left(\frac{n(n-1)}{2}\right) + n^2 = n^2 - n + n^2 = 2n^2 - n.$$

So the two expressions are equal since both equal $2n^2 - n$.

30.4. We have

$$\binom{r}{s}\binom{s}{t} = \frac{r!}{s!\,(r-s)!}\cdot\frac{s!}{t!\,(s-t)!} = \frac{r!}{(r-s)!\,t!\,(s-t)!},$$

and

$$\binom{r}{t}\binom{r-t}{s-t} = \frac{r!}{t!\,(r-t)!}\cdot\frac{(r-t)!}{(s-t)!\,[(r-t)-s-t)]!} = \frac{r!\,(r-t)!}{t!\,(r-t)!\,(s-t)!\,(r-s)!} = \frac{r!}{(r-s)!\,t!\,(s-t)!}$$

So the two expressions are equal.

30.5. Scenario: We want to pick a committee of s people from a company with r employees, and a subcommittee of t of those s to act as the committee's board. In how many ways can that be done?

Method 1: Select the s people from the total of all r ($\binom{r}{s}$ ways to do that), and then select t of those s to be the board ($\binom{s}{t}$ ways to do that). So according to the product rule, there are $\binom{r}{s}\binom{s}{t}$ ways to form the committee and its board.

Method 2: First select the t people to serve on the board from the r people available ($\binom{r}{t}$ ways to do that). To fill out the committee, we need to pick $s-t$ more people from the remaining $r-t$ people ($\binom{r-t}{s-t}$ ways to do that). So, according to the product rule, there are $\binom{r}{t}\binom{r-t}{s-t}$ ways to form the committee and its board.

Since the two counting methods must give the same answer, we get $\binom{r}{s}\binom{s}{t} = \binom{r}{t}\binom{r-t}{s-t}$.

30.6. $(x+y+z)^{15} = (x+y+z)(x+y+z)(x+y+z)\ldots(x+y+z)(x+y+z)$ where there are 15 of the trinomials $x+y+z$. When this is expanded, we will get a term of the form $x^4y^5z^6$ by selecting 4 of the 15 trinomials to take the x from, and then 5 of the remaining 11 trinomials to take the y from (and, by default, taking z from the remaining six trinomials). So the number of terms of the form $x^4y^5z^6$ (before terms are combined) will be

$$\binom{15}{4}\binom{11}{5}\binom{6}{6} = \frac{15!}{4!\,11!}\cdot\frac{11!}{5!\,6!}\cdot\frac{6!}{6!\,0!} = \frac{11!}{4!\,5!\,6!}.$$

30.7. Suppose p is a prime and k is an integer with $1 < k < p$. Let $\binom{p}{k} = n$ (note that n is an integer). Expanding the binomial coefficient we get $\frac{p!}{k!(p-k)!} = n$. Rewrite that equation as $p! = n\cdot k!\,(p-k)!$. We want to show that p divides n.

Since the prime p divides the left side of that equation, it must divide the right side, and one of the properties of primes we proved is that if a prime divides a product of integers, it must divide one of those integers. So we can conclude p divides n or $k!$, or $(p-k)!$. Since $k! = (1)(2)\cdots(k)$ and $k < p$, there is no factor of p in $k!$. So p does not divide $k!$. Likewise, $(p-k)! = (1)(2)\cdots(p-k)$ and $p-k < p$, so p does not divide $(p-k)!$. We conclude p must divide n as we wanted to prove. □

Chapter 31

31.1. Let M be the set of students with a math major, C: chemistry majors, B: biology majors, G: geology majors, P: physics majors, and A: anthropology majors.

The total number of students is (taking advantage of the fact each student has at most two majors, and some double majors have no students)

$$|M \cup C \cup B \cup G \cup P \cup A| = |M| + |C| + |B| + |G| + |P| + |A|$$
$$-|M \cap P| - |M \cap C| - |M \cap B| - |B \cap C| - |B \cap A|$$
$$= 70 + 160 + 230 + 56 + 24 + 35 - 12 - 10 - 4 - 53 - 5 = 491.$$

31.2. Let A be the length 15 bit strings that start with 1111 (they look like 1111...........). Let B be the set of length 15 bit strings that end with 1000 (they look like ....1000), and let C be the set of length 15 bit strings with bits 4 through 7 equal to 1010 (they look like ...1010 .... We need to compute $|A \cup B \cup C|$.

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|$$
$$= 2^{11} + 2^{11} + 2^{11} - 2^7 - 2^8 - 2^7 + 2^4 = 5648.$$

31.3. Let A be the set of integers between 1000 and 9999 (inclusive) that are multiples of 4. To count the number of integers, n, in A, we want to solve $1000 \le 4n \le 9999$. In other words, solve $250 \le n \le 2499.75$. The number of integers in the range $[250, 2499]$ is $2499 - 250 + 1 = 2250$. So $|A| = 2250$. Likewise, letting B be the set of multiples of 10 in the range 1000 to 9999, we get $|B| = 900$, and with C being the set of multiples of 25 in that range, we get $|C| = 360$. Now things get a little trickier: For $A \cap B$ we want to count the integers that are both multiples of 4 and 10. But that is the same as the multiples of 20, so $|A \cap B| = 450$. Likewise $A \cap C$ (multiples of 100) has $|A \cap C| = 90$, and $B \cap C$ (multiples of 50) has $|B \cap C| = 180$. Finally, $A \cap B \cap C$ (also multiples of 100) has $|A \cap B \cap C| = 90$. So

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|$$
$$= 2250 + 900 + 360 - 450 - 100 - 180 + 100 = 2880.$$

31.4. Letting A_1 be the set of permutations of $1, 2, 3, 4, 5$ with 1 in spot 1, A_2 be the set of permutations of $1, 2, 3, 4, 5$ with 2 in spot 2, and so on, until A_5 is the set of permutations of $1, 2, 3, 4, 5$ with 5 in spot 5, we want to count in the number of elements in $A_1 \cup A_2 \cup A_3 \cup A_4 \cup A_5$.

