Florida State University
Course Notes
MAD 2104 Discrete Mathematics I

Florida State University

Tallahassee, Florida 32306-4510

Contents

CHAPTER 1

Introduction to Sets and Functions

1. Introduction to Sets

1.1. Basic Terminology. We begin with a refresher in the basics of set theory. Our treatment will be an informal one rather than taking an axiomatic approach at this time. Later in the semester we will revisit sets with a more formal approach.

A **set** is a collection or group of objects or **elements** or **members**. (Cantor 1895)

- A set is said to **contain** its elements.
- In each situation or context, there must be an underlying **universal set** U, either specifically stated or understood.

Notation:

- If x is a member or element of the set S, we write $x \in S$.
- If x is not an element of S we write $x \notin S$.

1.2. Notation for Describing a Set.

EXAMPLE 1.2.1. *List the elements between braces:*

- $S = \{a, b, c, d\} = \{b, c, a, d, d\}$

Specify by attributes:

- $S = \{x | x \geq 5 \text{ or } x < 0\}$, *where the universe is the set of real numbers.*

Use brace notation with ellipses:

- $S = \{\ldots, -3, -2, -1\}$, *the set of negative integers.*

Discussion

Sets can be written in a variety of ways. One can, of course, simply list the elements if there are only a few. Another way is to use set builder notation, which specifies the sets using a predicate to indicate the attributes of the elements of the set. For example, the set of even integers is

$$\{x|x = 2n, n \in \mathbb{Z}\}$$

or

$$\{\dots, -2, 0, 2, 4, 6, \dots\}.$$

The first set could be read as "the set of all x's such that x is twice an integer." The symbol | stands for "such that." A colon is often used for "such that" as well, so the set of even integers could also be written

$$\{x : x = 2n, n \in \mathbb{Z}\}.$$

1.3. Common Universal Sets. The following notation will be used throughout these notes.

- $\mathbb{R}$ = the real numbers
- $\mathbb{N}$ = the natural numbers = $\{0, 1, 2, 3, \dots\}$
- $\mathbb{Z}$ = the integers = $\{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}$
- $\mathbb{Z}^+$ = the positive integers = $\{1, 2, 3, \dots\}$

Discussion

The real numbers, natural numbers, rational numbers, and integers have special notation which is understood to stand for these sets of numbers. Corresponding bold face letters are also a common notation for these sets of numbers. Some authors do not include 0 in the set of natural numbers. We will include zero.

1.4. Complements and Subsets.

DEFINITION 1.4.1. *The* **complement** *of A*

$$\overline{A} = \{x \in U|x \notin A\}.$$

DEFINITION 1.4.2. *A set A is a* **subset** *of a set B, denoted*

$A \subseteq B$, *if and only if every element of A is also an element of B.*

DEFINITION 1.4.3. *If $A \subseteq B$ but $A \neq B$ then we say A is a* **proper subset** *of B and denote it by*

$$A \subset B.$$

DEFINITION 1.4.4. *The* **null** *set, or* **empty** *set, denoted* $\emptyset$*, is the set with no members.*

Note:

- $\emptyset$ *is a subset of every set.*
- *A set is always a subset of itself.*

Discussion

Please study the notation for elements, subsets, proper subsets, and the empty set. Two other common notations for the complement of a set, A, is A^c and A'. Notice that we make a notational distinction between subsets in general and proper subsets. Not all texts and/or instructors make this distinction, and you should check in other courses whether or not the notation $\subset$ really does mean proper as it does here.

1.5. Element v.s. Subsets. Sets can be *subsets* and *elements* of other sets.

EXAMPLE 1.5.1. *Let* $A = \{\emptyset, \{\emptyset\}\}$*. Then* A *has two elements*

$$\emptyset \text{ and } \{\emptyset\}$$

and the four subsets

$$\emptyset, \{\emptyset\}, \{\{\emptyset\}\}, \{\emptyset, \{\emptyset\}\}.$$

EXAMPLE 1.5.2. *Pay close attention to whether the symbols means "element" or "subset" in this example*

If $S = \{2, 3, \{2\}, \{4\}\}$*, then*

- $2 \in S$
- $3 \in S$
- $4 \notin S$
- $\{2\} \in S$
- $\{3\} \notin S$
- $\{4\} \in S$
- $\{2\} \subset S$
- $\{3\} \subset S$
- $\{4\} \not\subset S$
- $\{\{2\}\} \subset S$
- $\{\{3\}\} \not\subset S$
- $\{\{4\}\} \subset S$

EXERCISE 1.5.1. *Let* $A = \{1, 2, \{1\}, \{1, 2\}\}$*. True or false?*

(a) $\{1\} \in A$ *(b)* $\{1\} \subseteq A$ *(c)* $2 \in A$
(d) $2 \subseteq A$ *(e)* $\{2\} \in A$ *(f)* $\{2\} \subseteq A$

1.6. Cardinality.

DEFINITION 1.6.1. *The number of (distinct) elements in a set A is called the* **cardinality** *of A and is written $|A|$.*

If the cardinality is a natural number, then the set is called **finite***, otherwise it is called* **infinite***.*

EXAMPLE 1.6.1. *Suppose $A = \{a, b\}$. Then*

$$|A| = 2,$$

EXAMPLE 1.6.2. *The cardinality of $\emptyset$ is 0, but the cardinality of $\{\emptyset, \{\emptyset\}\}$ is 2.*

EXAMPLE 1.6.3. *The set of natural numbers is infinite since its cardinality is not a natural number. The cardinality of the natural numbers is a* **transfinite cardinal number***.*

Discussion

Notice that the real numbers, natural numbers, integers, rational numbers, and irrational numbers are all infinite. Not all infinite sets are considered to be the same "size." The set of real numbers is considered to be a much larger set than the set of integers. In fact, this set is so large that we cannot possibly list all its elements in any organized manner the way the integers can be listed. We call a set like the real numbers that has too many elements to list *uncountable* and a set like the integers that can be listed is called *countable*. We will not delve any deeper than this into the study of the relative sizes of infinite sets in this course, but you may find it interesting to read further on this topic.

EXERCISE 1.6.1. *Let $A = \{1, 2, \{1\}, \{1, 2\}\}$, $B = \{1, \{2\}\}$, and $C = \{1, 2, 2, 2\}$. Find the cardinality of each set.*

1.7. Set Operations.

DEFINITION 1.7.1. *The* **union** *of sets A and B, denoted by $A \cup B$ (read "A union B"), is the set consisting of all elements that belong to either A or B or both. In symbols*

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

DEFINITION 1.7.2. *The* **intersection** *of sets A and B, denoted by $A \cap B$ (read "A intersection B"), is the set consisting of all elements that belong both A and B. In symbols*

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

DEFINITION 1.7.3. *The* **difference** *or* **relative compliment** *of two sets A and B, denoted by $A - B$ is the set of all elements in A that are not in B.*

$$A - B = \{x | x \in A \text{ and } x \notin B\}$$

Discussion

The operations of union and intersection are the basic operations used to combine two sets to form a third. Notice that we always have $A \subseteq A \cup B$ and $A \cap B \subseteq A$ for arbitrary sets A and B.

1.8. Example 1.8.1.

EXAMPLE 1.8.1. *Suppose*
$A = \{1, 3, 5, 7, 9, 11\}$,
$B = \{3, 4, 5, 6, 7\}$ *and*
$C = \{2, 4, 6, 8, 10\}$.
Then

(a) $A \cup B = \{1, 3, 4, 5, 6, 7, 9, 11\}$
(b) $A \cap B = \{3, 5, 7\}$
(c) $A \cup C = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11\}$
(d) $A \cap C = \emptyset$
(e) $A - B = \{1, 9, 11\}$

EXERCISE 1.8.1. *Let* $A = \{1, 2, \{1\}, \{1, 2\}\}$ *and* $B = \{1, \{2\}\}$. *True or false:*

(a) $2 \in A \cap B$ *(b)* $2 \in A \cup B$ *(c)* $2 \in A - B$
(d) $\{2\} \in A \cap B$ *(e)* $\{2\} \in A \cup B$ *(f)* $\{2\} \in A - B$

1.9. Product.

DEFINITION 1.9.1. *The* **(Cartesian) Product** *of two sets, A and B, is denoted $A \times B$ and is defined by*

$$A \times B = \{(a, b) | a \in A \text{ and } b \in B\}$$

Discussion

A cartesian product you have used in previous classes is $\mathbb{R} \times \mathbb{R}$. This is the same as the real plane and is shortened to $\mathbb{R}^2$. Elements of $\mathbb{R}^2$ are the points in the plane.

Notice the notation for and element in $\mathbb{R} \times \mathbb{R}$ is the same as the notation for an open interval of real numbers. In other words, $(3, 5)$ could mean the ordered pair in $\mathbb{R} \times \mathbb{R}$ or it could mean the interval $\{x \in \mathbb{R} | 3 < x < 5\}$. If the context does not make it clear which $(3, 5)$ stands for you should make it clear.

EXAMPLE 1.9.1. *Let* $A = \{a, b, c, d, e\}$ *and let* $B = \{1, 2\}$. *Then*

(1) $A \times B = \{(a, 1), (b, 1), (c, 1), (d, 1), (e, 1), (a, 2), (b, 2), (c, 2), (d, 2), (e, 2)\}$.

(2) $|A \times B| = 10$
(3) $\{a, 2\} \notin A \times B$
(4) $(a, 2) \notin A \cup B$

EXERCISE 1.9.1. *Let* $A = \{a, b, c, d, e\}$ *and let* $B = \{1, 2\}$. *Find*

(1) $B \times A$.
(2) $|B \times A|$
(3) Is $(a, 2) \in B \times A$*?*
(4) Is $(2, a) \in B \times A$*?*
(5) Is $2a \in B \times A$*?*

2. Introduction to Functions

2.1. Function.

DEFINITION 2.1.1. *Let A and B be sets. A* **function**

$$f\colon A \to B$$

is a rule which assigns to every element in A exactly one element in B.

If f assigns $a \in A$ to the element $b \in B$, then we write

$$f(a) = b,$$

and we call b the **image** *or* **value** *of f at a.*

Discussion

This is the familiar definition of a function f from a set A to a set B as a rule that assigns each element of A to exactly one element B. This is probably quite familiar to you from your courses in algebra and calculus. In the context of those subjects, the sets A and B are usually subsets of real numbers $\mathbb{R}$, and the *rule* usually refers to some concatenation of algebraic or transcendental operations which, when applied to a number in the set A, give a number in the set B. For example, we may define a function $f\colon \mathbb{R} \to \mathbb{R}$ by the formula (rule) $f(x) = \sqrt{1 + \sin x}$. We can then compute values of f – for example, $f(0) = 1$, $f(\pi/2) = \sqrt{2}$, $f(3\pi/2) = 0$, $f(1) = 1.357$ (approximately) – using knowledge of the sine function at special values of x and/or a calculator. Sometimes the rule may vary depending on which part of the set A the element x belongs. For example, the absolute value function $f\colon \mathbb{R} \to \mathbb{R}$ is defined by

$$f(x) = |x| = \begin{cases} x, & \text{if } x \geq 0, \\ -x, & \text{if } x < 0. \end{cases}$$

The *rule* that defines a function, however, need not be given by a formula of such as those above. For example, the rule that assigns to each resident of the state of Florida his or her last name defines a function from the set of Florida residents to the set of all possible names. There is certainly nothing formulaic about about the rule that defines this function. At the extreme we could randomly assign everyone in this class one of the digits 0 or 1, and we would have defined a function from the set of students in the class to the set $\{0, 1\}$. We will see a more formal definition of a function later on that avoids the use of the term *rule* but for now it will serve us reasonably well. We will instead concentrate on terminology related to the concept of a function, including special properties a function may possess.

2.2. Terminology Related to Functions. Let A and B be sets and suppose $f : A \to B$.

- The set A is called the **domain** of f.
- The set B is called the **codomain** of f.
- If $f(x) = y$, then x is a **preimage** of y. Note, there may be more than one preimage of y, but only one image (or value) of x.
- The set $f(A) = \{f(x)|x \in A\}$ is called the **range** of f.
- If $S \subseteq A$, then the **image** of S under f is the set
$$f(S) = \{f(s)|s \in S\}.$$
- If $T \subseteq B$, then the **preimage** of T under f is the set
$$f^{-1}(T) = \{x \in A|f(x) \in T\}.$$

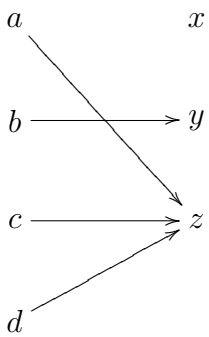
Discussion

Some of the fine points to remember:

- Every element in the domain must be assigned to exactly one element in the codomain.
- Not every element in the codomain is necessarily assigned to one of the elements in the domain.
- The range is the subset of the codomain consisting of all those elements that are the image of at least one element in the domain. It is the *image* of the domain.
- If a subset T of the codomain consists of a single element, say, $T = \{b\}$, then we usually write $f^{-1}(b)$ instead of $f^{-1}(\{b\})$. Regardless, $f^{-1}(b)$ is still a *subset* of A.

2.3. Example 2.3.1.

EXAMPLE 2.3.1. *Let $A = \{a, b, c, d\}$ and $B = \{x, y, z\}$. The function f is defined by the relation pictured below:*



- $f(a) = z$
- *the image of* d *is* z
- *the domain of* f *is* $A = \{a, b, c, d\}$
- *the codomain is* $B = \{x, y, z\}$
- $f(A) = \{y, z\}$
- $f(\{c, d\}) = \{z\}$
- $f^{-1}(y) = \{b\}$.
- $f^{-1}(z) = \{a, c, d\}$
- $f^{-1}(\{y, z\}) = \{a, b, c, d\}$
- $f^{-1}(x) = \emptyset$.

Discussion

This example helps illustrate some of the differences between the codomain and the range. $f(A) = \{y, z\}$ is the range, while the codomain is all of $B = \{x, y, z\}$.

Notice also that the image of a single element is a single element, but the preimage of a single element may be more than one element. Here is another example.

EXAMPLE 2.3.2. *Let* $f\colon \mathbb{N} \to \mathbb{R}$ *be defined by* $f(n) = \sqrt{n}$.

- *The domain is the set of natural numbers.*
- *The codomain is the set of real numbers.*
- *The range is* $\{0, 1, \sqrt{2}, \sqrt{3}, 2, \sqrt{5}, \dots\}$.
- *The image of* 5 *is* $\sqrt{5}$.
- *The preimage of* 5 *is* 25.
- *The preimage of* $\mathbb{N}$ *is the set of all perfect squares in* $\mathbb{N}$.

EXERCISE 2.3.1. *Suppose* $f\colon \mathbb{R} \to \mathbb{R}$ *is defined by* $f(x) = |x|$. *Find*

(1) the range of f.
(2) the image of $\mathbb{Z}$, *the set of integers.*
(3) $f^{-1}(\pi)$.
(4) $f^{-1}(-1)$.
(5) $f^{-1}(\mathbb{Q})$, *where* $\mathbb{Q}$ *is the set of rational numbers.*

2.4. Floor and Ceiling Functions.

DEFINITIONS 2.4.1. *Floor and ceiling functions:*

- *The* **floor function**, *denoted*

$$f(x) = \lfloor x \rfloor$$

or

$$f(x) = \mathit{floor}(x),$$

is the function that assigns to x the greatest integer less than or equal to x.

- *The* **ceiling function**, *denoted*

$$f(x) = \lceil x \rceil$$

or

$$f(x) = \mathit{ceiling}(x),$$

is the function that assigns to x the smallest integer greater than or equal to x.

Discussion

These two functions may be new to you. The floor function, $\lfloor x \rfloor$, also known as the "greatest integer function", and the ceiling function, $\lceil x \rceil$, are assumed to have domain the set of all reals, unless otherwise specified, and range is then the set of integers.

EXAMPLE 2.4.1. *(a)* $\lfloor 3.5 \rfloor = 3$
(b) $\lceil 3.5 \rceil = 4$
(c) $\lfloor -3.5 \rfloor = -4$
(d) $\lceil -3.5 \rceil = -3$
(e) notice that the floor function is the same as truncation for positive numbers.

EXERCISE 2.4.1. *Suppose x is a real number. Do you see any relationships among the values $\lfloor -x \rfloor$, $-\lfloor x \rfloor$, $\lceil -x \rceil$, and $-\lceil x \rceil$?*

EXERCISE 2.4.2. *Suppose $f\colon \mathbb{R} \to \mathbb{R}$ is defined by $f(x) = \lfloor x \rfloor$. Find*

(1) the range of f.
(2) the image of $\mathbb{Z}$, the set of integers.
(3) $f^{-1}(\pi)$.
(4) $f^{-1}(-1.5)$.
(5) $f^{-1}(\mathbb{N})$, where $\mathbb{N}$ is the set of natural numbers.
(6) $f^{-1}([2.5, 5.5])$.
(7) $f([2.5, 5.5])$.
(8) $f(f^{-1}([2.5, 5.5]))$.
(9) $f^{-1}(f([2.5, 5.5]))$.

2.5. Characteristic Function.

Definition: Let U be a universal set and $A \subseteq U$. The **Characteristic Function** of A is defined by

$$\chi_A(s) = \begin{cases} 1 & \text{if } s \in A \\ 0 & \text{if } s \notin A \end{cases}$$

Discussion

The Characteristic function is another function that may be new to you.

EXAMPLE 2.5.1. *Consider the set of integers as a subset of the real numbers. Then*

$$\chi_{\mathbb{Z}}(y)$$

will be 1 when y is an integer and will be zero otherwise.

EXERCISE 2.5.1. *Graph the function*

$$\chi_{\mathbb{Z}}$$

given in the previous example in the plane.

EXERCISE 2.5.2. *Find*

(1) $\chi_{\mathbb{Z}}(0)$
(2) $\chi_{\mathbb{Z}}^{-1}(0)$
(3) $\chi_{\mathbb{Z}}([3,5])$
(4) $\chi_{\mathbb{Z}}^{-1}([3,5])$

CHAPTER 2

Logic

1. Logic Definitions

1.1. Propositions.

DEFINITION 1.1.1. *A* **proposition** *is a declarative sentence that is either*

true (denoted either T or 1) or

false (denoted either F or 0).

Notation: Variables are used to represent propositions. The most common variables used are p, q, and r.

Discussion

Logic has been studied since the classical Greek period (600-300BC). The Greeks, most notably Thales, were the first to formally analyze the reasoning process. Aristotle (384-322BC), the "father of logic", and many other Greeks searched for universal truths that were irrefutable. A second great period for logic came with the use of symbols to simplify complicated logical arguments. Gottfried Leibniz (1646-1716) began this work at age 14, but failed to provide a workable foundation for symbolic logic. George Boole (1815-1864) is considered the "father of symbolic logic". He developed logic as an abstract mathematical system consisting of defined terms (propositions), operations (conjunction, disjunction, and negation), and rules for using the operations. It is this system that we will study in the first section.

Boole's basic idea was that if simple propositions could be represented by precise symbols, the relation between the propositions could be read as precisely as an algebraic equation. Boole developed an "algebra of logic" in which certain types of reasoning were reduced to manipulations of symbols.

1.2. Examples.

EXAMPLE 1.2.1. *"Drilling for oil caused dinosaurs to become extinct." is a proposition.*

EXAMPLE 1.2.2. *"Look out!" is not a proposition.*

EXAMPLE 1.2.3. *"How far is it to the next town?" is not a proposition.*

EXAMPLE 1.2.4. *"$x + 2 = 2x$" is not a proposition.*

EXAMPLE 1.2.5. *"$x + 2 = 2x$ when $x = -2$" is a proposition.*

Recall a *proposition* is a declarative sentence that is either true or false. Here are some further examples of propositions:

EXAMPLE 1.2.6. *All cows are brown.*

EXAMPLE 1.2.7. *The Earth is further from the sun than Venus.*

EXAMPLE 1.2.8. *There is life on Mars.*

EXAMPLE 1.2.9. $2 \times 2 = 5$.

Here are some sentences that are not propositions.

EXAMPLE 1.2.10. *"Do you want to go to the movies?" Since a question is not a declarative sentence, it fails to be a proposition.*

EXAMPLE 1.2.11. *"Clean up your room." Likewise, an imperative is not a declarative sentence; hence, fails to be a proposition.*

EXAMPLE 1.2.12. *"$2x = 2 + x$." This is a declarative sentence, but unless x is assigned a value or is otherwise prescribed, the sentence neither true nor false, hence, not a proposition.*

EXAMPLE 1.2.13. *"This sentence is false." What happens if you assume this statement is true? false? This example is called a paradox and is not a proposition, because it is neither true nor false.*

Each proposition can be assigned one of two *truth values*. We use T or 1 for true and use F or 0 for false.

1.3. Logical Operators.

DEFINITION 1.3.1. *Unary Operator* **negation:** *"not p", $\neg p$.*

DEFINITIONS 1.3.1. *Binary Operators*

(a) **conjunction:** *"p and q", $p \wedge q$.*
(b) **disjunction:** *"p or q", $p \vee q$.*
(c) **exclusive or:** *"exactly one of p or q", "p xor q", $p \oplus q$.*
(d) **implication:** *"if p then q", $p \rightarrow q$.*
(e) **biconditional:** *"p if and only if q", $p \leftrightarrow q$.*

Discussion

A sentence like "I can jump and skip" can be thought of as a combination of the two sentences "I can jump" and "I can skip." When we analyze arguments or logical expression it is very helpful to break a sentence down to some composition of simpler statements.

We can create *compound propositions* using propositional variables, such as $p, q, r, s, ...$, and *connectives* or *logical operators*. A logical operator is either a *unary* operator, meaning it is applied to only a single proposition; or a *binary* operator, meaning it is applied to two propositions. *Truth tables* are used to exhibit the relationship between the truth values of a compound proposition and the truth values of its component propositions.

1.4. Negation. Negation Operator, "not", has symbol $\neg$.

EXAMPLE 1.4.1. *p: This book is interesting.*

$\neg p$ can be read as:

(i.) This book is not interesting.
(ii.) This book is uninteresting.
(iii.) It is not the case that this book is interesting.

Truth Table:

p	$\neg\ p$
T	F
F	T

Discussion

The *negation* operator the a unary operator which, when applied to a proposition p, changes the truth value of p. That is, the negation of a proposition p, denoted by $\neg p$, is the proposition that is false when p is true and true when p is false. For example, if p is the statement "I understand this", then its negation would be "I do not understand this" or "It is not the case that I understand this". Another notation commonly used for the negation of p is $\sim p$.

Generally, an appropriately inserted "not" or removed "not" is sufficient to negate a simple statement. Negating a compound statement may be a bit more complicated as we will see later on.

1.5. Conjunction. Conjunction Operator, "and", has symbol $\wedge$.

EXAMPLE 1.5.1. *p: This book is interesting. q: I am staying at home.*

$p \wedge q$*: This book is interesting, and I am staying at home.*

Truth Table:

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

Discussion

The *conjunction* operator is the binary operator which, when applied to two propositions p and q, yields the proposition "p and q", denoted $p \wedge q$. The conjunction $p \wedge q$ of p and q is the proposition that is true when both p and q are true and false otherwise.

1.6. Disjunction. Disjunction Operator, inclusive "or", has symbol $\vee$.

EXAMPLE 1.6.1. *p: This book is interesting. q: I am staying at home.*

$p \vee q$*: This book is interesting, or I am staying at home.*

Truth Table:

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

Discussion

The *disjunction* operator is the binary operator which, when applied to two propositions p and q, yields the proposition "p or q", denoted $p \vee q$. The disjunction $p \vee q$ of p and q is the proposition that is true when either p is true, q is true, or *both* are true, and is false otherwise. Thus, the "or" intended here is the *inclusive or*. In fact, the symbol $\vee$ is the abbreviation of the Latin word *vel* for the inclusive "or".

1.7. Exclusive Or. Exclusive Or Operator, "xor", has symbol $\oplus$.

EXAMPLE 1.7.1. *p: This book is interesting. q: I am staying at home.*

$p \oplus q$*: Either this book is interesting, or I am staying at home, but not both.*

Truth Table:

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

Discussion

The *exclusive or* is the binary operator which, when applied to two propositions p and q yields the proposition "p xor q", denoted $p \oplus q$, which is true if exactly one of p or q is true, but not both. It is false if both are true or if both are false.

Many times in our every day language we use "or" in the exclusive sense. In logic, however, we always mean the inclusive or when we simply use "or" as a connective in a proposition. If we mean the exclusive or it must be specified. For example, in a restaurant a menu may say there is a choice of soup or salad with a meal. In logic this would mean that a customer may choose both a soup and salad with their meal. The logical implication of this statement, however, is probably not what is intended. To create a sentence that logically states the intent the menu could say that there is a choice of *either* soup or salad (but not both). The phrase "either ... or ..." is normally indicates the exclusive or.

1.8. Implications. Implication Operator, "if...then...", has symbol $\rightarrow$.

EXAMPLE 1.8.1. *p: This book is interesting. q: I am staying at home.*

$p \rightarrow q$*: If this book is interesting, then I am staying at home.*

Truth Table:

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

Equivalent Forms of "If p then q":

- p implies q
- If p, q
- p only if q
- p is a sufficient condition for q
- q if p
- q whenever p
- q is a necessary condition for p

Discussion

The *implication* $p \to q$ is the proposition that is often read "if p then q." "If p then q" is false precisely when p is true but q is false. There are many ways to say this connective in English. You should study the various forms as shown above.

One way to think of the meaning of $p \to q$ is to consider it a contract that says if the first condition is satisfied, then the second will also be satisfied. If the first condition, p, is not satisfied, then the condition of the contract is null and void. In this case, it does not matter if the second condition is satisfied or not, the contract is still upheld.

For example, suppose your friend tells you that if you meet her for lunch, she will give you a book she wants you to read. According to this statement, you would expect her to give you a book if you do go to meet her for lunch. But what if you do not meet her for lunch? She did not say anything about that possible situation, so she would not be breaking any kind of promise if she dropped the book off at your house that night or if she just decided not to give you the book at all. If either of these last two possibilities happens we would still say the implication stated was true, because she did not break her promise.

1.9. Terminology. For the compound statement $p \to q$

- p is called the **premise, hypothesis,** or the **antecedent**.
- q is called the **conclusion** or **consequent**.
- $q \to p$ is the **converse** of $p \to q$.
- $\neg p \to \neg q$ is the **inverse** of $p \to q$.
- $\neg q \to \neg p$ is the **contrapositive** of $p \to q$.

Discussion

We will see later that the converse and the inverse are not equivalent to the original implication, but the contrapositive $\neg q \to \neg p$ is. In other words, $p \to q$ and its contrapositive have the exact same truth values.

1.10. Example.

EXAMPLE 1.10.1. *Implication: If this book is interesting, then I am staying at home.*

- **Converse:** *If I am staying at home, then this book is interesting.*
- **Inverse:** *If this book is not interesting, then I am not staying at home.*
- **Contrapositive:** *If I am not staying at home, then this book is not interesting.*

Discussion

The converse of your friend's promise given above would be "if she gives you a book she wants you to read, then you will meet her for lunch," and the inverse would be "If you do not meet her for lunch, then she will not give you the book." We can see from the discussion about this statement that neither of these are the same as the original promise. The contrapositive of the statement is "if she does not give you the book, then you do not meet her for lunch." This is, in fact, equivalent to the original promise. Think about when would this promise be broken. It should be the exact same situation where the original promise is broken.

1.11. Biconditional. Biconditional Operator, "if and only if", has symbol $\leftrightarrow$.

EXAMPLE 1.11.1. *p: This book is interesting. q: I am staying at home.*

$p \leftrightarrow q$: *This book is interesting if and only if I am staying at home.*

Truth Table:

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

Discussion

The biconditional statement is equivalent to $(p \rightarrow q) \wedge (q \rightarrow p)$. In other words, for $p \leftrightarrow q$ to be true we must have both p and q true or both false. The difference between the implication and biconditional operators can often be confusing, because in our every day language we sometimes say an "if...then" statement, $p \rightarrow q$, when we actually mean the *biconditional* statement $p \leftrightarrow q$. Consider the statement you may have heard from your mother (or may have said to your children): "If you eat your

broccoli, then you may have some ice cream." Following the strict logical meaning of the first statement, the child still may or may not have ice cream even if the broccoli isn't eaten. The "if...then" construction does not indicate what would happen in the case when the hypothesis is not true. The intent of this statement, however, is most likely that the child *must* eat the broccoli in order to get the ice cream.

When we set out to prove a biconditional statement, we often break the proof down into two parts. First we prove the implication $p \to q$, and then we prove the converse $q \to p$.

Another type of "if...then" statement you may have already encountered is the one used in computer languages. In this "if...then" statement, the premise is a condition to be tested, and if it is true then the conclusion is a procedure that will be performed. If the premise is not true, then the procedure will not be performed. Notice this is different from "if...then" in logic. It is actually closer to the biconditional in logic. However, it is not actually a logical statement at all since the "conclusion" is really a list of commands, not a proposition.

1.12. NAND and NOR Operators.

DEFINITION 1.12.1. *The* **NAND Operator**, *which has symbol* $|$ *("Sheffer Stroke"), is defined by the truth table*

p	q	$p \mid q$
T	T	F
T	F	T
F	T	T
F	F	T

DEFINITION 1.12.2. *The* **NOR Operator**, *which has symbol* $\downarrow$ *("Peirce Arrow"), is defined by the truth table*

p	q	$p \downarrow q$
T	T	F
T	F	F
F	T	F
F	F	T

Discussion

These two additional operators are very useful as logical gates in a combinatorial circuit, a topic we will discuss later.

1.13. Example.

EXAMPLE 1.13.1. *Write the following statement symbolically, and then make a truth table for the statement. "If I go to the mall or go to the movies, then I will not go to the gym."*

Solution. *Suppose we set*

- $p = I$ *go to the mall*
- $q = I$ *go to the movies*
- $r = I$ *will go to the gym*

The proposition can then be expressed as "If p *or* q*, then not* r*," or* $(p \vee q) \rightarrow \neg r$.

p	q	r	$(p \vee q)$	$\neg r$	$(p \vee q) \rightarrow \neg r$
T	T	T	T	F	F
T	T	F	T	T	T
T	F	T	T	F	F
T	F	F	T	T	T
F	T	T	T	F	F
F	T	F	T	T	T
F	F	T	F	F	T
F	F	F	F	T	T

Discussion

When building a truth table for a compound proposition, you need a row for every possible combination of T's and F's for the component propositions. Notice if there is only one proposition involved, there are 2 rows. If there are two propositions, there are 4 rows, if there are 3 propositions there are 8 rows.

EXERCISE 1.13.1. *How many rows should a truth table have for a statement involving* n *different propositions?*

propositions p, q, and r, you would, in theory, only need four columns: one for each of p, q, and r, and one for the compound proposition under discussion, which is $(p \vee q) \rightarrow \neg r$ in this example. In practice, however, you will probably want to have a column for each of the successive intermediate propositions used to build the final one. In this example it is convenient to have a column for $p \vee q$ and a column for $\neg r$, so that the truth value in each row in the column for $(p \vee q) \rightarrow \neg r$ is easily supplied from the truth values for $p \vee q$ and $\neg r$ in that row.

Another reason why you should show the intermediate columns in your truth table is for grading purposes. If you make an error in a truth table and do not give this

extra information, it will be difficult to evaluate your error and give you partial credit.

EXAMPLE 1.13.2. *Suppose p is the proposition "the apple is delicious" and q is the proposition "I ate the apple." Notice the difference between the two statements below.*

(a) $\neg p \wedge q$ *= The apple is not delicious, and I ate the apple.*
(b) $\neg(p \wedge q)$ *= It is not the case that: the apple is delicious and I ate the apple.*

EXERCISE 1.13.2. *Find another way to express Example 1.13.2 Part b without using the phrase "It is not the case."*

EXAMPLE 1.13.3. *Express the proposition "If you work hard and do not get distracted, then you can finish the job" symbolically as a compound proposition in terms of simple propositions and logical operators.*

Set

- p = you work hard
- q = you get distracted
- r = you can finish the job

In terms of p, q, and r, the given proposition can be written

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

The comma in Example 1.13.3 is not necessary to distinguish the order of the operators, but consider the sentence "If the fish is cooked then dinner is ready and I am hungry." Should this sentence be interpreted as $f \rightarrow (r \wedge h)$ or $(f \rightarrow r) \wedge h$, where f, r, and h are the natural choices for the simple propositions? A comma needs to be inserted in this sentence to make the meaning clear or rearranging the sentence could make the meaning clear.

EXERCISE 1.13.3. *Insert a comma into the sentence "If the fish is cooked then dinner is ready and I am hungry." to make the sentence mean*

(a) $f \rightarrow (r \wedge h)$
(b) $(f \rightarrow r) \wedge h$

EXAMPLE 1.13.4. *Here we build a truth table for $p \rightarrow (q \rightarrow r)$ and $(p \wedge q) \rightarrow r$. When creating a table for more than one proposition, we may simply add the necessary columns to a single truth table.*

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

EXERCISE 1.13.4. *Build one truth table for $f \to (r \wedge h)$ and $(f \to r) \wedge h$.*

1.14. Bit Strings.

DEFINITION 1.14.1. *A* **bit** *is a 0 or a 1 and a* **bit string** *is a list or string of bits.*

The logical operators can be turned into **bit operators** by thinking of 0 as false and 1 as true. The obvious substitutions then give the table

$\overline{0} = 1$	$\overline{1} = 0$	
$0 \vee 0 = 0$	$0 \wedge 0 = 0$	$0 \oplus 0 = 0$
$0 \vee 1 = 1$	$0 \wedge 1 = 0$	$0 \oplus 1 = 1$
$1 \vee 0 = 1$	$1 \wedge 0 = 0$	$1 \oplus 0 = 1$
$1 \vee 1 = 1$	$1 \wedge 1 = 1$	$1 \oplus 1 = 0$

Discussion

We can define the *bitwise NEGATION* of a string and *bitwise OR, bitwise AND,* and *bitwise XOR* of two bit strings of the same length by applying the logical operators to the corresponding bits in the natural way.

EXAMPLE 1.14.1.

(a) $\overline{11010} = 00101$

(b) $11010 \vee 10001 = 11011$

(c) $11010 \wedge 10001 = 10000$

(d) $11010 \oplus 10001 = 01011$

2. Propositional Equivalences

2.1. Tautology/Contradiction/Contingency.

DEFINITION 2.1.1. *A* **tautology** *is a proposition that is always true.*

EXAMPLE 2.1.1. $p \vee \neg p$

DEFINITION 2.1.2. *A* **contradiction** *is a proposition that is always false.*

EXAMPLE 2.1.2. $p \wedge \neg p$

DEFINITION 2.1.3. *A* **contigency** *is a proposition that is neither a tautology nor a contradiction.*

EXAMPLE 2.1.3. $p \vee q \rightarrow \neg r$

Discussion

One of the important techniques used in proving theorems is to replace, or substitute, one proposition by another one that is equivalent to it. In this section we will list some of the basic propositional equivalences and show how they can be used to prove other equivalences.

Let us look at the classic example of a tautology, $p \vee \neg p$. The truth table

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

shows that $p \vee \neg p$ is true no matter the truth value of p.

> [*Side Note.* This tautology, called the *law of excluded middle*, is a direct consequence of our basic assumption that a proposition is a statement that is either true or false. Thus, the logic we will discuss here, so-called Aristotelian logic, might be described as a "2-valued" logic, and it is the logical basis for most of the theory of modern mathematics, at least as it has developed in western culture. There is, however, a consistent logical system, known as constructivist, or intuitionistic, logic which does not assume the law of excluded middle. This results in a 3-valued logic in which one allows for

a third possibility, namely, "other." In this system proving that a statement is "not true" is not the same as proving that it is "false," so that indirect proofs, which we shall soon discuss, would not be valid. If you are tempted to dismiss this concept, you should be aware that there are those who believe that in many ways this type of logic is much closer to the logic used in computer science than Aristotelian logic. You are encouraged to explore this idea: there is plenty of material to be found in your library or through the worldwide web.]

The proposition $p \vee \neg(p \wedge q)$ is also a tautology as the following the truth table illustrates.

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

EXERCISE 2.1.1. *Build a truth table to verify that the proposition* $(p \leftrightarrow q) \wedge (\neg p \wedge q)$ *is a contradiction.*

2.2. Logically Equivalent.

DEFINITION 2.2.1. *Propositions* r *and* s *are* **logically equivalent** *if the statement* $r \leftrightarrow s$ *is a tautology.*

Notation: If r and s are logically equivalent, we write

$$r \Leftrightarrow s.$$

Discussion

A second notation often used to mean statements r and s are logically equivalent is $r \equiv s$. You can determine whether compound propositions r and s are logically equivalent by building a single truth table for both propositions and checking to see that they have exactly the same truth values.

Notice the new symbol $r \Leftrightarrow s$, which is used to denote that r and s are logically equivalent, is defined to mean the statement $r \leftrightarrow s$ is a tautology. In a sense the

symbols $\leftrightarrow$ and $\Leftrightarrow$ convey similar information when used in a sentence. However, $r \Leftrightarrow s$ is generally used to assert that the statement $r \leftrightarrow s$ is, in fact, true while the statement $r \leftrightarrow s$ alone does not imply any particular truth value. The symbol $\Leftrightarrow$ is the preferred shorthand for "is equivalent to."

2.3. Examples.

EXAMPLE 2.3.1. *Show that $(p \to q) \wedge (q \to p)$ is logically equivalent to $p \leftrightarrow q$.*

Solution 1. *Show the truth values of both propositions are identical.*

Truth Table:

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

Solution 2. *Examine every possible case in which the statement $(p \to q) \wedge (q \to p)$ may not have the same truth value as $p \leftrightarrow q$*

Case 1. Suppose $(p \to q) \wedge (q \to p)$ is false and $p \leftrightarrow q$ is true.
- *Assume $p \to q$ is false. Then p is true and q is false. But if this is the case, the $p \leftrightarrow q$ is false.*
- *Assume $q \to p$ is false. Then q is true and p is false. But if this is the case, the $p \leftrightarrow q$ is false.*

Case 2. Suppose $(p \to q) \wedge (q \to p)$ is true and $p \leftrightarrow q$ is false. If the latter is false, the p and q do not have the same truth value.
- *Assume p is true and q is false. Then $p \to q$ is false, the the conjunction is also must be false.*
- *Assume p is false and q is true. Then $q \to p$ is false, the the conjunction is also must be false.*

We exhausted all the possibilities, so the two propositions must be logically equivalent.

Discussion

This example illustrates an alternative to using truth tables to establish the equivalence of two propositions. An alternative proof is obtained by excluding all possible ways in which the propositions may fail to be equivalent. Here is another example.

EXAMPLE 2.3.2. *Show* $\neg(p \to q)$ *is equivalent to* $p \wedge \neg q$.

Solution 1. Build a truth table containing each of the statements.

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

Since the truth values for $\neg(p \to q)$ and $p \wedge \neg q$ are exactly the same for all possible combinations of truth values of p and q, the two propositions are equivalent.

Solution 2. We consider how the two propositions could *fail* be equivalent. This can happen only if the first is true and the second is false or vice versa.

Case 1. Suppose $\neg(p \to q)$ is true and $p \wedge \neg q$ is false.
$\neg(p \to q)$ would be true if $p \to q$ if false. Now this only occurs if p is true and q is false. However, if p is true and q is false, then $p \wedge \neg q$ will be true. Hence this case is not possible.

Case 2. Suppose $\neg(p \to q)$ is false and $p \wedge \neg q$ is true.
$p \wedge \neg q$ is true only if p is true and q is false. But in this case, $\neg(p \to q)$ will be true. So this case is not possible either.

Since it is not possible for the two propositions to have different truth values, they must be equivalent.