All five of the counts $|A_1|, |A_2|, \ldots, |A_5|$ will be be 4! (place one number in its spot, and arrange the other four in any way at all). Likewise, all ten counts like $|A_1 \cap A_2|$ will be 3! (place two numbers in the correct spots, and arrange the remaining three in any way at all. Continuing, counts for $A_1 \cap A_2 \cap A_3$ type will be 2!, and for the $A_1 \cap A_2 \cap A_3 \cap A_4$ type will be 1!. Finally, $|A_1 \cap A_2 \cap A_3 \cap A_4 \cap A_5| = 1$ (every number is its correct spot). So,

$$|A_1 \cup A_2 \cup A_3 \cup A_4 \cup A_5|$$
$$= 5(4!) - 10(3!) + 10(2!) - 5(1!) + 1 = 120 - 60 + 20 - 5 + 1 = 76.$$

31.5. Using the *good = total - bad* method, the number of permutations of $1, 2, 3, 4, 5$ with no digit in its correct spot will be the total number of permutations of 1, 2, 3, 4, 5 minus the number of those permutations with at least one numbers in its correct spot. From the problem above, that is $5! - 76 = 120 - 76 = 44$.

Chapter 32

32.1. If we distribute eight objects (people) into seven piles (days of the week), there will be at least one pile (day) with $\lceil 8/7 \rceil = 2$ objects (people).

32.2. If we distribute 100 objects (people) into seven piles (days of the week), there will be at least least one pile (day) with $\lceil 100/7 \rceil = 15$ objects (people).

32.3. Since there are four suits (piles), we need the least value of n so that $\lceil n/4 \rceil = 6$. That will be $5 \cdot 4 + 1 = 21$.

32.4. Let $a_1, a_2, \ldots, a_n$ be n integers. They are each equivalent to one of $0, 1, 2, \ldots, n-2$ modulo $n-1$. Since there are n values, but only $n-1$ options modulo $n-1$, some two of those numbers must be the same modulo $n-1$. Say $a_k \equiv a_j \pmod{n-1}$. Thus $a_k - a_j \equiv 0 \pmod{n-1}$, and that means $n-1$ divides $a_k - a_j$.

32.5. Let t_i denote the total number of hours studied from day 1 to day i.
We are told

$$0 < t_1 < t_2 < \cdots < t_{75} \leq 125.$$

Adding 24 to each of those numbers gives

$$24 < t_1 + 24 < t_2 + 24 < \cdots < t_{75} + 24 \leq 149.$$

So we have 150 numbers (namely $t_1, t_2, \ldots, t_{75}$ and $t_1 + 24, t_2 + 24, \ldots, t_{75} + 24$) all between 1 and 149. By the Pigeonhole Principle some two of those must be equal. Since the numbers in the first list are all different, and the numbers in the second list are also all different, it must be that $t_i = t_j + 24$ for some i and j. That means $t_i - t_j = 24$.
That tells us that on days $j + 1, j + 2, \cdots, i$, Al studied exactly a total of 24 hours.

32.6. Since there are only 216 different values modulo 216, the pigeonhole principle says some two of the 217 numbers, say m and n, must have the same value modulo 216. So $m \equiv n \pmod{216}$. That means $216 \mid (m - n)$. So m and n have a difference that is a multiple of 216.

Chapter 33

33.1. There are four choices for each of the n positions in the string. So there are 4^n such strings.

33.2. Including the string of length 0, there are $1 + 4 + 4^2 + \cdots + 4^7 = \frac{4^8-1}{4-1} = \frac{4^8-1}{3}$ such strings.

33.3. (a donut shop problem) Ask for 3 x_1's, 4 x_2's, 5 x_3's, and 6 each of x_4's, x_5's, x_6's, and x_7's. That gives us 36 donuts so far. The remaining 18 can be selected in any way at all. So there are $\binom{18+7-1}{18} = \binom{25}{18}$ acceptable solutions to the equation.

33.4. For ternary strings, each position is a 0, 1, or 2. If the ternary string begins 0101 and ends 212 (and must be of length 7 or more!). There will be $n - 7$ positions left to fill, and there are three choices for each position, so there are 3^{n-7} such strings (assuming $n \geq 7$).

33.5. When you build your order, tell the clerk to start with four jalapeño, six cherry, and eight strawberry. That accounts for 18 donuts, and so you need 18 more, and any combination is okay for those last 18. So there are $\binom{8+18-1}{18} = \binom{25}{18}$ ways to form the donut order.

33.6a. 26^{10} (There are 26 choices for each of the middle ten spots.)

33.6b. Pick a spot for the x (12 options). Fill in the 11 empty spots (25 choices for each spot since we can't use the x again): Answer: $(12)(25^{11})$.

33.6c. Pick a spot for the x (12 choices), then pick a spot for the y (11 choices), then fill in the remaining 10 spots (24 choices for each spot): Answer: $(12)(11)(24^{10})$.

33.6d. (good = total - bad method) There are 13 letters in the second half of the alphabet, and so 13^{12} twelve letter words made up of only letters from the second half of the alphabet. These are all *bad* for this problem. There are 26^{12} words of length twelve. So, there are $26^{12} - 13^{12}$ twelve letters words with at least one letter from the first half of the alphabet.

33.7. Let A be the length 19 bit strings of the form 0101................
Let B be the length 19 bit strings of the form ...101..............
Let C be the length 19 bit strings of the form................. 1010.

(inclusion/exclusion)

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|$$
$$= 2^{15} + 2^{16} + 2^{15} - 2^{13} - 2^{11} - 2^{12} + 2^{9}.$$

33.8. (Task 1) Pick three of the fifteen spots for 0's: $\binom{15}{3}$ ways to do that.
(Task 2) Pick four of the remaining twelve spots for 1's: $\binom{12}{4}$ ways.
(Task 3) Pick three of the remaining eight spots for 2's: $\binom{8}{3}$ ways.
(Task 4) Pick four of the remaining five spots for 3's: $\binom{5}{4}$ ways.
(Task 5) Pick one of the remaining one spots for 4: $\binom{1}{1}$ ways.

The number of good strings is

$$\binom{15}{3}\binom{12}{4}\binom{8}{3}\binom{5}{4} = \frac{15!}{3!\,12!}\cdot\frac{12!}{4!\,8!}\cdot\frac{8!}{3!\,5!}\cdot\frac{5!}{4!\,1!} = \frac{15!}{3!\,4!\,3!\,4!\,1!}.$$

33.9a. As usual, we assume people are distinguishable. We can pair 7, 6, 5 or 4 lecturers with 0, 1, 2 or 3 professors respectively.

$$\binom{7}{7}\binom{14}{0} + \binom{7}{6}\binom{14}{1} + \binom{7}{5}\binom{14}{2} + \binom{7}{4}\binom{14}{3}.$$

33.9b. We can pair 7, 6, or 5 professors with 0, 1 or 2 lecturers respectively.

$$\binom{14}{7}\binom{7}{0} + \binom{14}{6}\binom{7}{1} + \binom{14}{5}\binom{7}{2}.$$

33.9c. The final size of the committee isn't specified so we will assume any size (five or more) is ok We will pick 5, 6, or 7 lecturers, and pair each selection with any subset of the professors.

$$\binom{7}{7}2^{14} + \binom{7}{6}2^{14} + \binom{7}{5}2^{14}.$$

33.10. (good = total - bad method) There are 20! ways to form a line of the 20 people. If we tie Hans and Brunhilda together, there are 19 items, and so there are 19! ways to line those 19 items up. Of course Hans and Brunhilda could be in either order, so there are 2(19!) bad lines. The number of good lines is 20! − 2(19!).

33.11. Here is a proof by induction:

(basis) For a set of one element, $\{a\}$, the two subsets are $\{\}$ and $\{a\}$. The first has an even number of elements, and the second has an odd number of elements, so we are okay in this case.

(inductive step) Suppose that for some $n \geq 1$, an n element set has the same number of subsets of even cardinality as odd cardinality. Now consider a set with $n + 1$ elements. Say the set, A, consists of n elements along with one additional element e (for extra). List all the subsets of A. By

the inductive hypothesis, there will be some number t with even cardinality, and the same number t with odd cardinality. Adding the element e to the subsets of A with even cardinality will produce t subsets of $A \cup \{e\}$ with odd cardinality, and adding the e to the subsets of A with odd cardinality will produce t subsets of $A \cup \{e\}$ with even cardinality. Conclusion: $A \cup \{e\}$ has the same number ($2t$ in fact) of subsets with even and odd cardinality.
An easier proof: By the binomial theorem

$$0 = 0^n = (1 + (-1)^n) = \binom{n}{0} - \binom{n}{1} + \binom{n}{2} - \binom{n}{3} + \cdots + (-1)^n \binom{n}{n}.$$

Chapter 34

There are usually many different recursive formulas that will be correct answers to the problems below. If your answer does not agree with the answer provided, you should use both formulas to generate a numbers of terms, say six or eight or so. If the numbers do not agree, your work is wrong. If they do agree, then likely your work is okay. You could make sure of that by using induction, for example, to show the two recursive formulas are equivalent.

34.1. Let p_n be the number of pennies in the bank on day n. The initial value is $p_0 = 0$. The recursive relation is $p_n = p_{n-1} + n$, for $n \geq 1$.

34.2. For $n \geq 0$, let c_n be the number of different ways Al can climb n steps.

The initial conditions are $c_0 = 1$, and $c_1 = 1$.

For the recursive formula: When climbing $n \geq 2$ steps, Sal can start with one step and finish the climb in c_{n-1} ways, or start with two steps and finish the climb in c_{n-2} ways.
So, for $n \geq 2$, $c_n = c_{n-1} + c_{n-2}$.

34.3. Let a_n be the number of bit strings of length n with an even number of 0's. For an initial condition, we have $a_1 = 1$ since the only good length 1 bit string is 1. If the empty bit string doesn't bother you, we could use initial condition $a_0 = 1$. Now we think recursively: If we have a good bit string of length $n - 1$, we can add a 1 to the end to get a good bit string of length n. That accounts for all the good length n bit strings that end with 1. We get the good bit strings of length n that end with 0 by adding 0 to the end of a *bad* length $n-1$ bit string. Using *good = total - bad* (well, actually *bad = total - good* in this case), we see there are $2^{n-1} - a_{n-1}$ bad bit strings. So, the solution is $a_n = a_{n-1} + (2^{n-1} - a_{n-1}) = 2^{n-1}$, for $n \geq 1$, with $a_n = 1$

34.4. We can be sneaky about this and use the example in the text: A recursive relation for the number of bit strings with no adjacent 0's is given by $a_0 = 1$, $a_1 = 2$, and $a_n = a_{n-1} + a_{n-2}$ for $n \geq 2$. Let b_n be the number of bit strings that do contain the pattern 00, the ones we are really interested in. Now, using *good* = *total* - *bad*, $a_n = 2^n - b_n$. So $b_0 = 0$ and $b_1 = 0$, and for $n \geq 2$ we get

$$2^n - b_n = (2^{n-1} - b_{n-1}) + (2^{n-2} - b_{n-2})$$

which can be rearranged as

$$b_n = b_{n-1} + b_{n-2} + 2^n - 2^{n-1} - 2^{n-2} = b_{n-1} + b_{n-2} + 2^{n-2}.$$

The sequence, for $n \geq 0$, begins: $0, 0, 1, 3, 8, 19, 43$, which agrees with a few brute force computations.

But that was sort of an unsportsmanlike solution since we didn't really reason recursively. So, let's try it again. After getting $b_0 = 0$ and $b_1 = 0$, let's think about how to build length $n \geq 2$ *good* bit strings. We could add 00 to the right end of any of the 2^{n-2} bit strings of length $n - 2$. Or, we could add 10 to any of the b_{n-2} good bit strings of length $n - 2$. Or we could add 1 to any of the b_{n-1} good bit strings of length $n - 1$. These three options account for all the good length n bit strings, the first two count the length n bit strings that end with 0, and the last counts the good bit strings that end with 1. So, the recursive part of the solution is, for $n \geq 2$,

$$b_n = b_{n-1} + b_{n-2} + 2^{n-2}$$

as before.