EXERCISE 2.3.1. *Use a truth table to show that the propositions* $p \leftrightarrow q$ *and* $\neg(p \oplus q)$ *are equivalent.*

EXERCISE 2.3.2. *Use the method of Solution 2 in Example 2.3.2 to show that the propositions* $p \leftrightarrow q$ *and* $\neg(p \oplus q)$ *are equivalent.*

2.4. Important Logical Equivalences. The logical equivalences below are important equivalences that should be memorized.

Identity Laws:	$p \wedge T \Leftrightarrow p$
	$p \vee F \Leftrightarrow p$
Domination Laws:	$p \vee T \Leftrightarrow T$
	$p \wedge F \Leftrightarrow F$
Idempotent Laws:	$p \vee p \Leftrightarrow p$
	$p \wedge p \Leftrightarrow p$
Double Negation Law:	$\neg(\neg p) \Leftrightarrow p$
Commutative Laws:	$p \vee q \Leftrightarrow q \vee p$
	$p \wedge q \Leftrightarrow q \wedge p$
Associative Laws:	$(p \vee q) \vee r \Leftrightarrow p \vee (q \vee r)$
	$(p \wedge q) \wedge r \Leftrightarrow p \wedge (q \wedge r)$
Distributive Laws:	$p \vee (q \wedge r) \Leftrightarrow (p \vee q) \wedge (p \vee r)$
	$p \wedge (q \vee r) \Leftrightarrow (p \wedge q) \vee (p \wedge r)$
De Morgan's Laws:	$\neg(p \wedge q) \Leftrightarrow \neg p \vee \neg q$
	$\neg(p \vee q) \Leftrightarrow \neg p \wedge \neg q$
Absorption Laws:	$p \wedge (p \vee q) \Leftrightarrow p$
	$p \vee (p \wedge q) \Leftrightarrow p$

Implication Law: $(p \to q) \Leftrightarrow (\neg p \vee q)$

Contrapositive Law: $(p \to q) \Leftrightarrow (\neg q \to \neg p)$

Tautology: $p \vee \neg p \Leftrightarrow T$

Contradiction: $p \wedge \neg p \Leftrightarrow F$

Equivalence: $(p \to q) \wedge (q \to p) \Leftrightarrow (p \leftrightarrow q)$

Discussion

Study carefully what each of these equivalences is saying. With the possible exceptions of the De Morgan Laws, they are fairly straight-forward to understand. The main difficulty you might have with these equivalences is remembering their names.

EXAMPLE 2.4.1. *Use the logical equivalences above and substitution to establish the equivalence of the statements in Example 2.3.2.*

Solution.

$$\begin{aligned} \neg(p \to q) &\Leftrightarrow \neg(\neg p \vee q) && \textit{Implication Law} \\ &\Leftrightarrow \neg\neg p \wedge \neg q && \textit{De Morgan's Law} \\ &\Leftrightarrow p \wedge \neg q && \textit{Double Negation Law} \end{aligned}$$

This method is very similar to simplifying an algebraic expression. You are using the basic equivalences in somewhat the same way you use algebraic rules like $2x - 3x = -x$ or $\dfrac{(x+1)(x-3)}{x-3} = x + 1$.

EXERCISE 2.4.1. *Use the propositional equivalences in the list of important logical equivalences above to prove* $[\neg(s \wedge t) \wedge (\neg w \to t)] \to (s \to w)$ *is a tautology.*

EXERCISE 2.4.2. *Use truth tables to verify De Morgan's Laws.*

2.5. Simplifying Propositions.

EXAMPLE 2.5.1. *Use the logical equivalences above to show that* $\neg(p \vee \neg(p \wedge q))$ *is a contradiction.*

Solution.

$$\begin{aligned}
&\neg(p \vee \neg(p \wedge q)) && \\
&\Leftrightarrow \neg p \wedge \neg(\neg(p \wedge q)) && \textit{De Morgan's Law} \\
&\Leftrightarrow \neg p \wedge (p \wedge q) && \textit{Double Negation Law} \\
&\Leftrightarrow (\neg p \wedge p) \wedge q && \textit{Associative Law} \\
&\Leftrightarrow F \wedge q && \textit{Contradiction} \\
&\Leftrightarrow F && \textit{Domination Law}
\end{aligned}$$

EXAMPLE 2.5.2. *Find a simple form for the negation of the proposition "If the sun is shining, then I am going to the ball game."*

Solution. *This proposition is of the form* $p \rightarrow q$*. As we showed in Example 2.3.2 its negation,* $\neg(p \rightarrow q)$*, is equivalent to* $p \wedge \neg q$*. This is the proposition*

"The sun is shining, and I am not going to the ball game."

Discussion

The main thing we should learn from Examples 2.3.2 and 2.5.2 is that the negation of an implication is *not* equivalent to another implication, such as "If the sun is shining, then I am not going to the ball game" or "If the sun is not shining, I am going to the ball game." This may be seen by comparing the corresponding truth tables:

p	q	$p \rightarrow \neg q$	$\neg(p \rightarrow q) \Leftrightarrow (p \wedge \neg q)$	$\neg p \rightarrow q$
T	T	F	F	T
T	F	T	T	T
F	T	T	F	T
F	F	T	F	F

If you were to construct truth tables for all of the other possible implications of the form $r \rightarrow s$, where each of r and s is one of p, $\neg p$, q, or $\neg q$, you will observe that none of these propositions is equivalent to $\neg(p \rightarrow q)$.

The rule $\neg(p \to q) \Leftrightarrow p \wedge \neg q$ should be memorized. One way to memorize this equivalence is to keep in mind that the negation of $p \to q$ is the statement that describes the only case in which $p \to q$ is false.

2.6. Implication.

DEFINITION 2.6.1. *We say the proposition r implies the proposition s and write $r \Rightarrow s$ if $r \to s$ is a tautology.*

This is very similar to the ideas previously discussed regarding the $\Leftrightarrow$ verses $\leftrightarrow$. We use $r \Rightarrow s$ to imply that the statement $r \to s$ is true, while that statement $r \to s$ alone does not imply any particular truth value. The symbol $\Rightarrow$ is often used in proofs as a shorthand for "implies."

EXERCISE 2.6.1. *Prove $(p \to q) \wedge \neg q \Rightarrow \neg p$.*

2.7. Normal or Canonical Forms.

DEFINITION 2.7.1. *Every compound proposition in the propositional variables p, q, r, ..., is uniquely equivalent to a proposition that is formed by taking the disjunction of conjunctions of some combination of the variables $p, q, r, ...$ or their negations. This is called the* **disjunctive normal form** *of a proposition.*

Discussion

The *disjunctive normal form* of a compound proposition is a natural and useful choice for representing the proposition from among all equivalent forms, although it may not be the simplest representative. We will find this concept useful when we arrive at the module on Boolean algebra.

2.8. Examples.

EXAMPLE 2.8.1. *Construct a proposition in disjunctive normal form that is true precisely when*

(1) p is true and q is false
Solution. $p \wedge \neg q$
(2) p is true and q is false or when p is true and q is true.

Solution. $(p \wedge \neg q) \vee (p \wedge q)$
(3) either p is true or q is true, and r is false
Solution. $(p \vee q) \wedge \neg r \Leftrightarrow (p \wedge \neg r) \vee (q \wedge \neg r)$ *(Distributive Law)*

(Notice that the second example could be simplified to just p.)

Discussion

The methods by which we arrived at the disjunctive normal form in these examples may seem a little *ad hoc*. We now demonstrate, through further examples, a sure-fire method for its construction.

2.9. Constructing Disjunctive Normal Forms.

EXAMPLE 2.9.1. *Find the disjunctive normal form for the proposition* $p \to q$.

Solution. *Construct a truth table for* $p \to q$*:*

p	q	$p \to q$	
T	T	T	$\leftarrow$
T	F	F	
F	T	T	$\leftarrow$
F	F	T	$\leftarrow$

$p \to q$ is true when either
p is true and q is true, or
p is false and q is true, or
p is false and q is false.
The disjunctive normal form is then

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

Discussion

This example shows how a truth table can be used in a systematic way to construct the disjunctive normal forms. Here is another example.

EXAMPLE 2.9.2. *Construct the disjunctive normal form of the proposition*

$$(p \to q) \wedge \neg r$$

Solution. *Write out the truth table for* $(p \to q) \wedge \neg r$*:*

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

The disjunctive normal form will be a disjunction of three conjunctions, one for each row in the truth table that gives the truth value T for $(p \to q) \wedge \neg r$. *These rows have been boxed. In each conjunction we will use* p *if the truth value of* p *in that row is* T *and* $\neg p$ *if the truth value of* p *is* F, q *if the truth value of* q *in that row is* T *and* $\neg q$ *if the truth value of* q *is* F, *etc. The disjunctive normal form for* $(p \to q) \wedge \neg r$ *is then*

$$(p \wedge q \wedge \neg r) \vee (\neg p \wedge q \wedge \neg r) \vee (\neg p \wedge \neg q \wedge \neg r),$$

because each of these conjunctions is true only for the combination of truth values of p, q, *and* r *found in the corresponding row. That is,* $(p \wedge q \wedge \neg r)$ *has truth value* T *only for the combination of truth values in row 2,* $(\neg p \wedge q \wedge \neg r)$ *has truth value* T *only for the combination of truth values in row 6, etc. Their disjunction will be true for precisely the three combinations of truth values of* p, q, *and* r *for which* $(p \to q) \wedge \neg r$ *is also true.*

Terminology. The individual conjunctions that make up the disjunctive normal form are called **minterms**. In the previous example, the disjunctive normal form for the proposition $(p \to q) \wedge \neg r$ has three minterms, $(p \wedge q \wedge \neg r)$, $(\neg p \wedge q \wedge \neg r)$, and $(\neg p \wedge \neg q \wedge \neg r)$.

2.10. Conjunctive Normal Form. The **conjunctive normal form** of a proposition is another "canonical form" that may occasionally be useful, but not to the same degree as the disjunctive normal form. As the name should suggests after our discussion above, the conjunctive normal form of a proposition is the equivalent form that consists of a "conjunction of disjunctions." It is easily constructed indirectly using disjunctive normal forms by observing that if you negate a disjunctive normal form you get a conjunctive normal form. For example, three applications of De Morgan's Laws gives

$$\neg[(p \wedge \neg q) \vee (\neg p \wedge \neg q)] \Leftrightarrow (\neg p \vee q) \wedge (p \vee q).$$

Thus, if you want to get the conjunctive normal form of a proposition, construct the disjunctive normal form of its *negation* and then negate again and apply De Morgan's Laws.

EXAMPLE 2.10.1. *Find the conjunctive normal form of the proposition* $(p \wedge \neg q) \vee r$.

Solution.

(1) Negate: $\neg[(p \wedge \neg q) \vee r] \Leftrightarrow (\neg p \vee q) \wedge \neg r$.

(2) Find the disjunctive normal form of $(\neg p \vee q) \wedge \neg r$:

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

The disjunctive normal form for $(\neg p \vee q) \wedge \neg r$ is

$$(p \wedge q \wedge \neg r) \vee (\neg p \wedge q \wedge \neg r) \vee (\neg p \wedge \neg q \wedge \neg r).$$

(3) The conjunctive normal form for $(p \wedge \neg q) \vee r$ is then the negation of this last expression, which, by De Morgan's Laws, is

$$(\neg p \vee \neg q \vee r) \wedge (p \vee \neg q \vee r) \wedge (p \vee q \vee r).$$

3. Predicates and Quantifiers

3.1. Predicates and Quantifiers.

DEFINITION 3.1.1. *A* **predicate** *or* **propositional function** *is a description of the property (or properties) a variable or subject may have. A proposition may be created from a propositional function by either assigning a value to the variable or by quantification.*

DEFINITION 3.1.2. *The independent variable of a propositional function must have a* **universe of discourse***, which is a set from which the variable can take values.*

Discussion

Recall from the introduction to logic that the sentence "$x + 2 = 2x$" is not a proposition, but if we assign a value for x then it becomes a proposition. The phrase "$x + 2 = 2x$" can be treated as a function for which the input is a value of x and the output is a proposition.

Another way we could turn this sentence into a proposition is to quantify its variable. For example, "for every real number x, $x + 2 = 2x$" is a proposition (which is, in fact, false, since it fails to be true for the number $x = 0$).

This is the idea behind *propositional functions* or *predicates*. As stated above a *predicate* is a property or attribute assigned to elements of a particular set, called the *universe of discourse*. For example, the predicate "$x + 2 = 2x$", where the universe for the variable x is the set of all real numbers, is a property that some, but not all, real numbers possess.

In general, the set of all x in the universe of discourse having the attibute $P(x)$ is called the **truth set of** $P(x)$. That is, the truth set of $P(x)$ is

$$\{x \in U | P(x)\}.$$

3.2. Example of a Propositional Function.

EXAMPLE 3.2.1. *The propositional function $P(x)$ is given by "$x > 0$" and the universe of discourse for x is the set of integers. To create a proposition from P, we may assign a value for x. For example,*

- *setting $x = -3$, we get $P(-3)$: "$-3 > 0$", which is false.*
- *setting $x = 2$, we get $P(2)$: "$2 > 0$", which is true.*

Discussion

In this example we created propositions by choosing particular values for x.

Here are two more examples:

EXAMPLE 3.2.2. *Suppose $P(x)$ is the sentence "x has fur" and the universe of discourse for x is the set of all animals. In this example $P(x)$ is a true statement if x is a cat. It is false, though, is x is an alligator.*

EXAMPLE 3.2.3. *Suppose $Q(y)$ is the predicate "y holds a world record," and the universe of discourse for y is the set of all competitive swimmers. Notice that the universe of discourse must be defined for predicates. This would be a different predicate if the universe of discourse is changed to the set of all competitive runners.*

Moral: Be very careful in your homework to specify the universe of discourse precisely!

3.3. Quantifiers. A **quantifier** turns a propositional function into a proposition without assigning specific values for the variable. There are primarily two quantifiers, the

universal quantifier

and the

existential quantifier.

DEFINITION 3.3.1. *The* **universal quantification** *of $P(x)$ is the proposition*

"$P(x)$ is true for all values x in the universe of discourse."

Notation: "For all x $P(x)$" or "For every x $P(x)$" is written

$$\forall x P(x).$$

DEFINITION 3.3.2. *The* **existential quantification** *of $P(x)$ is the proposition*

"There exists an element x in the universe of discourse such that $P(x)$ is true."

Notation: "There exists x such that $P(x)$" or "There is at least one x such that $P(x)$" is written

$$\exists x P(x).$$

Discussion

As an alternative to assigning particular values to the variable in a propositional function, we can turn it into a proposition by *quantifying* its variable. Here we see the two primary ways in which this can be done, the universal quantifier and the existential quantifier.

In each instance we have created a proposition from a propositional function by *binding its variable.*

3.4. Example 3.4.1.

EXAMPLE 3.4.1. *Suppose $P(x)$ is the predicate $x + 2 = 2x$, and the universe of discourse for x is the set $\{1, 2, 3\}$. Then...*

- $\forall x P(x)$ *is the proposition "For every x in $\{1, 2, 3\}$ $x + 2 = 2x$." This proposition is false.*
- $\exists x P(x)$ *is the proposition "There exists x in $\{1, 2, 3\}$ such that $x + 2 = 2x$." This proposition is true.*

3.5. Converting from English.

EXAMPLE 3.5.1. *Assume*

$F(x)$*: x is a fox.*

$S(x)$*: x is sly.*

$T(x)$*: x is trustworthy.*

and the universe of discourse for all three functions is the set of all animals.

- *Everything is a fox:* $\forall x F(x)$
- *All foxes are sly:* $\forall x[F(x) \to S(x)]$
- *If any fox is sly, then it is not trustworthy:*
  $\forall x[(F(x) \wedge S(x) \to \neg T(x)] \Leftrightarrow \neg\exists x[F(x) \wedge S(x) \wedge T(x)]$

Discussion

Notice that in this example the last proposition may be written symbolically in the two ways given. Think about the how you could show they are the same using the logical equivalences in Module 2.2.

3.6. Additional Definitions.

- An assertion involving predicates is **valid** if it is true for every element in the universe of discourse.
- An assertion involving predicates is **satisfiable** if there is a universe and an interpretation for which the assertion is true. Otherwise it is **unsatisfiable**.
- The **scope** of a quantifier is the part of an assertion in which the variable is bound by the quantifier.

Discussion

You would not be asked to state the definitions of the terminology given, but you would be expected to know what is meant if you are asked a question like "Which of the following assertions are satisfiable?"

3.7. Examples.

Example 3.7.1.

If the universe of discourse is $U = \{1, 2, 3\}$*, then*

(1) $\forall x P(x) \Leftrightarrow P(1) \wedge P(2) \wedge P(3)$
(2) $\exists x P(x) \Leftrightarrow P(1) \vee P(2) \vee P(3)$

Suppose the universe of discourse U *is the set of real numbers.*

(1) If $P(x)$ *is the predicate* $x^2 > 0$*, then* $\forall x P(x)$ *is false, since* $P(0)$ *is false.*
(2) If $P(x)$ *is the predicate* $x^2 - 3x - 4 = 0$*, then* $\exists x P(x)$ *is true, since* $P(-1)$ *is true.*
(3) If $P(x)$ *is the predicate* $x^2 + x + 1 = 0$*, then* $\exists x P(x)$ *is false, since there are no real solutions to the equation* $x^2 + x + 1 = 0$*.*
(4) If $P(x)$ *is the predicate "If* $x \neq 0$*, then* $x^2 \geq 1$*', then* $\forall x P(x)$ *is false, since* $P(0.5)$ *is false.*

Exercise 3.7.1. *In each of the cases above give the truth value for the statement if each of the* $\forall$ *and* $\exists$ *quantifiers are reversed.*

3.8. Multiple Quantifiers.

Multiple quantifiers are read from left to right.

Example 3.8.1. *Suppose* $P(x, y)$ *is "*$xy = 1$*", the universe of discourse for* x *is the set of positive integers, and the universe of discourse for* y *is the set of real numbers.*

(1) $\forall x \forall y P(x, y)$ *may be read "For every positive integer* x *and for every real number* y*,* $xy = 1$*. This proposition is false.*

(2) $\forall x \exists y P(x, y)$ *may be read "For every positive integer* x *there is a real number* y *such that* $xy = 1$*. This proposition is true.*
(3) $\exists y \forall x P(x, y)$ *may be read "There exists a real number* y *such that, for every positive integer* x*,* $xy = 1$*. This proposition is false.*

Discussion

Study the syntax used in these examples. It takes a little practice to make it come out right.

3.9. Ordering Quantifiers. The order of quantifiers is important; they may not commute.

For example,

(1) $\forall x \forall y P(x, y) \Leftrightarrow \forall y \forall x P(x, y)$, and
(2) $\exists x \exists y P(x, y) \Leftrightarrow \exists y \exists x P(x, y)$,
but
(3) $\forall x \exists y P(x, y) \not\Leftrightarrow \exists y \forall x P(x, y)$.

Discussion

The lesson here is that you have to pay careful attention to the order of the quantifiers. The only cases in which commutativity holds are the cases in which both quantifiers are the same. In the one case in which equivalence does not hold,

$$\forall x \exists y P(x, y) \not\Leftrightarrow \exists y \forall x P(x, y),$$

there is an implication in one direction. Notice that if $\exists y \forall x P(x, y)$ is true, then there is an element c in the universe of discourse for y such that $P(x, c)$ is true for all x in the universe of discourse for x. Thus, for all x there exists a y, namely c, such that $P(x, y)$. That is, $\forall x \exists y P(x, y)$. Thus,

$$\exists y \forall x P(x, y) \Rightarrow \forall x \exists y P(x, y).$$

Notice predicates use function notation and recall that the variable in function notation is really a place holder. The statement $\forall x \exists y P(x, y)$ means the same as $\forall s \exists t P(s, t)$. Now if this seems clear, go a step further and notice this will also mean the same as $\forall y \exists x P(y, x)$. When the domain of discourse for a variable is defined it is in fact defining the domain for the place that variable is holding at that time.

Here are some additional examples:

EXAMPLE 3.9.1. *$P(x, y)$ is "x is a citizen of y." $Q(x, y)$ is "x lives in y." The universe of discourse of x is the set of all people and the universe of discourse for y is the set of US states.*

(1) All people who live in Florida are citizens of Florida.

$$\forall x(Q(x, Florida) \rightarrow P(x, Florida))$$

(2) Every state has a citizen that does not live in that state.

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

EXAMPLE 3.9.2. *Suppose $R(x, y)$ is the predicate "x understands y," the universe of discourse for x is the set of students in your discrete class, and the universe of discourse for y is the set of examples in these lecture notes. Pay attention to the differences in the following propositions.*

(1) $\exists x \forall y R(x, y)$ is the proposition "There exists a student in this class who understands every example in these lecture notes."
(2) $\forall y \exists x R(x, y)$ is the proposition "For every example in these lecture notes there is a student in the class who understands that example."
(3) $\forall x \exists y R(x, y)$ is the proposition "Every student in this class understands at least one example in these notes."
(4) $\exists y \forall x R(x, y)$ is the proposition "There is an example in these notes that every student in this class understands."

EXERCISE 3.9.1. *Each of the propositions in Example 3.9.2 has a slightly different meaning. To illustrate this, set up the following diagrams: Write the five letters A, B, C, D, E on one side of a page, and put the numbers 1 through 6 on the other side. The letters represent students in the class and the numbers represent examples. For each of the propositions above draw the minimal number of lines connecting people to examples so as to construct a diagram representing a scenario in which the given proposition is true.*

Notice that for any chosen pair of propositions above you can draw diagrams that would represent situations where the two propositions have opposite truth values.

EXERCISE 3.9.2. *Give a scenario where parts 1 and 2 in Example 3.9.2 have opposite truth values.*

3.10. Unique Existential.

DEFINITION 3.10.1. *The* **unique existential quantification** *of $P(x)$ is the proposition "There exists a unique element x in the universe of discourse such that $P(x)$ is true."*

Notation: *"There exists unique x such that $P(x)$" or "There is exactly one x $P(x)$" is written*

$$\exists! x P(x).$$

Discussion

Continuing with Example 3.9.2, the proposition $\forall x \exists! y R(x, y)$ is the proposition "Every student in this class understands exactly one example in these notes (but not necessarily the same example for all students)."

EXERCISE 3.10.1. *Repeat Exercise 3.9.1 for the four propositions $\forall x \exists! y R(x, y)$, $\exists! y \forall x R(x, y)$, $\exists! x \forall y R(x, y)$, and $\forall y \exists! x R(x, y)$.*

Remember: A predicate is *not* a proposition until <u>all</u> variables have been bound either by quantification or by assignment of a value!

3.11. De Morgan's Laws for Quantifiers.

- $\neg \forall x P(x) \Leftrightarrow \exists x \neg P(x)$
- $\neg \exists x P(x) \Leftrightarrow \forall x \neg P(x)$

Discussion

The negation of a quantified statement are obtained from the De Morgan's Laws in Module 2.1.

So the negation of the proposition "Every fish in the sea has gills," is the proposition "there is at least one fish in the sea that does not have gills."

If there is more than one quantifier, then the negation operator should be passed from left to right across one quantifier at a time, using the appropriate De Morgan's Law at each step. Continuing further with Example 3.9.2, suppose we wish to negate the proposition "Every student in this class understands at least one example in these notes." Apply De Morgan's Laws to negate the symbolic form of the proposition:

$$\begin{aligned} \neg(\forall x \exists y R(x, y)) &\Leftrightarrow \exists x(\neg \exists y R(x, y)) \\ &\Leftrightarrow \exists x \forall y \neg R(x, y) \end{aligned}$$

The first proposition could be read "It is not the case that every student in this class understands at least one example in these notes." The goal, however, is to find an expression for the negation in which the verb in each predicate in the scope of the

quantifiers is negated, and this is the intent in any exercise, quiz, or test problem that asks you to "negate the proposition ... ." Thus, a correct response to the instruction to negate the proposition "Every student in this class understands at least one example in these notes" is the proposition "There is at least one student in this class that does not understand any of the examples in these notes."

EXERCISE 3.11.1. *Negate the rest of the statements in Example 3.9.2.*

It is easy to see why each of these rules of negation is just another form of De Morgan's Law, if you assume that the universe of discourse is finite: $U = \{x_1, x_2, ..., x_n\}$. For example,

$$\forall x P(x) \Leftrightarrow P(x_1) \wedge P(x_2) \wedge \cdots \wedge P(x_n)$$

so that

$$\begin{aligned} \neg\forall x P(x) &\Leftrightarrow \neg[P(x_1) \wedge P(x_2) \wedge \cdots \wedge P(x_n)] \\ &\Leftrightarrow [\neg P(x_1) \vee \neg P(x_1) \vee \cdots \vee \neg P(x_1)] \\ &\Leftrightarrow \exists x \neg P(x) \end{aligned}$$

If U is an arbitrary universe of discouse, we must argue a little differently: Suppose $\neg\forall x P(x)$ is true. Then $\forall x P(x)$ is false. This is true if and only if there is some c in U such that $P(c)$ is false. This is true if and only if there is some c in U such that $\neg P(c)$ is true. But this is true if and only if $\exists x \neg P(x)$.

The argument for the other equivalence is similar.

Here is a formidable example from the calculus. Suppose a and L are fixed real numbers, and f is a real-valued function of the real variable x. Recall the rigorous definition of what it means to say "the limit of $f(x)$ as x tends to a is L":

$$\lim_{x \to a} f(x) = L \Leftrightarrow$$

for every $\epsilon > 0$ there exists $\delta > 0$ such that, for every x,
if $0 < |x - a| < \delta$, then $|f(x) - L| < \epsilon$.

Here, the universe of discourse for the variables ϵ, δ, and x is understood to be the set of all real numbers.

What does it mean to say that $\lim_{x \to a} f(x) \neq L$? In order to figure this out, it is useful to convert this proposition into a symbolic proposition. So, let $P(\epsilon, \delta, x)$ be the predicate "$0 < |x - a| < \delta$" and let $Q(\epsilon, \delta, x)$ be the predicate "$|f(x) - L| < \epsilon$."

(It is perfectly OK to list a variable in the argument of a predicate even though it doesn't actually appear!) We can simplify the proposition somewhat by restricting the universe of discourse for the variables ϵ and δ to be the set of *positive* real numbers. The definition then becomes

$$\forall\epsilon\exists\delta\forall x[P(\epsilon,\delta,x)\to Q(\epsilon,\delta,x)].$$

Use De Morgan's Law to negate:

$$\neg[\forall\epsilon\exists\delta\forall x[P(\epsilon,\delta,x)\to Q(\epsilon,\delta,x)]]\Leftrightarrow\exists\epsilon\forall\delta\exists x[P(\epsilon,\delta,x)\wedge\neg Q(\epsilon,\delta,x)],$$

and convert back into words:

There exists $\epsilon>0$ such that, for every $\delta>0$ there exists x such that,
$0<|x-a|<\delta$ and $|f(x)-L|\geq\epsilon$.

3.12. Distributing Quantifiers over Operators.

(1) $\forall x[P(x)\wedge Q(x)]\Leftrightarrow\forall xP(x)\wedge\forall xQ(x)$, but
(2) $\forall x[P(x)\vee Q(x)]\not\Leftrightarrow\forall xP(x)\vee\forall xQ(x)$.
(3) $\exists x[P(x)\vee Q(x)]\Leftrightarrow\exists xP(x)\vee\exists xQ(x)$, but
(4) $\exists x[P(x)\wedge Q(x)]\not\Leftrightarrow\exists xP(x)\wedge\exists xQ(x)$.

Discussion

Here we see that in only half of the four basic cases does a quantifier distribute over an operator, in the sense that doing so produces an equivalent proposition.

Exercise 3.12.1. *In each of the two cases in which the statements are not equivalent, there is an implication in one direction. Which direction? In order to help you analyze these two cases, consider the predicates* $P(x)=[x\geq 0]$ *and* $Q(x)=[x<0]$*, where the universe of discourse is the set of all real numbers.*

CHAPTER 3

Methods of Proofs

1. Logical Arguments and Formal Proofs

1.1. Basic Terminology.

- An **axiom** is a statement that is given to be true.
- A **rule of inference** is a logical rule that is used to deduce one statement from others.
- A **theorem** is a proposition that can be proved using definitions, axioms, other theorems, and rules of inference.

Discussion

In most of the mathematics classes that are prerequisites to this course, such as calculus, the main emphasis is on using facts and theorems to solve problems. Theorems were often stated, and you were probably shown a few proofs. But it is very possible you have never been asked to prove a theorem on your own. In this module we introduce the basic structures involved in a mathematical proof. One of our main objectives from here on out is to have you develop skills in recognizing a valid argument and in constructing valid mathematical proofs.

When you are first shown a proof that seemed rather complex you may think to yourself "How on earth did someone figure out how to go about it that way?" As we will see in this chapter and the next, a proof must follow certain *rules of inference*, and there are certain strategies and methods of proof that are best to use for proving certain types of assertions. It is impossible, however, to give an exhaustive list of strategies that will cover all possible situations, and this is what makes mathematics so interesting. Indeed, there are conjectures that mathematicians have spent much of their professional lives trying to prove (or disprove) with little or no success.

1.2. More Terminology.

- A **lemma** is a "pre-theorem" or a result which is needed to prove a theorem.
- A **corollary** is a "post-theorem" or a result which follows from a theorem (or lemma or another corollary).

Discussion

The terms "lemma" and "corollary" are just names given to theorems that play particular roles in a theory. Most people tend to think of a theorem as the main result, a lemma a smaller result needed to get to the main result, and a corollary as a theorem which follows relatively easily from the main theorem, perhaps as a special case. For example, suppose we have proved the Theorem: "If the product of two integers m and n is even, then either m is even or n is even." Then we have the Corollary: "If n is an integer and n^2 is even, then n is even." Notice that the Corollary follows from the Theorem by applying the Theorem to the special case in which $m = n$. There are no firm rules for the use of this terminology; in practice, what one person may call a lemma another may call a theorem.

Any mathematical theory *must* begin with a collection of undefined terms and axioms that give the properties the undefined terms are assumed to satisfy. This may seem rather arbitrary and capricious, but any mathematical theory you will likely encounter in a serious setting is based on concrete ideas that have been developed and refined to fit into this setting. To justify this necessity, see what happens if you try to define every term. You define a in terms of b, and then you define b in terms of c, etc. If a, b, c, ... are all different terms, you are lead to an infinite chain of definitions; otherwise, one of them is repeated and you are left with a circular chain of definitions. Neither of these alternatives is logically acceptable. A similar criticism can be made for any attempt to prove every assertion. Here are a few important examples of mathematical systems and their basic ingredients.

In plane geometry one takes "point" and "line" as undefined terms and assumes the five axioms Euclidean geometry.

In set theory, the concept of a "set" and the relation "is an element of," or "$\in$", are left undefined. There are five basic axioms of set theory, the so-called Zermelo-Fraenkel axioms, which we will use informally in this course, rather than giving them a rigorous exposition. In particular, these axioms justify the "set builder" notation we discussed in *Module 1.1: Sets* and the existence of the "power set" of a set, which we shall discuss later in *Module 4.1: Set Operations*.

The real number system begins with the four Peano Postulates for the positive integers, taking the elements, "numbers," in the set of positive integers as undefined, as well as the relation "is a successor of" between positive integers. (To say "x is a successor of y" turns out to mean that $x = y + 1$.) The fourth Peano Postulate is the Principle of Mathematical Induction, which we shall use extensively in the next module. From these modest beginnings, and with a little help from set theory, one can construct the entire set of real numbers, including its order and completeness properties. As with our treatment of set theory, we shall, with the one exception mentioned above, use these axioms informally, assuming the familiar model of the real

number line together with its important subsets, the natural numbers, the integers, and the rational numbers.

Once we have the undefined terms and axioms for a mathematical system, we can begin defining new terms and proving theorems (or lemmas, or corollaries) within the system.

1.3. Formal Proofs. To prove an argument is valid:

- Assume the hypotheses are true.
- Use the rules of inference and logical equivalences to show that the conclusion is true.

Discussion

What is a proof?

A proof is a demonstration, or argument, that shows beyond a shadow of a doubt that a given assertion is a logical consequence of our axioms and definitions. Thus, in any problem in which you are asked to provide a proof, your solution will not simply be a short answer that you circle. There are certain rules that must be followed (which we will get to shortly), and certain basic knowledge must be assumed. For example, one may assume the axioms and any previously stated theorems (unless the instructions state otherwise). A large number of proofs simply involve showing that a certain definition is satisfied.

In almost every case, the assertions we will be proving are of the form "if p, then q", where p and q are (possibly compound) propositions. The proposition p is the *hypothesis* and q is the *conclusion*. It is almost always useful to translate a statement that must be proved into an "if ..., then ..." statement if it is not already in that form. To begin a proof we assume the hypotheses. For example, consider the argument

Every dog will have his day.
Fido is a dog.
Therefore, Fido will have his day.

The hypotheses of this argument are "Every dog will have his day" and "Fido is a dog." The conclusion is "Fido will have his day."

1.4. Rules of Inference.

Modus Ponens or the Law of Detachment	p $p \to q$ $\therefore q$
Disjunction Introduction	p $\therefore p \vee q$
Conjunction Elimination	$p \wedge q$ $\therefore p$
Modus Tollens	$\neg q$ $p \to q$ $\therefore \neg p$
Hypothetical Syllogism	$p \to q$ $q \to r$ $\therefore p \to r$
Disjunctive Syllogism	$p \vee q$ $\neg p$ $\therefore q$
Conjunction Introduction	p q $\therefore p \wedge q$
Constructive Dilemma	$(p \to q) \wedge (r \to s)$ $p \vee r$ $\therefore q \vee s$

Discussion

An argument is **valid** if it is uses only the given hypotheses together with the axioms, definitions, previously proven assertions, and the *rules of inference*, which are listed above. In those rules in which there is more than one hypothesis, the order

of the hypotheses is not important. For example, *modus tollens* could be just as well stated:

$$\begin{array}{l} p \to q \\ \neg q \\ \hline \therefore \neg p \end{array}$$

The notation used in these slides is commonly used in logic to express an argument symbolically. The proposition(s) before the horizontal line are the hypotheses and the proposition below the line is the conclusion. The symbol $\therefore$ is a common shorthand for "therefore."

Each of the rules of inference is a tautology expressed in a different form. For example, the rule of *modus ponens*, when stated as a propositional form, is the tautology

$$[p \wedge (p \to q)] \to q.$$

(This can be verified using a truth table.)

REMARK 1.4.1. *An argument of the form*

$$\begin{array}{l} h_1 \\ h_2 \\ \vdots \\ h_n \\ \hline \therefore c \end{array}$$

is valid if and only if the proposition $[h_1 \wedge h_2 \wedge \cdots \wedge h_n] \to c$ *is a tautology.*

1.5. Example 1.5.1.

EXAMPLE 1.5.1. *The following is a valid logical argument:*

1. *If the dog eats the cat food or scratches at the door, then the parrot will bark.*
2. *If the cat eats the parrot, then the parrot will not bark.*
3. *If the cat does not eat the parrot, then it will eat the cat food.*
4. *The cat did not eat the cat food.*
5. *Therefore, the dog does not eat the cat food either.*

Discussion

Here is how the hypotheses give us the conclusion:

1. Assign propositional variables to the component propositions in the argument:
   - d – the dog eats the cat food
   - s – the dog scratches at the door
   - p – the parrot will bark
   - c – the cat eats the parrot
   - e – the cat eats the cat food
2. Represent the formal argument using the variables:
   $(d \vee s) \rightarrow p$
   $c \rightarrow \neg p$
   $\neg c \rightarrow e$
   $\neg e$
   $\therefore \neg d$
3. Use the hypotheses, the rules of inference, and any logical equivalences to prove that the argument is valid:

	Assertion	Reason
1.	$\neg c \rightarrow e$	hypothesis 3
2.	$\neg e$	hypothesis 4
3.	c	steps 1 and 2 and *modus tollens*
4.	$c \rightarrow \neg p$	hypothesis 2
5.	$\neg p$	steps 3 and 4 and *modus ponens*
6.	$(d \vee s) \rightarrow p$	hypothesis 1
7.	$\neg(d \vee s)$	steps 5 and 6 and *modus tollens*
8.	$\neg d \wedge \neg s$	step 7 and De Morgan's law
9.	$\neg d$	step 8 and conjunction elimination

We could also determine if the argument is valid by checking if the proposition $[((d \vee s) \rightarrow p) \wedge (c \rightarrow \neg p) \wedge (\neg c \rightarrow e) \wedge (\neg e)] \rightarrow (\neg d)$ is a tautology. In practice, though, it is more useful to recognize if the rules of inference have been applied appropriately or if one of the common fallacies have been used to determine if an argument is valid or not. It will serve you better later on to understand the two column proof of a valid argument and to recognize how the rules of inference are applied.

EXERCISE 1.5.1. *Give a formal proof that the following argument is valid. Provide reasons.*

$a \vee b$

$\neg c \rightarrow \neg b$

$\neg a$

$\therefore c$

1.6. Rules of Inference for Quantifiers.

Universal Instantiation	$\dfrac{\forall x P(x)}{\therefore P(c)}$
Universal Generalization	$\dfrac{P(x)}{\therefore \forall x P(x)}$
Existantial Generalization	$\dfrac{P(c)}{\therefore \exists x P(x)}$
Existential Instantiation	$\dfrac{\exists x P(x)}{\therefore P(c)}$

Discussion

Here is the list of additional rules of inference related to quantifiers. The symbol c represents some particular element from the universe of discourse for the variable x.

In Universal Instantiation, c may be any element from the universe of discourse for x. For example, suppose the universe of discourse is the set of real numbers, and

$P(x)$ is the predicate $x^2 \geq 0$. Since $x^2 \geq 0$ for all x, we may conclude $(a-b)^2 \geq 0$ for arbitrary real numbers a and b. Here, $c = a - b$. We may also conclude $(-\pi)^2 \geq 0$.

In Existential Instantiation, c must be chosen so that $P(c)$ is true. For example, suppose the universe of discourse is the set of integers, and let $P(x)$ be the predicate, "x is a divisor of 17283 and $1 < x < 17283$." Then $\exists x P(x)$ is a true statement (e.g., $P(3)$). We may then assume c is a divisor of 17283 and $1 < c < 17283$ for some integer c.

Sometimes we may know a statement of the form $\exists x P(x)$ is true, but we may not know exactly for what x in the domain of discourse gives us that this is true. In a proof when we know the truth of $\exists x P(x)$ we can define a variable, say c, to stand for a fixed element of the domain where $P(c)$ is true. This is what Existential Instantiation gives you. An example in which we have this situation is by using the Intermediate Value Theorem from algebra.

Let $f : \mathbb{R} \to \mathbb{R}$ be the polynomial $f(x) = -3x^4 + x^3 + 2x + 1$. Since $f(1) = 1$, $f(2) = -35$, and f is continuous, there must be a solution to $f(x) = 0$ in the interval $[1, 2]$. It may not possible to find this solution algebraically, though, and may only be possible to numerically approximate the root. However, if we needed to use the solution for some purpose we could simply say let $c \in [1, 2]$ be such that $f(c) = 0$ and this fixes c as the solution we know exists in $[1, 2]$.

Universal Generalization is a subtle and very useful rule and the meaning may not be clear to you yet. The variable x stands for any arbitrary element of the universe of discourse. You only assume x is a member of the universe and do not place any further restrictions on x. If you can show $P(x)$ is true, then it will also be true for any other object satisfying the same properties you've claimed for x. In other words, $P(x)$ is true for all the members of the universe, $\forall x P(x)$. You will see a standard approach in proving statements about sets is to use Universal Generalization.

1.7. Example 1.7.1.

EXAMPLE 1.7.1. *Here is a simple argument using quantifiers.*

1. *Every dog will have his day.*
2. *Fido is a dog.*
3. *Therefore, Fido will have his day.*

Discussion

To verify this is a valid argument we use the same technique as before.

Define the predicates

- $M(x)$: x is a dog
- $D(x)$: x has his day

and let F represent Fido, a member of the universe of discourse.