34.5. Again, there is an easy way to do this counting using the *good* = *total* - *bad* method. The only *bad* bit strings have to look like a number of 1's followed on the right by a number of 0's. Examples: (length n) $000 \cdots 0$, $100 \cdots 0$,$110 \cdots 0$,$111 \cdots 0$, and so on, until we get to $111 \cdots 1$. That is a total of $n + 1$ *bad* strings. So the number of *good* length n bit strings must be $2^n - (n + 1)$.

The first few terms of the sequence, starting at $n = 0$, are : $0, 0, 1, 4, 11, 26, 57$.

But, again, that wasn't recursive counting. So let's try again. Letting g_n be the number of good strings of length n, we get $g_0 = 0$, $g_1 = 0$, and $g_2 = 1$. That's a good start. Now let's think recursively. For good strings of length $n \geq 1$, we can make good strings of length $n+1$ by adding a 0 to the right end, and that will account for all the good length $n+1$ strings ending with 0. Next, let's count the number of good length $n+1$ strings ending with 1. Here there are several choices depending on the number of 1's that end the bit string (*any* will mean any bit string of appropriate length):

- (any)01 (2^{n-2} of these)
- (any)011 (2^{n-3} of these)
- (any) 0111 (2^{n-4} of these)
- and so on until we reach $01111\cdots 1$ (1 of these)

So, we get

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

This doesn't look exactly like our first solution, so let's do a bit of testing. Checking this recursive formula against the terms computed above, using initial value $g_0 = 0$, we get

$$0, 0, 1, 4, 11, 26, 57$$

So things look pretty good.

34.6. Let g_n be the number of ternary strings of length n that contain 00. A little trial-and-error gives the values $g_0 = 0$, $g_1 = 0$, $g_2 = 1$, and $g_3 = 5$. For larger values of n it is already too much trouble writing down the good strings without some sort of organized plan.

Let's break the problem of finding longer strings into a number of cases:

- Length n strings ending with 1. (g_{n-1} of these)
- Length n strings ending with 2. (g_{n-1} of these)

- Length n strings ending with 0.
    - ending with 00. (3^{n-2} of these)
    - end with 10. (g_{n-2} of these)
    - end with 20. (g_{n-2} of these)

That accounts for all the good strings of length n, so the recursive formula is

$$g_0 = 0 \qquad g_1 = 0$$
$$g_n = 2g_{n-1} + 2g_{n-2} + 3^{n-2}.$$

The first few values are 0, 0, 1, 5, 21, 79, 281, 963, 3217.

An alternative recursive answer (as given in sequence A186244 in the *The On-Line Encyclopedia of Integer Sequences*) (Google it!) is,

$$g_n = 3g_{n-1} + 2(3^{n-3} - g_{n-3}) \text{ for } n \geq 3$$
$$\text{with } g_0 = 0 \quad g_1 = 0 \quad g_2 = 1$$

Reasoning: The recursive formula is based on adding any of 0, 1, 2 to strings of length $n-1$ which already have 00 in them, or 100, 200 to strings of length $n-3$ which do not.

34.7. Let $A_n = \{1, 2, 3, \ldots, n\}$. A subset B of A_n is good if B does not contain any two consecutive integers. Let g_n be the number of good subsets of A_n. Split the good subsets of A_n into two groups:

(1) good subsets of A_n that contain n

and

(2) good subsets of A_n that do not contain n.

Good subsets in group (1) cannot contain $n-1$, and so those good subsets of A_n are produced by adding n to a good subset of A_{n-2} (at least if $n \geq 2$). That accounts for all the good subsets of A_n that contain n. That shows there are g_{n-2} good subsets of A_n that contain n.

Next, let's count the number of good subsets of A_n that do not contain n. But that is easy: these are just the good subsets of A_{n-1}, and so there are g_{n-1} of these.

Conclusion: for $n \ge 2$, $g_n = g_{n-1} + g_{n-2}$ (the Fibonacci recurrence!). We need initial terms: $g_0 = 1$ and $g_1 = 2$.

The first few values are $1, 2, 3, 5, 8, 13, 21, 34$. This is the Fibonacci sequence with the first two terms discarded.

34.8. Maybe a bit surprisingly, this is a very difficult problem. A conjectured answer was given in 1941, and evidently the conjecture was proved correct in 2014. No one has been able to solve the problem for more than four pegs, though there are suspicions that remain unproven.

Chapter 35

35.1. Using the given recursive formula we see the sequence begins $2, 4, 10, 28, 82, 244$, and those values look a lot like powers of 3: $1, 3, 8, 27, 81, 243$, so it looks like a good guess is $a_n = 3^n + 1$.

The basis for the induction is the case $n = 0$. We are given $a_0 = 2$ while $3^0 + 1 = 1 + 1 = 2$, and $a_1 = 4$ while $3^1 + 1 = 4$. So $a_n = 3^n + 1$ is correct for $n = 1, 2$ works out okay. For the inductive step, suppose $a_k = 3^k + 1$ for all values of $k \le n$ for some $n \ge 1$. Then

$$a_{n+1} = 4a_n - 3a_{n-1} = 4(3^n + 1) - 3(3^{n-1} + 1)$$
$$= 4 \cdot 3^n - 3^n + 4 - 3 = 3 \cdot 3^n + 1 = 3^{n+1} + 1.$$

35.2.

$$a_n = 5a_{n-1} = 5(5a_{n-2}) = 5^2 a_{n-2} = 5^2(5a_{n-3}) = 5^3 a_{n-3}.$$

As we continue to unfold, eventually we will reach a_0. Notice that the exponent on the 5 and the subscript on the a always add up to n. That makes sense since at each step the exponent goes up 1 and the subscript goes down 1, and so the exponent and the subscript always add up to the n they started at in the first step. That means we eventually reach$a_n = 5^n a_0$. Since $a_0 = 2$, we conclude $a_n = 2 \cdot 5^n$.

35.3.

$$
\begin{aligned}
a_n &= 3 + 5a_{n-1} = 3 + 5(3 + 5a_{n-2}) \\
&= (3 + 3 \cdot 5) + 5^2 a_{n-2} = (3 + 3 \cdot 5) + 5^2(3 + 5a_{n-3})) \\
&= (3 + 3 \cdot 5 + 3 \cdot 5^2) + 5^3 a_{n-3}.
\end{aligned}
$$

As we continue to unfold, the first group in parentheses will continue gaining one term at each step and the last term will have the exponent of the 5 going up one at a time while the subscript on the a will decrease by one at a time. Eventually we will reach the expression

$$
\begin{aligned}
a_n &= (3 + 3 \cdot 5 + 3 \cdot 5^2 + \cdots + 3 \cdot 5^{n-1}) + 5^n a_0 \\
&= 3(1 + 5 + 5^2 \cdots + 5^{n-1}) + 5^n \cdot 2 \\
&= 3\frac{5^n - 1}{5 - 1} + 2 \cdot 5^n = \frac{11 \cdot 5^n - 3}{4}.
\end{aligned}
$$

As a check of our work, we can use the recursive formula and the closed form formula to generate six or so terms to see if they produce the same values (or, if we are really ambitious, we can use induction to verify the closed form formula is correct). In any case, the recursive formula and the closed form formula both give

2, 13, 68, 343, 1718, 8593

for the first six terms, and so we can be reasonably confident our work is okay.

Chapter 36

36.1 a. $a_0 = 3$ and for $n \geq 1$, $a_n = a_{n-1} + 2$. But also $a_0 = 3, a_1 = 5$ and $a_n = 2a_{n-1} - a_{n-2}$ for $n \geq 2$.

36.1 b. $a_1 = 6$ and for $n \geq 2$, $a_n = 2a_{n-1}$.

36.1 c. $a_1 = 1$ and for $n \geq 2$, $a_n = a_{n-1} + 2n - 1$. (among others)

36.1 d. $a_0 = 1$ and for $n \geq 1$, $a_n = a_{n-1} + 2(-1)^n + 1$. (among others)

36.2. The characteristic equation is

$$\chi(x) = x^2 - x - 6 = (x+2)(x-3) = 0.$$

The characteristic roots are $x = -2, 3$.

The general solution is $a_n = \alpha(-2)^n + \beta 3^n$.

The initial values ($n = 0, 1$) produce the linear system

$$\begin{aligned} \alpha + \beta &= 3 \\ -2\alpha + 3\beta &= 6 \end{aligned}$$

With solution $\alpha = 3/5$ and $\beta = 12/5$.
So the close form formula is

$$a_n = \frac{3}{5}(-2)^n + \frac{12}{5}3^n, \text{ for } n \geq 0.$$

36.3. The characteristic equation is

$$\chi(x) = x^2 - 5x + 6 = (x-2)(x-3) = 0.$$

The characteristic roots are $x = 2, 3$.

The general solution is $a_n = \alpha 2^n + \beta 3^n$.

The initial values ($n = 0, 1$) produce the linear system

$$\alpha + \beta = 4$$
$$2\alpha + 3\beta = 7$$

with solution $\alpha = 5$ and $\beta = -1$.

So the closed form formula is $a_n = 5 \cdot 2^n - 3^n$, for $n \geq 0$.

36.4. The characteristic equation is

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

The characteristic roots are $x = 2, 5$.

The general solution is $a_n = \alpha 2^n + \beta 5^n$.

The initial values (be careful! $n = 2, 3$) produce the linear system

$$4\alpha + 25\beta = 5$$
$$8\alpha + 125\beta = 13$$

With solution $\alpha = 1$ and $\beta = 1/25$.
So the closed form formula is

$$a_n = 2^n + \frac{5^n}{25} = 2^n + 5^{n-2}, \text{ for } n \geq 2.$$

36.5. The characteristic equation is

$$\chi(x) = x^2 - 4x + 4 = (x-2)^2.$$

The characteristic roots are $x = 2, 2$.

The general solution is $a_n = \alpha 2^n + \beta n 2^n$.

The initial values (be careful! $n = 1, 2$) produce the linear system

$$2\alpha + 2\beta = 3$$
$$4\alpha + 8\beta = 5$$

with solution $\alpha = 7/4$ and $\beta = -1/4$.
So the closed form formula is

$$a_n = \frac{7 \cdot 2^n}{4} - \frac{n2^n}{4} = 2^{n-2}(7 - n), \text{ for } n \geq 1.$$

36.6. The characteristic equation is

$$\chi(x) = x^2 - 6x - 9 = (x - 3)^2.$$

The characteristic roots are $x = 3, 3$.

The general solution is $a_n = \alpha 3^n + \beta n 3^n$.

The initial values ($n = 0, 1$) produce the linear system

$$\alpha = 1$$
$$3\alpha + 3\beta = 6$$

with solution $\alpha = 1$ and $\beta = 1$.

So, the closed form formula is

$$a_n = 3^n + n3^n = 3^n(n + 1), \text{ for } n \geq 0.$$

36.7. This one is easily solved by inspection (that is, it is easy to guess the solution), but let's use the characteristic equation method for the practice.

The characteristic equation is

$$\chi(x) = x^2 - 1 = (x+1)(x-1).$$

The characteristic roots are $x = -1,1$.

The general solution is $a_n = \alpha(-1)^n + \beta$.

The initial values (n = 1, 2) produce the linear system

$$-\alpha + \beta = 2$$
$$\alpha + \beta = 8$$

with solution $\alpha = 3$ and $\beta = 5$.