The argument becomes

$$\begin{array}{l} \forall x[M(x) \to D(x)] \\ M(F) \\ \hline \therefore D(F) \end{array}$$

The proof is

1.	$\forall x[M(x) \to D(x)]$	hypothesis 1
2.	$M(F) \to D(F)$	step 1 and universal instantiation
3.	$M(F)$	hypothesis 2
4.	$D(F)$	steps 2 and 3 and *modus ponens*

1.8. Fallacies. The following are **not valid** argument forms.

Affirming the Consequent	$\begin{array}{l} p \to q \\ q \\ \hline \therefore p \end{array}$
Denying the Antecedent	$\begin{array}{l} p \to q \\ \neg p \\ \hline \therefore \neg q \end{array}$
Begging the Question or Circular Reasoning	Use the truth of the consequent in the argument

Discussion

There are several common mistakes made in trying to create a proof. Here we list three of the most common *fallacies* or errors in logic. Since they are not valid arguments, obviously you should *not* use them in a proof. Just as important, you should be able to recognize one of them if you were to encounter it in someone else's argument.

The fallacy of affirming the consequent occurs when the converse of a premise is used to prove a statement. For example, here is an "argument" using the fallacy of affirming the consequent.

EXAMPLE 1.8.1. *If Jack lands the new account, then he will get a raise. Jack got a raise. Therefore, he landed the new account.*

Note that $[(p \to q) \wedge q] \to p]$ is *not* a tautology, so this is not a valid argument. The "if ..., then ..." statement is not equivalent to its converse. In the above example, just because Jack got a raise, you can't conclude from the hypothesis that he landed the new account.

The fallacy of denying the antecedent comes from the fact that an implication is not equivalent to its inverse. Here is an example of incorrect reasoning using the fallacy of denying the antecedent:

EXAMPLE 1.8.2. *If the cat is purring, then he ate the canary. The cat is not purring. Therefore, the cat didn't eat the canary.*

In this example, the hypothesis does not allow you to conclude anything if the cat is not purring, only if he *is* purring. The fallacy results from the fact that the propositional form $[(p \to q) \wedge \neg p] \to \neg q$ is not a tautology.

Begging the question, or circular reasoning, occurs when the conclusion itself is used in the proof. Here is an example of this type of fallacy:

EXAMPLE 1.8.3. *Prove: If xy is divisible by 5, then x is divisible by 5 or y is divisible by 5.*

> *Incorrect Proof: If xy is divisible by 5, then $xy = 5k$ for some k. Then $x = 5\ell$ for some ℓ, or $y = 5\ell$ for some ℓ. Hence, x is divisible by 5 or y is divisible by 5.*

This argument breaks down once we assert, without justification, that either $x = 5\ell$ for some ℓ, or $y = 5\ell$ for some ℓ. This, of course, is what we are trying to prove, and it doesn't follow directly from $xy = 5k$.

EXERCISE 1.8.1. *Give a careful proof of the statement: For all integers m and n, if m is odd and n is even, then $m + n$ is odd.*

EXAMPLE 1.8.4. *Prove: for all real x, $x < x + 1$.*

PROOF. First we fix an arbitrary real number: Let $x \in \mathbb{R}$. We wish to show $x < x+1$. This inequality is equivalent to $0 < (x+1)-x$. But by the commutative and associative properties of real numbers this inequality is equivalent to $0 < 1 + (x - x)$ or equivalently, $0 < 1$. We know the last inequality is true and so the equivalent expression $x < x + 1$ is also true.

□

In the previous example we took the expression we wished to show was true and rewrote it several times until we reached an expression that we knew to be true. This is a useful tool but one must be extremely cautious in using this technique. Notice we did not actually assume what we wished to prove. Instead, we used equivalences to rephrase what we needed to show.

EXAMPLE 1.8.5. *Now, here is an incorrect "proof" of the same statement in Exercise 1.8.4. This proof would be marked wrong.*

INCORRECT "PROOF" OF EXERCISE 1.8.4. Let $x \in \mathbb{R}$. Then

$$\begin{aligned} & x < x + 1 \\ \Rightarrow\ & 0 < (x + 1) - x \\ \Rightarrow\ & 0 < (x - x) + 1 \quad \text{by the associative and commutative} \\ & \qquad\qquad\ \ \text{properties of real numbers} \\ \Rightarrow\ & 0 < 1 \end{aligned}$$

We know $0 < 1$. □

EXERCISE 1.8.2. *Consider the following hypotheses: If the car does not start today, then I will not go to class. If I go to class today then I will take the quiz. If I do not take the quiz today then I will ask the teacher for an extra credit assignment. I asked the teacher for an extra credit assignment.*

Determine whether each of the following are valid or invalid conclusions of the above hypotheses. Why or why not?

(1) *I did not go to class today.*
(2) *Remove the hypothesis "I asked the teacher for an extra credit assignment" from the above assumptions. Can one now conclude "If the car does not start today, then I will ask the teacher for an extra credit assignment" for the remaining assumptions?*

EXERCISE 1.8.3. *Find the error in the proof of the following statement.*

Suppose x is a positive real number. Claim: the sum of x and its reciprocal is greater than or equal to 2.

INCORRECT "PROOF". Multiplying by x we get $x^2 + 1 \geq 2x$. By algebra, $x^2 - 2x + 1 \geq 0$. Thus $(x-1)^2 \geq 0$. Any real number squared is greater than or equal to 0, so $\frac{x^2+1}{x} \geq 2$ is true. □

EXERCISE 1.8.4. *Find the fallacy associated with the following:*

Problem: *Solve for x given the equation $\sqrt{x} + \sqrt{x-a} = 2$, where a is a real number.*

Incorrect "Solution": *The given equation also implies that*

$$\frac{1}{\sqrt{x} + \sqrt{x-a}} = \frac{1}{2},$$

so

$$\sqrt{x} - \sqrt{x-a} = \frac{a}{\sqrt{x} + \sqrt{x-a}} = \frac{a}{2}.$$

Adding the original equation with this one gives

$$2\sqrt{x} = 2 + (a/2)$$

and thus

$$x = (1 + \frac{a}{4})^2.$$

Notice, however, if $a = 8$ then $x = 9$ according to the solution, but this does not satisfy the original equation.

2. Methods of Proof

2.1. Types of Proofs. Suppose we wish to prove an implication $p \to q$. Here are some strategies we have available to try.

- **Trivial Proof:** If we know q is true then $p \to q$ is true regardless of the truth value of p.
- **Vacuous Proof:** If p is a conjunction of other hypotheses and we know one or more of these hypotheses is false, then p is false and so $p \to q$ is vacuously true regardless of the truth value of q.
- **Direct Proof:** Assume p, and then use the rules of inference, axioms, definitions, and logical equivalences to prove q.
- **Indirect Proof or Proof by Contradiction:** Assume p and $\neg q$ and derive a contradiction $r \wedge \neg r$.
- **Proof by Contrapositive:** (Special case of Proof by Contradiction.) Give a direct proof of $\neg q \to \neg p$. (Can be thought of as a proof by contradiction in which you assume p and $\neg q$ and arrive at the contradiction $p \wedge \neg p$.)
- **Proof by Cases:** If the hypothesis p can be separated into cases $p_1 \vee p_2 \vee \cdots \vee p_k$, prove each of the propositions,
  $p_1 \to q$, $p_2 \to q$, $\ldots$, $p_k \to q$, separately.
  (You may use different methods of proof for different cases.)

Discussion

We are now getting to the heart of this course: methods you can use to write proofs. Let's investigate the strategies given above is some detail.

2.2. Trivial Proof/Vacuous Proof.

EXAMPLE 2.2.1. *Prove the statement: If there are 100 students enrolled in this course this semester, then* $6^2 = 36$.

PROOF. The assertion is *trivially* true, since the conclusion is true, independent of the hypothesis (which, may of may not be true depending on the enrollment).

□

EXAMPLE 2.2.2. *Prove the statement. If 6 is a prime number, then* $6^2 = 30$.

PROOF. The hypothesis is false, therefore the statement is *vacuously* true (even though the conclusion is also false).

□

Discussion

The first two methods of proof, the "Trivial Proof" and the "Vacuous Proof" are certainly the easiest when they work. Notice that the form of the "Trivial Proof", $q \to (p \to q)$, is, in fact, a tautology. This follows from disjunction introduction, since $p \to q$ is equivalent to $\neg p \vee q$. Likewise, the "Vacuous Proof" is based on the tautology $\neg p \to (p \to q)$.

EXERCISE 2.2.1. *Fill in the reasons for the following proof of the tautology* $\neg p \to (p \to q)$.

$$\begin{aligned}[\neg p \to (p \to q)] &\Leftrightarrow [p \vee (\neg p \vee q)] \\ &\Leftrightarrow [(p \vee \neg p) \vee q] \\ &\Leftrightarrow T \vee q \\ &\Leftrightarrow T\end{aligned}$$

EXERCISE 2.2.2. *Let* $A = \{1, 2, 3\}$ *and* $R = \{(2, 3), (2, 1)\} (\subseteq A \times A)$. *Prove: if* $a, b, c \in A$ *are such that* $(a, b) \in R$ *and* $(b, c) \in R$ *then* $(a, c) \in R$.

Since it is a rare occasion when we are able to get by with one of these two methods of proof, we turn to some we are more likely to need. In most of the following examples the underlying "theorem" may be a fact that is well known to you. The purpose in presenting them, however, is not to surprise you with new mathematical facts, but to get you thinking about the correct way to set up and carry out a mathematical argument, and you should read them carefully with this in mind.

2.3. Direct Proof.

EXAMPLE 2.3.1. *Prove the statement: For all integers* m *and* n, *if* m *and* n *are odd integers, then* $m + n$ *is an even integer.*

PROOF. Assume m and n are arbitrary odd integers. Then m and n can be written in the form

$$m = 2a + 1 \text{ and } n = 2b + 1,$$

where a and b are also integers. Then

$$\begin{aligned}m + n &= (2a + 1) + (2b + 1) && \text{(substitution)} \\ &= 2a + 2b + 2 && \text{(associative and commutative} \\ & && \quad \text{laws of addition)} \\ &= 2(a + b + 1) && \text{(distributive law)}\end{aligned}$$

Since $m+n$ is twice another integer, namely, $a+b+1$, $m+n$ is an even integer. □

Discussion

The first strategy you should try when attempting to prove any assertion is to give a direct proof. That is, assume the hypotheses that are given and try to argue directly that the conclusion follows. This is often the best approach when the hypotheses can be translated into algebraic expressions (equations or inequalities) that can be manipulated to give other algebraic expressions, which are useful in verifying the conclusion.

Example 2.3.1 shows a simple direct proof of a very familiar result. We are using the familiar definitions of what it means for an integer to be even or odd: An integer n is *even* if $n = 2k$ for some integer k; an integer n is *odd* if $n = 2k + 1$ for some integer k. Study the *form* of this proof. There are two hypotheses, "m is an odd integer," and "n is an odd integer"; and the conclusion is the statement "$m + n$ is an even integer." This "theorem" is a quantified statement ("for all integers m and n", or "for all odd integers m and n"). In the proof we assumed the hypotheses held for arbitrarily integers m and n, and then we wrote down equations that follow from the definition of what it means for these integers to be odd. Although this looks like a pretty obvious thing to do, at least when you see someone else do it, this step, in which you bring your knowledge to the problem, may seem like a big one to take, and you may find yourself stalling out at this point.

One possible reason this may happen is that you may be trying to do too much at once. The cure for this is to be patient: take small steps, using the appropriate definitions and previously proven facts, and see where they lead. When we wrote down $m = 2a + 1$ and $n = 2b + 1$, we did a number of fairly sophisticated things. First, we used our knowledge (definitions) of what it means for an integer to be odd. Second, in order for this information to be useful, we needed to translate this knowledge into a mathematical expression, or expressions in this case, that are subject to manipulation. And third, in setting up these expressions, we needed to use *appropriate* mathematical notation, so that we did not introduce any subtle or hidden relationships into the picture that are unwarranted by the hypotheses.

A common mistake of this type might arise as follows:

> "Well, m is an odd integer, so I can write $m = 2k + 1$, where k is an integer. Since n is also an odd integer, I can write $n = 2k + 1$, where k is an integer."

Do you see the mistake? By allowing the same letter k to represent what might be different integers, we have inadvertently added another assumption, namely, that $m =$

n! Of course, we didn't mean to do this, but, unfortunately, our intentions haven't been carried out, and so our proof breaks down at this point. In order to maintain the "arbitrariness" of m and n, we must allow, at the least, that they be different. We accomplish this by choosing different letters a and b in our representations of m and n as "twice an integer plus one." There is nothing sacred about a and b; we could have used k and ℓ, or x and y, or α and β, or any pair of symbols that have not been appropriated for some other use.

Upon closer scrutiny, this first step now starts to seem like a big one indeed! Especially if we may not be sure just where it will lead. The rest of the proof, however, proceeds fairly routinely. We add m and n and observe that the resulting expression has a factor of 2. We now only have to get past the *recognition problem*: observing that the resulting expression gives us what we were looking for. Since we have expressed $m + n$ as twice another integer, $m + n$ is, by definition, an even integer. By Universal Generalization we may now confidently declare "Q.E.D." (the abbreviation of *quod erat demonstrandum* or "which was to be demonstrated"). Often a box at the end of a proof or the abbrviation "Q.E.D." is used at the end of a proof to indicate it is finished.

EXERCISE 2.3.1. *Give a careful proof of the statement: For all integers m and n, if m is odd and n is even, then $m + n$ is odd.*

2.4. Proof by Contrapositive.

EXAMPLE 2.4.1. *Prove the statement: For all integers m and n, if the product of m and n is even, then m is even or n is even.*

We prove the contrapositive of the statement: If m and n are both odd integers, then mn is odd.

PROOF. Suppose that m and n are arbitrary odd integers. Then $m = 2a + 1$ and $n = 2b + 1$, where a and b are integers. Then

$$\begin{aligned} mn &= (2a + 1)(2b + 1) && \text{(substitution)} \\ &= 4ab + 2a + 2b + 1 && \text{(associative, commutative, and distributive laws)} \\ &= 2(2ab + a + b) + 1 && \text{(distributive law)} \end{aligned}$$

Since mn is twice an integer (namely, $2ab + a + b$) plus 1, mn is odd. □

Discussion

If a direct proof of an assertion appears problematic, the next most natural strategy to try is a proof of the contrapositive. In Example 2.4.1 we use this method to prove that if the product of two integers, m and n, is even, then m or n is even. This statement has the form $p \to (r \vee s)$. If you take our advice above, you will first try to give a direct proof of this statement: assume mn is even and try to prove m is even or n is even. Next, you would use the definition of "even" to write $mn = 2k$, where k is an integer. You would now like to conclude that m or n has the factor 2. This can, in fact, be proved directly, but it requires more knowledge of number theory than we have available at this point. Thus, we seem to have reached a dead-end with the direct approach, and we decide to try an indirect approach instead.

The contrapositive of $p \to (r \vee s)$ is $\neg(r \vee s) \to \neg p$, or, by De Morgan's Law,

$$(\neg r \wedge \neg s) \to \neg p.$$

This translates into the statement

"If m and n are odd, then mn is odd"

(where "not even" translates to "odd"). This is a good illustration of how the symbolic form of a proposition can be helpful in finding the correct statement we wish to prove. In this particular example, the necessity of De Morgan's Law may be more evident in the symbolic form than in the "English version."

Now we give a *direct* proof of the contrapositive: we assume m and n are arbitrary odd integers and deduce mn is odd. This proof is carried out in very much the same way as the direct proof in Example 2.3.1. The main difficulty we encounter with the problem of proving the original assertion is to realize that a direct proof should be abandoned in favor of some other strategy.

EXERCISE 2.4.1. *The following statement is a special case of the proposition proved in Example 2.4.1. Give a careful proof of this statement without assuming the result in Example 2.4.1.*

For every integer n, if n^2 is even, then n is even.

2.5. Proof by Contradiction.

EXAMPLE 2.5.1. *Prove the statement: If $5x + 25y = 1723$, then x or y is not an integer.*

PROOF. Assume $5x + 25y = 1723$, and assume that both x and y are integers. By the distributive law,

$$5(x + 5y) = 1723.$$

Since x and y are integers, this implies 1723 is divisible by 5. The integer 1723, however, is clearly not divisible by 5. This contradiction establishes the result. □

Discussion

If we have tried unsuccessfully to find a direct proof of a statement or its contrapositive, we might next try to give a proof by contradiction. In this method of proof we assume the hypotheses are true and the conclusion is false and try to arrive at a contradiction. The validity of proof by contradiction follows from the fact that $\neg(p \wedge \neg q)$ is equivalent to $p \to q$: if we can show that $p \wedge \neg q$ is false, then $\neg(p \wedge \neg q)$ is true, so that the equivalent proposition $p \to q$ is also true.

In Example 2.5.1 we are asked to prove that if $5x + 25y = 1723$, then x is not an integer or y is not an integer. This has the same propositional form as the example in Example 2.4.1:

$$p \to (r \vee s).$$

If we try to give a direct proof of this statement, then we are forced to "prove a negative," which can be difficult. If we try to prove the contrapositive, then knowing that x and y are integers doesn't seem to be helpful in trying to show directly that $5x + 25y \neq 1723$, since we are again trying to prove a negative.

One the other hand, if we assume p and $\neg(r \vee s)$, which is equivalent to $\neg r \wedge \neg s$, then we have two positive statements to work with: $5x+25y = 1723$, and x and y are integers. After a couple of observations we arrive at the contradiction $r \wedge \neg r$, where r is the statement "1723 is divisible by 5." This contradiction establishes the truth of the statement, and we are through.

EXERCISE 2.5.1. *Prove: For all real numbers x and y, if $35x + 14y = 253$, then x is not an integer or y is not an integer.*

Here is another example of a proposition that is best proved by contradiction.

EXAMPLE 2.5.2. *For all positive real numbers a, b, and c, if $ab = c$, then $a \leq \sqrt{c}$ or $b \leq \sqrt{c}$.*

PROOF. Suppose a, b, and c are positive real numbers such that $ab = c$, and suppose $a > \sqrt{c}$ and $b > \sqrt{c}$. (Notice the use of De Morgan's Law again. Also, recall that the symbol $\sqrt{c}$ represents the *positive* square root of c, not $\pm\sqrt{c}$.) By order properties of the real numbers,

$$b > \sqrt{c} \Leftrightarrow ab > a\sqrt{c}, \text{ since } a > 0,$$

and

$$a > \sqrt{c} \Leftrightarrow a\sqrt{c} > \sqrt{c} \cdot \sqrt{c} = c, \text{ since } \sqrt{c} > 0.$$

Thus, $ab > a\sqrt{c} > \sqrt{c} \cdot \sqrt{c} = c$ implies

$$ab > c.$$

But $ab = c$; hence, ab is not greater than c, a contradiction.

This proves our assuption $a > \sqrt{c}$ and $b > \sqrt{c}$ cannot be true when a, b, and c are positive real numbers such that $ab = c$. Therefore $a \leq \sqrt{c}$ or $b \leq \sqrt{c}$. □

EXERCISE 2.5.2. *Consider the statement: For all nonnegative real numbers a, b, and c, if $a^2 + b^2 = c^2$, then $a + b \geq c$.*

(a) Give a proof by contradiction.
(b) Give a direct proof. [Hint: The extra idea needed for a direct proof should emerge naturally from a proof by contradiction.]

Let's step back and compare direct proof, proof by contrapositive, and proof by contradiction.

EXERCISE 2.5.3. *Fill in the blanks.*

If we are proving the implication $p \to q$ we assume...

(1) p for a direct proof.
(2) _____ for a proof by contrapositive
(3) _____ for a proof by contradiction.

We are then allowed to use the truth of the assumption in 1, 2, or 3 in the proof.

After the initial assumption, we prove $p \to q$ by showing

(4) q must follow from the assumptions for a direct proof.
(5) _____ must follow the assumptions for a proof by contrapositive.
(6) _____ must follow the assumptions for a proof by contradiction.

2.6. Proof by Cases.

EXAMPLE 2.6.1. *If x is a real number such that $\dfrac{x^2 - 1}{x + 2} > 0$, then either $x > 1$ or $-2 < x < -1$.*

PROOF. Assume x is a real number for which the inequality

$$\frac{x^2 - 1}{x + 2} > 0$$

holds. Factor the numerator of the fraction to get the inequality

$$\frac{(x + 1)(x - 1)}{x + 2} > 0.$$

For this combination of $x + 1, x - 1$, and $x + 2$ to be positive, either all are positive or two are negative and the other is positive. This gives four cases to consider:

Case 1. $x+1>0, x-1>0$, and $x+2>0$. In this case $x>-1, x>1$, and $x>-2$, which implies $x>1$.
Case 2. $x+1>0, x-1<0$, and $x+2<0$. In this case $x>-1, x<1$, and $x<-2$, and there is no x satisfying all three inequalities simultaneously.
Case 3. $x+1<0, x-1>0$, and $x+2<0$. In this case $x<-1, x>1$, and $x<-2$, and there is no x satisfying all three inequalities simultaneously.
Case 4. $x+1<0, x-1<0$, and $x+2>0$. In this case $x<-1, x<1$, and $x>-2$, which implies that $-2<x<-1$.

Thus, either $x>1$ (Case 1) or $-2<x<-1$ (Case 4). □

Discussion

Sometimes the hypothesis of a statement can be broken down into simpler cases that may be investigated separately. The validity of a *proof by cases* rests on the equivalence

$$[(p_1 \vee \cdots \vee p_n) \to q] \Leftrightarrow [(p_1 \to q) \vee \cdots \vee (p_n \to q)].$$

In Example 2.6.1 this method is used to verify the "solution" to the inequality, $\dfrac{x^2-1}{x+2}>0$.

EXERCISE 2.6.1. *Prove: For every real number x, $\sqrt{x^2}=|x|$. [Hint: Recall as above that $\sqrt{x^2}$ represents the positive square of x^2, and look at two cases: $x\geq 0$ and $x<0$.]*

A proof by cases can tend to be a little tedious. Here is an extreme example of such a proof.

EXAMPLE 2.6.2. *Prove that if n is a natural number less than 41, then n^2-n+41 is a prime number.*

PROOF. Recall that a prime number is an integer greater than 1 that is only divisible by itself and 1. It would be nice if there was some general line of argument that would work, but, unfortunately, there doesn't seem to be an obvious one. As a result, the proof must be broken down into 41 cases corresponding to $n=0,1,2,...,40$. In each case we examine the integer n^2-n+41 to see if it is prime. For example, we can observe:

$n=0$: $0^2-0+41=41$ is prime.

$n=1$: $1^2-1+41=41$ is prime.

$n=2$: $2^2-2+41=43$ is prime.

$n = 3$: $3^2 - 3 + 41 = 47$ is prime.

$n = 4$: $4^2 - 4 + 41 = 53$ is prime.

As n increases, it becomes increasingly more time-consuming to show that $n^2 - n + 41$ is, indeed, prime. For example, when $n = 40$, $40^2 - 40 + 41 = 1601$. The simplest way to show that 1601 is prime is to show that every prime number $\leq \sqrt{1601}$ fails to be a divisor of 1601. There are 12 such numbers to try, and you might as well check them on your calculator. Alternatively, you could write a computer program or use a symbolic program such as Maple or Mathematica that has a routine to test a number for primality. □

2.7. Existence Proofs. An **existence proof** is a proof of a statement of the form $\exists x P(x)$. Existence proofs generally fall into one of the following two types:

Constructive Proof: Establish $P(c)$ for some c in the universe of discourse.

Nonconstructive Proof: Assume no c exists that makes $P(c)$ true and derive a contradiction.

2.8. Constructive Proof.

EXAMPLE 2.8.1. *Prove the statement: There exists a triple (a, b, c) of positive integers such that $a^2 + b^2 = c^2$.*

PROOF. Choose $a = 3$, $b = 4$ and $c = 5$. □

Discussion

In a constructive proof one finds an explicit example in the universe of discourse for which the statement is true.

Here is another example.

EXAMPLE 2.8.2. *Prove: If $f(x) = x^3 + x - 5$, then there exists a positive real number c such that $f'(c) = 7$.*

PROOF. Calculate the derivative of f: $f'(x) = 3x^2 + 1$. Then we want to find a positive number c such that $f'(c) = 3c^2 + 1 = 7$. Solving for c:

$$\begin{aligned} 3c^2 &= 6 \\ c^2 &= 2 \\ c &= \pm\sqrt{2} \end{aligned}$$

Then $c = \sqrt{2}$ is a positive real number and $f'(\sqrt{2}) = 3(\sqrt{2})^2 + 1 = 7$. □

2.9. Nonconstructive Proof.

EXAMPLE 2.9.1. **Pigeon Hole Principle:** *If $n + 1$ objects (pigeons) are distributed into n boxes (pigeon holes), then some box must contain at least 2 of the objects.*

PROOF. Suppose the boxes are labeled
$B_1, B_2, ..., B_n$, and assume that no box contains more than 1 object. Let k_i denote the number of objects placed in B_i. Then $k_i \leq 1$ for $i = 1, ..., n$, and so

$$k_1 + k_2 + \cdots + k_n \leq \underbrace{1 + 1 + \cdots + 1}_{n \text{ terms}} \leq n.$$

But this contradicts the fact that $k_1 + k_2 + \cdots + k_n = n + 1$, the total number of objects we started with. □

Discussion

Sometimes, constructing an example may be difficult, if not impossible, due to the nature of the problem. If you suspect this is the case, you should try a proof by contradiction: Assume there is no such example and show that this leads to a contradiction. If you are successful, you have established existence, but you have not exhibited a specific example. After you have studied the proof of the basic pigeon hole principal in Example 2.9.1, try your hand at the following variations.

EXERCISE 2.9.1. *Prove: If $2n + 1$ objects are distributed into n boxes, then some box must contain at least 3 of the objects.*

EXERCISE 2.9.2. *Fill in the blank in the following statement and then give a proof.*

Suppose k is a positive integer. If $kn + 1$ objects are distributed into n boxes, then some box must contain at least _____ of the objects.

EXERCISE 2.9.3. *Suppose that 88 chairs are arranged in a rectangular array of 8 rows and 11 columns, and suppose 50 students are seated in this array (1 student per chair).*

(a) Prove that some row must have at least 7 students.
(b) Prove that some column must have at most 4 students.

2.10. Nonexistence Proofs. Suppose we wish to establish the truth of the statement $\neg\exists x P(x)$, which is equivalent to $\forall x \neg P(x)$. One way is to assume there is an member, c, of the universe of discourse for which $P(c)$ is true, and try to arrive at a contradiction.

EXAMPLE 2.10.1. *Prove there does not exist an integer k such that $4k + 3$ is a perfect square.*

PROOF. Proof by Contradiction: Assume there is an integer k such that $4k + 3$ is a perfect square. That is, $4k + 3 = m^2$, where m is an integer. Since the square of an even integer is even and $4k + 3$ is odd, m must be odd. Then $m = 2a + 1$ for some integer a. Thus,

$$\begin{aligned} 4k + 3 &= m^2 \\ 4k + 3 &= (2a + 1)^2 \\ 4k + 3 &= 4a^2 + 4a + 1 \\ 4k + 3 &= 4(a^2 + a) + 1 \\ 3 - 1 &= 4(a^2 + a) - 4k \\ 2 &= 4(a^2 + a - k) \\ 1 &= 2(a^2 + a - k) \end{aligned}$$

But this contradicts the fact that 1 is an odd integer. □

Discussion

In order to show some property is false for every member of the universe of discourse it is almost always best to try to use a proof by contradiction. Example 2.10.1 illustrates a property of the integers that can be easily proved in this way.

EXERCISE 2.10.1. *Prove: There does not exist a positive real number a such that* $a + \dfrac{1}{a} < 2$.

2.11. The Halting Problem.

EXAMPLE 2.11.1. **The Halting Problem:** *There does not exist a program which will always determine if an arbitrary program P halts. We say the Halting Problem is undecidable. Note that this is not the same as determining if a specific program or finite set of programs halts. This is decidable.*

PROOF. We simplify the proof by only considering input-free programs (which may call other procedures). Assume there is a program called Halt which will determine if any input-free program P halts.

Halt(P) prints "yes" and halts if P halts. Halt(P) prints "no" and halts otherwise.

Now we construct a new procedure.

procedure Absurd
if Halt(Absurd) = "yes" **then**
    while true **do** print "ha"

Notice that the procedure Absurd is input-free. Now we consider two cases.

Case 1 If Absurd halts then we execute the loop which prints unending gales of laughter and thus the procedure does not halt – a contradiction.

Case 2 If Absurd does not halt then we will exit the program and halt. Again, this is a contradiction.

Now the only assumption we made was that a program exists which determines if any program will halt. Thus this assumption must be false. There is no such program.

□

2.12. Counterexample. Counterexample to $\forall x P(x)$: We may disprove a statement of the form $\forall x P(x)$ by finding a counterexample. That is, use the equivalence $\neg\forall x P(x) \Leftrightarrow \exists x \neg P(x)$, and find a c in the universe of discourse for which $P(x)$ is false.

Discussion

From Example 2.6.1 one might be led to think that $n^2 - n + 41$ is a prime number for every natural number n. After all, it worked for the first 41 natural numbers. (Or so, you were led to believe. Did you finish the remaining 35 cases?) Showing that a predicate $P(x)$ is true for a few, perhaps many millions of x's in its universe of discourse, however, does not constitute a *proof* of $\forall x P(x)$, unless you were able to exhaust all possibilities. This, of course, is not possible if the universe of discourse is

an infinite set, such as the set of natural numbers or the set of real numbers. Since the negation of $\forall x P(x)$ is $\neg\forall x P(x) \Leftrightarrow \exists x \neg P(x)$, it only takes *one* x for which $P(x)$ is false, a *counterexample*, to disprove $\forall x P(x)$. The assertion "for every natural number n, $n^2 - n + 41$ is prime" is, in fact, false.

EXERCISE 2.12.1. *Find a counterexample to the statement: For every natural number n, $n^2 - n + 41$ is prime.*

2.13. Biconditional. In order to establish the truth of the statement $p \leftrightarrow q$, use the fact that $(p \leftrightarrow q)$ is equivalent to $(p \rightarrow q) \wedge (q \rightarrow p)$, and prove both implications using any of the previous methods.

Discussion

We conclude this module with a discussion on proving a biconditional or "if and only if" statement. As pointed out above, a proof of a biconditional requires two proofs: the proof an implication *and* a proof of its converse. Our example below is very similar to theorems we have proved earlier. The point here is that the two implications may be proved independently of each other, and the decision on the best strategy to use should be made for each one separately.

EXAMPLE 2.13.1. *Prove: For any integer n, n is odd if and only if n^2 is odd.*

In order to prove this statement, we must prove two implications:

(a) If n is odd, then n^2 is odd.
(b) If n^2 is odd, then n is odd.

PROOF OF (a): We give a direct proof of this statement. Assume n is an odd integer. Then $n = 2a + 1$ for some integer a. Then $n^2 = (2a + 1)^2 = 4a^2 + 4a + 1 = 2(2a^2 + 2a) + 1$, which is twice an integer plus 1. Thus, n^2 is odd. □

PROOF OF (b): We give a proof of the contrapositive of this statement: "If n is even (not odd), then n^2 is even (not odd). Assume n is an even integer. Then $n = 2a$ for some integer a. Then $n^2 = (2a)^2 = 4a^2 = 2(2a^2)$, which is an even integer.

□

3. Mathematical Induction

3.1. First Principle of Mathematical Induction. Let $P(n)$ be a predicate with domain of discourse (over) the natural numbers $\mathbb{N} = \{0, 1, 2, ...\}$. If

(1) $P(0)$, and
(2) $P(n) \to P(n+1)$

then $\forall n P(n)$.

Terminology: The hypothesis $P(0)$ is called the **basis step** and the hypothesis, $P(n) \to P(n+1)$, is called the **induction (or inductive) step**.

Discussion

The Principle of Mathematical Induction is an axiom of the system of natural numbers that may be used to prove a quantified statement of the form $\forall n P(n)$, where the universe of discourse is the set of natural numbers. The principle of induction has a number of equivalent forms and is based on the last of the four Peano Axioms we alluded to in *Module 3.1 Introduction to Proofs*. The axiom of induction states that if S is a set of natural numbers such that (i) $0 \in S$ and (ii) if $n \in S$, then $n + 1 \in S$, then $S = \mathbb{N}$. This is a fairly complicated statement: Not only is it an "if ..., then ..." statement, but its hypotheses *also* contains an "if ..., then ..." statement (if $n \in S$, then $n + 1 \in S$). When we apply the axiom to the truth set of a predicate $P(n)$, we arrive at the *first principle of mathematical induction* stated above. More generally, we may apply the principle of induction whenever the universe of discourse is a set of integers of the form $\{k, k+1, k+2, \dots\}$ where k is some fixed integer. In this case it would be stated as follows:

Let $P(n)$ be a predicate over $\{k, k+1, k+2, k+3, \dots\}$, where $k \in \mathbb{Z}$. If

(1) $P(k)$, and
(2) $P(n) \to P(n+1)$

then $\forall n P(n)$.

In this context the "for all n", of course, means for all $n \geq k$.

REMARK 3.1.1. *While the principle of induction is a very useful technique for proving propositions about the natural numbers, it isn't always necessary. There were a number of examples of such statements in Module 3.2 Methods of Proof that were proved without the use of mathematical induction.*

Why does the principle of induction work? This is essentially the domino effect. Assume you have shown the premises. In other words you know $P(0)$ is true and you know that $P(n)$ implies $P(n+1)$ for any integer $n \geq 0$.

Since you know $P(0)$ from the basis step and $P(0) \rightarrow P(1)$ from the inductive step, we have $P(1)$ (by *modus ponens*).

Since you now know $P(1)$ and $P(1) \rightarrow P(2)$ from the inductive step, you have $P(2)$.

Since you now know $P(2)$ and $P(2) \rightarrow P(3)$ from the inductive step, you have $P(3)$.

And so on ad infinitum (or ad nauseum).

3.2. Using Mathematical Induction. Steps

1. Prove the basis step.
2. Prove the inductive step
   (a) Assume $P(n)$ for arbitrary n in the universe. This is called the **induction hypothesis**.
   (b) Prove $P(n+1)$ follows from the previous steps.

Discussion

Proving a theorem using induction requires two steps. First prove the basis step. This is often easy, if not trivial. Very often the basis step is $P(0)$, but sometimes, when the universal set has k as its least element, the basis step is $P(k)$. Be careful to start at the correct place.

Next prove the inductive step. *Assume* the induction hypothesis $P(n)$ is true. You do *not try to prove the induction hypothesis.* Now you prove that $P(n+1)$ *follows* from $P(n)$. In other words, you will use the truth of $P(n)$ to show that $P(n+1)$ must also be true.

Indeed, it may be possible to prove the implication $P(n) \rightarrow P(n+1)$ even though the predicate $P(n)$ is actually false for every natural number n. For example, suppose

$P(n)$ is the statement $n = n - 1$, which is certainly false for all n. Nevertheless, it is possible to show that *if* you assume $P(n)$, then you can correctly deduce $P(n+1)$ by the following simple argument:

PROOF. If $n = n - 1$, then, after adding 1 to both sides, $n + 1 = (n - 1) + 1 = (n + 1) - 1$. Thus $P(n) \to P(n + 1)$. □

It is easy at this point to think you are assuming what you have to prove (circular reasoning). You must keep in mind, however, that when you are proving the implication $P(n) \to P(n + 1)$ in the induction step, you are not proving $P(n)$ directly, as the example above makes clear, so this is not a case of circular reasoning. To prove an implication, all you need to show is that *if the premise is true* then the conclusion is true. Whether the premise is actually true at this point of an induction argument is completely irrelevant.

EXERCISE 3.2.1. *Notice in the above example that, while we proved* $\forall n[P(n) \to P(n + 1)]$*, we did not prove* $\forall n P(n)$*. Why?*

3.3. Example 3.3.1.

EXAMPLE 3.3.1. *Prove:* $\sum_{i=0}^{n} i = \frac{n(n+1)}{2}$ *for* $n = 0, 1, 2, 3, \ldots$

PROOF. Let $P(n)$ be the statement $\sum_{i=0}^{n} i = \frac{n(n+1)}{2}$.

1. Basis Step, $n = 0$:
   Prove $\sum_{i=0}^{0} i = 0(0 + 1)/2$.
   Proof: $\sum_{i=0}^{0} i = 0$ and $0(0 + 1)/2 = 0$
   Thus, $P(0)$.

2. Induction Step: Let $n \in \mathbb{N}$. At this step we are fixing an arbitrary integer $n \geq 0$ and making the following assumption for this fixed n. We then show the statement $P(n+1)$ must also be true. In general, we assume the induction hypothesis for an integer at least as large as the integer used in the basis case.
   (i) Assume $P(n)$: $\sum_{i=0}^{n} i = n(n + 1)/2$, for some integer $n \geq 0$.

(ii) Use the induction hypothesis to prove
$\sum_{i=0}^{n+1} i = (n+1)((n+1)+1)/2.$

Proof: Write out the sum on the left hand side of the statement to be proven.

$$\begin{aligned}
\sum_{i=0}^{n+1} i &= 0+1+2+3+\cdots+n+(n+1) \\
&= (0+1+2+3+\cdots+n)+(n+1) \\
&= \underbrace{\left(\sum_{i=0}^{n} i\right)}_{\text{equal by the induction hypothesis}} +(n+1) \\
&= \overbrace{\left[\frac{n(n+1)}{2}\right]}^{\text{equal by the induction hypothesis}} +(n+1) \\
&= \frac{n^2+n+2n+2}{2} = \frac{n^2+3n+2}{2} \\
&= \frac{(n+1)(n+2)}{2} \\
&= \frac{(n+1)((n+1)+1)}{2}
\end{aligned}$$

By the principle of mathematical induction it follows that $\sum_{i=0}^{n} i = \frac{n(n+1)}{2}$ for all natural numbers n.

□

Discussion

Example 3.3.1 is a classic example of a proof by mathematical induction. In this example the predicate $P(n)$ is the statement

$$\sum_{i=0}^{n} i = n(n+1)/2.$$

[Recall the "Sigma-notation": $\sum_{i=k}^{n} a_i = a_k + a_{k+1} + \cdots + a_n$.]

It may be helpful to state a few cases of the predicate so you get a feeling for whether you believe it's true and for the differences that occur when you change n. But keep in mind that exhibiting a few cases *does not constitute a proof.* Here are a few cases for Example 3.3.1. Notice what happens to the summation (left-hand side) as you increase n.

$P(0)$: $\sum_{i=0}^{0} i = 0 = 0(0+1)/2.$

$P(1)$: $\sum_{i=0}^{1} i = 0 + 1 = 1(1+1)/2.$

$P(2)$: $\sum_{i=0}^{2} i = 0 + 1 + 2 = 2(2+1)/2.$

$P(3)$: $\sum_{i=0}^{3} i = 0 + 1 + 2 + 3 = 3(3+1)/2.$

In the basis step of an induction proof you only need to prove the first statement above, but not the rest.

In the induction step you assume the induction hypothesis, $P(n)$, for some arbitrary integer $n \geq 0$. Write it out so you know what you have to work with. Then write out $P(n+1)$ so you can see what you need to prove. It will be easier to see how to proceed if you write both of these down. (A common mistake students make is to think of $P(n)$ as a particular expression (say, $P(n) = \sum_{i=0}^{n} i$) instead of as a *sentence*: $\sum_{i=0}^{n} i = \frac{n(n+1)}{2}$.) Once you have written down the induction hypothesis and what you need to prove, look for a way to express part of $P(n+1)$ using $P(n)$. In this example we use the summation notation $\sum_{i=0}^{n+1} a_i = (\sum_{i=0}^{n} a_i) + a_{n+1}$. This is a typical step when proving a summation formula of this type. After rewriting $\sum_{i=0}^{n+1} i$ this way,

we can apply the induction hypothesis to substitute $n(n+1)/2$ for $\sum_{i=0}^{n} i$. Note that you should use the induction hypothesis at some point in the proof. Otherwise, it is not really an induction proof.