So the closed form formula is

$$a_n = 3(-1)^n + 5, \text{ for } n \geq 1.$$

36.8. The characteristic equation is

$$\chi(x) = x^3 - 6x^2 + 11x - 6 = (x-1)(x-2)(x-3).$$

(You might need to review the topic (*finding rational roots of polynomials* in a college algebra text or via an internet search) to refresh your memory about finding that factorization.)

The characteristic roots are $x = 1, 2, 3$.

The general solution is $a_n = \alpha 1^n + \beta 2^n + \gamma 3^n$.

The initial values ($n = 0, 1, 2$) produce the linear system

$$\alpha + \beta + \gamma = 2$$
$$\alpha + 2\beta + 3\gamma = 5$$
$$\alpha + 4\beta + 9\gamma = 15$$

with solution $\alpha = 1$, $\beta = -1$, and $\gamma = 2$.

So the closed form formula is

$$a_n = 1 - 2^n + 2 \cdot 3^n, \text{ for } n \geq 0.$$

36.9. The characteristic equation is

$$\chi(x) = x^2 - x + 1.$$

The characteristic roots (use the quadratic formula) are $r_1 = \frac{1+\sqrt{5}}{2}$ and $r_2 = \frac{1-\sqrt{5}}{2}$.

The general solution is $a_n = \alpha r_1^n + \beta r_2^n$.

The initial values ($n = 0,1$) produce the linear system

$$\begin{cases} \alpha \quad + \quad \beta = 0 \\ \alpha r_1 + \beta r_2 = 1 \end{cases}$$

with solution $\alpha = \frac{1}{r_1 - r_2} = \frac{1}{\sqrt{5}}$ and $\beta = -\frac{1}{\sqrt{5}}$..

So the closed form formula is

$$a_n = \frac{1}{\sqrt{5}} r_1^n - \frac{1}{\sqrt{5}} r_2^n = \frac{1}{\sqrt{5}} \left(\frac{1+\sqrt{5}}{2} \right)^n - \frac{1}{\sqrt{5}} \left(\frac{1-\sqrt{5}}{2} \right)^n, \text{ for } n \geq 0.$$

This closed form formula for the Fibonacci numbers is called Binet's Formula.

Chapter 37

37.1. For exercise 36.2, we know the general solution to the related homogeneous recursion is

$$a_n^{(h)} = \alpha(-2)^n + \beta 3^n.$$

That general solution of the related homogeneous recursion needs to be paired up with a particular solution of the original recursive formula

$$a_n = a_{n-1} + 6a_{n-2} + 1.$$

Since the nonhomogeneous part of the original recursion is the constant 1, our first guess should be that there will be a particular solution of the form $a_n = A$, a constant. Putting that guess in the recursive formula, we get

$$A = A + 6A + 1 \text{ with solution } A = -1/6.$$

So that general solution to the nonhomogeneous recursion is

$$a_n = \alpha(-2)^n + \beta 3^n - \frac{1}{6}.$$

Using the initial conditions produces the system

$$\begin{cases} \alpha + \beta - \dfrac{1}{6} = 3 \\ -2\alpha + 3\beta - \dfrac{1}{6} = 6 \end{cases}$$

with solution $\alpha = 2/3$ and $\beta = 5/2$.

So the solution to the original recursive formula is

$$a_n = \frac{2}{3}(-2)^n + \frac{5}{2}3^n - \frac{1}{6}.$$

37.2. For exercise 36.4, we know the general solution to the related homogeneous recursion is

$$a_n^{(h)} = \alpha 2^n + \beta 5^n.$$

That general solution of the related homogeneous recursion needs to be paired up with a particular solution of the original recursive formula

$$a_n = 7a_{n-1} - 10a_{n-2} + \text{n}.$$

Since the nonhomogeneous part of the original recursion is *n,* our first guess should be that there will be a particular solution of the form $a_n = An + B$, a general first degree expression. Putting that guess in the recursive formula, we get

$$An + B = 7(A(n-1) + B) - 10(A(n-2) + B) + n.$$

Gathering all the term on the left side of the equation gives

$$(4A - 1)n + (-13A + 4B) = 0.$$

So $4A - 1 = 0$ and $-13A + 4B = 0$. Thus, $A = 1/4$ and $B = 13/16$.

We finally have the general solution to the original problem:

$$a_n = \alpha 2^n + \beta 5^n + \frac{n}{4} + \frac{13}{16}.$$

The last step is using the initial conditions to determine α and β. Remember that the given initial conditions were for $n = 2, 3$: $a_2 = 5$ and $a_3 = 13$. The system to solve is

$$\begin{cases} 4\alpha + 25\beta + \frac{1}{4}(2) + \frac{13}{16} = 5 \\ 8\alpha + 125\beta + \frac{1}{4}(3) + \frac{13}{16} = 13 \end{cases}$$

with solution $\alpha = 7/12$ and $\beta = 13/240$.

Assembling the pieces, the solution is

$$a_n = \frac{7}{12}2^n + \frac{13}{240}3^n + \frac{n}{4} + \frac{13}{16}.$$

37.3. For exercise 36.5, we know the general solution to the related homogeneous recursion is

$$a_n^{(h)} = \alpha 2^n + \beta n 2^n.$$

That general solution of the related homogeneous recursion needs to be paired up with a particular solution of the original recursive formula

$$a_n = 4a_{n-1} - 4a_{n-2} + 2^n.$$

Since the nonhomogeneous part of the original recursion is 2^n, our first guess should be that there will be a particular solution of the form $a_n = A2^n$ a multiple of that exponential expression. Putting that guess in the recursive formula, we get

$$A2^n = 4(A2^{n-1}) - 4(A2^{n-2}) + 2^n.$$

Moving all the terms involving A to the left side of the equation gives

$$A(2^n - 4 \cdot 2^{n-1} + 4 \cdot 2^{n-2}) = 2^n$$

which can be rewritten as

$$A(0) = 2^n, \text{which isn't possible.}$$

Now, hold on: that makes some sense because our general solution to the homogeneous equation already has a term of the form $A2^n$, so we won't need any more of those. Likewise, we are already accounting for terms like $An2^n$. So let's take our guess for a particular solution up two notches to $a_n = An^2 2^n$. Putting that guess in the recursive formula, we get

$$An^2 2^n = 4(A(n-1)^2 2^{n-1}) - 4(A(n-2)^2 2^{n-2}) + 2^n.$$

Moving all the terms involving A to the left side of the equation and combining terms gives

$$A2^{n+1} = 2^n \text{ and so } A = 1/2.$$

So, a particular solution is $a_n^{(p)} = \frac{1}{2} n^2 2^n = n^2 2^{n-1}$.

The general solution to the original problem is

$$a_n = \alpha 2^n + \beta n 2^n + n^2 2^{n-1}.$$

The last step is using the initial conditions to determine α and β. Remember that the given initial conditions were for $n = 1, 2$: $a_1 = 3$ and $a_2 = 5$. The system to solve is

$$2\alpha + 2\beta + 1 = 3$$
$$4\alpha + 8\beta + 8 = 5$$

with solution $\alpha = 11/4$ and $\beta = -7/4$. Assembling the pieces, the solution is

$$a_n = \frac{11}{4} 2^n - \frac{7}{4} n 2^n + n^2 2^{n-1} = 2^{n-2}(11 - 7n + 2n^2).$$

37.4. For exercise 36.6, we know the general solution to the related homogeneous recursion is

$$a_n^{(h)} = \alpha 3^n + \beta n 3^n.$$

That general solution of the related homogeneous recursion needs to be paired up with a particular solution of the original recursive formula

$$a_n = 6a_{n-1} - 9a_{n-2} + n.$$

Since the nonhomogeneous part of the original recursion is n, our first guess should be that there will be a particular solution of the form $a_n = An + B$, the general first degree polynomial in n. Putting that guess in the recursive formula, we get

$$An + B = 6(A(n-1) + B) - 9(A(n-2) + B) + n.$$

Moving all the terms to the left side of the equation and combining terms gives

$$(4A - 1)n + (4B - 12A) = 0 \text{ which implies } A = \frac{1}{4} \text{ and } B = \frac{3}{4}.$$

So, a particular solution is $a_n^{(p)} = \frac{1}{4}n + \frac{3}{4}$.

The general solution to the original problem is

$$a_n = \alpha 3^n + \beta n 3^n + \frac{1}{4}n + \frac{3}{4}.$$

The last step is using the initial conditions to determine α and β. The given initial conditions are $a_0 = 1$ and $a_1 = 6$. The system to solve is

$$\begin{cases} \alpha + \frac{3}{4} = 1 \\ 3\alpha + 3\beta + 1 = 6 \end{cases}$$

With solution $\alpha = \frac{1}{4}$ and $\beta = \frac{17}{12}$. Assembling the pieces, the solution is

$$a_n = \frac{1}{4}(3^n) + \frac{17}{12}(n3^n) + \frac{1}{4}n + \frac{3}{4}, \text{ for } n \geq 0.$$

37.5. For exercise 36.8, we know the general solution to the related homogeneous recursion is

$$a_n^{(h)} = \alpha + \beta 2^n + \gamma 3^n.$$

That general solution of the related homogeneous recursion needs to be paired up with a particular solution of the original recursive formula

$$a_n = 6a_{n-1} - 11a_{n-2} + 6a_{n-3} + 2n + 1.$$

Since the nonhomogeneous part of the original recursion is the linear polynomial $2n + 1$, our first guess should be that there will be a particular solution of the form $a_n = An + B$, the general linear polynomial. Putting that guess in the recursive formula, we get

$$An + B = 6(A(n-1) + B) - 11(A(n-2) + B) + 6(A(n-3) + B) + 2n + 1.$$

Moving all the terms to the left side of the equation and combining terms gives

$$-2n + (2A - 1) = 0.$$

Well, that's not possible, so we will need to lift the guess up a bit by multiplying by n: our new guess for a particular solution is $An^2 + Bn$. Putting that guess in the recursive formula, we get

$$An^2 + Bn = 6(A(n-1)^2 + B(n-1)) - 11(A(n-2)^2 + B(n-2)) + 6(A(n-3)^2 + B(n-3)) + 2n + 1.$$

Moving all the terms to the left side of the equation and combining terms gives

$$(4a - 2)n + (2B - 16A - 1) = 0 \text{ which means } A = \frac{1}{2} \text{ and } B = \frac{9}{2}.$$

So, we have a particular solution $a_n^{(p)} = \frac{1}{2}n^2 + \frac{9}{2}n$. Adding that to the general solution of the related homogeneous recursion, we see that the general solution to the original problem is

$$a_n = \alpha + \beta 2^n + \gamma 3^n + \frac{1}{2}n^2 + \frac{9}{2}n.$$

The last step is using the initial conditions to determine α, β, and γ. The given initial conditions are $a_0 = 2$, $a_1 = 5$, and $a_2 = 15$. The system to solve is

$$\begin{cases} \alpha + \beta + \gamma = 2 \\ \alpha + 2\beta + 3\gamma + 5 = 5 \\ \alpha + 4\beta + 9\gamma + 10 = 15 \end{cases}$$

with solution $\alpha = 8, \beta = -10$, and $\gamma = 4$. Assembling the pieces, the solution is

$$a_n = 8 - 10 \cdot 2^n + 4 \cdot 3^n + \frac{1}{2}n^2 + \frac{9}{2}n, \text{ for } n \geq 0.$$

Whew: I'm relieved there were only five of these annoying, tedious, tiresome, monotonous problems.

Chapter 38

38.1. There are many possible isomorphisms. One is

$$\begin{array}{llll} a \to s & b \to t & c \to u & d \to v \\ e \to z & f \to y & g \to x & h \to w \end{array}$$

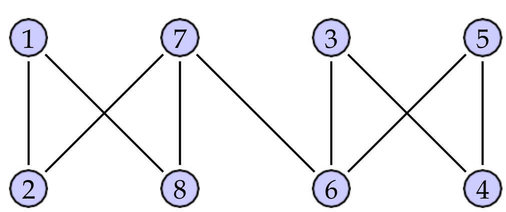
The adjacency matrix for G using vertex ordering a, b, c, d, e, f, g, h is

$$A_G = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \end{bmatrix}$$

The adjacency matrix of H, using the order s, t, u, v, z, y, x, w is identical to A_G.