EXERCISE 3.3.1. *Prove:* $\sum_{i=1}^{n}(2i-1) = 1+3+5+\cdots+(2n-1) = n^2$, *for all* $n \geq 1$.

EXERCISE 3.3.2. *Prove:* $\sum_{i=1}^{n} \frac{1}{i(i+1)} = \frac{n}{n+1}$

EXERCISE 3.3.3. *Prove* $\sum_{k=1}^{n} 3 \cdot 2^{k-1} = 3(2^n - 1)$

3.4. Example 3.4.1.

EXAMPLE 3.4.1. *Prove:* $5n+5 \leq n^2$ *for all integers* $n \geq 6$.

PROOF. Let $P(n)$ be the statement $5n+5 \leq n^2$.

1. Basis Step, $n = 6$: Since $5(6)+5 = 35$ and $6^2 = 36$ this is clear. Thus, $P(6)$.
2. Induction Step: Assume $P(n)$: $5n+5 \leq n^2$, for some integer $n \geq 6$. Use the induction hypothesis to prove $5(n+1)+5 \leq (n+1)^2$.
   First rewrite $5(n+1)+5$ so that it uses $5n+5$:
   $$5(n+1)+5 = (5n+5)+5.$$
   By the induction hypothesis we know
   $$(5n+5)+5 \leq n^2+5.$$
   Now we need to show
   $$n^2+5 \leq (n+1)^2 = n^2+2n+1.$$
   To see this we note that when $n \geq 2$,
   $$2n+1 \geq 2 \cdot 2+1 = 5$$
   (and so this is also valid for $n \geq 6$).
   Thus, when $n \geq 6$,
   $$\begin{aligned} 5(n+1)+5 &= (5n+5)+5 \\ (5n+5)+5 &\leq n^2+5 \\ n^2+5 &\leq n^2+2n+1 \\ n^2+2n+1 &= (n+1)^2. \end{aligned}$$

Which shows $5(n+1)+5 \leq (n+1)^2$. By the principle of mathematical induction it follows that $5n+5 \leq n^2$ for all integers $n \geq 6$.

□

Discussion

In Example 3.4.1, the predicate, $P(n)$, is $5n+5 \leq n^2$, and the universe of discourse is the set of integers $n \geq 6$. Notice that the basis step is to prove $P(6)$. You might also observe that the statement $P(5)$ is false, so that we can't start the induction any sooner.

In this example we are proving an *inequality* instead of an equality. This actually allows you more "fudge room", but sometimes that extra freedom can make it a bit more difficult to see what to do next. In this example, the hardest part, conceptually, is recognize that we need *another* inequality, $5 \leq 2n+1$, which holds whenever $n \geq 2$. A good approach to showing $f(n+1) \leq g(n+1)$ is to start with $f(n+1)$, think of a way express $f(n+1)$ in terms of $f(n)$ so that you can use the induction hypothesis, then find ways to get to $g(n+1)$ using further equalities or inequalities (that go in the right direction!).

In the induction step we use the fact that if you know $a \leq b$, then $a+5 \leq b+5$. The induction hypothesis gives us an inequality. Then we add 5 to both sides of that inequality prove $P(n+1)$.

Remark 3.4.1. *In proving an identity or inequality, you don't always have to start with the left side and work toward the right. In Example 3.4.1 you might try to complete the induction step by starting with $(n+1)^2$ and showing that it is greater than or equal to $5(n+1)+5$. The steps would go as follows:*

$$\begin{aligned}
(n+1)^2 &= n^2+2n+1 \\
n^2+2n+1 &\geq (5n+5)+2n+1 && \textit{by the induction hypothesis} \\
(5n+5)+2n+1 &= 5(n+1)+2n+1 \\
5(n+1)+2n+1 &\geq 5(n+1)+5 && \textit{if } n \geq 2
\end{aligned}$$

With this approach the place where the induction hypothesis comes in as well as the fact that we need the inequality $2n+1 \geq 5$ for $n \geq 2$ are, perhaps, a little more transparent.

Exercise 3.4.1. *Prove: $2n+1 \leq 2^n$, for all $n \geq 3$. Establish the induction step in two ways, as suggested in the remark above. [Hint: $2^n + 2^n = 2 \cdot 2^n = 2^{n+1}$.]*

EXERCISE 3.4.2. *Prove: $n^2 + 3 \leq 2^n$, for all $n \geq 5$. [Hint: Look for a place to use the inequality in the exercise above in the induction step.]*

3.5. Example 3.5.1.

EXAMPLE 3.5.1. *Prove: A set with n elements, $n \geq 0$, has exactly 2^n subsets.*

PROOF. Let $P(n)$ be the statement "A set with n elements has 2^n subsets."

1. Basis Step, $n = 0$: The only set with 0 elements is the empty set, $\emptyset$, which has exactly one subset, namely, $\emptyset$. We also have $2^0 = 1$, therefore a set with 0 elements has exactly 2^0 subsets. Thus $P(0)$.
2. Induction Step: Let $n \in \mathbb{N}$. Assume $P(n)$: every set with n elements has 2^n subsets. Use the induction hypothesis to prove a set with $n+1$ elements has 2^{n+1} subsets.

   Suppose A is a set with $n+1$ elements, say, $A = \{a_1, a_2, \ldots, a_n, a_{n+1}\}$. Let B be the subset $\{a_1, a_2, \ldots, a_n\}$ of A. Since B has n elements, we can apply the induction hypothesis to B, which says that B has exactly 2^n subsets. Each subset S of B corresponds to exactly two subsets of A, namely, S and $S \cup \{a_{n+1}\}$. But every subset of A is of one of these two forms; hence, A has exactly twice as many subsets as B. Thus, A has exactly $2 \cdot 2^n = 2^{n+1}$ subsets.

By the principle of mathematical induction it follows that a set with n elements has exactly 2^n subsets for all $n \geq 0$. □

Discussion

EXERCISE 3.5.1. *Let $A = \{a, b, c, d, e\}$ and $B = \{a, b, c, d\}$. List all the subsets of A in one column and all the subsets of B in another column. Draw a line connecting every subset of A to a subset from B to demonstrate the 2 to 1 correspondence used in the previous proof. Note that an example such as this does not prove the previous Theorem, but it does help to illustrate the tools used.*

Induction is used in a variety of situations. In Example 3.5.1 induction is used to establish a formula for the number of subsets of a set with n elements. In this case we are not trying to prove an equality in the sense of establishing an identity, as with the summation examples. The induction step involves more "pure reasoning" than algebraic manipulation. We have to devise a strategy to count the number of subsets of a set with $n+1$ elements, given that we know a formula for the number of subsets of a set with n elements. Having devised a strategy, we then need to show that the formula works for a set with $n+1$ elements as well. Once you begin to be proficient in constructing inductive proofs of this type, you are well on your way to a complete understanding of the induction process.

EXERCISE 3.5.2. *Prove: For all $n \geq 0$, a set with n elements has $\dfrac{n(n-1)}{2}$ subsets with exactly two elements. [Hint: In order to complete the induction step try to devise a strategy similar to the one used in the example in Example 3.5.1. It is interesting to observe that the formula works for sets with fewer than 2 elements.]*

Here is another type of problem from number theory that is amenable to induction.

EXAMPLE 3.5.2. *Prove: For every natural number n, $n(n^2+5)$ is a multiple of 6 (i.e. $n(n^2+5)$ equals 6 times some integer).*

PROOF. Let $P(n)$ be the statement $n(n^2+5)$ is a multiple of 6.

1. Basis Step, $n = 0$: $0(0^2+5) = 0 = 0 \cdot 6$. Thus $P(0)$.
2. Induction Step: Suppose $n \in \mathbb{N}$, and suppose $n(n^2+5)$ is divisible by 6. Show that this implies $(n+1)((n+1)^2+5)$ is divisible by 6. In order to use the inductive hypothesis, we need to extract the expression $n(n^2+5)$ out of the expression $(n+1)((n+1)^2+5)$.

$$\begin{aligned}(n+1)((n+1)^2+5) &= n((n+1)^2+5) + 1\cdot((n+1)^2+5)\\ &= n(n^2+2n+1+5) + (n^2+2n+1+5)\\ &= n(n^2+5) + n(2n+1) + (n^2+2n+6)\\ &= n(n^2+5) + 2n^2 + n + n^2 + 2n + 6\\ &= n(n^2+5) + 3n^2 + 3n + 6\\ &= n(n^2+5) + 3n(n+1) + 6\end{aligned}$$

By the induction hypothesis, the first term on the right-hand side, $n(n^2+5)$, is a multiple of 6. Notice that n and $n+1$ are consecutive integers; hence, one of them is even. Thus, $n(n+1)$ is a multiple of 2, and so $3n(n+1)$ is a multiple of 6. If we write $n(n^2+5) = 6k$ and $3n(n+1) = 6\ell$, then

$$(n+1)((n+1)^2+5) = n(n^2+5) + 3n(n+1) + 6 = 6k + 6\ell + 6 = 6(k+\ell+1)$$

so $(n+1)((n+1)^2+5)$ is a multiple of 6. Thus, we have shown $P(n) \to P(n+1)$.

By the principle of mathematical induction, $n(n^2+5)$ is a multiple of 6 for every $n \geq 0$. □

You may have noticed that in order to make the inductive step work in most of the examples and exercises we have seen so far, the restriction placed on n is actually

used, either implicitly or explicitly, whereas in the previous example it was not. (At no place in the inductive step above did we need the assumption that $n \geq 0$.) This leaves open the possibility that $n(n^2+5)$ is a multiple of 6 for (some/all?) integers $n < 0$ as well. Checking some cases, we see for

$n = -1$: $n(n^2+5) = (-1)((-1)^2+5) = -6$ is a multiple of 6,

$n = -2$: $n(n^2+5) = (-2)((-2)^2+5) = -18$ is a multiple of 6,

$n = -3$: $n(n^2+5) = (-3)((-3)^2+5) = -42$ is a multiple of 6,

$n = -4$: $n(n^2+5) = (-4)((-4)^2+5) = -84$ is a multiple of 6.

EXERCISE 3.5.3. *Use mathematical induction to prove that $n(n^2+5)$ is a multiple of 6 for all $n \leq 0$. [Hint: You will have to find the appropriate predicate $P(k)$.]*

EXERCISE 3.5.4. *Prove $5^{2n-1}+1$ is divisible by 6 for $n \in \mathbb{Z}^+$.*

EXERCISE 3.5.5. *Prove $a-b$ is a factor of $a^n - b^n$. Hint: $a^{k+1} - b^{k+1} = a(a^k - b^k) + b^k(a-b)$.*

EXERCISE 3.5.6. *The following is an incorrect "proof" that any group of n horses are the same color. What is the error in the proof?*

PROOF. The basis case is certainly true since any group of 1 horse is the same color. Now, let $n \in \mathbb{Z}^+$ and assume any group of n horses are the same color. We need to show any group of $n+1$ horses is the same color. Let $\{h_1, h_2, \ldots, h_{n+1}\}$ be a set of $n+1$ horses. The set $\{h_1, h_2, \ldots, h_n\}$ is a set of n horses and so these horses are the same color. Moreover, the set $\{h_2, h_3, \ldots, h_{n+1}\}$ is a set of n horses, so they are all the same color. Therefore the set of horses $\{h_1, h_2, \ldots h_{n+1}\}$ must all be the same color.

□

3.6. The Second Principle of Mathematical Induction. Let k be an integer, and let $P(n)$ be a predicate whose universe of discourse is the set of integers $\{k, k+1, k+2, ...\}$. Suppose

1. $P(k)$, and
2. $P(j)$ for $k \leq j \leq n$ implies $P(n+1)$.

Then $\forall n P(n)$.

Discussion

The second principle of induction differs from the first only in the form of the induction hypothesis. Here we assume not just $P(n)$, but $P(j)$ for all the integers j between k and n (inclusive). We use this assumption to show $P(n+1)$. This method of induction is also called **strong mathematical induction**. It is used in computer science in a variety of settings such as proving recursive formulas and estimating the number of operations involved in so-called "divide-and-conquer" procedures.

EXERCISE 3.6.1. *Prove the first principle of mathematical induction is equivalent to the second principle of mathematical induction.*

EXAMPLE 3.6.1. *Prove: Every integer $n \geq 2$ can be expressed as a product of one or more prime numbers. A prime number is defined to be an integer greater than one that is only divisible by itself and one.*

PROOF. Recall that a prime number is an integer ≥ 2 that is only divisible by itself and 1. (The number 1 is not considered to be prime.)

Let $P(n)$ be the predicate "n can be expressed as a product of prime numbers."

1. Basis Step, $n = 2$: Since 2 is prime, 2 can be expressed as a product of prime numbers in a trivial way (just one factor). Thus, $P(2)$ is true.
2. Induction Step: Let n be an integer with $n \geq 2$. Suppose that every integer j, $2 \leq j \leq n$, can be expressed as a product of prime numbers.
   The integer $n + 1$ is either a prime number or it is not.
   Case 1. If $n + 1$ is a prime number, then it is a product of prime numbers in a trivial way.
   Case 2. If $n + 1$ is not a prime number, then $n + 1 = a \cdot b$ where a and b are positive integers, both different from $n + 1$ and 1. Thus, $2 \leq a \leq n$ and $2 \leq b \leq n$. By the induction hypothesis, a and b can each be expressed as a product of prime numbers, say $a = p_1 p_2, \cdots p_r$ and $b = q_1 q_2 \cdots q_s$. Since $n + 1 = a \cdot b = p_1 p_2, \cdots p_r q_1 q_2 \cdots q_s$, $n + 1$ can also be expressed as a product of prime numbers, namely, the product of primes that multiply to give a times the product of primes that multiply to give b.

By the second principle of mathematical induction, every $n \geq 2$ can be expressed as a product of prime numbers. □

Discussion

In this example, the first principle of induction would be virtually impossible to apply, since the integer n is not a factor of $n + 1$ when $n \geq 2$. That is, knowing the factors of n doesn't tell us anything about the factors of $n + 1$.

3.7. Well-Ordered Sets.

DEFINITION 3.7.1. *A set S is* **well-ordered** *if every subset has a least element.*

Well-ordering Principle. The set $\mathbb{N}$ of natural numbers forms a well-ordered set.

Discussion

As we prove below, the principle of induction is equivalent to the *well-ordering principle*.

EXAMPLE 3.7.1. *The set S of integers greater than -5 is a well-ordered set.*

EXAMPLE 3.7.2. *The set P of rational numbers greater than or equal to zero is not a well-ordered set.*

EXAMPLE 3.7.3. $[0, 1]$ *is not well-ordered. The subset* $(0, 1]$ *does not have a least element in the set. (You may have to think about this for a moment.)*

EXAMPLE 3.7.4. *The set $\mathbb{Z}$ of integers is not well-ordered, since $\mathbb{Z}$, itself, does not have a least element.*

Study the proof of the following theorem carefully. Although it uses methods of proof discussed in *Module 3.2*, its level of abstraction may make it a bit difficult to absorb at first.

THEOREM 3.7.1. *The second principle of mathematical induction is equivalent to the well-ordering principle.*

PROOF. We must show that each principle implies the other.

1. Suppose $\mathbb{N}$ satisfies the principle of mathematical induction, and suppose that A is a nonempty subset of $\mathbb{N}$. We will give a proof by contradiction that A has a least element. Suppose A does not have a least element. Let $P(n)$ be the predicate $n \notin A$. Then
   (i) $0 \notin A$. Otherwise, 0 would be the least element of A. Thus $P(0)$.
   (ii) Let $n \in \mathbb{N}$. Suppose $P(k)$ for $0 \leq k \leq n$. Then $0, \ldots, n \notin A$. If $n + 1$ were in A, then $n + 1$ would be the least element of A. Thus, $n + 1 \notin A$, and so $P(n + 1)$. This proves that $P(0) \wedge \cdots \wedge P(n) \rightarrow P(n + 1)$.

   By the First Principle of Mathematical Induction, $\forall n P(n) = \forall n[n \notin A]$. But this means that A is empty, a contradiction. Thus $\mathbb{N}$ is well-ordered.
2. Suppose $\mathbb{N}$ is well-ordered, and suppose $P(n)$ is a predicate over $\mathbb{N}$ that satisfies the hypotheses of the First Principle of Mathematical Induction. That is,
   (i) $P(0)$, and
   (ii) $P(0) \wedge \cdots \wedge P(n) \rightarrow P(n + 1)$.

We will prove $\forall n P(n)$ by contradiction. Suppose $\neg\forall n P(n)$. Let A be the set of all $n \in \mathbb{N}$ such that $P(n)$ is false (i.e., $\neg P(n)$) Then A is nonempty, since $\neg\forall n P(n) \Leftrightarrow \exists n \neg P(n)$. Since $\mathbb{N}$ is well-ordered and A is a nonempty subset of $\mathbb{N}$, A has a least element k. In other words, if $P(n)$ fails to be true for all n, then there is a smallest natural number k for which $P(k)$ is false. By (i), $k \neq 0$, hence, $k > 0$, which implies $k - 1$ is a natural number. Since $k - 1 < k$, and k is the least element of A, $k - 1 \notin A$, so that $P(k - 1)$. But by (ii) $P(k - 1)$ implies $P(k)$, or $k \notin A$, which contradicts $k \in A$. Therefore, $\forall n P(n)$, and so $\mathbb{N}$ satisfies the principle of mathematical induction.

□

CHAPTER 4

Applications of Methods of Proof

1. Set Operations

1.1. Set Operations. The set-theoretic operations, intersection, union, and complementation, defined in *Module 1.1 Introduction to Sets* are are analgous to the operations $\wedge$, $\vee$, and $\neg$, respectively, that were defined for propositions. Indeed, each set operation was defined in terms of the corresponding operator from logic. We will discuss these operations in some detail in this section and learn methods to prove some of their basic properties.

Recall that in any discussion about sets and set operations there must be a set, called a *universal set*, that contains all others sets to be considered. This term is a bit of a misnomer: logic prohibits the existence of a "set of all sets," so that there is no one set that is "universal" in this sense. Thus the choice of a universal set will depend on the problem at hand, but even then it will in no way be unique. As a rule we usually choose one that is minimal to suit our needs. For example, if a discussion involves the sets $\{1, 2, 3, 4\}$ and $\{2, 4, 6, 8, 10\}$, we could consider our universe to be the set of natural numbers or the set of integers. On the other hand, we might be able to restrict it to the set of numbers $\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$.

We now restate the operations of set theory using the formal language of logic.

1.2. Equality and Containment.

Definition 1.2.1. *Sets A and B are* **equal***, denoted $A = B$, if*

$$\forall x[x \in A \leftrightarrow x \in B]$$

Note: This is equivalent to

$$\forall x[(x \in A \rightarrow x \in B) \wedge (x \in B \rightarrow x \in A)].$$

Definition 1.2.2. *Set A is contained in set B (or A is a subset of B), denoted $A \subseteq B$,if*

$\forall x[x \in A \rightarrow x \in B]$.

The above note shows that

$$A = B$$
$$\text{iff}$$
$$A \subseteq B \text{ and } B \subseteq A.$$

1.3. Union and Intersection.

DEFINITIONS 1.3.1.

- *The* **union** *of A and B,*
$$A \cup B = \{x | (x \in A) \vee (x \in B)\}.$$
- *The* **intersection** *of A and B,*
$$A \cap B = \{x | (x \in A) \wedge (x \in B)\}.$$
- *If $A \cap B = \emptyset$, then A and B are said to be* **disjoint**.

1.4. Complement.

DEFINITION 1.4.1. *The* **complement** *of A*
$$\overline{A} = \{x \in U | \neg(x \in A)\} = \{x \in U | x \notin A\}.$$

Discussion

There are several common notations used for the complement of a set. For example, A^c is often used to denote the compliment of A. You may find it easier to type A^c than $\overline{A}$, and you may use this notation in your homework.

1.5. Difference.

DEFINITION 1.5.1. *The* **difference** *of A and B, or the* **complement** *of B* **relative** *to A,*
$$A - B = A \cap \overline{B}.$$

DEFINITION 1.5.2. *The* **symmetric difference** *of A and B,*
$$\begin{aligned} A \oplus B &= (A - B) \cup (B - A) \\ &= (A \cup B) - (A \cap B). \end{aligned}$$

Discussion

The difference and symmetric difference of two sets are new operations, which were not defined in *Module 1.1*. Notice that B does *not* have to be a subset of A for

the difference to be defined. This gives us another way to represent the complement of a set A; namely, $\overline{A} = U - A$, where U is the universal set.

The definition of the difference of two sets A and B in some universal set, U, is equivalent to $A - B = \{x \in U | (x \in A) \wedge \neg(x \in B)\}$.

Many authors use the notation $A \setminus B$ for the difference $A - B$.

The symmetric difference of two sets corresponds to the logical operation $\otimes$, the exclusive "or".

The definition of the symmetric difference of two sets A and B in some universal set, U, is equivalent to

$$A \oplus B = \{x \in U | [(x \in A) \wedge \neg(x \in B)] \vee [\neg(x \in A) \wedge (x \in B)]\}.$$

1.6. Product.

DEFINITION 1.6.1. *The* **(Cartesian) Product** *of two sets, A and B, is denoted $A \times B$ and is defined by*

$$A \times B = \{(a, b) | a \in A \wedge b \in B\}$$

1.7. Power Set.

DEFINITION 1.7.1. *Let S be a set. The* **power set** *of S, denoted $\mathcal{P}(S)$ is defined to be the set of all subsets of S.*

Discussion

Keep in mind the power set is a set where all the *elements* are actually sets and the power set should include the empty set and itself as one of its elements.

1.8. Examples.

EXAMPLE 1.8.1. *Assume: $U = \{a, b, c, d, e, f, g, h\}$, $A = \{a, b, c, d, e\}$, $B = \{c, d, e, f\}$, and $C = \{a, b, c, g, h\}$. Then*

(a) $A \cup B = \{a, b, c, d, e, f\}$
(b) $A \cap B = \{c, d, e\}$
(c) $\overline{A} = \{f, g, h\}$
(d) $\overline{B} = \{a, b, g, h\}$
(e) $A - B = \{a, b\}$
(f) $B - A = \{f\}$
(g) $A \oplus B = \{a, b, f\}$

(h) $(A \cup B) \cap C = \{a, b, c\}$
(i) $A \times B = \{(a, c), (a, d), (a, e), (a, f), (b, c), (b, d), (b, e), (b, f), (c, c), (c, d), (c, e), (c, f), (d, c), (d, d), (d, e), (d, f), (e, c), (e, d), (e, e), (e, f)\}$
(j) $\mathcal{P}(A) = \{\emptyset, \{a\}, \{b\}, \{c\}, \{d\}, \{e\}, \{a, b\}, \{a, c\}, \{a, d\}, \{a, e\}, \{b, c\}, \{b, d\}, \{b, e\}, \{c, d\}, \{c, e\}, \{d, e\}, \{a, b, c\}, \{a, b, d\}, \{a, b, e\}, \{a, c, d\}, \{a, c, e\}, \{a, d, e\}, \{b, c, d\}, \{b, c, e\}, \{b, d, e\}, \{c, d, e\}, \{a, b, c, d\}, \{a, b, c, e\}, \{a, b, d, e\}, \{a, c, d, e\}, \{b, c, d, e\}, \{a, b, c, d, e\}\}$
(k) $|\mathcal{P}(A)| = 32$

EXERCISE 1.8.1. *Use the sets given in Example 1.8.1 to find*

(1) $B \times A$
(2) $\mathcal{P}(B)$
(3) $|\mathcal{P}(U)|$

EXAMPLE 1.8.2. *Let the universal set be $U = \mathbb{Z}^+$ the set of all positive integers, let P be the set of all prime (positive) integers, and let E be the set of all positive even integers. Then*

(a) $P \cup E = \{n \in \mathbb{Z}^+ | n \text{ is prime or even}\}$,
(b) $P \cap E = \{2\}$,
(c) $\overline{P}$ *is the set of all positive composite integers,*
(d) $\overline{E}$ *is the set of all positive odd integers,* $\{2n + 1 | n \in \mathbb{N}\}$,
(e) $P - E$ *is the set of all positive odd prime numbers (all prime numbers except 2),*
(f) $E - P = \{4, 6, 8, 10, \dots\} = \{2n | n \in \mathbb{Z}^+ \wedge n \geq 2\}$,
(g) $E \oplus P = \{n \in \mathbb{Z}^+ | (n \text{ is prime or even}) \wedge n \neq 2\}$

EXERCISE 1.8.2.
(1) If $|A| = n$ and $|B| = m$, how many elements are in $A \times B$?
(2) If S is a set with $|S| = n$, what is $|\mathcal{P}(S)|$?

EXERCISE 1.8.3. *Does $A \times B = B \times A$? Prove your answer.*

1.9. Venn Diagrams. A **Venn Diagram** is a useful geometric visualization tool when dealing with three or fewer sets. The Venn Diagram is generally set up as follows:

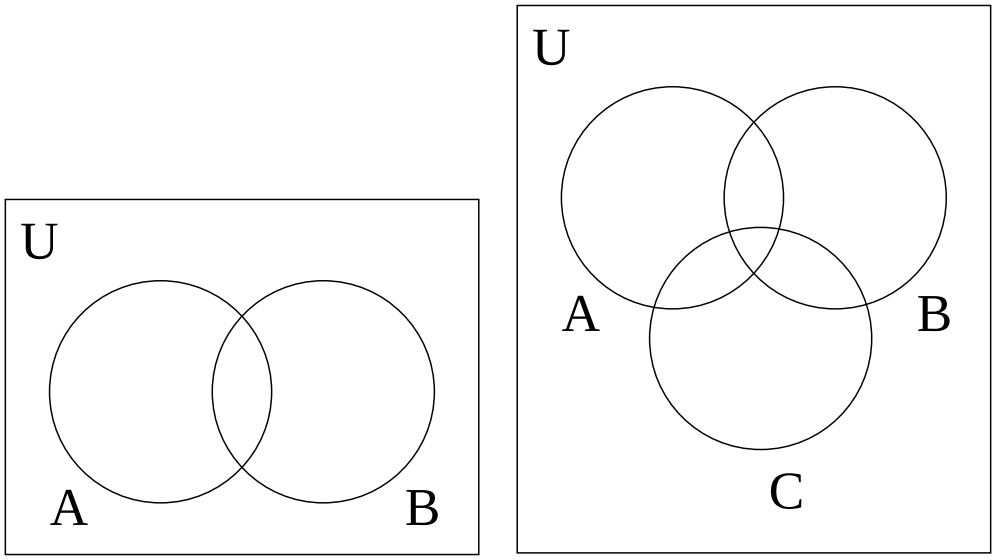
- The Universe U is the rectangular box.
- A set are represented by a circle and its interior.
- In the absence of specific knowledge about the relationships among the sets being represented, the most generic relationships should be depicted.

Discussion

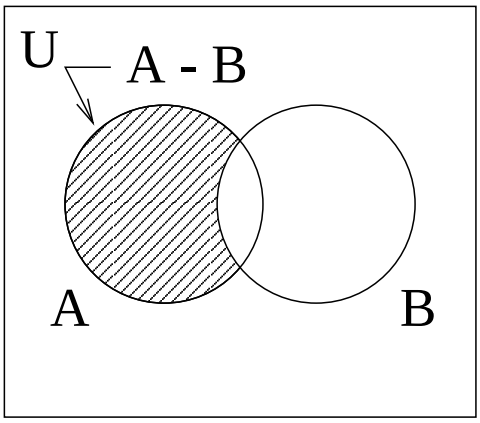
Venn Diagrams can be very helpful in visualizing set operations when you are dealing with three or fewer sets (not including the universal set). They tend not to be as useful, however, when considering more than three sets. Although Venn diagrams may be helpful in visualizing sets and set operations, they will *not* be used for proving set theoretic identities.

1.10. Examples.

EXAMPLE 1.10.1. *The following Venn Diagrams illustrate generic relationships between two and three sets, respectively.*



EXAMPLE 1.10.2. *This Venn Diagram represents the difference $A - B$ (the shaded region).*



The figures in the examples above show the way you might draw the Venn diagram if you aren't given any particular relations among the sets. On the other hand, if you knew, for example, that $A \subseteq B$, then you would draw the set A inside of B.

1.11. Set Identities.

EXAMPLE 1.11.1. *Prove that the complement of the union is the intersection of the complements:*

$$\overline{A \cup B} = \overline{A} \cap \overline{B}.$$

Proof 1. One way to show the two sets are equal is to use the fact that

$$\overline{A \cup B} = \overline{A} \cap \overline{B}$$

iff

$$\overline{A \cup B} \subseteq \overline{A} \cap \overline{B} \text{ and } \overline{A} \cap \overline{B} \subseteq \overline{A \cup B}.$$

Step 1. Show $\overline{A \cup B} \subseteq \overline{A} \cap \overline{B}$.
Assume x is an arbitrary element of $\overline{A \cup B}$ (and show $x \in \overline{A} \cap \overline{B}$). Since $x \in \overline{A \cup B}$, $x \notin A \cup B$. This means $x \notin A$ **and** $x \notin B$ (De Morgan's Law). Hence $x \in \overline{A} \cap \overline{B}$. Thus, by Universal Generalization,

$$\forall x[x \in (\overline{A \cup B}) \rightarrow x \in (\overline{A} \cap \overline{B})]$$

so that, by definition,

$$\overline{A \cup B} \subseteq \overline{A} \cap \overline{B}.$$

Step 2. Show $\overline{A} \cap \overline{B} \subseteq \overline{A \cup B}$.
Suppose x is an arbitrary element of $\overline{A} \cap \overline{B}$. Then $x \notin A$ and $x \notin B$. Therefore, $x \notin A \cup B$ (De Morgan's Law). This shows $x \in \overline{A \cup B}$. Thus, by Universal Generalization,

$$\forall x[x \in (\overline{A} \cap \overline{B}) \rightarrow x \in (\overline{A \cup B})]$$

so that, by definition,

$$\overline{A} \cap \overline{B} \subseteq \overline{A \cup B}.$$

□

Proof 2. The following is a second proof of the same result, which emphasizes more clearly the role of the definitions and laws of logic. We will show

$$\forall x[x \in \overline{A \cup B} \leftrightarrow x \in \overline{A} \cap \overline{B}].$$

Assertion	Reason
$\forall x$:	
$x \in \overline{A \cup B} \Leftrightarrow x \notin [A \cup B]$	Definition of complement
$\Leftrightarrow \neg[x \in A \cup B]$	Definition of $\notin$
$\Leftrightarrow \neg[(x \in A) \vee (x \in B)]$	Definition of union
$\Leftrightarrow \neg(x \in A) \wedge \neg(x \in B)$	De Morgan's Law
$\Leftrightarrow (x \in \overline{A}) \wedge (x \in \overline{B})$	Definition of complement
$\Leftrightarrow x \in \overline{A} \cap \overline{B}$	Definition of intersection

Hence $\forall x[x \in \overline{A \cup B} \leftrightarrow x \in \overline{A} \cap \overline{B}]$ is a tautology. □

(In practice we usually omit the formality of writing $\forall x$ in the initial line of the proof and assume that x is an arbitrary element of the universe of discourse.)

Proof 3. A third way to prove this identity is to build a **membership table** for the sets $\overline{A \cup B}$ and $\overline{A} \cap \overline{B}$, and show the membership relations for the two sets are the same. The 1's represent membership in a set and the 0's represent nonmembership.

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

Compare this table to the truth table for the proof of De Morgan's Law:

$$\neg(p \vee q) \leftrightarrow (\neg p \wedge \neg q)$$

□

Discussion

A set identity is an equation involving sets and set operations that is true for all possible choices of sets represented by symbols in the identity. These are analgous to identities such as

$$(a + b)(a - b) = a^2 - b^2$$

that you encounter in an elementary algebra course.

There are various ways to prove an identity, and three methods are covered here. This is a good place to be reminded that when you are proving an identity, you must show that it holds in *all* possible cases. Remember, giving an example does *not* prove an identity. On the other hand, if you are trying to show that an expression is *not* an identity, then you need only provide one counterexample. (Recall the negation of $\forall x P(x)$ is $\exists x \neg P(x)$).

Proof 1 establishes equality by showing each set is a subset of the other. This method can be used in just about any situation.

Notice that in Proof 1 we start with the assumption, x is in $\overline{A \cup B}$, where x is otherwise an arbitrary element in some universal set. If we can show that x must then be in $\overline{A} \cap \overline{B}$, then we will have established

$$\forall x, [x \in \overline{A \cup B}] \rightarrow [x \in \overline{A} \cap \overline{B}].$$

That is, the *modus operandi* is to prove the implications hold for an arbitrary element x of the universe, concluding, by Universal Generalization, that the implications hold for all such x.

Notice the way De Morgan's Laws are used here. For example, in the first part of Proof 1, we are given that $x \notin (A \cup B)$. This means

$$\neg[x \in (A \cup B)] \Leftrightarrow \neg[(x \in A) \vee (x \in B)] \Leftrightarrow [(x \notin A) \wedge (x \notin B)].$$

Proof 2 more clearly exposes the role of De Morgan's Laws. Here we prove the identity by using propositional equivalences in conjunction with Universal Generalization. When using this method, as well as any other, you must be careful to provide reasons.

Proof 3 provides a nice alternative when the identity only involves a small number of sets. Here we show two sets are equal by building a member table for the sets. The member table has a 1 to represent the case in which an element is a member of the set and a 0 to represent the case when it is not. The set operations correspond to a logical connective and one can build up to the column for the set desired.

You will have proved equality if you demonstrate that the two columns for the sets in question have the exact same entries. Notice that all possible membership relations of an element in the universal set for the sets A and B are listed in the first two columns of the membership table. For example, if an element is in both A and B in our example, then it satisfies the conditions in the first row of the table. Such an element ends up in neither of the two sets $\overline{A \cup B}$ nor $\overline{A} \cap \overline{B}$.

This is very straight forward method to use for proving a set identity. It may also be used to prove containment. If you are only trying to show the containment $M \subseteq N$, you would build the membership table for M and N as above. Then you would look in every row where M has a 1 to see that

N also has a 1. However, you will see examples in later modules where a membership table cannot be created. It is not always possible to represent all the different possibilities with a membership table.

EXAMPLE 1.11.2. *Prove the identity* $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$

Proof 1. Suppose x is an arbitrary element of the universe.

Assertion	Reason
$x \in (A \cup B) \cap C$	
$\Leftrightarrow [x \in (A \cup B)] \wedge [x \in C]$	definition of intersection
$\Leftrightarrow [(x \in A) \vee (x \in B)] \wedge [x \in C]$	definition of union
$\Leftrightarrow [(x \in A) \wedge (x \in C)] \vee [(x \in B) \wedge (x \in C)]$	distributive law of "and" over "or"
$\Leftrightarrow [x \in (A \cap C)] \vee [(x \in (B \cap C)]$	definition of intersection
$\Leftrightarrow x \in [(A \cap C) \cup (B \cap C)]$	definition of union

□

Proof 2. Build a membership table:

A	B	C	$A \cup B$	$(A \cap C)$	$(B \cap C)$	$(A \cup B) \cap C$	$(A \cap C) \cup (B \cap C)$
1	1	1	1	1	1	1	1
1	1	0	1	0	0	0	0
1	0	1	1	1	0	1	1
0	1	1	1	0	1	1	1
1	0	0	1	0	0	0	0
0	1	0	1	0	0	0	0
0	0	1	0	0	0	0	0
0	0	0	0	0	0	0	0

Since the columns corresponding to $(A \cup B) \cap C$ and $(A \cap C) \cup (B \cap C)$ are identical, the two sets are equal. □

EXAMPLE 1.11.3. *Prove* $(A - B) - C \subseteq A - (B - C)$.

PROOF. Consider the membership table:

A	B	C	$A - B$	$B - C$	$(A - B) - C$	$A - (B - C)$
1	1	1	0	0	0	1
1	1	0	0	1	0	0
1	0	1	1	0	0	1
1	0	0	1	0	1	1
0	1	1	0	0	0	0
0	1	0	0	1	0	0
0	0	1	0	0	0	0
0	0	0	0	0	0	0

Notice the only 1 in the column for $(A-B)-C$ is the fourth row. The entry in the same row in the column for $A-(B-C)$ is also a 1, so $(A-B)-C \subseteq A-(B-C)$. □

EXERCISE 1.11.1. *Prove the identitiy* $A - B = A \cap \overline{B}$ *using the method of Proof 2 in Example 1.11.1.*

EXERCISE 1.11.2. *Prove the identitiy* $A - B = A \cap \overline{B}$ *using the method of Proof 3 in Example 1.11.1.*

EXERCISE 1.11.3. *Prove the identity* $(A \cup B) - C = (A - C) \cup (B - C)$ *using the method of Proof 1 in Example 1.11.1.*

EXERCISE 1.11.4. *Prove the identity* $(A \cup B) - C = (A - C) \cup (B - C)$ *using the method of Proof 2 in Example 1.11.1.*

EXERCISE 1.11.5. *Prove the identity* $(A \cup B) - C = (A - C) \cup (B - C)$ *using the method of Proof 3 in Example 1.11.1.*

1.12. Union and Intersection of Indexed Collections.

DEFINITION 1.12.1. *The the union and intersection of an indexed collection of sets*

$$\{A_1, A_2, A_3, \ldots, A_n\}$$

can be written as

$$\bigcup_{i=1}^{n} A_i = A_1 \cup A_2 \cup A_3 \cup \cdots \cup A_n$$

and

$$\bigcap_{i=1}^{n} A_i = A_1 \cap A_2 \cap A_3 \cap \cdots \cap A_n,$$

respectively.

1.13. Infinite Unions and Intersections.

DEFINITION 1.13.1. *The the union and intersection of an indexed collection of infinitely many sets*

$$\{A_1, A_2, A_3, \ldots, \}$$

can be written as

$$\bigcup_{i=1}^{\infty} A_i = \{a | a \in A_i \text{ for some } i \text{ in } \mathbf{Z}^+\}$$

and

$$\bigcap_{i=1}^{\infty} A_i = \{a | a \in A_i \text{ for all } i \text{ in } \mathbf{Z}^+\}$$

Discussion

If you have a collection of more than two sets, you can define the intersection and the union of the sets as above. (Since the operations are associative, it isn't necessary to clutter the picture with parentheses.) The notation is similar to the Σ notation used for summations. The subscript is called an *index* and the collection of sets is said to be *indexed* by the set of indices. In the example, the collection of sets is $\{A_1, A_2, ..., A_n\}$, and the set of indices is the set $\{1, 2, ..., n\}$. There is no requirement that sets with different indices be different. In fact, they could all be the same set. This convention is very useful when the each of the sets in the collection is naturally described in terms of the index (usually a number) it has been assigned.

An equivalent definition of the union and intersection of an indexed collection of sets is as follows:

$$\bigcup_{i=1}^{n} A_i = \{x | \exists i \in \{1, 2, \ldots, n\} \text{ such that } x \in A_i\}$$

and

$$\bigcap_{i=1}^{n} A_i = \{x | \forall i \in \{1, 2, \ldots, n\}, x \in A_i\}.$$

Another standard notation for unions over collections of indices is

$$\bigcup_{i\in\mathbb{Z}^+} A_i = \bigcup_{i=1}^{\infty} A_i.$$

More generally, if $\mathcal{I}$ is any set of indices, we can define

$$\bigcup_{i\in\mathcal{I}} A_i = \{x | \exists i \in \mathcal{I} \text{ such that } x \in A_i\}.$$

1.14. Example 1.14.1.

EXAMPLE 1.14.1. *Let $A_i = [i, i+1)$, where i is a positive integer. Then*

- $\bigcup_{i=1}^{n} A_i = [1, n+1)$, *and*
- $\bigcap_{i=1}^{n} A_i = \emptyset$, *if* $n > 1$.
- $\bigcup_{i=1}^{\infty} A_i = [1, \infty)$
- $\bigcap_{i=1}^{\infty} A_i = \emptyset$

Discussion

This is an example of a collection of subsets of the real numbers that is naturally indexed. If $A_i = [i, i+1)$, then $A_1 = [1, 2)$, $A_2 = [2, 3)$, $A_3 = [3, 4)$, etc. It may help when dealing with an indexed collection of sets to explicitly write out a few of the sets as we have done here.

EXAMPLE 1.14.2. *Suppose $C_i = \{i-2, i-1, i, i+1, i+2\}$, where i denotes and arbitrary natural number. Then*

- $C_0 = \{-2, -1, 0, 1, 2\}$,
- $C_1 = \{-1, 0, 1, 2, 3\}$,
- $C_2 = \{0, 1, 2, 3, 4\}$,
- $\bigcup_{i=0}^{n} C_i = \{-2, -1, 0, 1, \ldots, n, n+1, n+2\}$
- $\bigcap_{i=0}^{4} C_i = \{2\}$

- $\bigcap_{i=0}^{n} C_i = \emptyset$ *if* $n > 4$.
- $\bigcup_{i=0}^{\infty} C_i = \{-2, -1, 0, 1, 2, 3, \dots\}$
- $\bigcap_{i=0}^{\infty} C_i = \emptyset$

EXERCISE 1.14.1. *For each positive integer* k*, let* $A_k = \{kn | n \in \mathbb{Z}\}$*. For example,*

- $A_1 = \{n | n \in \mathbb{Z}\} = \mathbb{Z}$
- $A_2 = \{2n | n \in \mathbb{Z}\} = \{..., -2, 0, 2, 4, 6, ...\}$
- $A_3 = \{3n | n \in \mathbb{Z}\} = \{..., -3, 0, 3, 6, 9, ...\}$

Find

1. $\bigcap_{k=1}^{10} A_k$
2. $\bigcap_{k=1}^{m} A_k$*, where* m *is an arbitrary positive integer.*

EXERCISE 1.14.2. *Use the definition for* A_k *in exercise 1.14.1 to answer the following questions.*

(1) $\bigcap_{i=1}^{\infty} A_i$

(2) $\bigcup_{i=1}^{\infty} A_i$

1.15. Computer Representation of a Set. Here is a method for storing subsets of a given, finite universal set:

Order the elements of the universal set and then assign a bit number to each subset A as follows. A bit is 1 if the element corresponding to the position of the bit in the universal set is in A, and 0 otherwise.

EXAMPLE 1.15.1. *Suppose* $U = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$*, with the obvious ordering. Then*

- *The bit string corresponding to* $A = \{2, 4, 6, 8, 10\}$ *is* 0101010101.
- *The bit string corresponding to* $B = \{1, 2, 3, 4\}$ *is* 1111000000.

Discussion

There are many ways sets may be represented and stored in a computer. One such method is presented here. Notice that this method depends not only on the universal set, but on the *order* of the universal set as well. If we rearrange the order of the universal set given in Example 1.15.1 to $U = \{10, 9, 8, 7, 6, 5, 4, 3, 2, 1\}$, then the bit string corresponding to the subset $\{1, 2, 3, 4\}$ would be 0000001111.

The beauty of this representation is that set operations on subsets of U can be carried out formally using the corresponding bitstring operations on the bit strings representing the individual sets.

EXAMPLE 1.15.2. *For the sets $A, B \subseteq U$ in Example 1.15.1:*

$$\begin{aligned} \overline{A} &= \overline{(0101010101)} \\ &= 1010101010 \\ &= \{1, 3, 5, 7, 9\} \end{aligned}$$

$$\begin{aligned} A \cup B &= 0101010101 \vee 1111000000 \\ &= 1111010101 \\ &= \{1, 2, 3, 4, 6, 8, 10\} \end{aligned}$$

$$\begin{aligned} A \cap B &= 0101010101 \wedge 1111000000 \\ &= 0101000000 \\ &= \{2, 4\} \end{aligned}$$

$$\begin{aligned} A \oplus B &= 0101010101 \oplus 1111000000 \\ &= 1010010101 \\ &= \{1, 3, 6, 8, 10\} \end{aligned}$$

EXERCISE 1.15.1. *Let $U = \{a, b, c, d, e, f, g, h, i, j\}$ with the given alphabetical order. Let $A = \{a, e, i\}$, $B = \{a, b, d, e, g, h, j\}$, and $C = \{a, c, e, g, i\}$.*

(1) Write out the bit string representations for A, B, and C.
(2) Use these representations to find
  (a) $\overline{C}$

(b) $A \cup B$
(c) $A \cap B \cap C$
(d) $B - C$

2. Properties of Functions

2.1. Injections, Surjections, and Bijections.

DEFINITION 2.1.1. *Given* $f\colon A \to B$

1. f *is* **one-to-one** *(short hand is* $1-1$*) or* **injective** *if preimages are unique. In this case,* $(a \neq b) \to (f(a) \neq f(b))$.
2. f *is* **onto** *or* **surjective** *if every* $y \in B$ *has a preimage. In this case, the range of* f *is equal to the codomain.*
3. f *is* **bijective** *if it is surjective and injective (one-to-one and onto).*

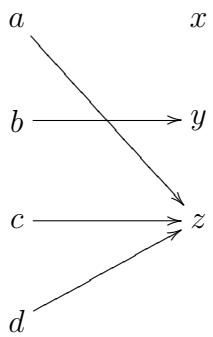
Discussion

We begin by discussing three very important properties functions defined above.

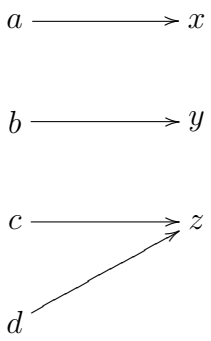
1. A function is *injective* or *one-to-one* if the preimages of elements of the range are unique. In other words, if every element in the range is assigned to exactly one element in the domain. For exampe, if a function is defined from a subset of the real numbers to the real numbers and is given by a formula $y = f(x)$, then the function is one-to-one if the equation $f(x) = b$ has *at most one solution* for every number b.
2. A function is *surjective* or *onto* if the range is equal to the codomain. In other words, if every element in the codomain is assigned to at least one value in the domain. For exampe, if, as above, a function is defined from a subset of the real numbers to the real numbers and is given by a formula $y = f(x)$, then the function is onto if the equation $f(x) = b$ has *at least one solution* for every number b.
3. A function is a *bijection* if it is both injective and surjective.

2.2. Examples.

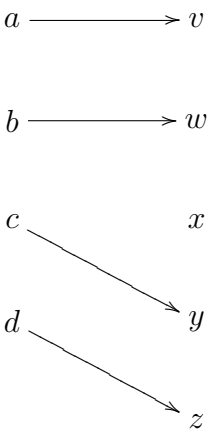
EXAMPLE 2.2.1. *Let* $A = \{a, b, c, d\}$ *and* $B = \{x, y, z\}$. *The function* f *is defined by the relation pictured below. This function is neither injective nor surjective.*



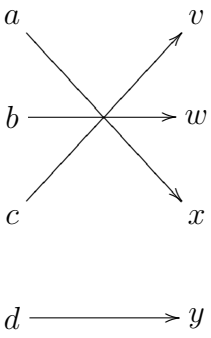
EXAMPLE 2.2.2. *$f\colon A \to B$ where $A = \{a, b, c, d\}$ and $B = \{x, y, z\}$ defined by the relation below is a surjection, but not an injection.*



EXAMPLE 2.2.3. *$f\colon A \to B$ where $A = \{a, b, c, d\}$ and $B = \{v, w, x, y, z\}$ defined by the relation below is an injection, but not a surjection.*



EXAMPLE 2.2.4. *$f\colon A \to B$ where $A = \{a, b, c, d\}$ and $B = \{v, w, x, y\}$ defined by the relation below both a surjection and an injection, and therefore a bijection. Notice that for a function to be a bijection, the domain and codomain must have the same cardinality.*



Discussion

The examples illustrate functions that are injective, surjective, and bijective. Here are further examples.

EXAMPLE 2.2.5. *Let $f\colon [0,\infty) \to [0,\infty)$ be defined by $f(x) = \sqrt{x}$. This function is an injection and a surjection and so it is also a bijection.*

EXAMPLE 2.2.6. *Suppose $f(x) = x^2$. If the domain and codomain for this function is the set of real numbers, then this function would be neither a surjection nor an injection. It is not a surjection because the range is not equal to the codomain. For example, there is no number in the domain with image -1 which is an element of the codomain. It is not an injection since more than one distinct element in the domain is mapped to the same element in the codomain. For example, $f(-1) = f(1)$ but $-1 \neq 1$.*

EXERCISE 2.2.1. *What if we say the domain of the function in Example 2.2.6 is the set of all reals and the codomain is $[0,\infty)$. Which properties would the function have (injective and/or surjective)? Explain.*

EXERCISE 2.2.2. *Now, if we say the domain and the codomain are both $[0,\infty)$. What properties does the function in Example 2.2.6 have? Explain.*

2.3. Example 2.3.1.

EXAMPLE 2.3.1. *Prove that the function $f\colon \mathbb{N} \to \mathbb{N}$ be defined by $f(n) = n^2$ is injective.*

PROOF. Let $a, b \in \mathbb{N}$ be such that $f(a) = f(b)$. This implies

$a^2 = b^2$ by the definition of f.

Thus $a = b$ or $a = -b$. Since the domain of f is the set of natural numbers, both a and b must be nonnegative. Thus $a = b$.

This shows $\forall a \forall b[f(a) = f(b) \to a = b]$, which shows f is injective. □

Discussion

In Example 2.3.1 we prove a function is injective, or one-to-one. Notice that to prove a function, $f\colon A \to B$ is one-to-one we must show the following:

$$(\forall x \in A)(\forall y \in A)[(x \neq y) \to (f(x) \neq f(y))].$$

This is equivalent to showing

$$(\forall x \in A)(\forall y \in A)[(f(x) = f(y)) \to (x = y)].$$

To prove this statement (which actually uses the contrapositive of the definition) we begin by choosing two arbitrary elements of the domain and assume the hypothesis

of the implication, i.e. we begin with "Let $x, y \in A$ and assume $f(x) = f(y)$." We then use the rules of algebra to show that the conclusion must follow.

2.4. Example 2.4.1.

EXAMPLE 2.4.1. *Prove that the function $g\colon \mathbb{N} \to \mathbb{N}$, defined by $g(n) = \lfloor n/3 \rfloor$, is surjective.*

PROOF. Let $n \in \mathbb{N}$. Notice that $g(3n) = \lfloor (3n)/3 \rfloor = \lfloor (3n)/3 \rfloor = n$. Since $3n \in \mathbb{N}$, this shows n is in the range of g. Hence g is surjective. □

Discussion

To prove a function, $f\colon A \to B$ is surjective, or onto, we must show $f(A) = B$. In other words, we must show the two sets, $f(A)$ and B, are equal. We already know that $f(A) \subseteq B$ if f is a well-defined function. While most functions encountered in a course using algebraic functions are well-defined, this should not be an automatic assumption in general. With that said, though, we will usually assume the functions given to us are well defined, so all that must be shown is that $B \subseteq f(A)$. To do this we may use the definition of a subset: show every element of B is also an element of $f(A)$. Thus we begin the proof by fixing an arbitrary element of B. We then use the tools at our disposal (definition of the function, algebra, any other known information) to show that this arbitrary element must in fact be the image of some element of A.

2.5. Example 2.5.1.

EXAMPLE 2.5.1. *Prove that the function $g\colon \mathbb{N} \to \mathbb{N}$, defined by $g(n) = \lfloor n/3 \rfloor$, is not injective.*

PROOF. The numbers 1 and 2 are in the domain of g and are not equal, but $g(1) = g(2) = 0$. Thus g is not injective. □

Discussion

To show a function is *not* injective we must show

$$\neg[(\forall x \in A)(\forall y \in A)[(x \neq y) \to (f(x) \neq f(y))]].$$

This is equivalent to

$$(\exists x \in A)(\exists y \in A)[(x \neq y) \wedge (f(x) = f(y))].$$

Thus when we show a function is not injective it is enough to find an example of two different elements in the domain that have the same image.

2.6. Example 2.6.1.

EXAMPLE 2.6.1. *Prove that the function* $f\colon \mathbb{N} \to \mathbb{N}$ *be defined by* $f(n) = n^2$*, is not surjective.*

PROOF. The number 3 is an element of the codomain, $\mathbb{N}$. However, 3 is not the square of any integer. Therefore, there is no element of the domain that maps to the number 3, so f is not surjective. □

Discussion

To show a function is not surjective we must show $f(A) \neq B$. Since a well-defined function must have $f(A) \subseteq B$, we should show $B \not\subseteq f(A)$. Thus to show a function is not surjective it is enough to find an element in the codomain that is not the image of any element of the domain. You may assume the familiar properties of numbers in this module as done in the previous examples.

2.7. Inverse Functions.

DEFINITION 2.7.1. *Suppose* $f\colon A \to B$ *is a bijection. Then the* **inverse** *of* f*, denoted*

$$f^{-1}\colon B \to A,$$

is the function defined by the rule

$$f^{-1}(y) = x \text{ if and only if } f(x) = y.$$

Discussion

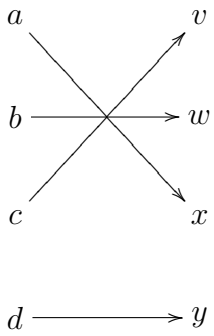
If a function f is a bijection, then it makes sense to define a new function that reverses the roles of the domain and the codomain, but uses the same rule that defines f. This function is called the inverse of the f. If the function is not a bijection, it does *not* have an inverse.

You have seen many functions in algebra and calculus that are defined as inverses of other functions. For example, the square-root function $\sqrt{x}$ is defined by the rule $y = \sqrt{x}$ if and only if $y \geq 0$ and $y^2 = x$. That is, the square-root function is the inverse of the square function. Before we can invert the square function, however, its usual domain, the set of all real numbers, must be restricted to the set of nonnegative real numbers in order to have a bijection. That is, if $A = B = \{x \in \mathbb{R} : x \geq 0\}$,

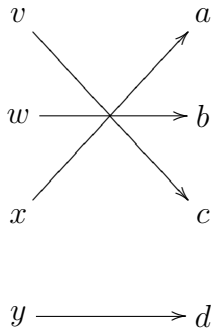
then the function $f\colon A \to B$ defined by $f(x) = x^2$ is a bijection, and its inverse $f^{-1}\colon B \to A$ is the square-root function, $f^{-1}(x) = \sqrt{x}$.

Another important example from algebra is the logarithm function. If a is a positive real number, different from 1, and $\mathbb{R}^+ = \{x \in \mathbb{R} : x > 0\}$, the function $f\colon \mathbb{R} \to \mathbb{R}^+$ defined by $f(x) = a^x$ is a bijection. Its inverse, $f^{-1}\colon \mathbb{R}^+ \to \mathbb{R}$, is the logarithm function with base a: $f^{-1}(x) = \log_a x$. In other words $y = \log_a x$ if and only if $a^y = x$.

EXAMPLE 2.7.1. *Let $f\colon A \to B$, where $A = \{a, b, c, d\}$ and $B = \{v, w, x, y\}$, be defined as follows*



Then the inverse function is $f^{-1}\colon B \to A$ defined as follows



EXAMPLE 2.7.2. *Suppose $f\colon \mathbb{R} - \{2\} \to \mathbb{R} - \{1\}$ is defined by $f(x) = \dfrac{x}{x-2}$. Then the function $f^{-1}(x) = \dfrac{2x}{x-1}$ is the inverse of f. This is a good example of an algebraically defined function whose inverse has a nice formula specifying its rule. You may recall that the formula defining f^{-1} can be obtained by setting $y = f(x) = \dfrac{x}{x-2}$, interchanging x and y, and solving algebraically for y.*

EXERCISE 2.7.1. *Find the inverse of the function $f\colon \mathbb{R} - \{-2\} \to \mathbb{R} - \{1\}$ defined by $f(x) = \dfrac{x-1}{x+2}$.*

EXERCISE 2.7.2. *Find the inverse of the function* $f\colon \mathbb{R} \to (-\infty, 1)$ *defined by* $f(x) = 1 - e^{-x}$.

THEOREM 2.7.1. *If a function is a bijection, then its inverse is also a bijection.*

PROOF. Let $f : A \to B$ be a bijection and let $f^{-1} : B \to A$ be its inverse. To show f^{-1} is a bijection we must show it is an injection and a surjection.

Let $x_1, x_2 \in B$ be such that $f^{-1}(x_1) = f^{-1}(x_2)$. Then by the definition of the inverse we have $x_1 = f(f^{-1}(x_2) = x_2$. This shows f^{-1} is injective.

We leave the proof that f^{-1} is surjective as an exercise for the reader. □

EXERCISE 2.7.3. *Finish the proof of Theorem 2.7.1.*

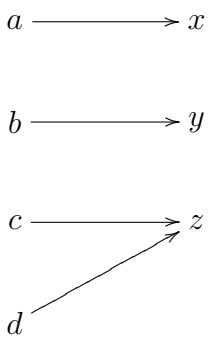
2.8. Inverse Image.

DEFINITION 2.8.1. *Let* $f\colon A \to B$ *and let* S *be a subset of* B. *Then the* **inverse image** *of* S *under* f *is the set*

$$f^{-1}(S) = \{x \in A | f(x) \in S\}.$$

There is no requirement for f *to be injective or surjective for this definition to hold.*

EXAMPLE 2.8.1. *Let* $f\colon A \to B$, *where* $A = \{a, b, c, d\}$ *and* $B = \{x, y, z\}$ *be defined as follows*



- $f^{-1}(\{z\}) = \{c, d\}$
- $f^{-1}(\{x, y\}) = \{a, b\}$

Discussion

We revisit the definition of the inverse image of a set to emphasis the difference between the *inverse image of a subset of the codomain* and the *inverse of a function*. The inverse image of a set is the set of all elements that map into that set. The function does not have to be injective or surjective to find the inverse image of a set. For example, the function $f(n) = 1$ with domain and codomain all natural numbers

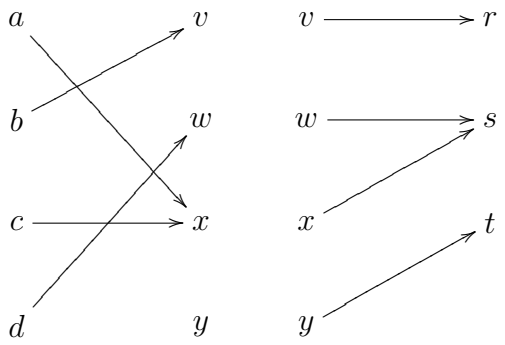
would have the following inverse images: $f^{-1}(\{1\}) = \mathbb{N}$ and $f^{-1}(\{5,6,7,8,9\}) = \emptyset$. This function does not have an inverse, however.

The context of the use of the notation f^{-1} usually indicates if the inverse image or the inverse function is intended. If A is a subset of the codomain we would always assume $f^{-1}(A)$ is the inverse image of A. When discussing a bijection the distinction between the inverse image and inverse function is often blurred.

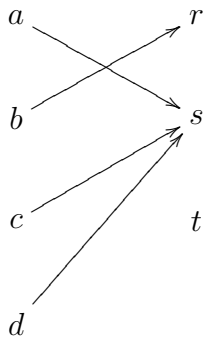
2.9. Composition.

DEFINITION 2.9.1. *Let $f\colon A \to B$ and $g\colon B \to C$. The* **composition** *of g with f, denoted $g \circ f$, is the function from A to C defined by $(g \circ f)(x) = g(f(x))$.*

EXAMPLE 2.9.1. *Let $A = \{a, b, c, d\}$, $B = \{v, w, x, y\}$, and $C = \{r, s, t\}$; and let $f\colon A \to B$ and $g\colon B \to C$ by defined as follows*



Then the composition $g \circ f$ is as follows



Discussion

The composition of two functions is defined by following one function by another. To define the composition $g \circ f$ we must have the range of f contained in the domain of g.

2.10. Example 2.10.1.

EXAMPLE 2.10.1. *Prove that the composition of two injective functions is injective.*

PROOF. Let A, B, and C be sets and let $f\colon A \to B$ and $g\colon B \to C$ be two injections.

Suppose x and y are elements of A such that $(g \circ f)(x) = (g \circ f)(y)$. This means $g(f(x)) = g(f(y))$. Since the codomain of f is B, $f(x) \in B$ and $f(y) \in B$. Thus we have two elements of B, $f(x)$ and $f(y)$, such that $g(f(x)) = g(f(y))$. Since g is injective, we must have $f(x) = f(y)$. But now we may use the fact that f is injective to conclude $x = y$.

This shows that $g \circ f$ is an injection. □

Discussion

Sometimes we have to prove properties about functions without any specific formula for the functions. In Example 2.10.1 we prove that the composition of two injections is again an injection. We cannot use specific examples of function because that would not prove this more general statement. We have to use the tools given by the assumptions, namely, that we know the two functions that make up the composition are known to be injections.

EXERCISE 2.10.1. *Prove the composition of two surjections is a surjection.*

THEOREM 2.10.1 (Corollary to Example 2.10.1 and Exercise 2.10.1). *The composition of two bijections is a bijection.*

EXERCISE 2.10.2. *Prove or disprove: if the composition of two functions is an injection then the two original functions must be injections too.*

EXERCISE 2.10.3. *Prove or disprove: if the composition of two functions is an surjection then the two original functions must be surjections too.*

3. Recurrence

3.1. Recursive Definitions.

To construct a **recursively defined function:**

1. Initial Condition(s) (or basis): Prescribe initial value(s) of the function.
2. Recursion: Use a fixed procedure (rule) to compute the value of the function at the integer $n+1$ using one or more values of the function for integers $\leq n$.

To construct a **recursively defined set:**

1. Initial Condition(s) (or basis): Prescribe one or more elements of the set.
2. Recursion: Give a rule for generating elements of the set in terms of previously prescribed elements.

Discussion

In computer programming evaluating a function or executing a procedure is often accomplished by using a **recursion**. A recursive process is one in which one or more initial stages of the process are specified, and the nth stage of the process is defined in terms of previous stages, using some fixed procedure. In a computer program this is usually accomplished by using a subroutine, such as a "For" loop, in which the same procedure is repeated a specified number of times; that is, the procedure *calls itself.*

EXAMPLE 3.1.1. *The function $f(n) = 2^n$, where n is a natural number, can be defined recursively as follows:*

1. Initial Condition: $f(0) = 1$,

2. Recursion: $f(n+1) = 2 \cdot f(n)$, *for* $n \geq 0$.

For any particular n, this procedure could be programmed by first initializing $F = 1$, and then executing a loop "For $i = 1$ to n, $2 * F = F$."

Here is how the definition gives us the first few powers of 2:

$2^1 = 2^{0+1} = 2^0 \cdot 2 = 2$

$2^2 = 2^{1+1} = 2^1 \cdot 2 = 2 \cdot 2 = 4$

$2^3 = 2^{2+1} = 2^2 \cdot 2 = 4 \cdot 2 = 8$

3.2. Recursive definition of the function $f(n) = n!$.

EXAMPLE 3.2.1. *The factorial function $f(n) = n!$ is defined recursively as follows:*

1. *Initial Condition:* $f(0) = 1$
2. *Recursion:* $f(n+1) = (n+1)f(n)$

Discussion

Starting with the initial condition, $f(0) = 1$, the recurrence relation, $f(n+1) = (n+1)f(n)$, tells us how to build new values of f from old values. For example,

$1! = f(1) = 1 \cdot f(0) = 1,$

$2! = f(2) = 2 \cdot f(1) = 2,$

$3! = f(3) = 3 \cdot f(2) = 6,$ etc.

When a function $f(n)$, such as the ones in the previous examples, is defined recursively, the equation giving $f(n+1)$ in terms of previous values of f is called a **recurrence relation**.

3.3. Recursive definition of the natural numbers.

DEFINITION 3.3.1. *The set of natural numbers may be defined recursively as follows.*

1. *Initial Condition:* $0 \in \mathbb{N}$
2. *Recursion: If $n \in \mathbb{N}$, then $n+1 \in \mathbb{N}$.*

Discussion

There are a number of ways of defining the set $\mathbb{N}$ of natural numbers recursively. The simplest definition is given above. Here is another recursive definition for $\mathbb{N}$.

EXAMPLE 3.3.1. *Suppose the set S is defined recursively as follows:*

1. *Initial Condition:* $0, 1 \in S$,
2. *Recursion: If $x, y \in S$, then $x + y \in S$.*

Then $S = \mathbb{N}$.

Notice that, in the recursive step, x *and* y *don't have to represent different numbers. Thus, having* $x = y = 1 \in S$*, we get* $1 + 1 = 2 \in S$*. Then we get* $1 + 2 = 3 \in S$*. And so on.*

It should be noted that there is an *extremal clause* in recursively defined sets. If you cannot build a given element in a finite number of applications of the recursion then it is not in the set built from the recursive definition. To prove an element is in a recursively defined set, you must show that element can be built in a finite number of steps.

EXAMPLE 3.3.2. *Prove that the set* S *recursively in Example 3.3.1 is equal to the set* $\mathbb{N}$ *of natural numbers.*

PROOF. We will show that $S = \mathbb{N}$ by showing separately that $S \subseteq \mathbb{N}$ and $\mathbb{N} \subseteq S$.

1. First we show $\mathbb{N} \subseteq S$. Prove by induction that $n \in S$ for every natural number $n \geq 0$.
   (a) Basis Step. $0, 1 \in S$ by definition.
   (b) Induction Step. Suppose $n \in S$, for some $n \geq 1$. Then, by the recursion step, $n \in S$ and $1 \in S$ imply $n + 1 \in S$.

   Thus, by the first principle of mathematical induction, $n \in S$ for every natural number $n \geq 0$.
2. Now we show $S \subseteq \mathbb{N}$. This time we apply the second principle of mathematical induction on n to show that if $s \in S$ is produced by applying n steps (1 initial condition and $n - 1$ recursive steps), then $s \in \mathbb{N}$.
   (a) Basis Step. After one step the only elements produced are 0 and 1, each of which is in $\mathbb{N}$.
   (b) Induction Step. Suppose $n \geq 1$ and assume any element in S produced by applying n or fewer steps is also an element of $\mathbb{N}$. Suppose $s \in S$ is produced by applying $n + 1$ steps. Since $n + 1 \geq 2$, there must be two elements x and y in S, such that $s = x + y$. Both x and y must have been produced by applying *fewer* than $n + 1$ steps, since s is produced by applying $n + 1$ steps, and we use one step to obtain s from x and y. By the induction hypothesis both x and y are element of $\mathbb{N}$. Since the sum of two natural numbers is again a natural number we have $s \in \mathbb{N}$.

□

3.4. Proving assertions about recursively defined objects. Assertions about recursively defined objects are usually proved by mathematical induction. Here are three useful versions of induction. In particular, note the third version which we introduce here.

Version 1. *Second Principle of Induction*

a. Basis Step: Prove the assertion for the initial conditions. (The assertion may have to be verified for more than one particular value.)

b. Induction Step: Assume the assertion is true for integers $\leq n$, and use the recurrence relation to prove the assertion for the integer $n+1$.

Version 2. a. Basis Step: Prove the assertion for the initial conditions. (The assertion may have to be verified for more than one particular value.)

b. Induction Step: Assume the assertion is true when the recursive definition has been applied less than or equal to n times for some integer $n \geq 0$, and use the recurrence relation to prove the assertion when the recursive definition is applied $n+1$ times.

Version 3. *Generalized or Structural Principle of Induction*: Use to prove an assertion about a set S defined recursively by using a set X given in the basis and a set of rules using $s_1, s_2, \ldots, s_k \in S$ for producing new members in the recursive set.

a. Basis Step: Prove the assertion for every $s \in X$.

b. Induction Step: Let $s_1, s_2, \ldots, s_k \in S$ be arbitrary and assume the assertion for these elements (this is the induction hypothesis). Prove all elements of S produced using the recursive definition and $s_1, s_2, \ldots, s_k$ satisfies the assertion.

Discussion

EXAMPLE 3.4.1. *Let S, a subset of $\mathbb{N} \times \mathbb{N}$, be defined recursively by*

1. *Initial Condition:* $(0,0) \in S$
2. *Recursion: If $(m,n) \in S$, then $(m+2, n+3) \in S$.*

Prove that if $(m,n) \in S$, then $m+n$ is a multiple of 5.

PROOF. We use the Generalized Principle of Induction.

1. Basis Step: Show the statement for $(0,0)$: $0+0$ is a multiple of 5. Since $0 = 0 \cdot 5$ this is clear.
2. Inductive Step: Let $(m,n) \in S$ and assume $m+n$ is a multiple of 5. Show the statement is true for $(m+2, n+3)$. In other words, show $(m+2)+(n+3)$ is a multiple of 5.

$$(m+2)+(n+3) = (m+n)+5$$

We know $m+n$ is a multiple of 5 and clearly 5 is a multiple of 5, so the sum must also be a multiple of 5. This proves the induction step.

Therefore, by the first principle of mathematical induction if $(m, n) \in S$, then $m + n$ is a multiple of 5. □

EXERCISE 3.4.1. *Suppose a S is defined recursively as follows:*

1. $1 \in S$,
2. *If $x \in S$, then $2 \cdot x \in S$.*

Prove that $S = \{2^n | n \geq 0\}$.

EXERCISE 3.4.2. *Suppose a S is defined recursively as follows:*

1. $0, 1 \in S$,
2. *If $x, y \in S$, then $x \cdot y \in S$.*

What are the elements of S? Prove that your answer is correct.

EXAMPLE 3.4.2. *Suppose $f(n)$ is defined by*

1. *Initial Condition:* $f(0) = 0$
2. *Recursion:* $f(n + 1) = f(n) + (n + 1)$, *for* $n \geq 0$.

Then $f(n) = \dfrac{n(n+1)}{2}$ *for all* $n \geq 0$

PROOF. 1. Basis Step ($n = 0$): $f(0) = 0$, by definition. On the other hand $\dfrac{0(0+1)}{2} = 0$. Thus, $f(0) = \dfrac{0(0+1)}{2}$.

2. Inductive Step: Suppose $f(n) = \dfrac{n(n+1)}{2}$ for some $n \geq 0$. We must prove $f(n+1) = \dfrac{(n+1)((n+1)+1)}{2}$.

$$\begin{aligned} f(n+1) &= f(n) + (n+1) && \text{(recurrence relation)} \\ &= \frac{n(n+1)}{2} + (n+1) && \text{(by the induction hypothesis)} \\ &= (n+1)\left[\frac{n}{2} + 1\right] \\ &= (n+1)\left[\frac{n+2}{2}\right] \\ &= \frac{(n+1)((n+1)+1)}{2} \end{aligned}$$

Therefore, by the first principle of mathematical induction $f(n) = \dfrac{n(n+1)}{2}$ for all $n \geq 0$. □

EXERCISE 3.4.3. *Suppose $f(n)$, $n \geq 0$, is defined recursively as follows:*

1. $f(0) = 0$,
2. $f(n+1) = f(n) + (2n+1)$, *for* $n \geq 0$.

Prove that $f(n) = n^2$ for all $n \geq 0$.

3.5. Definition of f^n.

DEFINITION 3.5.1. *Let $f : A \to A$ be function. Then we define f^n recursively as follows*

1. *Initial Condition:* $f^1 = f$
2. *Recursion:* $f^{n+1} = f \circ f^n$, *for* $n \geq 1$.

Discussion

EXAMPLE 3.5.1. *Prove that if f is injective, then f^n is injective for $n \geq 1$.*

PROOF. Assume f is injective.

1. Basis Step: $f^1 = f$ is injective by our assumption.
2. Inductive Step: Let $n \geq 1$ and assume f^n is injective. Prove f^{n+1} is injective.
   Recall that to prove f^{n+1} is injective we must show $\forall a, b \in A[(f^{n+1}(a) = f^{n+1}(b)) \to (a = b)]$
   Assume $a, b \in A$ and $f^{n+1}(a) = f^{n+1}(b)$.

$$\begin{aligned}
f^{n+1}(a) &= f^{n+1}(b) && \text{(recurrence relation)}\\
(f \circ f^n)(a) &= (f \circ f^n)(b) && \text{(recurrence relation)}\\
f(f^n(a)) &= f(f^n(b)) && \text{(by the definition of composition)}\\
f^n(a) &= f^n(b) && \text{(since f is injective)}\\
a &= b && \text{(by the induction hypothesis, } f^n \text{ is injective)}
\end{aligned}$$

Therefore, by the first principle of mathematical induction f^n is injective for all positive integers.

□

EXERCISE 3.5.1. *Prove that if f is surjective that f^n is surjective.*

3.6. Example 3.6.1.

EXAMPLE 3.6.1. *Given a real number $a \neq 0$, define a^n for all natural numbers, n, inductively by*

1. *Initial Condition:* $a^0 = 1$
2. *Recursion:* $a^{(n+1)} = a^n a$

THEOREM 3.6.1. $\forall m \forall n[a^m a^n = a^{m+n}]$ *where m, n are natural numbers.*

PROOF. Proof that $(\forall m)(\forall n)[a^m a^n = a^{m+n}]$, where m, n are natural numbers. We accomplish this by assuming m is an arbitrary natural number and proving $\forall n[a^m a^n = a^{m+n}]$ by induction on n.

1. Basis Step ($n = 0$): Show $a^m a^0 = a^{m+0}$.
   This follows directly from the initial condition of the definition: $a^0 = 1$, therefore $a^m a^0 = a^m(1) = a^m = a^{m+0}$.
2. Induction Step:
   Induction hypothesis: Let n be a natural number and assume $a^m a^n = a^{n+m}$. Prove $a^m a^{n+1} = a^{m+(n+1)}$.

$$\begin{aligned} a^m a^{n+1} &= a^m a^n a && \text{by the recursive part of the definition: } a^{n+1} = a^n a \\ &= a^{m+n} a && \text{by the induction hypothesis} \\ &= a^{(m+n)+1} = a^{(m+(n+1))} && \text{by the recursive part of the definition} \end{aligned}$$

   By induction, $\forall n[a^m a^n = a^{m+n}]$.

Since m was an arbitrary natural number the statement is true for all natural numbers m. □

Discussion

Here we see a recursive definition for the function $f(n) = a^n$, where n is a natural number and a is an arbitrary nonzero real number followed by a proof of one of the laws of exponents, $a^{m+n} = a^m a^n$. This proof uses both mathematical induction and Universal Generalization. We fix m as some arbitrary natural number and then proceed to use induction on n. We do not need to use induction on both m and n simultaneously. When there are two different variables, this is a standard strategy to try. There are circumstances, however, when this strategy doesn't work so that you

would need to use induction on *both* of the variables (*double induction*). We will not encounter these kinds of problems in this course, however.

3.7. Fibonacci Sequence.

DEFINITION 3.7.1. *The Fibonacci sequence may be defined recursively as follows:*

1. *Initial Conditions:* $F(0) = 0, F(1) = 1$

2. *Recursion:* $F(n+1) = F(n) + F(n-1)$ *for* $n \geq 1$

Discussion

The famous *Fibonacci sequence* is defined here using a recursively defined function. The definition of the Fibonacci sequence requires two initial conditions. There are no limits on the number of initial conditions on a recursively defined object – only that there be a fixed finite number in each instance.

EXAMPLE 3.7.1. *Suppose* $F(n)$, $n \geq 0$, *denotes the Fibonacci sequence. Prove that* $1 < \dfrac{F(n+1)}{F(n)} < 2$ *for all* $n \geq 3$.

PROOF. Let $R(n) = \dfrac{F(n+1)}{F(n)}$ for $n \geq 1$. We will prove by induction that

$$1 < R(n) < 2$$

for all $n \geq 3$.

1. Basis Step ($n = 3$): $R(3) = \dfrac{F(4)}{F(3)} = \dfrac{3}{2}$, and $1 < \frac{3}{2} < 2$.

2. Induction Step: Suppose $1 < R(n) < 2$ for some $n \geq 3$. We need to prove $1 < R(n+1) < 2$.

$$\begin{aligned} R(n+1) &= \frac{F(n+2)}{F(n+1)} \\ &= \frac{F(n+1)+F(n)}{F(n+1)} \quad \text{by the recursive definition of } F(n) \\ &= 1 + \frac{F(n)}{F(n+1)} \qquad \text{or} \\ R(n+1) &= 1 + \frac{1}{R(n)} \end{aligned}$$

By the inductive hypothesis $1 < R(n) < 2$; and so

$$1 > \frac{1}{R(n)} > \frac{1}{2}.$$

Thus,

$$2 > 1 + \frac{1}{R(n)} > \frac{3}{2} > 1,$$

or

$$1 < 1 + \frac{1}{R(n)} < 2.$$

Substituting from above, we have

$$1 < R(n+1) < 2.$$

By the first principle of mathematical induction, $1 < \dfrac{F(n+1)}{F(n)} < 2$ for all $n \geq 3$.

□

EXAMPLE 3.7.2. *(calculus required) Prove that* $\lim\limits_{n\to\infty} \dfrac{F(n+1)}{F(n)} = \dfrac{1+\sqrt{5}}{2}$, *if the limit exists.*

PROOF. Let $R(n) = \dfrac{F(n+1)}{F(n)}$ as in Example 3.7.1. Then, from the induction step in Example 3.7.1, we see that $R(n+1) = 1 + \dfrac{1}{R(n)}$. Assume $\lim\limits_{n\to\infty} \dfrac{F(n+1)}{F(n)}$ exists and let $L = \lim\limits_{n\to\infty} \dfrac{F(n+1)}{F(n)} = \lim\limits_{n\to\infty} R(n)$. Notice that $\lim\limits_{n\to\infty} R(n) = \lim\limits_{n\to\infty} R(n+1)$. Therefore, $L = \lim\limits_{n\to\infty} \big(1 + \dfrac{1}{R(n)}\big) = 1 + \dfrac{1}{L}$.

Since L is a real number, the equation

$$L = 1 + \frac{1}{L}$$

is equivalent to

$$L^2 - L - 1 = 0.$$

By the quadratic formula

$$L = \frac{1 \pm \sqrt{5}}{2}.$$

Since $L > 0$ and since $\sqrt{5} > 1$,

$$L = \frac{1 + \sqrt{5}}{2}.$$

□

EXERCISE 3.7.1. *Prove that* $F(n) > (\frac{3}{2})^{n-1}$ *for* $n \geq 6$. *[Hint: Show that the statement is true for* $n = 6$ *and 7 (basis step), and then show the induction step works for* $n \geq 7$.*]*

Here is another "Fibonacci-like" sequence.

EXAMPLE 3.7.3. *Suppose* $F(n)$, $n \geq 0$, *is defined recursively as follows:*

1. $F(0) = 1$ *and* $F(1) = 2$,
2. $F(n+1) = F(n) + 2F(n-1)$, *for* $n \geq 1$.

Prove that $F(n) = 2^n$ *for all* $n \geq 0$.

PROOF. (Using the *second* principle of mathematical induction)

1. Basis Step ($n = 0$ and $n = 1$): $F(0) = 1 = 2^0$ and $F(1) = 2 = 2^1$.
2. Induction Step: Let $n \geq 1$. Suppose $F(k) = 2^k$ for $0 \leq k \leq n$. Then

$$\begin{aligned} F(n+1) &= F(n) + 2F(n-1) \\ &= 2^n + 2 \cdot 2^{n-1} \\ &= 2^n + 2^n \\ &= 2 \cdot 2^n \\ &= 2^{n+1} \end{aligned}$$

Thus, by the second principle of mathematical induction, $F(n) = 2^n$ for all $n \geq 0$. □

REMARKS 3.7.1. *(1) Here and in the previous exercise we see the slight variation in the basis step from the ones encountered in Module 3.3 Induction; there may be more than one initial condition to verify before proceeding to the induction step.*

(2) Notice that in this example as well as in some of the other examples and exercises, we have been asked to prove that a function defined recursively is also given by a relatively simple formula. The problem of "solving" recurrence relations for these nice formulas (so-called closed form) is an interesting subject in its own right, but it will not be discussed in this course.