38.2. Let $\bar{G}$ be the complement of G with respect to the graph K_8. Likewise, $\overline{H}$ will be K_8 with the edges of H erased. It is easy to see that if G and H are isomorphic, then so are $\bar{G}$ and $\overline{H}$. But $\bar{G}$ is just an 8-cycle, while $\overline{H}$ is two disjoint 4-cycles, and so they are not isomorphic. So G and H are not isomorphic.

38.3. There are several possible solutions. One is



38.4. A Hamiltonian circuit is $a, b, c, d, e, f, g, h, i, j, a$.

The graph G has four vertices of odd degree, so it is not Eulerian.

38.5. An Eulerian circuit is $a, b, c, d, e, f, g, h, a, d, g, b, e, h, c, f, a$.

38.6. In a Hamiltonian circuit, exactly two edges must be used at each vertex. Therefore a Hamiltonian circuit would include the edges $\{a, b\}, \{a, j\}, \{b, c\}, \{c, d\}, \{d, g\}$, and $\{g, h\}$. But then the edges $\{b, i\}, \{b, e\}$, and $\{e, d\}$ would be forbidden. This leaves only the single edge $\{e, f\}$ to include e in the circuit. Conclusion: the graph G is not Hamiltonian.

Appendix B

GNU Free Documentation License

Appendix: GNU Free Documentation License
Version 1.2, November 2002

B.1 PREAMBLE

The purpose of this License is to make a manual, textbook, or other functional and useful document ″free″ in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondarily, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others. This License is a kind of ″copyleft″, which means

that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software. We have

designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

1. APPLICABILITY AND DEFINITIONS This License applies to any manual or other work,

in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The ”Document”, below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as ”you”. You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law. A ”Modified Version” of the Document means any work containing the Document or a portion

of it, either copied verbatim, or with modifications and/or translated into another language. A

”Secondary Section” is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document’s overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them. The ”Invariant Sections” are certain Secondary Sections whose titles are

designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none. The ”Cover Texts” are certain short

passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words. A ”Transparent” copy of the Document means a

machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not "Transparent" is called "Opaque". Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only. The "Title Page" means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, "Title Page" means the text near the most prominent appearance of the work's title, preceding the beginning of the body of the text. A section "Entitled XYZ" means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as "Acknowledgements", "Dedications", "Endorsements", or "History".) To "Preserve the Title" of such a section when you modify the Document means that it remains a section "Entitled XYZ" according to this definition. The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

2. VERBATIM COPYING You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3. You may also lend copies, under the same conditions stated above, and you may publicly display copies.

3. COPYING IN QUANTITY If you publish printed copies (or copies in media that commonly

have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects. If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public. It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

4. MODIFICATIONS You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

(1) A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.

(2) B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors

of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.

(3) C. State on the Title page the name of the publisher of the Modified Version, as the publisher.

(4) D. Preserve all the copyright notices of the Document.

(5) E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.

(6) F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.

(7) G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.

(8) H. Include an unaltered copy of this License.

(9) I. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.

(10) J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.

(11) K. For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.

(12) L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.

(13) M. Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version.

(14) N. Do not re-title any existing section to be Entitled "Endorsements" or to conflict in title with any Invariant Section.

(15) O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles. You may add a section Entitled "Endorsements", provided it contains nothing but endorsements

of your Modified Version by various parties–for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard. You may add a

passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one. The author(s) and publisher(s) of the Document do not by this License give permission to

use their names for publicity for or to assert or imply endorsement of any Modified Version.

5. COMBINING DOCUMENTS You may combine the Document with other documents released

under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers. The combined work need only contain one copy of

this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work. In the combination,

you must combine any sections Entitled "History" in the various original documents, forming one section Entitled "History"; likewise combine any sections Entitled "Acknowledgements", and any sections Entitled "Dedications". You must delete all sections Entitled "Endorsements".

6. COLLECTIONS OF DOCUMENTS You may make a collection consisting of the Document

and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects. You may extract a single document from such a collection, and distribute it individually under this

License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

7. AGGREGATION WITH INDEPENDENT WORKS A compilation of the Document or its

derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an ”aggregate” if the copyright resulting from the compilation is not used to limit the legal rights of the compilation’s users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document. If the Cover

Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document’s Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

8. TRANSLATION Translation is considered a kind of modification, so you may distribute transla-

tions of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail. If a section in the Document is Entitled ”Acknowledgements”, ”Dedications”, or ”History”, the

requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

9. TERMINATION You may not copy, modify, sublicense, or distribute the Document except

as expressly provided for under this License. Any other attempt to copy, modify, sublicense or distribute the Document is void, and will automatically terminate your rights under this License.

However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

10. FUTURE REVISIONS OF THIS LICENSE The Free Software Foundation may publish new,

revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns.
See http://www.gnu.org/copyleft/. Each version of the License is given a distinguishing version

number. If the Document specifies that a particular numbered version of this License ”or any later version” applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation.

ADDENDUM: How to use this License for your documents To use this License in a document you have written, include a copy of the License in the document and put the following copyright and license notices just after the title page:

Copyright (c) YEAR YOUR NAME.
Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.2 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled "GNU Free Documentation License".

If you have Invariant Sections, Front-Cover Texts and Back-Cover Texts, replace the "with...Texts." line with this: with the Invariant Sections being LIST THEIR TITLES, with the Front-Cover Texts being LIST, and with the Back-Cover Texts being LIST.

If you have Invariant Sections without Cover Texts, or some other combination of the three, merge those two alternatives to suit the situation. If your document contains nontrivial examples of program code, we recommend releasing these examples in parallel under your choice of free software license, such as the GNU General Public License, to permit their use in free software.