EXERCISE 3.7.2. *Let $f : \mathbb{N} \to \mathbb{Z}$ be the function defined recursively by*

1. *$f(0) = 1$ and $f(1) = -4$,*
2. *$f(n) = -3f(n-1) + 4f(n-2)$, for $n \geq 2$.*

Prove $f(n) = (-4)^n$

EXERCISE 3.7.3. *Let $f : \mathbb{N} \to \mathbb{Z}$ be the function defined recursively by*

1. *$f(0) = 2$ and $f(1) = 7$,*
2. *$f(n) = f(n-1) + 2f(n-2)$, for $n \geq 2$.*

Prove $f(n) = 3 \cdot 2^n - (-1)^n$

3.8. Strings.

DEFINITION 3.8.1. *Given a finite set of symbols, Σ, the set of strings, denoted Σ^*, over Σ is defined recursively as follows:*

1. *Initial Condition: The empty string λ is in Σ^*.*
2. *Recursion: If w is a string in Σ^* and a is a symbol in Σ, then wa is in Σ^*.*

Discussion

The set Σ is usually called an **alphabet**, and strings in Σ^* are called **words** in the alphabet Σ. Strings ("words") on a finite alphabet, Σ, are defined recursively using *right concatenation*. In other words, every string of symbols from Σ can be built from a smaller string by applying new letters to the right.

REMARK 3.8.1. *When a is a symbol from an alphabet, we use the notation a^n, where $n \in \mathbb{N}$, to represent a concatenated with itself n times. In particular, a^0 is understood to represent the empty string, λ.*

EXERCISE 3.8.1. *Suppose the alphabet Σ is the set $\{a, b, c, d\}$. Then Σ^* is the set of all words on Σ. Use right concatenation to build the bit string daabc starting with the empty string, λ (use $\lambda a = a$ for any element, a, in Σ).*

EXERCISE 3.8.2. *Now take the same bit string, daabc, and build it using left concatenation. Notice that your steps are not the same; that is, concatenation is not commutative. Regardless, we arrive at the same set of strings, Σ^*.*

3.9. Bit Strings.

EXAMPLE 3.9.1. *The set S of all bit strings with no more than a single 1 can be defined recursively as follows:*

1. Initial Condition: $\lambda, 1 \in S$

2. Recursion: If w is a string in S, then so are $0w$ and $w0$.

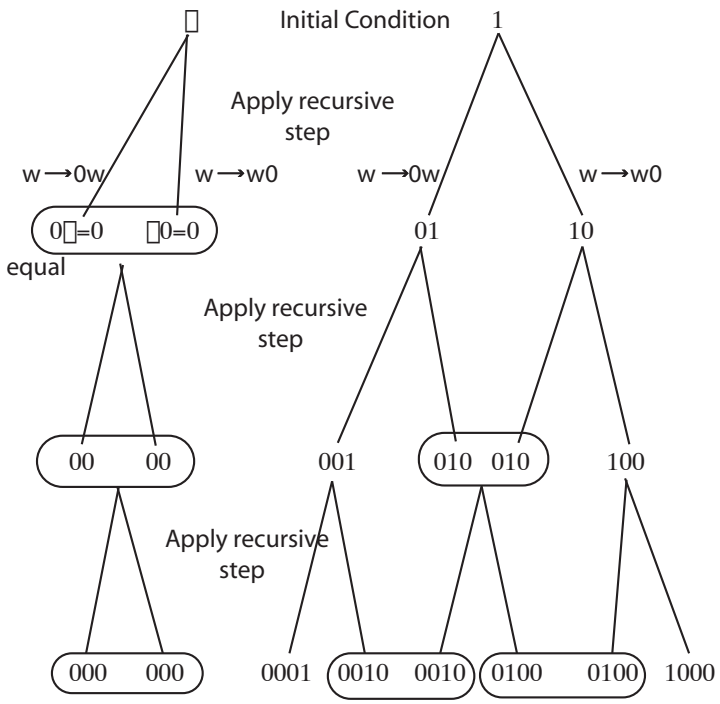
Discussion

In this definition we must start with two objects in the set. Then we can build all the bit strings that have *at most* one "1". We can define various subsets of bit strings using recursively defined sets.

EXAMPLE 3.9.2. *This example is an extension of example 3.9.1. Let the set S be defined recursively by*

Basis: $\lambda, 1 \in S$
Recursion: *If $w \in S$ then $0w \in S$ and $w0 \in S$.*

In creating a set recursively, it can help to use a tree diagram to develop more of an intuitive understanding of how this set is built. The following diagram shows how the applying the above definition 4 times gives the elements in the diagram. Some key ideas to keep in mind is that all strings in the tree and all strings that would be in the tree if you kept going are in the set. When a string is repeated, that means there is more than one way to get that element but there is no need to see what is produced from it more than one time.



Prove S is the set of all bit strings with no more than one 1.

PROOF. Let A denote the set of all bit strings with no more than one 1 in the string. Then we need to show $A = S$.

First we will show $A \subseteq S$. Let $a \in A$. Then a is a bit string with either no 1's or it is a bit string with exactly one 1.

Case 1 Suppose a has no 1's. Then $a = 0^n$ where n is some natural number. We can build a using the recursive definition by starting with λ and applying the recursive step n times. (If we apply the recursive step 0 times, we get λ).

Case 2 Suppose a has exactly on 1. Then $a = 0^n 10^m$ for some $n, m \in \mathbb{N}$. We can build a by starting with 1, which is given by the initial condition, and applying the recursive step $(w \in S) \to (0w \in S)$ n times and applying $(w \in S) \to (w0 \in S)$ m times.

This shows we can apply the recursive definition given for S finitely many times to obtain any element of A. Therefore $A \subseteq S$.

Now we show $S \subseteq A$ by general induction.

basis: The elements given by the basis (initial condition) of the defintion of S are both in A since λ has no 1's and 1 has one 1.

induction step: Let $x \in S$ and assume $x \in A$. We need to show any elements obtained by appling the recursive step one time will also be in A.

Notice we obtain $0x$ and $x0$ when we apply the recursive step one time to x. Since x is in A we know the string x has either no ones or a single one. $0x$ and $x0$ do not add any more 1's to the bit string, so they are also in A.

Thus by the principle of mathematical induction $\forall x \in S(x \in A)$.

This completes the proof that $S = A$. □

EXAMPLE 3.9.3. *Here's an example of proving a recursively defined set of bit strings is the same as set of bit strings defined in a non-recursive manner:*

Let A be the set of all bit strings of the form 0^n1^n, where n can be any natural number. Note that this is the same as $A = \{0^n1^n | n \in \mathbb{N}\} = \{\lambda, 01, 0011, 000111, 00001111, ...\}$ and is slightly different from the set described in Exercise 3.9.3.

Now define B by the following recursive definition:

Basis: $\lambda \in B$
Recursive Step: *If $w \in B$ then $0w1 \in B$*

Prove that $A = B$.

PROOF. First we prove $A \subseteq B$. Let $a \in A$. Then $a = 1^n0^n$ for some $n \in \mathbb{N}$. If we use the recursive definition of B we see $\lambda = 0^01^0$ by the basis step and if we apply the recursive step n times to λ we will build to the element 0^n1^n. This demonstrates that we can apply the recursive definition to find a using a finite number of steps. Thus $a \in B$.

Now we need to prove $B \subseteq A$. We will do this using generalized induction, which gives us a formal proof of this statement.

Basis: Notice the element, λ, created by the initial step (or basis step) in the definition of B is also an element of A ($\lambda = 0^01^0$).

Induction Step: Let $x \in B$ be such that $x \in A$ as well. Show that any element of B obtained from x by applying the recursive step one time is also an element of A.

If we apply the recursive step to x one time the only element we get $0x1$. Since x is an element of A we know $x = 0^n1^n$ for some $n \in \mathbb{N}$. So then $0x1 = 0(0^n1^n)1 = 0^{n+1}1^{n+1}$ which we see is also an element of A.

Thus by the principle of generalized induction $\forall x \in B(x \in A)$.

This completes that proof that $A = B$. □

EXERCISE 3.9.1. *What kinds of bit strings would we have if the initial condition in Example 3.9.1 is changed to* $1 \in S$ *only? So the definition would be*

1. *Initial Condition:* $1 \in S$,
2. *Recursion: If* w *is a string in* S*, then so are* $0w$ *and* $w0$.

EXERCISE 3.9.2. *What kinds of strings do we get from the following recursive definition?*

1. *Initial Conditions:* $\lambda, 1 \in S$,
2. *Recursion: If* w *is a string in* S*, then so is* $0w$.

EXERCISE 3.9.3. *Find a recursive definition for the set of bit strings* $T = \{0^r 1^s | r, s \in \mathbb{N}\}$.

EXERCISE 3.9.4. *Prove your answer for Exercise 3.9.3 is correct.*

EXERCISE 3.9.5. *Find a recursive definition for the set of all bit strings containing no adjacent 1's. (For example, 1001010 is allowed, but 0011010 is not.)*

4. Growth of Functions

4.1. Growth of Functions. Given functions f and g, we wish to show how to quantify the statement:

"g grows as fast as f".

The growth of functions is directly related to the complexity of algorithms. We are guided by the following principles.

- We only care about the behavior for "large" problems.
- We may ignore implementation details such as loop counter incrementation.

Discussion

When studying the complexity of an algorithm, we are concerned with the growth in the number of operations required by the algorithm as the size of the problem increases. In order to get a handle on its complexity, we first look for a function that gives the number of operations in terms of the size of the problem, usually measured by a positive integer n, to which the algorithm is applied. We then try to compare values of this function, for large n, to the values of some known function, such as a power function, exponential function, or logarithm function. Thus, the *growth of functions* refers to the relative size of the values of two functions for large values of the independent variable. This is one of the main areas in this course in which experience with the concept of a limit from calculus will be of great help.

Before we begin, one comment concerning notation for logarithm functions is in order. Most algebra and calculus texts use $\log x$ to denote $\log_{10} x$ (or, perhaps, $\log_e x$), but in computer science base 2 is used more prevalently. So we shall use $\log x$ to denote $\log_2 x$. As we shall see, in the context of this module it actually doesn't matter which base you use, since $\log_a x = \frac{\log_b x}{\log_b a}$ for any acceptable bases a and b.

EXERCISE 4.1.1. *Prove that* $\log_a x = \frac{\log_b x}{\log_b a}$ *for arbitrary positive real numbers* a *and* b *different from 1.*

4.2. The Big-O Notation.

DEFINITION 4.2.1. *Let* f *and* g *be functions from the natural numbers to the real numbers. Then* g **asymptotically dominates** f*, or*

f is big-O of g

if there are positive constants C *and* k *such that*

$$|f(x)| \leq C|g(x)| \text{ for } x \geq k.$$

If f is big-O of g, then we write

$$f(x) \text{ is } O(g(x))$$
or
$$f \in O(g).$$

THEOREM 4.2.1. *If* $\lim_{x\to\infty} \dfrac{|f(x)|}{|g(x)|} = L$, *where* $L \geq 0$, *then* $f \in O(g)$.

THEOREM 4.2.2. *If* $\lim_{x\to\infty} \dfrac{|f(x)|}{|g(x)|} = \infty$, *then* f *is* **not** $O(g)$ $(f \notin O(g))$.

Discussion

The most basic concept concerning the growth of functions is *big-O*. The statement that f is big-O of g expresses the fact that for large enough x, f will be bounded above by some constant multiple of g. Theorem 4.2.1 gives a necessary condition for f to be big-O of g in terms of limits. The two notions aren't equivalent since there are examples where the definition holds, but the limit fails to exist. For the functions we will be dealing with, however, this will not happen.

When working the problems in the module you may find it helpful to use a graphing calculator or other graphing tool to graph the functions involved. For example, if you graph the functions $x^2 + 10$ and $3x^2$, then you will see that $x^2 + 10 \leq 3x^2$ when $x \geq 3$. (Actually, when $x \geq \sqrt{5}$.) This seems to imply that $f(x) = x^2 + 10$ is big-O of $g(x) = x^2$. This is NOT a proof, but it can give you some ideas as to what to look for. In particular, you wouldn't try to show that $f(x) \leq 3g(x)$ for $x \geq 2$. It isn't necessary that you find the best bound, k, for x, however, as long as you find one that works. Also, there is nothing unique about the choice of C.

EXAMPLE 4.2.1. *Show that* $x^2 + 10$ *is* $O(x^2)$.

Proof 1 (using Definition of Big-O). Let $C = 3$ and $k = 3$. Then, if $x \geq 3$,

$$3x^2 = x^2 + 2x^2 \geq x^2 + 2 \cdot 3^2 \geq x^2 + 10. \qquad \square$$

Proof 2 (using Definition of Big-O). Let $C = 2$ and $k = 4$. Then, if $x \geq 4$,

$$2x^2 = x^2 + x^2 \geq x^2 + 4^2 \geq x^2 + 10. \qquad \square$$

Proof 3 (using Theorem 4.2.1). $\lim_{x\to\infty} \dfrac{x^2 + 10}{x^2} = \lim_{x\to\infty} \left(1 + \dfrac{10}{x^2}\right) = 1 + 0 = 1.$
So, by Theorem 1, $x^2 + 10 \in O(x^2)$. $\square$

EXERCISE 4.2.1. *Let $a, b \in \mathbb{R}^+ - \{1\}$. Prove $\log_a x$ is $O(\log_b x)$. Hint: recall exercise 4.1.1.*

4.3. Proofs of Theorems 4.2.1 and 4.2.2.

Proof of Theorem 4.2.1. Suppose $\lim_{x\to\infty} \frac{|f(x)|}{|g(x)|} = L$, where L is a nonnegative real number. Then, by the definition of limit, we can make $\frac{|f(x)|}{|g(x)|}$ as close to L as we wish by choosing x large enough. In particular, we can ensure that $\frac{|f(x)|}{|g(x)|}$ is within a distance 1 of L by choosing $x \geq k$ for some positive number k. That is, there is a number $k \geq 0$ such that if $x \geq k$, then

$$\left| \frac{|f(x)|}{|g(x)|} - L \right| \leq 1.$$

In particular,

$$\frac{|f(x)|}{|g(x)|} - L \leq 1$$

$$\frac{|f(x)|}{|g(x)|} \leq L + 1$$

$$|f(x)| \leq (L+1)|g(x)|$$

So, we can choose $C = L + 1$. Thus $f \in O(g)$. □

Proof of Theorem 4.2.2. Suppose $\lim_{x\to\infty} \frac{|f(x)|}{|g(x)|} = \infty$. This means that for every positive number C, there is a positive number N such that

$$\frac{|f(x)|}{|g(x)|} > C$$

if $x \geq N$. Thus, for all positive numbers C and k there is an $x \geq k$ (take x greater than the larger of k and N) such that

$$\frac{|f(x)|}{|g(x)|} > C$$

or

$$|f(x)| > C|g(x)|.$$

Thus $f \notin O(g)$. □

Discussion

How do you interpret the statement $f \notin O(g)$? That is, how do you negate the definition? Let's apply principles of logic from Module 2.3. The definition says:

> $f \in O(g)$ if and only if there exist constants C and k such that, for all x, if $x \geq k$, then $|f(x)| \leq C|g(x)|$.

The negation would then read:

> $f \notin O(g)$ if and only if for all constants C and k, there exist x such that $x \geq k$ and $|f(x)| > C|g(x)|$.

EXAMPLE 4.3.1. *Show that x^2 is not $O(x)$.*

Proof 1 (using the Definition of big-O). As we have just seen, the definition requires us to show that no matter how we choose positive constants C and k, there will be a number $x \geq k$ such that $x^2 > Cx$. So, suppose C and k are arbitrary positive constants. Choose x so that $x \geq k$ and $x > C$. Then $x^2 = x \cdot x > C \cdot x$. (We don't have to use the absolute value symbol, since $x > 0$.) □

Proof 2 (using Theorem 4.2.2). $\lim_{x\to\infty} \frac{x^2}{x} = \lim_{x\to\infty} x = \infty$. So, by Theorem 4.2.2, $x^2 \notin O(x)$. □

While it is true that most of the functions f and g that measure complexity have domain $\mathbb{N}$, they are often defined on the set of all positive real numbers, and, as we see, this is where the calculus can come in handy.

4.4. Example 4.4.1.

EXAMPLE 4.4.1. *Show that $2x^3 + x^2 - 3x + 2$ is $O(x^3)$.*

Proof 1 (using the Definition of big-O). By the triangle inequality,

$$\begin{aligned} |2x^3 + x^2 - 3x + 2| &\leq |2x^3| + |x^2| + |3x| + 2 \\ &= 2|x^3| + |x^2| + 3|x| + 2. \end{aligned}$$

Now, if $x \geq 2$, then $x^2 \leq x^3$, $x \leq x^3$, and $2 \leq x^3$.

Thus

$$|2x^3| + |x^2| + |3x| + 2 \leq 2|x^3| + |x^3| + 3|x^3| + |x^3| = 7|x^3|$$

Using these inequalities, $C = 7$, and $k = 2$, we see that f is $O(x^3)$. □

Proof 2 (using Theorem 4.2.2).

$$\lim_{x\to\infty} \frac{2x^3 + x^2 - 3x + 2}{x^3}$$

$$= \lim_{x\to\infty} \frac{2 + 1/x - 3/x^2 + 2/x^3}{1} = \frac{2}{1}$$

By Theorem 4.2.1, $2x^3 + x^2 - 3x + 2$ is $O(x^3)$.

□

Discussion

In the first proof in Example 4.4.1 we used *the triangle inequality*, which is proved in the Appendix at the end of this module. We also need to use the fact $|ab| = |a||b|$.

Notice the strategy employed here. We did not try to decide what C and k were until after using the triangle inequality. The first constant we dealt with was k. After separating the function into the sum of absolute values we thought about what part of this function would be the biggest for large values of x and then thought about how large x needed to be in order for all the terms to be bounded by that largest term. This led to the choice of k. In general, the constant C depends on the choice of k and the two functions you are working with.

EXERCISE 4.4.1. *Use the definition to show that* $5x^3 - 3x^2 + 2x - 8 \in O(x^3)$.

EXERCISE 4.4.2. *Use Theorem 4.2.1 to show that* $10x^3 - 7x^2 + 5 \in O(x^3)$

EXERCISE 4.4.3. *Use Theorem 4.2.2 to show that* $x^5 \notin O(100x^4)$.

4.5. Calculus Definition.

DEFINITION 4.5.1. *If f and g are such that*

$$\lim_{n\to\infty} \frac{f(n)}{g(n)} = 0$$

then we say f is **little-o** *of g, written*

$$f \in o(g).$$

As a corollary to Theorem 4.2.1, we have

THEOREM 4.5.1. *If f is $o(g)$, then f is $O(g)$.*

Discussion

As Theorem 4.5.1 indicates, the *little-o* relation is stronger than big-O. Two of the most important examples of this relation are

(1) $\log_a x \in o(x)$, where a is a positive number different from 1, and
(2) $x^n \in o(a^x)$ if $a > 1$.

These are most easily seen using a version of l'Hôpital's rule from calculus:

l'Hôpital's Rule. If $\lim_{x\to\infty} f(x) = \lim_{x\to\infty} g(x) = \infty$, and if

$$\lim_{x\to\infty} \frac{f'(x)}{g'(x)} = L,$$

then

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = L.$$

(f' and g' denote the derivatives of f and g, respectively.)

EXAMPLE 4.5.1. *Show that* $\log_a x \in o(x)$*, where* a *is a positive number different from 1.*

PROOF. First observe that $\lim_{x\to\infty} \log_a x = \lim_{x\to\infty} x = \infty$. Recall that $\frac{d}{dx} \log_a x = \frac{1}{x \ln a}$, where $\ln x = \log_e x$. By l'Hôpital's rule,

$$\lim_{x\to\infty} \frac{\log_a x}{x} = \lim_{x\to\infty} \frac{\frac{1}{x \ln a}}{1} = 0.$$

□

EXERCISE 4.5.1. *Show that* $(\log_a x)^2 \in o(x)$.

EXAMPLE 4.5.2. *Show that, if* $a > 1$*, then* $x \in o(a^x)$.

PROOF. First observe that $\lim_{x\to\infty} x = \lim_{x\to\infty} a^x = \infty$. By l'Hôpital's rule,

$$\lim_{x\to\infty} \frac{x}{a^x} = \lim_{x\to\infty} \frac{1}{a^x \ln a} = 0,$$

since $a > 1$. □

EXERCISE 4.5.2. *Show that, if* $a > 1$*, then* $x^2 \in o(a^x)$.

EXERCISE 4.5.3. *Use mathematical induction to show that, if $a > 1$, then $x^n \in o(a^x)$ for every positive integer n.*

4.6. Basic Properties of Big-O. The following theorems and facts will be helpful in determining big-O.

THEOREM 4.6.1. *A polynomial of degree n is $O(x^n)$.*

Fact: Theorem 4.6.1. can be extended to functions with non-integral exponents (like $x^{1/2}$).

THEOREM 4.6.2. *If f_1 is $O(g_1)$ and f_2 is $O(g_2)$, then (f_1+f_2) is $O(max\{|g_1|, |g_2|\})$.*

COROLLARY 4.6.2.1. *If f_1 and f_2 are both $O(g)$, then $(f_1 + f_2)$ is $O(g)$.*

THEOREM 4.6.3. *If f_1 is $O(g_1)$ and f_2 is $O(g_2)$, then $(f_1 f_2)$ is $O(g_1 g_2)$.*

THEOREM 4.6.4. *If f_1 is $O(f_2)$ and f_2 is $O(f_3)$, then f_1 is $O(f_3)$.*

THEOREM 4.6.5. *If f is $O(g)$, then (af) is $O(g)$ for any constant a.*

Discussion

Use these theorems when working the homework problems for this module.

EXAMPLE 4.6.1. *Find the least integer n such that $(x^4 + 5\log x)/(x^3 + 1)$ is $O(x^n)$*

Solution: First we consider $\frac{x^4}{x^3+1}$. If you think back to calculus and consider which part of this function "takes over" when x gets large, that provides the clue that this function should be $O(x)$. To see this, we take the following limit;

$$\lim_{x\to\infty} \frac{(x^4)/(x^3+1)}{x} = \lim_{x\to\infty} \frac{x^3}{x^3+1} = 1.$$

Since that limit is 1, we have verified $\frac{x^4}{x^3+1}$ is $O(x)$. Theorem 4.2.2 can be used to show that $\frac{x^4}{x^3+1}$ is not $O(x^0) = O(1)$:

$$\lim_{x\to\infty} \frac{(x^4)/(x^3+1)}{1} = \lim_{x\to\infty} \frac{x}{1+1/x^3} = \infty.$$

Now consider $\frac{5\log x}{x^3+1}$. Since $\log x$ is $O(x)$, $\frac{5\log x}{x^3+1}$ is $O(\frac{5x}{x^3+1})$, and, by taking a limit as above, $\frac{5x}{x^3+1}$ is $o(x)$, hence, $O(x)$.

Since the original function is the sum of the two functions, each of which is $O(x)$, the sum $(x^4 + 5\log x)/(x^3 + 1)$ is $O(x)$, by Corollary 4.6.2.1.

4.7. Proof of Theorem 4.6.3.

PROOF OF THEOREM 4.6.3. Suppose f_1, f_2, g_1, g_2 are all functions with domain and codomain $\mathbb{R}$ such that f_1 is $O(g_1)$ and f_2 is $O(g_2)$.

Then by definition of big-O, there are positive constants C_1, k_1, C_2, k_2 such that

$$\forall x \geq k_1[|f_1(x)| \leq C_1|g_1(x)|] \text{ and } \forall x \geq k_2[|f_2(x)| \leq C_2|g_2(x)|].$$

Let $k = max\{k_1, k_2\}$ and $C = C_1C_2$. Then if $x \geq k$ we have

$$\begin{aligned} |(f_1f_2)(x)| &= |f_1(x)| \cdot |f_2(x)| \\ &\leq C_1|g_1(x) \cdot C_2|g_2(x)| \\ &= C_1C_2|(g_1g_2)(x)| \\ &= C|(g_1g_2)(x)| \end{aligned}$$

This shows f_1f_2 is $O(g_1g_2)$. □

4.8. Example 4.8.1.

EXAMPLE 4.8.1. *Suppose there are two computer algorithms such that*

- *Algorithm 1 has complexity $n^2 - n + 1$, and*
- *Algorithm 2 has complexity $n^2/2 + 3n + 2$.*

Then both are $O(n^2)$, but to indicate Algorithm 2 has a smaller leading coefficient, and hence would be faster, we write

- *Algorithm 1 has complexity $n^2 + O(n)$, and*
- *Algorithm 2 has complexity $n^2/2 + O(n)$.*

Discussion

Example 4.8.1 illustrates the way in which the big-O notation may be used to discuss complexity of algorithms.

4.9. Big-Omega.

DEFINITION 4.9.1. *f is* **big-Omega** *of g, written $f \in \Omega(g)$, if there are positive constants C and k such that*

$$|f(x)| \geq C|g(x)| \qquad \text{for } x > k.$$

Big-Omega is very similar to big-O. Big-Ω notation is used to indicate a lower bound on functions for large values of the independent variable. Notice that f is $\Omega(g)$ if and only if g is $O(f)$. Using this fact we see the properties for big-O give similar properties for big-Ω.

EXAMPLE 4.9.1. *x is $\Omega(\log x)$.*

EXAMPLE 4.9.2. *$2x^3 + x^2 - 3x + 2$ is $\Omega(x^3)$.*

PROOF USING THE DEFINITION OF BIG-Ω: Let $x \geq 3$. Then $x^2 - 3x \geq 0$ and so $x^2 - 3x + 2 \geq 0$ as well. Thus

$$|2x^3 + x^2 - 3x + 2| = 2x^3 + x^2 - 3x + 2 \geq 2x^3.$$

By choosing $C = 2$ and $k = 3$ in the definition of big-Ω the above work shows $2x^3 + x^2 - 3x + 2$ is $\Omega(x^3)$. □

EXERCISE 4.9.1. *Let $a, b \in \mathbb{R}^+ - \{1\}$. Prove $\log_a x$ is $\Omega(\log_b x)$.*

4.10. Big-Theta.

DEFINITION 4.10.1. *f is* **big-Theta** *of g, written $f \in \Theta(g)$, if f is both $O(g)$ and $\Omega(g)$.*

Discussion

The definition given for big-Θ is equivalent to the following:

THEOREM 4.10.1. *f is $\Theta(g)$ if and only if f is $O(g)$ and g is $O(f)$.*

EXERCISE 4.10.1. *Prove Theorem 4.10.1.*

EXAMPLE 4.10.1. $(2x^2-3)/(3x^4+x^3-2x^2-1)$ *is* $\Theta(x^{-2})$.

$$\begin{aligned}\frac{(2x^2-3)/(3x^4+x^3-2x^2-1)}{x^{-2}} &= \frac{2x^2-3}{3x^4+x^3-2x^2-1}\cdot x^2\\ &= \frac{2x^4-3x^2}{3x^4+x^3-2x^2-1}\\ &= \frac{2-3/x^2}{3+1/x-2/x^2-1/x^4}\end{aligned}$$

$$\lim_{x\to\infty}\frac{(2x^2-3)/(3x^4+x^3-2x^2-1)}{x^{-2}} = \frac{2}{3}$$

You now will show through the following exercise that any two logarithm functions have the same growth rate; hence, it doesn't matter what (acceptable) base is used.

EXERCISE 4.10.2. *If a and b are positive real numbers different from 1, show that* $\log_a x \in \Theta(\log_b x)$.

4.11. Summary. Suppose f and g are functions such that $\lim_{x\to\infty}\dfrac{|f(x)|}{|g(x)|} = L$, where $0 \le L \le \infty$.

1. If $L = 0$, then f is $o(g)$ (hence, $O(g)$), and g is $\Omega(f)$ (hence, not $O(f)$).
2. If $L = \infty$, then f is $\Omega(g)$ (hence, not $O(g)$), and g is $o(f)$ (hence, $O(f)$).
3. If $0 < L < \infty$, then f is $\Theta(g)$ (hence, $O(g)$), and g is $\Theta(f)$ (hence, $O(f)$).

4.12. Appendix. Proof of the Triangle Inequality. Recall the triangle inequality: for all real numbers a and b,

$$|a+b| \leq |a| + |b|.$$

PROOF. Recall from Module 1.2 that the absolute value function $f(x) = |x|$ is defined by

$$f(x) = |x| = \begin{cases} x, & \text{if } x \geq 0, \\ -x, & \text{if } x < 0. \end{cases}$$

We first observe that for any real numbers x and y, if $y \geq 0$, then $|x| \leq y$ if and only if $-y \leq x \leq y$. To see this, look at two cases:

Case 1. $x \geq 0$. Then $|x| = x$, and so $|x| \leq y$ if and only if $-y \leq 0 \leq x \leq y$, or $-y \leq x \leq y$.

Case 2. $x < 0$. Then $|x| = -x$, and so $|x| \leq y$ if and only if $-y \leq 0 \leq -x \leq y$. Multiplying through by -1 and reversing the inequalities, we get $y \geq x \geq -y$, or $-y \leq x \leq y$.

We now prove the triangle inequality. For arbitrary real numbers a and b, apply the above to $x = a$ and $y = |a|$, and then to $x = b$ and $y = |b|$, to get inequalities

$$-|a| \leq a \leq |a|$$
$$-|b| \leq b \leq |b|.$$

Then

$$-|a| - |b| \leq a + b \leq |a| + |b|$$

or

$$-(|a| + |b|) \leq a + b \leq |a| + |b|.$$

Now apply the assertion above to $x = a + b$ and $y = |a| + |b|$ to get:

$$|a+b| \leq |a| + |b|.$$

□

CHAPTER 5

Number Theory

1. Integers and Division

1.1. Divisibility.

DEFINITION 1.1.1. *Given two integers a and b, with $a \neq 0$, we say a* **divides** *b if there is an integer c such that $b = ac$. If a divides b, we write $a|b$. If a does not divide b, we write $a \nmid b$.*

Discussion

EXAMPLE 1.1.1. *The number 6 is divisible by 3, $3|6$, since $6 = 3 \cdot 2$.*

EXERCISE 1.1.1. *Let a, b, and c be integers with $a \neq 0$. Prove that if $ab|ac$, then $b|c$.*

Using this definition, we may define an integer to be *even* if it is divisible by 2 and *odd* if it is not divisible by 2. This concept is one of the simplest of properties of numbers to define, yet it is among the most complicated of all mathematical ideas. Keep in mind that we are talking about a very restricted notion of what it means for one number to "divide" another: we can certainly divide 7 by 3 and get the rational number $\frac{7}{3} = 2.3333\cdots$, but, since the result is not an integer, we say that 3 does not divide 7, or $3 \nmid 7$. For this reason, you should *avoid using fractions* in any discussion of integers and integer arithmetic.

1.2. Basic Properties of Divisibility.

THEOREM 1.2.1. *For all integers a, b, and c,*

1. *If $a|b$ and $a|c$, then $a|(b + c)$.*
2. *If $a|b$, then $a|(bc)$.*
3. *If $a|b$ and $b|c$, then $a|c$.*

Discussion

Theorem 1.2.1 states the most basic properties of division. Here is the proof of part 3:

Proof of part 3. Assume a, b, and c are integers such that $a|b$ and $b|c$. Then by definition, there must be integers m and n such that $b = am$ and $c = bn$. Thus

$$c = bn = (am)n = a(mn).$$

Since the product of two integers is again an integer, we have $a|c$. □

EXERCISE 1.2.1. *Prove part 1 of Theorem 1.2.1.*

EXERCISE 1.2.2. *Prove part 2 of Theorem 1.2.1.*

1.3. Theorem 1.3.1 - The Division Algorithm.

THEOREM 1.3.1. (Division Algorithm) *Given integers a and d, with $d > 0$, there exists unique integers q and r, with $0 \leq r < d$, such that $a = qd + r$.*

NOTATION 1.3.1. *We call a the* **dividend***, d the* **divisor***, q the* **quotient***, and r the* **remainder***.*

Discussion

The division algorithm is probably one of the first concepts you learned relative to the operation of division. It is not actually an algorithm, but this is this theorem's traditional name. For example, if we divide 26 by 3, then we get a quotient of 8 and remainder or 2. This can be expressed $26 = 3 \cdot 8 + 2$. It is a little trickier to see what q and r should be if $a < 0$. For example, if we divide -26 is by 3, then the remainder is *not* -2. We can, however, use the equation $26 = 3 \cdot 8 + 2$ to our advantage:

$$-26 = 3 \cdot (-8) - 2 = [3 \cdot (-8) - 3] - 2 + 3 = 3(-9) + 1$$

So dividing -26 by 3 gives a quotient of -9 and remainder 1. The condition $0 \leq r < d$ makes r and q unique for any given a and d.

1.4. Proof of Division Algorithm. Proof. Suppose a and d are integers, and $d > 0$. We will use the well-ordering principle to obtain the quotient q and remainder r. Since we can take $q = a$ if $d = 1$, we shall assume that $d > 1$.

Let S be the set of all **natural numbers** of the form $a - kd$, where k is an integer. In symbols

$$S = \{a - kd | k \in \mathbf{Z} \text{ and } a - kd \geq 0\}.$$

If we can show that S is nonempty, then the well-ordering principle will give us a least element of S, and this will be the remainder r we are looking for. There are two cases.

Case 1: $a \geq 0$. In this case, we can set $k = 0$ and get the element $a - 0 \cdot k = a \geq 0$ of S.

Case 2: $a < 0$. In this case, we can set $k = a$. Then $a - kd = a - ad = a(1 - d)$. Since $a < 0$ and $d > 1$, $a(1 - d) > 0$; hence is an element of S.

Thus, $S \neq \emptyset$, and so S has a least element $r = a - qd$ for some integer q. Thus, $a = qd + r$ and $r \geq 0$. We are left to show (i) $r < d$ and (ii) q and r are unique.

(i) Suppose $r \geq d$. Then $r = d + r'$, where $0 \leq r' < r$. Then $a = qd + r = qd + d + r' = (q + 1)d + r'$,
so that $r' = a - (q + 1)d$ is an element of S smaller than r. This contradicts the fact that r is the least element of S. Thus, $r < d$.

(ii) Suppose integers q' and r' satsify $a = q'd + r'$ and $0 \leq r' < d$. Without loss of generality, we may assume $r' \geq r$, so that $0 \leq r - r' < d$. Since $a = q'd + r' = qd + r$,

$$r - r' = d(q' - q).$$

This means that d divides $r - r'$, which implies either $r - r' \geq d$ or $r - r' = 0$. But but we know $0 \leq r - r' < d$. Thus, $r' = r$, which, in turn, implies $q' = q$. That is, q and r are unique.

1.5. Prime Numbers, Composites.

DEFINITION 1.5.1. *If p is an integer greater than 1, then p is a* **prime number** *if the only divisors of p are 1 and p.*

DEFINITION 1.5.2. *A positive integer greater than 1 that is not a prime number is called* **composite**.

Discussion

Prime numbers are the building blocks of arithmetic. At the moment there are no efficient methods (algorithms) known that will determine whether a given integer is prime or find its prime factors. This fact is the basis behind many of the cryptosystems currently in use. One problem is that there is no known procedure that will generate prime numbers, even recursively. In fact, there are many things about prime numbers that we don't know. For example, there is a conjecture, known as Goldbach's Conjecture, that there are infinitely many *prime pairs*, that is, consecutive odd prime numbers, such as 5 and 7, or 41 and 43, which no one so far has been able to prove or disprove. As the next theorem illustrates, it is possible, however, to prove that there are infinitely many prime numbers. Its proof, attributed to Euclid, is one of the most elegant in all of mathematics.

THEOREM 1.5.1. *There are infinitely many prime numbers.*

PROOF. We prove the theorem by contradiction. Suppose there are only finitely many prime numbers, say, $p_1, p_2, ..., p_n$. Let

$$N = p_1 p_2 \cdots p_n + 1.$$

Then N is an integer greater than each of $p_1, p_2, ..., p_n$, so N cannot be prime. In Example 9, Module 3.3, we showed that N can be written as a product of prime numbers; hence, some prime p divides N. We may assume, by reordering $p_1, p_2, ..., p_n$, if necessary, that $p = p_1$. Thus $N = p_1 a$ for some integer a. Substituting, we get

$$p_1 a = p_1 p_2 \cdots p_n + 1$$

$$p_1 a - p_1 p_2 \cdots p_n = 1$$

$$p_1(a - p_2 \cdots p_n) = 1.$$

Thus, $a - p_2 \cdots p_n$ is a positive integer. Since p_1 is a prime number, $p_1 > 1$, and so

$$p_1(a - p_2 \cdots p_n) > 1.$$

But this contradicts the equality above. □

1.6. Fundamental Theorem of Arithmetic.

THEOREM 1.6.1. (Fundamental Theorem of Arithmetic) *Every positive integer greater than one can be written uniquely as a product of primes, where the prime factors are written in nondecreasing order.*

Discussion

We have already given part of the proof Theorem 1.6.1 in an example of *Module 3.3 Induction*. There we showed that every positive integer greater than 1 can be written as a product of prime numbers. The *uniqueness* of the factors is important, and the proof that they are unique, which requires a few additional ideas, will be postponed until the next module.

The prime factorization of 140 is $2 \cdot 2 \cdot 5 \cdot 7$. You can see one reason why we do not want 1 to be prime: There is no limit to the number of times 1 may be repeated as a factor, and that would give us non-unique prime factorizations.

1.7. Factoring.

THEOREM 1.7.1. *If n is a composite integer, then n has a factor less that or equal to $\sqrt{n}$.*

Discussion

Theorem 1.7.1 can be helpful in narrowing down the list of possible prime factors of a number. It was proved in an example of *Module 3.2 Methods of Proof* and exploited in another example of that module. If the number 253 is composite, for example, it must have a factor less than or equal to 15. Thus we need only check the primes 2, 3, 5, 7, 11, and 13. It turns out $253 = 11 \cdot 23$.

1.8. Mersenne Primes.

DEFINITION 1.8.1. *A prime number of the form $2^p - 1$, where p is a prime number, is called a* **Mersenne prime**.

Discussion

Mersenne primes are a special class of primes, which lend themselves to a nice theoretical development. Not all primes are Mersenne, though, and not all numbers of the form $2^p - 1$ are prime. For example, $2^p - 1$ is prime for $p = 2$, 3, 5, and 7, but $2^{11} - 1 = 2047 = 23 \cdot 89$, which is not prime. On the other hand, the primes 5 and 11 cannot be written in this form.

1.9. Greatest Common Divisor and Least Common Multiple.

DEFINITIONS 1.9.1. *Given integers a and b*

(1) *The* **greatest common divisor** *of a and b, denoted* GCD (a, b)*, is the largest positive integer d such that $d|a$ and $d|b$.*
(2) *The* **least common multiple** *of a and b, denoted* LCM (a, b)*, is the smallest positive integer m such that $a|m$ and $b|m$.*
(3) *a and b are called* **relatively prime** *if* GCD $(a, b) = 1$.
(4) *The integers $a_1, a_2, a_3, \ldots, a_n$ are called* **pairwise relatively prime** *if* $\text{GCD}(a_i, a_j) = 1$ *for $1 \leq i < j \leq n$.*
(5) *The* **Euler ϕ function** *is the function $\phi : \mathbb{Z}^+ \to \mathbb{N}$ defined by $\phi(n) =$ the number of positive integers less than n that are relatively prime to n.*

LEMMA 1.9.1. *Suppose a and b are integers and $m = \text{LCM}(a, b)$. If c is a positive integer such that $a|c$ and $b|c$, then $m|c$.*

PROOF. Suppose $a|c$ and $b|c$, but $m \nmid c$. By the division algorithm there are (unique) positive integers q and r such that $c = mq + r$ and $0 \leq r < m$. Since $m \nmid c$, $r \neq 0$; that is, $r > 0$. Write $r = c - mq$. Since a and b both divide c and m, a and b both divide r. But this contradicts the fact that m is supposed to be the least positive integer with this property. Thus $m|c$. □

THEOREM 1.9.1. $ab = \text{GCD}(a, b) \cdot \text{LCM}(a, b)$.

Discussion

The proof of Theorem 1.9.1 will be discussed in the next module.

EXAMPLE 1.9.1. *Here are some examples to illustrate the definitions above.*

(1) $\text{GCD}(45, 60) = 15$*, since* $45 = 15 \cdot 3$ *and* $60 = 15 \cdot 4$ *and 15 is the largest number that divides both 45 and 60.*
(2) $\text{LCM}(45, 60) = 180$*, since* $180 = 45 \cdot 4 = 60 \cdot 3$ *and 180 is the smallest number that both 45 and 60 divide.*
(3) 45 and 60 are not relatively prime.
(4) 45 and 16 are relatively prime since $\text{GCD}(45, 16) = 1$.
(5) 4, 7, 13 and 55 are pairwise relatively prime.
(6) $\phi(15) = 8$

If we are given the prime factorizations of two integers, then it is easy to find their GCD and LCM. For example, $600 = 2^3 \cdot 3 \cdot 5^2$ and $220 = 2^2 \cdot 5 \cdot 11$ has greatest common divisor $2^2 \cdot 5 = 20$ and least common multiple $2^3 \cdot 3 \cdot 5^2 \cdot 11 = 6600$. Since prime factorizations can be difficult to find, however, this idea does not lead to an efficient way to compute GCD's. We will introduce an efficient algorithm in the next module that does not involve knowledge about prime factorizations.

EXERCISE 1.9.1. *Let* $F(n)$ *denote the n-th term of the Fibonacci Sequence. Prove using induction that* $\text{GCD}(F(n), F(n-1)) = 1$ *for all integers* $n \geq 2$.

1.10. Modular Arithmetic.

DEFINITION 1.10.1. *Given integers a and m, with* $m > 0$*,*
a **mod** m *is defined to be the remainder when a is divided by m.*

DEFINITION 1.10.2. $a \equiv b (mod\ m)$*, read "a is congruent to b modulo (or mod) m," if*
$m|(a - b)$*; that is,* $(a - b)$ **mod** $m = 0$.

THEOREM 1.10.1. *Given integers a, b, and m,*

1. $a \equiv b\ (mod\ m)$ *if and only if* a **mod** $m = b$ **mod** m.
2. $a \equiv b\ (mod\ m)$ *if and only if* $a = b + km$ *for some integer* k.
3. *If* $a \equiv b\ (mod\ m)$ *and* $c \equiv d\ (mod\ m)$ *then*
   (a) $a + c \equiv b + d\ (mod\ m)$
   (b) $a \cdot c \equiv b \cdot d\ (mod\ m)$

Discussion

The **mod** operation is derived from the Division Algorithm: If we divide the integer a by the positive integer m, we get a unique quotient q and remainder r satisfying $a = mq + r$ and $0 \leq r < m$. The remainder r is *defined* to be the value of a **mod** m. One of the notational aspects that may seem a little unusual is that we write $a + b (\text{mod } m)$ for $(a + b)(\text{mod } m)$. Also, the symbol $(\text{mod } m)$ may occasionally be omitted when it is understood.

EXAMPLE 1.10.1. *Here are some examples.*

(a) 12 **mod** $5 = 2$
(b) 139 **mod** $5 = 4$
(c) 1142 **mod** $5 = 2$
(d) $1142 \equiv 12 \equiv 2 (mod\ 5)$
(e) $1142 + 139 \equiv 2 + 4 \equiv 6 \equiv 1 (mod\ 5)$
(f) $1142 \cdot 139 \equiv 2 \cdot 4 \equiv 8 \equiv 3 (mod\ 5)$

One of the differences to note between the concept of congruence modulo m verses the **mod** operator is that an integer, k may be *congruent* to infinitely many other integers modulo m, however, k **mod** m is equal to one single integer. For example, 139 **mod** $5 = 4$, but 139 is *congruent* to all the elements of $\{\ldots, -6, -1, 4, 9, 14, 19, \ldots\}$.

EXERCISE 1.10.1. *Given a positive integer m, prove that the assignment $a \mapsto a$* **mod** *m defines a function $f : \mathbf{Z} \to \mathbf{Z}$. Is f one-to-one? onto? What is its range?*

$a \mapsto a$ **mod** *m is another way to write $f(a) = a$* **mod** *m.*

Here is a proof of part 3b of Theorem 1.10.1:

Proof of 3b. Since $a \equiv b (\text{mod } m)$ and $c \equiv d (\text{mod } m)$, there must be integers s and t such that $b = a + sm$ and $d = c + tm$ (part 2). Thus

$$\begin{aligned} bd &= (a + sm)(c + tm) \\ &= ac + atm + smc + stm^2 \\ &= ac + (at + sc + stm)m \end{aligned}$$

Since a, c, t, and s are all integers, $at + sc + st$ is as well. Thus, by part 2,

$$ac \equiv bd (\text{mod } m).$$

□

EXERCISE 1.10.2. *Prove part 3a of Theorem 1.10.1.*

1.11. Applications of Modular Arithmetic.

1. Hashing Functions
2. Pseudorandom Number Generators
3. Cryptology

Discussion

There are many applications of modular arithmetic in computer science. One such application is in the construction of pseudorandom number generators. Numbers that seem to be somewhat random may be produced using the **linear congruential method**. As you will see, it does not produce truly random numbers, but rather a sequence of numbers that will eventually repeat.

To generate a sequence we choose a **modulus** m, **multiplier** a, and an **increment** c. Then we start with a *seed* number x_0 and then construct a sequence of numbers recursively using the formula

$$x_{n+1} = (ax_n + c) \textbf{ mod } m.$$

EXAMPLE 1.11.1. *Suppose we choose* $m = 11$, $a = 7$, $c = 3$, *and* $x_0 = 1$. *Then we get*

$x_0 = 1$

$x_1 = (7 \cdot 1 + 3)\textbf{mod } 11 = 10$

$x_2 = (7 \cdot 10 + 3)\textbf{mod } 11 = 7$

$x_3 = (7 \cdot 7 + 3)\textbf{mod } 11 = 8$

$x_4 = (7 \cdot 10 + 8)\textbf{mod } 11 = 4$

$x_5 = (7 \cdot 4 + 3)\textbf{mod } 11 = 9$

$x_6 = (7 \cdot 9 + 3)\textbf{mod } 11 = 0$

$x_7 = (7 \cdot 0 + 3)\textbf{mod } 11 = 3$

etc.

The sequence will be 1, 10, 7, 8, 4, 9, 0, 3, etc. If we wanted a "random" sequence of bits, 0 and 1, we could then reduce each x_n **mod** *2. In practice, large Mersenne primes are often chosen for the modulus, and the repetition period for such sequences can be made to be quite large.*

EXERCISE 1.11.1. *Prove that for a given modulus m, and arbitrary multiplier a, increment c, and seed x_0, the sequence $x_0, x_1, x_2, ...$ must eventually repeat.*

2. Integers and Algorithms

2.1. Euclidean Algorithm. Euclidean Algorithm. Suppose a and b are integers with $a \geq b > 0$.

(1) Apply the division algorithm: $a = bq + r$, $0 \leq r < b$.
(2) Rename b as a and r as b and repeat until $r = 0$.

The last nonzero remainder is the greatest common divisor of a and b.

The Euclidean Algorithm depends upon the following lemma.

LEMMA 2.1.1. *If* $a = bq + r$, *then* $\mathrm{GCD}(a, b) = \mathrm{GCD}(b, r)$.

PROOF. We will show that if $a = bq + r$, then an integer d is a common divisor of a and b if, and only if, d is a common divisor of b and r.

Let d be a common divisor of a and b. Then $d|a$ and $d|b$. Thus $d|(a - bq)$, which means $d|r$, since $r = a - bq$. Thus d is a common divisor of b and r.

Now suppose d is a common divisor of b and r. Then $d|b$ and $d|r$. Thus $d|(bq + r)$, so $d|a$. Therefore, d must be a common divisor of a and b.

Thus, the set of common divisors of a and b are the same as the set of common divisors of b and r. It follows that d is the *greatest* common divisor of a and b if and only if d is the greatest common divisor of b and r. □

Discussion

The fact that the Euclidean algorithm actually gives the greatest common divisor of two integers follows from the division algorithm and the equality in Lemma 2.1.1. Applying the division algorithm repeatedly as indicated yields a sequence of remainders $r_1 > r_2 > \cdots > r_n > 0 = r_{n+1}$, where $r_1 < b$. Lemma 2.1.1 says that

$$\mathrm{GCD}(a, b) = \mathrm{GCD}(b, r_1) = \mathrm{GCD}(r_1, r_2) = \cdots = \mathrm{GCD}(r_{n-1}, r_n).$$

But, since $r_{n+1} = 0$, r_n divides r_{n-1}, so that

$$\mathrm{GCD}(r_{n-1}, r_n) = r_n.$$

Thus, the last nonzero remainder is the greatest common divisor of a and b.

EXAMPLE 2.1.1. *Find* GCD *(1317, 56).*

$$\begin{aligned} 1317 &= 56(23) + 29 \\ 56 &= 29(1) + 27 \\ 29 &= 27(1) + 2 \\ 27 &= 2(13) + 1 \\ 2 &= 1(2) + 0 \end{aligned}$$

GCD *(1317,56)=1*

Example 2.1.1 shows how to apply the Euclidean algorithm. Notice that when you proceed from one step to the next you make the new dividend the old divisor (replace a with b) and the new divisor becomes the old remainder (replace b with r). Recall that you can find the quotient q by dividing b into a on your calculator and rounding *down* to the nearest integer. (That is, $q = \lfloor a/b \rfloor$.) You can then solve for r. Alternatively, if your calculator has a **mod** operation, then $r = \mathbf{mod}(a, b)$ and $q = (a - r)/b$. Since you only need to know the remainders to find the greatest common divisor, you can proceed to find them recursively as follows:

Basis. $r_1 = a \,\mathbf{mod}\, b$, $r_2 = b \,\mathbf{mod}\, r_1$.

Recursion. $r_{k+1} = r_{k-1} \,\mathbf{mod}\, r_k$, for $k \geq 2$. (Continue until $r_{n+1} = 0$ for some n.)

2.2. GCD's and Linear Combinations.

THEOREM 2.2.1. *If $d = \mathrm{GCD}(a, b)$, then there are integers s and t such that*

$$d = as + bt.$$

Moreover, d is the smallest positive integer that can be expressed this way.

Discussion

Theorem 2.2.1 gives one of the most useful characterizations of the greatest common divisor of two integers. Given integers a and b, the expression $as + bt$, where s and t are also integers, is called a **linear combination** of a and b.

EXERCISE 2.2.1. *Prove that if a, b, s, t, and d are integers such that $d|a$ and $d|b$, then $d|(as + bt)$.*

The Euclidean Algorithm can, in fact, be used to provide the representation of the greatest common divisor of a and b as a linear combination of a and b. Here is how it would work for the example in Example 2.1.1.

EXAMPLE 2.2.1. *Express* $1 = \text{GCD}(1317, 56)$ *as a linear combination of 1317 and 56.*

Solution: We work backwards using the equations derived by applying the Euclidean algorithm in example 2.1.1, expressing each remainder as a linear combination of the associated divisor and dividend:

$$\begin{aligned}
1 &= \underline{27} - 13 \cdot \underline{2} && \textit{linear combination of 2 and 27}\\
1 &= \underline{27} - 13(\underline{29} - \underline{27} \cdot 1) && \textit{substitute } \underline{2} = \underline{29} - \underline{27}(1)\\
1 &= 14 \cdot \underline{27} - 13 \cdot \underline{29} && \textit{linear combination of 27 and 29}\\
1 &= 14(\underline{56} - 1 \cdot \underline{29}) - 13 \cdot \underline{29} && \textit{substitute } \underline{27} = \underline{56} - 1 \cdot \underline{29}\\
1 &= 14 \cdot \underline{56} - 27 \cdot \underline{29} && \textit{linear combination of 29 and 56}\\
1 &= 14 \cdot \underline{56} - 27(\underline{1317} - 23 \cdot \underline{56}) && \textit{substitute } \underline{29} = \underline{1317} - 23 \cdot \underline{56}\\
1 &= 635 \cdot \underline{56} - 27 \cdot \underline{1317} && \textit{linear combination of 56 and 1317}
\end{aligned}$$

(The dividends, divisors, and remainders have been underlined.)

So $\text{GCD}(1317, 56) = 1 = 1317(-27) + 56(635)$.

Theorem 2.2.1 can be proved by mathematical induction following the idea in the preceding example.

Proof of Theorem 2.2.1. Suppose a and b are integers. We may assume a and b are positive, since $\text{GCD}(a, b) = \text{GCD}(\pm a, \pm b)$. The Euclidean algorithm uses the division algorithm to produce a sequence of quotients and remainders as follows:

$$\begin{aligned}
a &= bq_1 + r_1\\
b &= r_1q_2 + r_2\\
r_1 &= r_2q_3 + r_3\\
&\vdots\\
r_{n-2} &= r_{n-1}q_n + r_n\\
r_{n-1} &= r_nq_{n+1} + 0
\end{aligned}$$

where r_n is the last nonzero remainder. We will use the second principle of mathematical induction to prove that r_k is a linear combination of a and b for $1 \leq k \leq n$.

1. Basis Step ($k = 1$). $r_1 = a - bq_1 = a(1) + b(-q_1)$.
2. Induction Step. Suppose r_i is a linear combination of a and b for $1 \leq i \leq k$. For $1 \leq k \leq n$ we have
$$r_{k+1} = r_{k-1} - r_kq_{k+1}$$

(where $r_0 = b$ when $k = 1$). By the inductive hypothesis r_{k-1} and r_k are linear combinations of a and b. (This works for $k = 1$, since $r_0 = b$ is trivially a linear combination of a and b.) Write

$$r_{k-1} = as_1 + bt_1$$

and

$$r_k = as_2 + bt_2$$

for integers s_1, t_1, s_2, t_2, and substitute into the equation above:

$$r_{k+1} = (as_1 + bt_1) - (as_2 + bt_2)q_{k+1} = a(s_1 - s_2q_{k+1}) + b(t_1 - t_2q_{k+1}).$$

Thus, r_{k+1} is a linear combination of a and b. By the second principle of mathematical induction, r_n is a linear combination of a and b. But r_n is the greatest common divisor of a and b. This proves the first part of the theorem.

Next, we show that d is the smallest positive integer expressible as a linear combination of a and b. Suppose a positive integer c can be expressed as a linear combination of a and b:

$$c = ax + by$$

for integers x and y. Since $d|a$ and $d|b$, then $d|c$, which implies $d \leq c$. □

Here is an alternative proof of Theorem 2.2.1 that does not use the Euclidean algorithm.

Second proof of Theorem 2.2.1. Let S be the set of all positive integers that can be expressed as a linear combination of the positive integers a and b. Clearly $S \neq \emptyset$, since $a, b \in S$. By the well-ordering principle S has a least element d. We will prove by contradiction that $d|a$ and $d|b$.

If $d \nmid a$, then use the division algorithm to get integers q and r such that

$$a = dq + r,$$

where $0 < r < d$. Since both a and d are linear combinations of a and b, then $r = a - dq$ is also. But this means that $r \in S$, contradicting the fact that d is the smallest member of S.

Similarly, one shows that $d|b$.

If c is a divisor of a and b, then c divides any linear combination of a and b; hence, $c|d$. Thus, $d = \text{GCD}(a, b)$. □

EXERCISE 2.2.2. *Prove that if p is a prime number and n is an integer that is not divisible by p, then there are integers s and t such that $ps + nt = 1$. [First show that* $\text{GCD}(p, n) = 1$.*]*

EXERCISE 2.2.3. *Prove that if 1 is a linear combination of a and b, then* $\mathrm{GCD}(a, b) = 1$.

2.3. Uniqueness of Prime Factorization.

LEMMA 2.3.1. *If* $\mathrm{GCD}(a, b) = 1$ *and* $a|bc$, *then* $a|c$.

PROOF. Write $1 = as + bt$ for integers s and t. Multiply both sides by c:

$$c = acs + bct.$$

Since $a|bc$, a divides this linear combination

$$a(cs) + (bc)t = c$$

of a and bc.

□

THEOREM 2.3.1. *Suppose a and b are integers and p is a prime number. If $p|ab$, then either $p|a$ or $p|b$.*

PROOF. We will prove the equivalent statement: if $p|ab$ and $p \nmid a$, then $p|b$. (You should convince yourself that the two propositional forms $P \to (Q \vee R)$ and $(P \wedge \neg Q) \to R$ are equivalent.)

Suppose $p|ab$ and $p \nmid a$. Then $\mathrm{GCD}(p, a) = 1$. By the Lemma 1, $p|b$. □

Discussion

Theorem 2.3.1 is very useful in deciding how prime factors are distributed in a product of two integers. For example, we gave an indirect proof in Module 3.2 that if the product of two integers x and y is even, then either x is even or y is even. As we hinted there, a direct proof is possible, and Theorem 2.3.1 provides just the right information to make it work.

EXERCISE 2.3.1. *Use Theorem 2.3.1 to give a direct proof that if the product of two integers x and y is even, then either x is even or y is even.*

EXERCISE 2.3.2. *Use mathematical induction to prove the following generalization of Theorem 2.3.1. Suppose $a_1, a_2, ..., a_n$ are integers and p is a prime number. If $p|a_1a_2 \cdots a_n$, then $p|a_i$ for some $i = 1, 2, ..., n$. [Hint: The induction step has two cases.]*

EXERCISE 2.3.3. *Use Lemma 2.3.1 to prove that if k, ℓ, and m are positive integers such that $k|m$, $\ell|m$, and k and ℓ are relatively prime, then the product $k\ell|m$.*

EXERCISE 2.3.4. *Suppose a and b are positive integers, $d = \mathrm{GCD}(a, b)$, $a = dk$, and $b = d\ell$. Prove that k and ℓ are relatively prime. [Hint: Show that 1 can be expressed as a linear combination of k and ℓ.]*

We can now give a proof of Theorem 6 of *Module 5.1 Integers and Division*: If a and b are positive integers, then $ab = \mathrm{GCD}(a, b) \cdot \mathrm{LCM}(a, b)$.

Proof of Theorem 6, Module 5.1. Let $d = \mathrm{GCD}(a, b)$. Write $a = dk$, $b = d\ell$, where k and ℓ are positive integers, which, by Exercise 2.3.4, are relatively prime. Then

$$ab = (dk)(d\ell) = d \cdot (k\ell d) = \mathrm{GCD}(a, b) \cdot (k\ell d).$$

We will succeed once we show that $k\ell d = \mathrm{LCM}(a, b)$. We will prove this by contradiction.

Suppose $m = \mathrm{LCM}(a, b)$ and $m < k\ell d$. Observe that $k\ell d = (dk)\ell = a\ell$ and $k\ell d = (d\ell)k = bk$. That is, both a and b divide $k\ell d$; hence, their least common multiple m does also.

Since $k|a$ and $\ell|b$, k and ℓ both divide m; hence, by Exercise 2.3.3, the product $k\ell|m$.

> [Aside: We also know that d divides m, so it is tempting to assert that $k\ell d$ also divides m. But we can't use Exercise 2.3.3 to conclude this, since d may not be relatively prime to either k or ℓ. Can you give an example where d divides both k and ℓ?]

Thus $m = k\ell x$ for some positive integer x, and $x < d$, by hypothesis. Since $m|k\ell d$, $x|d$. Write $d = xy$, where y is an integer > 1. Now:

$a = dk = xyk|m = k\ell x$, so $y|\ell$.

$b = d\ell = xy\ell|m = k\ell x$, so $y|k$.

This implies that k and ℓ are not relatively prime, since $y > 1$. Thus, the assumption $m < k\ell d$ is false, and so $m = k\ell d$. □

This generalization of Theorem 2.3.1 can be used to prove the uniqueness of prime factorizations asserted in the Fundamental Theorem of Arithmetic (Module 5.1): If n is a positive integer greater than 1, then n can be written uniquely as a product of prime numbers where the factors appear in nondecreasing order.

Proof of uniqueness of prime factorization. We have already shown that we can write any integer $n > 1$ as a product

$$n = p_1 p_2 \cdots p_k,$$

where each p_i is prime. By reordering the factors, if necessary, we can always assume that

$$p_1 \leq p_2 \leq \cdots \leq p_k.$$

We will prove by induction on k that if an integer $n > 1$ has a factorization into k primes, $k \geq 1$, then the factorization is unique.

1. Basis Step ($k = 1$). In this case $n = p_1$ is prime, and so it has no other factorization into primes.
2. Induction Step. Assume that every integer that can be factored into k primes has a unique factorization. Suppose

   $$n = p_1 p_2 \cdots p_k p_{k+1},$$

   where each p_i is prime and

   $$p_1 \leq p_2 \leq \cdots \leq p_k \leq p_{k+1}.$$

   Suppose n has another prime factorization

   $$n = q_1 q_2 \cdots q_\ell,$$

   where each q_j is prime (possibly, $\ell \neq k+1$) and

   $$q_1 \leq q_2 \leq \cdots \leq q_\ell.$$

   By the generalization of Theorem 2.3.1 in Exercise 2.3.2, since $p_1 | n = q_1 q_2 \cdots q_\ell$, then $p_1 | q_j$ for some j. But q_j is also prime, so

   $$p_1 = q_j \geq q_1.$$

   On the other hand, since $q_1 | p_1 p_2 \cdots p_k p_{k+1}$, then $q_1 | p_i$ for some i, and since p_i is prime,

   $$q_1 = p_i \geq p_1.$$

   But if $p_1 \geq q_1$ and $q_1 \geq p_1$, then $p_1 = q_1$. Thus we can cancel the first factor from both sides of the equation

   $$p_1 p_2 \cdots p_k p_{k+1} = q_1 q_2 \cdots q_\ell$$

   to get

   $$p_2 \cdots p_k p_{k+1} = q_2 \cdots q_\ell.$$

   The integer on the left-hand side of this equation has a prime factorization using k primes. By the induction hypothesis, this factorization is unique. This means that $\ell = k+1$ and

   $$p_2 = q_2,\ p_3 = q_3,\ \ldots,\ p_{k+1} = q_{k+1}.$$

   Thus, $p_i = q_i$ for $1 \leq i \leq k+1$, and the factorization of n is unique.

By the first principle of mathematical induction, every integer greater than one has a unique prime factorization. □

3. Applications of Number Theory

3.1. Representation of Integers.

THEOREM 3.1.1. *Given an integer $b > 1$, every positive integer n can be expresses uniquely as*

$$n = a_k b^k + a_{k-1} b^{k-1} + \cdots + a_1 b + a_0,$$

where $k \geq 0$, $0 \leq a_0, a_1, a_2, \ldots, a_k < b$, and are all integers.

DEFINITION 3.1.1. **Base b expansion of** n *is $(a_k a_{k-1} \cdots a_1 a_0)_b$ if the a_i are as described in Theorem 3.1.1.*

EXAMPLE 3.1.1. *Here are examples of common expansions other than the more familiar decimal expansion.*

- **Binary expansion** *is the base 2 expansion.*
- **Octal expansion** *is the base 8 expansion.*
- **Hexadecimal expansion** *is base 16 expansion. The symbols A through F are used to represent 10 through 15 in the expansion.*

Discussion

Theorem 3.1.1 asserts that each positive integer n can be expressed uniquely as a linear combination of powers of a fixed integer $b > 1$. The coefficients in the linear combination must be less than b and must be greater than or equal to zero. These coefficients are, by definition, the digits of the **base b expansion of** n, $n = (a_k a_{k-1} \ldots a_1 a_0)_b$.

3.2. Constructing Base b Expansion of n. Use the division algorithm to get the base b expansion of n:

1. $n = bq_1 + a_0$, $0 \leq a_0 < b$ and $q_1 < n$.
2. $q_1 = bq_2 + a_1$, $0 \leq a_1 < b$ and $q_2 < q_1$.
3. $q_2 = bq_3 + a_2$, $0 \leq a_2 < b$ and $q_3 < q_2$.
4. etc. until $q_i = 0$.

Then $n = (a_k a_{k-1} \ldots a_1 a_0)_b$.

EXAMPLE 3.2.1. *Find the binary expansion of 42.*

Solution: We can use the division algorithm to get the a_i's.

$$\begin{aligned} 42 &= 2(21) + 0 \\ 21 &= 2(10) + 1 \\ 10 &= 2(5) + 0 \\ 5 &= 2(2) + 1 \\ 2 &= 2(1) + 0 \\ 1 &= 2(0) + 1 \end{aligned}$$

This gives us $42 = (1)(2^5) + (0)(2^4) + (1)(2^3) + (0)(2^2) + (1)(2^1) + 0$. *Thus the binary expansion of 42 is* $(101010)_2$.

EXAMPLE 3.2.2. *Find the hexadecimal expansion of 42.*

Solution: This time we use 16 for b.

$$\begin{aligned} 42 &= 16(2) + 10 \\ 2 &= 16(0) + 2 \end{aligned}$$

So the hexadecimal expansion of 42 is $(2A)_{16}$ *(recall we use* $A = 10$ *in hexadecimal notation).*

EXAMPLE 3.2.3. *Find the decimal notation of the octal representation* $(1024)_8$.

$(1024)_8 = 1(8^3) + 0(8^2) + 2(8^1) + 4 = 532$

3.3. Cancellation in Congruences.

THEOREM 3.3.1. *Suppose* $\text{GCD}(c, m) = 1$ *and* $ac \equiv bc (mod\ m)$. *Then* $a \equiv b (mod\ m)$.

PROOF. Suppose $\text{GCD}(c, m) = 1$ and
$ac \equiv bc (\text{mod } m)$. (We may assume $m > 1$ so that $c \neq 0$.) Then

$$ac - bc = c(a - b) = km$$

for some integer k. This implies

$$c|km.$$

Since $\text{GCD}(c, m) = 1$, Lemma 2 from Module 5.2 asserts that if $c|km$, then

$$c|k.$$

Write $k = cd$ for some integer d, and substitute for k in the equation above:

$$c(a - b) = km = (cd)m = c(dm).$$

Since the cancellation law holds for integers and $c \neq 0$, we can cancel c to get

$$a - b = dm.$$

Thus, $a \equiv b (\text{mod } m)$. □

Discussion

Theorem 3.3.1 provides a criterion for being able to cancel when you have a congruence. Notice that in order to perform the cancellation, the modulus m and the factor to be cancelled must be relatively prime. Here is an example to illustrate why.

EXAMPLE 3.3.1. $3 \cdot 6 \equiv 1 \cdot 6 (mod\ 12)$, *but* $3 \not\equiv 1 (mod\ 12)$. *The reason cancellation fails is that 6 and 12 are not relatively prime.*

EXAMPLE 3.3.2. $3 \cdot 6 \equiv 8 \cdot 6 (mod\ 5)$. *Here 6 and 5 are relatively prime and we can easily check that* $3 \equiv 8 (mod\ 5)$.

3.4. Inverses mod m.

DEFINITION 3.4.1. *An integer a' is a* **(multiplicative) inverse to a modulo m** *if*

$$aa' \equiv 1 (mod\ m).$$

EXAMPLE 3.4.1. *The inverse of 14 modulo 9 is 2, since* $14 \cdot 2 \equiv 28 \equiv 1 (mod\ 9)$. *There is no inverse to 6 modulo 9, however.*

In general, an "inverse" refers to something that "undoes" another thing leaving something that is an "identity".

- With regular multiplication of real numbers, the inverse of x is $\frac{1}{x}$ since $x(\frac{1}{x}) = 1$. Inverses do not necessarily exist if we look only at integers.
- With regular addition of real numbers, the inverse of x is $-x$ since $x + (-x) = 0$,
- With matrices and matrix multiplication, the inverse of a matrix, A, is a matrix A^{-1}, such that $AA^{-1} = A^{-1}A = I$, where I is the identity matrix. Not all matrices have inverses.
- With functions and composition, the inverse of a function, f, is a function, f^{-1}, such that $(f \circ f^{-1})(x) = (f^{-1} \circ f)(x) = x = identity(x)$. Not all functions have inverses.
- Not all integers, even nonzero integers, have inverses modulo m. Moreover, if an inverse does exist it is not unique. This last part is different from all the other ones mentioned before! We shall see below, however, that if an integer a has an inverse modulo m, then it has a unique inverse lying between 0 and m.

3.5. Linear Congruence.

DEFINITION 3.5.1. *A* **linear congruence** *is a congruence of the form* $ax \equiv b(mod\ m)$*, where* a*,* b*, and* m *are fixed integers and* $m > 0$*. One may solve for* x *by finding an inverse of* a *modulo* m*, if an inverse exists.*

EXAMPLE 3.5.1. *Solve the linear congruence* $2x \equiv 7(mod\ 15)$ *for* x*.*

Solution: An inverse of 2 modulo 15 is 8. Thus

$$\begin{aligned}(8 \cdot 2)x &\equiv 8(7)(mod\ 15)\\ x &\equiv 56(mod\ 15)\\ x &\equiv 11(mod\ 15)\end{aligned}$$

Discussion

Solving a linear congruence, $ax \equiv b\,(\bmod m)$, is very similar to solving an ordinary linear equation $ax = b$. We can solve for x in the linear equation by multiplying through by the multiplicative inverse $1/a$ of a, provided $a \neq 0$. In a similar manner, we can solve a linear congruence, $ax \equiv b(\bmod m)$, provided a has a multiplicative inverse a' modulo m. Then $x \equiv a'ax \equiv a'b(\bmod m)$. To get a canonical choice for x, we would reduce $a'b$ modulo m.

Caution. DO NOT express the solution to a linear congruence $ax \equiv b(\bmod m)$ as $x = \frac{b}{a}$, as you would the solution to the linear equation $ax = b$. We have previously cautioned against using fractional notation when doing integer arithmetic, but, in the world of integers modulo m, they are expressly forbidden.

3.6. Criterion for Invertibility mod m.

THEOREM 3.6.1. *Suppose* a *and* m *are integers and* $m > 1$*. Then* a *has an inverse modulo* m *if and only if* $\mathrm{GCD}(a, m) = 1$*. Moreover, if* $\mathrm{GCD}(a, m) = 1$*, then* a *has a unique inverse,* a'*, with* $0 < a' < m$*.*

PROOF. $\mathrm{GCD}(a, m) = 1$ if and only if there are integers s and t such that $1 = as + mt$. This is true if and only if there is an integer s such that $1 \equiv as(\bmod m)$. By definition, this is true if and only if a has an inverse, namely s, modulo m. □

Discussion

Theorem 3.6.1 provides us with the conditions required for inverses modulo m to exist: For a to have an inverse modulo m, a and m must be relatively prime. The proof of the "moreover" part is complete once you solve the following exercise.

EXERCISE 3.6.1. *Prove that if* $ab \equiv 1(mod\ m)$ *and* $b \equiv c(mod\ m)$, *then* $ac \equiv 1(mod\ m)$.

3.7. Example 3.7.1. We can use the Euclidean Algorithm and the division algorithm to find the "unique" inverse of a modulo m.

EXAMPLE 3.7.1. *Find the inverse of 8 modulo 35.*

1. *Apply the Euclidean Algorithm.*
$$\begin{aligned} 35 &= 4(8) + 3 \\ 8 &= 2(3) + 2 \\ 3 &= 1(2) + 1 \\ 2 &= 2(1) + 0 \end{aligned}$$
2. *Find the linear combination of* 8 *and* 35 *that equals 1, the* GCD.
$$\begin{aligned} 1 &= 3 - 1(2) \\ &= [35 - 4(8)] - 1[8 - 2(3)] \\ &= [35 - 4(8)] - 1[8 - 2[35 - 4(8)]] \\ &= 3(35) - 13(8) \end{aligned}$$
3. *This gives*
$$-13(8) \equiv 1(mod\ 35),$$
so an inverse of 8 modulo 35 is -13.
4. *To find the inverse between 0 and 35 use the division algorithm*
$$-13 = -1(35) + 22.$$
The unique inverse of 8 modulo 35 between 0 and 35 is 22.
5. *Check:* $8 \cdot 22 = 176 = 5 \cdot 35 + 1 \equiv 1$ *(mod 35)*

3.8. Fermat's Little Theorem.

THEOREM 3.8.1. *If* p *is a prime that does not divide the integer* a, *then*
$$a^{p-1} \equiv 1(mod\ p).$$
and
$$a^p \equiv a(mod\ p).$$

EXAMPLE 3.8.1. *Find* 5^{158} **mod** 11.

Solution: Since $158 = 15(10) + 8$, *we have*

$$\begin{aligned} 5^{158} &= (5^{15})^{10}(5^8) \\ &\equiv 5^8 (mod\ 11), \end{aligned}$$

by Fermat's little theorem, applied to $a = 5^{15}$ *and* $p = 11$.

Now,

$$\begin{aligned} 5^8 &= (5^2)^4 \\ &= 25^4 \\ &\equiv 3^4 (mod\ 11)\ . \\ &= 81 \\ &\equiv 4 (mod\ 11). \end{aligned}$$

Thus 5^{158} **mod** $11 = 4$.

Discussion

The problem of determining whether a given integer is a prime may be very difficult. This fact is both interesting mathematically and useful in coding theory. Fermat's little theorem provides some help in working with prime numbers and provides the basis for many *probabilistic primality tests*. We will not give a proof of Fermat's theorem, since it involves concepts from the theory of groups that would take us too far afield. An elementary proof can be found in *Introduction to Modern Algebra*, Fourth Edition, by McCoy and Janusz (Allyn and Bacon, 1987).

The converse of Fermat's little theorem is not true. In particular, there are composite numbers n such that

$$2^{n-1} \equiv 1 (\text{mod } n).$$

These are called *pseudoprimes*. They are very rare, but 341 is a pseudoprime.

Fermat's little theorem can be used to reduce the problem of finding the remainder of a large power modulo a prime. In Example 3.8.1, we use the fact that 5^{15} and 11 are relatively prime and Fermat's little theorem to get $(5^{15})^{10} \equiv 1 (\text{mod } 11)$, thereby reducing 5^{158} to a smaller power of 5 modulo 11. One clearly has to be comfortable with the laws of exponents to carry out an exercise such as this.

3.9. RSA System. The RSA system is a public key cryptosystem based on modular exponentiation modulo the product of two large primes. This system, named after the researchers who introduced it: Rivest, Shamir, and Adleman, is a public key cryptosystem.

RSA Code

(1) Find p and q, large primes.
(2) Choose e so that $e < pq$ and $\text{GCD}(e, (p-1)(q-1)) = 1$. e must be odd, but not necessarily prime.
(3) Find d such that $de \equiv 1 (\text{mod } (p-1)(q-1))$.
(4) Encryption function $f(t) = t^e (\text{mod } pq)$.
(5) Decryption function $f^{-1}(c) = c^d (\text{mod } pq)$.

The *public keys* are (p, q, e) and the *private key* is d.

EXAMPLE 3.9.1. *Here is the routine, using* $p = 61$, $q = 53$, $e = 17$, *and* $d = 2753$.

1. *The first prime number (destroy after computing* e *and* d*):* $p = 61$
2. *The second prime number (destroy after computing* e *and* d*):* $q = 53$
3. *Modulus (give this to others):* $pq = 3233$
4. *Public exponent (give this to others):* $e = 17$
5. *Private exponent (keep this secret):* $d = 2753$
6. *Your public key is* $(pq, e) = (3233, 17)$.
7. *Your private key is* $d = 2753$.
8. *The encryption function is* $f(t) = (t^{17})(mod\ 3233)$.
9. *The decryption function is :* $f^{-1}(c) = (c^{2753})(mod\ 3233)$.

To encrypt the plaintext value 123, do this:

$f(123) = (123^{17})(\text{mod } 3233) = 337587917446653715596592958817679803(\text{mod } 3233) = 855$

To decrypt the ciphertext value 855, do this:

$$\begin{aligned} f^{-1}(855) &= (855^{2753})(\text{mod } 3233) \\ &= \text{ (an incredibly huge number goes here) (mod } 3233) \\ &= 123 \end{aligned}$$

The large exponential expressions can be reduced modulo 3233 in a piecemeal fashion, however, so that you don't actually have to calculate these large numbers.

4. Matrices

4.1. Definitions.

DEFINITION 4.1.1. *A* **matrix** *is a rectangular array of numbers. A matrix with m rows and n columns is said to have* **dimension** $m \times n$ *and may be represented as follows:*

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = [a_{ij}]$$

DEFINITION 4.1.2. *Matrices A and B are* **equal**, *$A = B$, if A and B have the same dimensions and each entry of A is equal to the corresponding entry of B.*

Discussion

Matrices have many applications in discrete mathematics. You have probably encountered them in a precalculus course. We present the basic definitions associated with matrices and matrix operations here as well as a few additional operations with which you might not be familiar.

We often use capital letters to represent matrices and enclose the array of numbers with brackets or parenthesis; e.g., $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ or $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. We do not use simply straight lines in place of brackets when writing matrices because the notation $\begin{vmatrix} a & b \\ c & d \end{vmatrix}$ has a special meaning in linear algebra. $A = [a_{ij}]$ is a shorthand notation often used when one wishes to specify how the elements are to be represented, where the first subscipt i denotes the row number and the subscript j denotes the column number of the entry a_{ij}. Thus, if one writes a_{34}, one is referring to the element in the 3rd row and 4th column. This notation, however, does not indicate the dimensions of the matrix. Using this notation, we can say that two $m \times n$ matrices $A = [a_{ij}]$ and $B = [b_{ij}]$ are equal if and only if $a_{ij} = b_{ij}$ for all i and j.

EXAMPLE 4.1.1. *The following matrix is a 1×3 matrix with $a_{11} = 2$, $a_{12} = 3$, and $a_{13} = -2$.*

$$\begin{bmatrix} 2 & 3 & -2 \end{bmatrix}$$

EXAMPLE 4.1.2. *The following matrix is a* 2×3 *matrix.*

$$\begin{bmatrix} 0 & \pi & -2 \\ 2 & 5 & 0 \end{bmatrix}$$

4.2. Matrix Arithmetic. Let α be a scalar, $A = [a_{ij}]$ and $B = [b_{ij}]$ be $m \times n$ matrices, and $C = [c_{ij}]$ a $n \times p$ matrix.

(1) Addition: $A + B = [a_{ij} + b_{ij}]$
(2) Subtraction: $A - B = [a_{ij} - b_{ij}]$
(3) Scalar Multiplication: $\alpha A = [\alpha a_{ij}]$
(4) Matrix Multiplication: $AC = \left[\sum_{k=1}^{n} a_{ik} c_{kj}\right]$

Discussion

Matrices may be added, subtracted, and multiplied, provided their dimensions satisfy certain restrictions. To add or subtract two matrices, the matrices must have the same dimensions.

Notice there are two types of multiplication. Scalar multiplication refers to the product of a matrix times a scalar (real number). A scalar may be multiplied by a matrix of any size. On the other hand, matrix multiplication refers to taking the product of two matrices. The definition of matrix multiplication may not seem very natural at first. It has a great many applications, however, some of which we shall see. Notice that in order for the product AC to be defined, the number of columns in A must equal the number of rows of C. Thus, it is possible for the product AC may be defined, while CA is not. When multiplying two matrices, the order is important. In general, AC is not necessarily the same as CA, even if both products AC and CA are defined. In other words, matrix multiplication is *not commutative*.

4.3. Example 4.3.1.

EXAMPLE 4.3.1. *Suppose*

$$A = \begin{bmatrix} 1 & -2 & 3 \\ 0 & 3 & 4 \end{bmatrix}, \; B = \begin{bmatrix} 0 & 1 & -2 \\ 3 & -4 & 5 \end{bmatrix}, \text{ and } C = \begin{bmatrix} 3 & 4 & -6 & 0 \\ 0 & -1 & 2 & 2 \\ 1 & -2 & 3 & 4 \end{bmatrix}$$

Then

$$A+B=\begin{bmatrix}1&-1&1\\3&-1&9\end{bmatrix}$$
$$A-B=\begin{bmatrix}1&-3&5\\-3&7&-1\end{bmatrix}$$
$$3A=\begin{bmatrix}3&-6&9\\0&9&12\end{bmatrix}$$
$$AC=\begin{bmatrix}6&0&-1&8\\4&-11&18&22\end{bmatrix}$$

Let us break down the multiplication of A and C in Example 4.3.1 down to smaller pieces.

$$\begin{bmatrix}1&-2&3\end{bmatrix}\cdot\begin{bmatrix}3\\0\\1\end{bmatrix}=\begin{bmatrix}3+0+3\end{bmatrix}=[6]$$

$$\begin{bmatrix}1&-2&3\end{bmatrix}\begin{bmatrix}3&4\\0&-1\\1&-2\end{bmatrix}=\begin{bmatrix}3+0+3&4+2-6\end{bmatrix}=\begin{bmatrix}6&0\end{bmatrix}$$

$$\begin{bmatrix}1&-2&3\end{bmatrix}\begin{bmatrix}3&4&-6&0\\0&-1&2&2\\1&-2&3&4\end{bmatrix}=\begin{bmatrix}3+0+3&4+2-6&-6-4+9&0-4+12\end{bmatrix}$$
$$=\begin{bmatrix}6&0&-1&8\end{bmatrix}$$

Now compute the second row to get

$$AC=\begin{bmatrix}6&0&-1&8\\4&-11&18&22\end{bmatrix}.$$

4.4. Special Matrices.

1. A **square matrix** is a matrix with the same number of rows as columns.
2. A **diagonal matrix** is a square matrix whose entries off the main diagonal are zero.
3. An **upper triangular matrix** is a matrix having all the entries below the main diagonal equal to zero.
4. A **lower triangular matrix** is a matrix having the entries above the main diagonal equal to zero.
5. The $n \times n$ **identity matrix**, I, is the $n \times n$ matrix with ones down the diagonal and zeros elsewhere.
6. The **inverse** of a square matrix, A, is the matrix A^{-1}, if it exists, such that $AA^{-1} = A^{-1}A = I$.
7. The **transpose** of a matrix $A = [a_{ij}]$ is $A^t = [a_{ji}]$.
8. A **symmetric matrix** is one that is equal to its transpose.

Discussion

Many matrices have special forms and special properties. Notice that, although a diagonal matrix must be square, no such condition is put on upper and lower triangular matrices.

The following matrix is a diagonal matrix (it is also upper and lower triangular).

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 6 \end{bmatrix}$$

The following matrix is upper triangular.

$$\begin{bmatrix} -1 & 0 & 3 & -2 \\ 0 & 1 & 2 & 5 \\ 0 & 0 & -3 & 3 \end{bmatrix}$$

The next matrix is the transpose of the previous matrix. Notice that it is lower triangular.

$$\begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 2 & -3 \\ -2 & 5 & 3 \end{bmatrix}$$

The identity matrix is a special matrix that is the multiplicative identity for any matrix multiplication. Another way to define the identity matrix is the square matrix $I = [a_{ij}]$ where $a_{ij} = 0$ if $i \neq j$ and $a_{ii} = 1$. The $n \times n$ identity I has the property that $IA = A$ and $AI = A$, whenever either is defined. For example,

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 3 & -4 & -2 \\ 2 & 7 & 0 \end{bmatrix} = \begin{bmatrix} 3 & -4 & -2 \\ 2 & 7 & 0 \end{bmatrix}$$

The inverse of a matrix A is a special matrix A^{-1} such that $AA^{-1} = A^{-1}A = I$. A matrix must be square to define the inverse. Moreover, the inverse of a matrix does not always exist.

Example 4.4.1.

$$\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

so that

$$\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}.$$

The transpose of a matrix is the matrix obtained by interchanging the rows for the columns. For example, the transpose of

$$A = \begin{bmatrix} 2 & 3 & -1 \\ -2 & 5 & 6 \end{bmatrix} \qquad \text{is} \qquad A^t = \begin{bmatrix} 2 & -2 \\ 3 & 5 \\ -1 & 6 \end{bmatrix}$$

If the transpose is the same as the original matrix, then the matrix is called symmetric. Notice a matrix must be square in order to be symmetric.

We will show here that matrix multiplication is distributive over matrix addition.

Let $A = [a_{ij}]$ and $B = [b_{ij}]$ be $m \times n$ matrices and let $C = [c_{ij}]$ be an $n \times p$ matrix. We use the definitions of addition and matrix multiplication and the distributive properties of the real numbers to show the distributive property of matrix multiplication. Let i and j be integers with $1 \leq i \leq m$ and $1 \leq j \leq p$. Then the element in the i-th row and the j-th column in $(A + B)C$ would be given by

$$\begin{aligned}\sum_{k=1}^{n}(a_{ik} + b_{ik})(c_{kj}) &= \sum_{k=1}^{n}(a_{ik}c_{kj} + b_{ik}c_{kj}) \\ &= \sum_{k=1}^{n} a_{ik}c_{kj} + \sum_{k=1}^{n} b_{ik}c_{kj} \\ &= \sum_{k=1}^{n} a_{ik}c_{kj} + \sum_{k=1}^{n} b_{ik}c_{kj}\end{aligned}$$

This last part corresponds to the form the element in the i-th row and j-th column of $AC + BC$. Thus the element in the i-th row and j-th column of $(A + B)C$ is the same as the corresponding element of $AC + BC$. Since i and j were arbitrary this shows $(A + B)C = AC + BC$.

The proof that $C(A + B) = CA + CB$ is similar. Notice that we must be careful, though, of the order of the multiplication. Matrix multiplication is *not* commutative.

4.5. Boolean Arithmetic. If a and b are binary digits (0 or 1), then

$$a \wedge b = \begin{cases} 1, & \text{if } a = b = 1 \\ 0, & \text{otherwise.} \end{cases}$$

$$a \vee b = \begin{cases} 0, & \text{if } a = b = 0 \\ 1, & \text{otherwise.} \end{cases}$$

Definitions 4.5.1. *Let A and B be $n \times m$ matrices.*

1. *The* **meet** *of A and B: $A \wedge B = [a_{ij} \wedge b_{ij}]$*
2. *The* **join** *of A and B: $A \vee B = [a_{ij} \vee b_{ij}]$*

Definition 4.5.1. *Let $A = [a_{ij}]$ be $m \times k$ and $B = [b_{ij}]$ be $k \times n$. The* **Boolean product** *of A and B, $A \odot B$, is the $m \times n$ matrix $C = [c_{ij}]$ defined by*

$$c_{ij} = (a_{i1} \wedge b_{1j}) \vee (a_{i2} \wedge b_{2j}) \vee (a_{i3} \wedge b_{3j}) \vee \cdots \vee (a_{ik} \wedge b_{kj}).$$

Discussion

Boolean operations on zero-one matrices is completely analogous to the standard operations, except we use the Boolean operators $\wedge$ and $\vee$ on the binary digits instead of ordinary multiplication and addition, respectively.

4.6. Example 4.6.1.

EXAMPLE 4.6.1. *Let* $A = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}$, $B = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{bmatrix}$, *and* $C = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix}$.

Then

1. $A \wedge B = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$

2. $A \vee B = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix}$

3. $A \odot C = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$

Here are more details of the Boolean product in Example 4.6.1:

$$A \odot C = \begin{bmatrix} (1\wedge1)\vee(1\wedge0)\vee(0\wedge0)\vee(1\wedge1) & (1\wedge1)\vee(1\wedge1)\vee(0\wedge1)\vee(1\wedge0) & (1\wedge0)\vee(1\wedge0)\vee(0\wedge1)\vee(1\wedge0) \\ (0\wedge1)\vee(1\wedge0)\vee(1\wedge0)\vee(0\wedge1) & (0\wedge1)\vee(1\wedge1)\vee(1\wedge1)\vee(0\wedge0) & (0\wedge0)\vee(1\wedge0)\vee(1\wedge1)\vee(0\wedge0) \end{bmatrix}$$

$$= \begin{bmatrix} 1\vee0\vee0\vee1 & 1\vee1\vee0\vee0 & 0\vee0\vee0\vee0 \\ 0\vee0\vee0\vee0 & 0\vee1\vee1\vee0 & 0\vee0\vee1\vee0 \end{bmatrix}$$

EXERCISE 4.6.1.

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \end{bmatrix}$$

Find

(1) $A \vee B$
(2) $A \wedge B$

Exercise 4.6.2.

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

Find

(1) $A \odot A$
(2) $A \odot A \odot A$
(3) $A \odot A \odot A \odot A$

Exercise 4.6.3.

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

Find $\odot_{k=1}^{n} A$*, the Boolean product of* A *with itself* n *times. Hint: Do exercise 4.6.2 first.*

CHAPTER 6

Introduction to Graph Theory

1. Introduction to Graphs

1.1. Simple Graphs.

DEFINITION 1.1.1. *A* **simple graph** (V, E) *consists of a set representing vertices,* V*, and a set of unordered pairs of elements of* V *representing edges,* E*. A simple graph has*

- *no arrows,*
- *no loops, and*
- *cannot have multiple edges joining vertices.*

Discussion

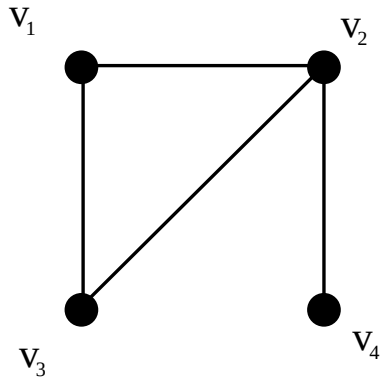
Graphs offer a convenient way to represent various kinds of mathematical objects. Essentially, any graph is made up of two sets, a set of vertices and a set of edges. Depending on the particular situation we are trying to represent, however, we may wish to impose restrictions on the type of edges we allow. For some problems we will want the edges to be *directed* from one vertex to another; whereas, in others the edges are undirected. We begin our discussion with **undirected graphs**.

The most basic graph is the **simple graph** as defined above. Since the edges of a simple graph are undirected, they are represented by **unordered pairs** of vertices rather than **ordered pairs**. For example, if $V = \{a, b, c\}$, then $\{a, b\} = \{b, a\}$ would represent the same edge.

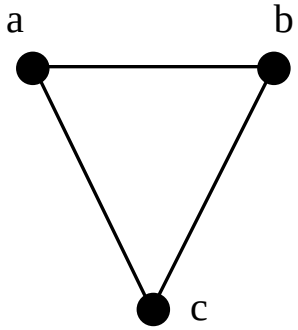
EXERCISE 1.1.1. *If a simple graph* G *has 5 vertices, what is the maximum number of edges that* G *can have?*

1.2. Examples.

EXAMPLE 1.2.1. $V = \{v_1, v_2, v_3, v_4\}$ $E = \{\{v_1v_2\}, \{v_1, v_3\}, \{v_2, v_3\}, \{v_2, v_4\}\}$



EXAMPLE 1.2.2. $V = \{a, b, c\}$, $E = \{\{a, b\}, \{b, c\}, \{a, c\}\}$



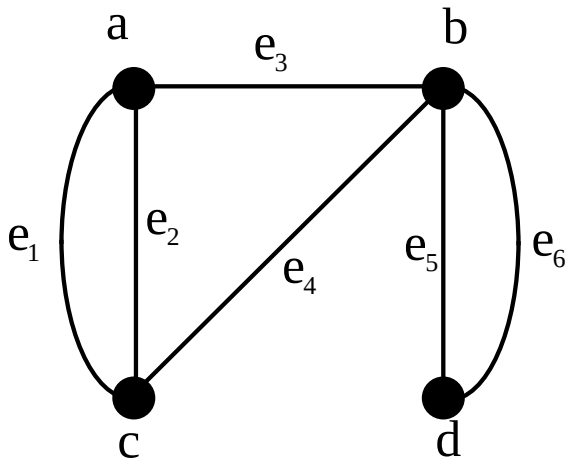
1.3. Multigraphs. Definition: A **multigraph** is a set of vertices, V, a set of edges, E, and a function

$$f : E \to \{\{u, v\} : u, v \in V \text{ and } u \neq v\}.$$

If $e_1, e_2 \in E$ are such that $f(e_1) = f(e_2)$, then we say e_1 and e_2 are **multiple** or **parallel edges**.

EXAMPLE 1.3.1. $V = \{a, b, c, d\}$, $E = \{e_1, e_2, \ldots, e_6\}$, $f : E \to \{\{u, v\} : u, v \in V \text{ and } u \neq v\}$ *is defined as follows.*

e	e_1	e_2	e_3	e_4	e_5	e_6
$f(e)$	$\{a, c\}$	$\{a, c\}$	$\{a, b\}$	$\{c, b\}$	$\{b, d\}$	$\{b, d\}$



Discussion

In Example 1.3.1 e_1 and e_2 are parallel edges, but the edges e_2 and e_5 are not called parallel edges.

EXERCISE 1.3.1. *Find all the parallel edges Example 3.*

Notice that a multigraph allows for multiple edges between a pair of vertices, but does not allow for loops. In some applications it may be desirable to illustrate all the connections between the vertices. Say for example, in a network there may be multiple wires connecting the same units.

1.4. Pseudograph.

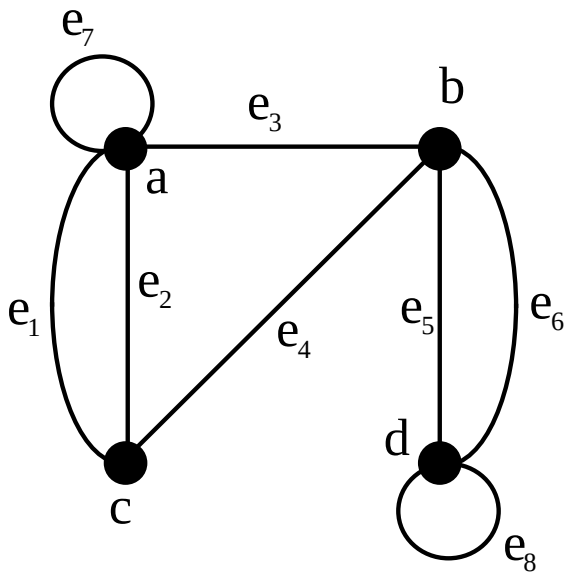
DEFINITION 1.4.1. *A* **pseudograph** *is a set of vertices, V, a set of edges, E, and a function $f : E \to \{\{u, v\} : u, v \in V\}$. If $e \in E$ is such that $f(e) = \{u, u\} = \{u\}$, then we say e is a* **loop**.

EXAMPLE 1.4.1. $V = \{a, b, c, d\}$, $E = \{e_1, e_2, \ldots, e_8\}$,

$$f : E \to \{\{u, v\} : u, v \in V\}$$

is defined as follows.

e	e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8
$f(e)$	$\{a, c\}$	$\{a, c\}$	$\{a, b\}$	$\{c, b\}$	$\{b, d\}$	$\{b, d\}$	$\{a\}$	$\{d\}$



Discussion

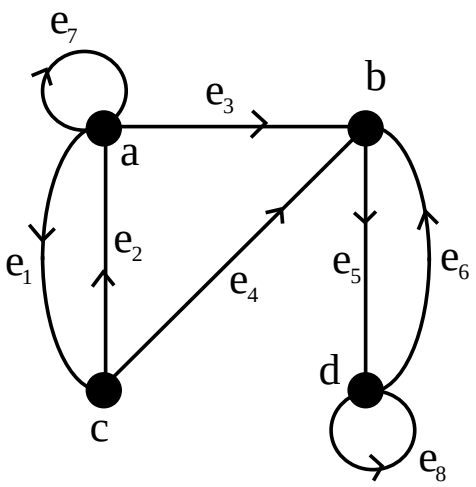
The psuedograph adds the possibility of loops. For example, a diagnostic line may be used in a network, which is a line connecting a computer to itself.

1.5. Directed Graph.

DEFINITION 1.5.1. *A* **directed graph** (V, E) *consists of a set of vertices,* V*, and a set of directed edges,* E*. The elements of* E *are* **ordered pairs** *of vertices.*

EXAMPLE 1.5.1. $V = \{a, b, c, d\}$,

$E = \{(a, c), (c, a), (a, b), (c, b), (b, d), (d, b), (a, a), (d, d)\}$,



Discussion

A directed graph, or **digraph**, allows loops and allows two edges joining the same vertex, but going in the opposite direction. More than one edge going in the same direction between vertices, however, is not allowed. A directed edge is defined by an **ordered pair** rather than an unordered pair. That is, the ordered pair (a, b) is different from the ordered pair (b, a), while the unordered pair $\{a, b\} = \{b, a\}$. Be careful of the notation you use when writing an edge.

EXERCISE 1.5.1. *If a directed graph G has 5 vertices, what is the maximum number of (directed) edges of G?*

1.6. Directed Multigraph.

DEFINITION 1.6.1. *A* **directed multigraph** (V, E) *consists of vertices, V, and edges, E, and a function*

$$f\colon E \to V \times V = \{(u, v) | u, v \in V\}.$$

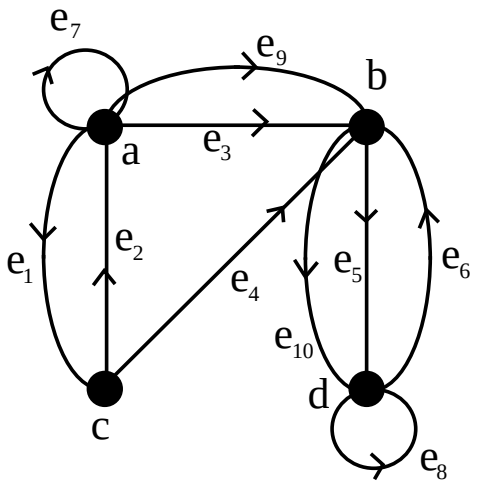
The edges e_1 and e_2 are **multiple edges** *if $f(e_1) = f(e_2)$*

EXAMPLE 1.6.1. $V = \{a, b, c, d\}$, $E = \{e_1, e_2, \ldots, e_{10}\}$,

$$f\colon E \to \{(u, v) : u, v \in V\}$$

is defined as follows.

e	e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8	e_9	e_{10}
$f(e)$	(a, c)	(c, a)	(a, b)	(c, b)	(b, d)	(d, b)	(a, a)	(d, d)	(a, b)	(b, d)



Discussion

Notice the difference between a directed graph and a directed multigraph: a directed graph allows more than one edged to connect the same two vertices as long as they have opposite directions; whereas, no such restriction is placed on the edges of a directed multigraph.

EXERCISE 1.6.1. *Give all the multiple edges in Example 1.6.1.*

1.7. Graph Isomorphism.

DEFINITION 1.7.1. *Let $G_1 = (V, E)$ and $G_2 = (U, F)$ be simple graphs. The graphs G_1 and G_2 are* **isomorphic** *if there exists a bijection*

$$f : V \to U$$

such that for all $v_1, v_2 \in V$, v_1 and v_2 are adjacent in G_1 if and only if $f(v_1)$ and $f(v_2)$ are adjacent in G_2.

DEFINITION 1.7.2. *If f is a bijection as described above, then f is called an* **isomorphism** *between G_1 and G_2, and we often write*

$$f : G_1 \to G_2.$$

Discussion

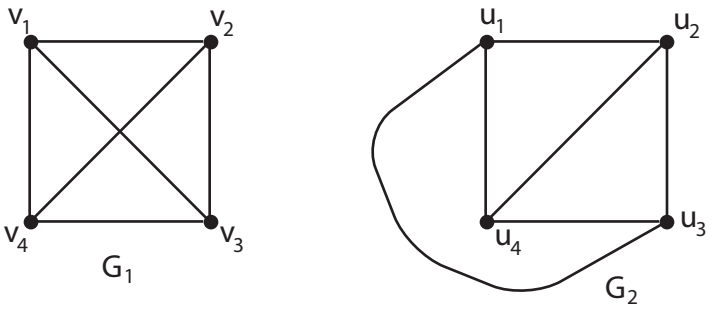
There are several notations that are used to represent an isomorphism. We will use a common notation $G_1 \simeq G_2$ to mean that G_1 is isomorphic to G_2.

Trying to construct an isomorphism between graphs can be a very difficult problem in general. If simple graphs G_1 and G_2 are isomorphic, then, clearly, they must have

the same number of vertices. As the next exercise shows, G_1 and G_2 must also have the same number of edges. **Having the same number of vertices and edges, however, is in no way sufficient for graphs G_1 and G_2 to be isomorphic.** Often to prove existence of an isomorphism between two graphs one must actually construct the isomorphism.

EXERCISE 1.7.1. *Prove that if simple graphs G_1 and G_2 are isomorphic, then G_1 and G_2 have the same number of edges.*

EXAMPLE 1.7.1. *The graphs G_1 and G_2 below are isomorphic. The bijection is defined by $f(v_i) = u_i$.*



Example 1.7.1 illustrates a situation in which it is very easy to construct an isomorphism. The graph G_2 is merely an alteration of G_1 obtained by moving one of the edges so it goes around rather than crossing over another edge and relabeling its vertices.

One way to visualize when two graphs are isomorphic is to imagine that all the vertices are beads and each edge is represented by a sting with each end tied to the beads that represents its endpoints. If you pick one or more beads up and place it in another location without untying the strings, you obtain a graph that is isomorphic to the original. In fact, if you can move the vertices to different positions keeping the edges attached to go from one graph to another, the two graphs are isomorphic. Edges are allowed to "pass through" each other, so a straight edge and a knotted edge would be considered the same edge.

When two graphs are isomorphic, they share many of the important properties of graphs. In many instances we do not differentiate between two graphs that are

isomorphic. Until we study isomorphism in detail, we will not differentiate between two isomorphic graphs. We will discuss graph isomorphisms further in Module 6.3.

EXERCISE 1.7.2. *Construct a definition for "isomorphism" between*

(a) two multigraphs.
(b) two pseudographs.
(c) two directed graphs.
(d) two directed multigraphs.

2. Graph Terminology

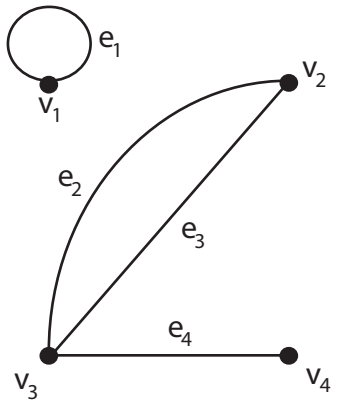
2.1. Undirected Graphs.

DEFINITIONS 2.1.1. *Suppose $G = (V, E)$ is an undirected graph.*

(1) Two vertices $u, v \in V$ are **adjacent** *or* **neighbors** *if there is an edge e between u and v.*
- *The edge e* **connects** *u and v.*
- *The vertices u and v are* **endpoints** *of e.*

(2) The **degree** *of a vertex v, denoted $deg(v)$, is the number of edges for which it is an endpoint. A loop contributes twice in an undirected graph.*
- *If $deg(v) = 0$, then v is called* **isolated**.
- *If $deg(v) = 1$, then v is called* **pendant**.

EXAMPLE 2.1.1. *$V = \{v_1, v_2, v_3, v_4\}$ and $E = \{e_1, e_2, e_3, e_4\}$.*



(1) v_2 and v_3 are adjacent.
(2) $deg(v_1) = 2$
(3) $deg(v_2) = 2$
(4) $deg(v_3) = 3$
(5) $deg(v_4) = 1$

Discussion

Notice that in computing the degree of a vertex in an undirected graph a loop contributes two to the degree. In this example, none of the vertices is isolated, but v_4 is pendant. In particular, the vertex v_1 is not isolated since its degree is 2.

2.2. The Handshaking Theorem.

THEOREM 2.2.1. **(The Handshaking Theorem)** *Let $G = (V, E)$ be an undirected graph. Then*

$$2|E| = \sum_{v \in V} deg(v)$$

PROOF. Each edge contributes twice to the sum of the degrees of all vertices. □

Discussion

Theorem 2.2.1 is one of the most basic and useful combinatorial formulas associated to a graph. It lets us conclude some facts about the numbers of vertices and the possible degrees of the vertices. Notice the immediate corollary.

COROLLARY 2.2.1.1. *The sum of the degrees of the vertices in any graph must be an even number.*

In other words, it is impossible to create a graph so that the sum of the degrees of its vertices is odd (try it!).

2.3. Example 2.3.1.

EXAMPLE 2.3.1. *Suppose a graph has 5 vertices. Can each vertex have degree 3? degree 4?*

1. *The sum of the degrees of the vertices would be $3 \cdot 5$ if the graph has 5 vertices of degree 3. This is an odd number, though, so this is not possible by the handshaking Theorem.*
2. *The sum of the degrees of the vertices if there are 5 vertices with degree 4 is 20. Since this is even it is possible for this to equal $2|E|$.*

Discussion

If the sum of the degrees of the vertices is an even number then the handshaking theorem is not contradicted. In fact, you can create a graph with any even degree you want if multiple edges are permitted. However, if you add more restrictions it may not be possible. Here are two typical questions the handshaking theorem may help you answer.

EXERCISE 2.3.1. *Is it possible to have a graph S with 5 vertices, each with degree 4, and 8 edges?*

EXERCISE 2.3.2. *A graph with 21 edges has 7 vertices of degree 1, three of degree 2, seven of degree 3, and the rest of degree 4. How many vertices does it have?*

THEOREM 2.3.1. *Every graph has an even number of vertices of odd degree.*

PROOF. Let V_o be the set of vertices of odd degree, and let V_e be the set of vertices of even degree. Since $V = V_o \cup V_e$ and $V_o \cap V_e = \emptyset$, the handshaking theorem gives us

$$2|E| = \sum_{v \in V} deg(v) = \sum_{v \in V_o} deg(v) + \sum_{v \in V_e} deg(v)$$

or

$$\sum_{v \in V_o} deg(v) = 2|E| - \sum_{v \in V_e} deg(v).$$

Since the sum of any number of even integers is again an even integer, the right-hand-side of this equations is an even integer. So the left-hand-side, which is the sum of a collection of odd integers, must also be even. The only way this can happen, however, is for there to be an even number of odd integers in the collection. That is, the number of vertices in V_o must be even. □

Theorem 2.3.1 goes a bit further than our initial corollary of the handshaking theorem. If you have difficulty with the last sentence of the proof, consider the following facts:

- odd + odd = even
- odd + even = odd
- even + even = even

If we add up an odd number of odd numbers the previous facts will imply we get an odd number. Thus to get an even number out of $\sum_{v \in V_o} deg(v)$ there must be an even number of vertices in V_o (the set of vertices of odd degree).

While there must be an even number of vertices of odd degree, there is no restrictions on the parity (even or odd) of the number of vertices of even degree.

This theorem makes it easy to see, for example, that it is not possible to have a graph with 3 vertices each of degree 1 and no other vertices of odd degree.

2.4. Directed Graphs.

DEFINITIONS 2.4.1. *Let $G = (V, E)$ be a directed graph.*

(1) Let (u, v) be an edge in G. Then u is an **initial vertex** *and is* **adjacent to** *v. The vertex v is a* **terminal vertex** *and is* **adjacent from** *u.*
(2) The **in-degree** *of a vertex v, denoted $deg^-(v)$ is the number of edges which terminate at v.*

(3) Similarly, the **out-degree** *of v, denoted $deg^+(v)$, is the number of edges which initiate at v.*

2.5. The Handshaking Theorem for Directed Graphs.

THEOREM 2.5.1. *For any directed graph $G = (V, E)$,*

$$|E| = \sum_{v \in V} deg^-(v) = \sum_{v \in V} deg^+(v).$$

Discussion

When considering directed graphs we differentiate between the number of edges going into a vertex verses the number of edges coming out from the vertex. These numbers are given by the in-degree and the out-degree.

Notice that each edge contributes one to the in-degree of some vertex and one to the out-degree of some vertex. This is essentially the proof of Theorem 2.5.1.

EXERCISE 2.5.1. *Prove Theorem 2.5.1.*

2.6. Underlying Undirected Graph.

DEFINITION 2.6.1. *If we take a directed graph and remove the arrows indicating the direction we get the* **underlying undirected graph**.

Discussion

There are applications in which you may start with a directed graph, but then later find it useful to consider the corresponding undirected graph obtained by removing the direction of the edges.

Notice that if you take a vertex, v, in a directed graph and add its in-degree and out-degree, you get the degree of this vertex in the underlying undirected graph.

2.7. New Graphs from Old.

DEFINITIONS 2.7.1. *1. (W, F) is a* **subgraph** *of $G = (V, E)$ if*

$$W \subseteq V \text{ and } F \subseteq E.$$

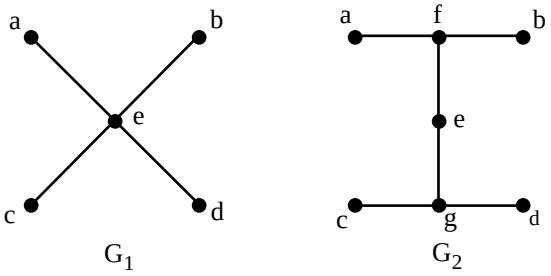
2. Given graphs G_1 and G_2, their union

$$G_1 \cup G_2 = (V_1 \cup V_2, E_1 \cup E_2).$$

3. *Given graphs G_1 and G_2, their* **join**, *denoted by $G_1 * G_2$, is the graph consisting of the union $G_1 \cup G_2$ together with all possible edges connecting a vertex of G_1 that is not in G_2 to a vertex of G_2 that is not in G_1.*

EXAMPLE 2.7.1. *Suppose G has vertex set $V = \{a, b\}$ and one edge $e = \{a, b\}$ connecting a and b, and H has a single vertex c and no edges. Then $G * H$ has vertex set $\{a, b, c\}$ and edges $\{a, b\}$, $\{a, c\}$, $\{b, c\}$.*

EXERCISE 2.7.1. *Find the union and join of the graphs G_1 and G_2 below.*



EXERCISE 2.7.2. *Prove that the union of two simple graphs is a simple graph.*

EXERCISE 2.7.3. *Prove that the join of two simple graphs is a simple graph.*

2.8. Complete Graphs.

DEFINITION 2.8.1. *The* **complete graph** *with n vertices, denoted K_n, is the simple graph with exactly one edge between each pair of distinct vertices.*

Discussion

There are a certain types of simple graphs that are important enough that they are given special names. The first of these are the **complete graphs**. These are the simple graphs that have the maximal number of edges for the given set of vertices. For example, if we were using graphs to represent a local area network, a complete graph would represent the maximum redundancy possible. In other words, each pair of computers would be directly connected. It is easy to see that any two complete graphs with n vertices are isomorphic, so that the symbol K_n is ambiguously used to denote any such graph.

Complete graphs also arise when considering the question as to whether a graph G is **planar**, that is, whether G can be drawn in a plane without having any two edges intersect. The complete graphs K_1, K_2, K_3, and K_4 are planar graphs, but K_n is not planar if $n \geq 5$. Draw K_4 without making the edges intersect, then try to draw K_5 without creating an unwanted intersection between edges. (Notice that K_{n+1} can be created from K_n by adding one new vertex and an edge from the new vertex to each vertex of K_n.)

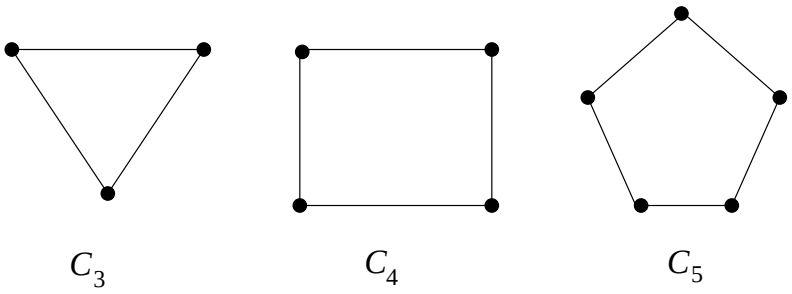
EXERCISE 2.8.1. *Prove that the complete graph K_n, $n \geq 1$, is the join $K_{n-1} * G$, where G is a graph with one vertex and no edges.*

2.9. Cycles.

DEFINITION 2.9.1. *A* **cycle** *with n vertices $\{v_1, v_2, ..., v_n\}$, denoted by C_n, is a simple graph with edges of the form $\{v_1, v_2\}, \{v_2, v_3\}, \{v_3, v_4\}, \ldots, \{v_{n-1}, v_n\}, \{v_n, v_1\}$.*

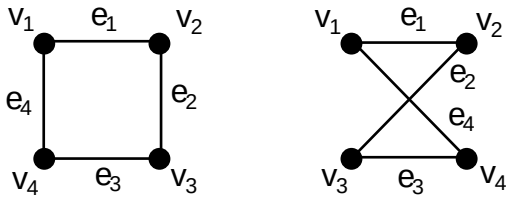
Discussion

Notice that a cycle must have at least 3 vertices. Here are examples of the first three possibilities:



Local area networks that are configured this way are often called **ring** networks.

Notice that the following two graphs are isomorphic. Pay close attention to the labels.



The point of the last illustration, is that sometimes you have to redraw the graph to see the ring shape. It also demonstrates that a graph may be planar even though this fact may not be evident from a given representation.

2.10. Wheels.

DEFINITION 2.10.1. *A* **wheel** *is a join $C_n * G$, where C_n is a cycle and G is a graph with one vertex and no edges. The wheel with $n + 1$ vertices is denoted W_n.*

Discussion

Notice that a wheel is obtained by starting with a cycle and then adding one new vertex and an edge from that vertex to each vertex of the cycle. Be careful! The index on the notation W_n is actually one less that the number of vertices. The n stands for the number of vertices in the cycle that we used to get the wheel.

2.11. n-Cubes.

DEFINITION 2.11.1. *The n-**cube**, denoted Q_n, is the graph with 2^n vertices represented by the vertices and edges of an n-dimensional cube.*

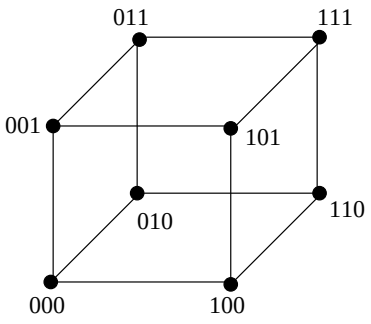
These graphs can be constructed recursively as follows:

1. Initial Condition. A 0-cube is a graph with one vertex and no edges.
2. Recursion. Let Q_n^1 and Q_n^2 be two disjoint n-cubes, $n \geq 0$, and let $f\colon Q_n^1 \to Q_n^2$ be an isomorphism. Q_{n+1} is the graph consisting of the union $Q_n^1 \cup Q_n^2$, together with all edges $\{v, f(v)\}$ joining a vertex v in Q_n^1 with its corresponding vertex $f(v)$ in Q_n^2.

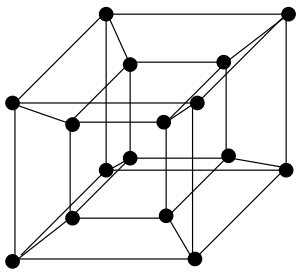
Q_n can also be represented as the graph whose vertices are the bit strings of length n, having an edge between each pair of vertices that differ by one bit.

Discussion

The n-Cube is a common way to connect processors in parallel machines. Here are the cubes Q_3 and Q_4.



3-cube



4-cube

EXERCISE 2.11.1. *Find all the subgraphs of Q_1, and Q_2.*

EXERCISE 2.11.2. *Label the vertices of Q_4 appropriately, using bit strings of length four.*

EXERCISE 2.11.3. *Use your labeling of the vertices of Q_4 from Exercise 2.11.2 to identify two disjoint Q_3's, and an isomorphism between them, from which Q_4 would be obtained in the recursive description above.*

EXERCISE 2.11.4. *Prove that $Q_{n+1} \subseteq Q_n^1 * Q_n^2$, where Q_n^1 and Q_n^2 are disjoint n-cubes, $n \geq 0$.*

EXERCISE 2.11.5. *Prove that the 2-cube is not (isomorphic to) the join of two 1-cubes.*

EXERCISE 2.11.6. *Draw the graph Q_5. [Hint: Abandon trying to use a "cube" shape. Put* 00000 *on the top of the page and* 11111 *on the bottom and look for an organized manner to put the rest in between.]*

2.12. Bipartite Graphs.

DEFINITION 2.12.1. *A simple graph G is* **bipartite** *if V can be partitioned into two nonempty disjoint subsets V_1 and V_2 such that every edge connects a vertex in V_1 and a vertex in V_2.*

DEFINITION 2.12.2. *A bipartite graph is* **complete** *if V can be partitioned into two disjoint subsets V_1 and V_2 such that there is an edge from every vertex V_1 to every vertex in V_2.*

$K_{m,n}$ denotes the complete bipartite graph when $m = |V_1|$ and $n = |V_2|$.

Discussion

The definition of a bipartite graph is not always consistent about the necessary size of $|V_1|$ and $|V_2|$. We will assume V_1 and V_2 must have at least one element each, so we will not consider the graph consisting of a single vertex bipartite.

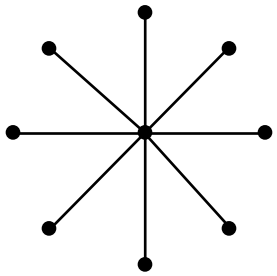
Note: There are no edges connecting vertices in V_1 in a bipartite graph. Neither are there edges connecting vertices in V_2.

EXERCISE 2.12.1. *How many edges are there in the graph $K_{m,n}$?*

EXERCISE 2.12.2. *Prove that a complete bipartite graph $K_{m,n}$ is the join $G_m * G_n$ of graphs G_m and G_n, where G_m has m vertices and no edges, and G_n has n vertices and no edges.*

2.13. Examples.

EXAMPLE 2.13.1. *A* **star network** *is a $K_{1,n}$ bipartite graph.*

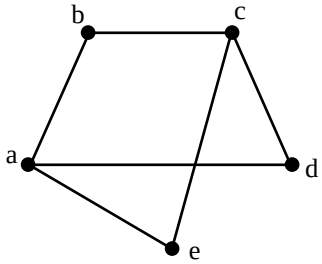


$K_{1,8}$

EXAMPLE 2.13.2. *C_k, for k even, is a bipartite graph: Label the vertices $\{v_1, v_2, ..., v_k\}$ cyclicly around C_k, and put the vertices with odd subscripts in V_1 and the vertices with even subscripts in V_2.*

(1) Suppose V_1 is a set of airlines and V_2 is a set of airports. Then the graph with vertex set $V = V_1 \cup V_2$, where there is an edge between a vertex of V_1 and a vertex of V_2 if the given airline serves the given airport, is bipartite. If every airline in V_1 serves every airport in V_2, then the graph would be a **complete bipartite graph**.
(2) Supplier, warehouse transportation models where an edge represents that a given supplier sends inventory to a given warehouse are bipartite.

EXERCISE 2.13.1. *Is the following graph bipartite?*



EXERCISE 2.13.2. *Prove that Q_n is bipartite. [Hint: You don't need mathematical induction; use the bit string model for the vertex set.]*

Bipartite graphs also arise when considering the question as to whether a graph is planar. It is easy to see that the graphs $K_{1,n}$ and $K_{2,n}$ are planar for every $n \geq 1$. The graphs $K_{m,n}$, however, are not planar if both m and n are greater than 2. In particular, $K_{3,3}$ is not planar. (Try it!) A theorem, which we shall not prove, states that *every* nonplanar graph contains (in some sense) a subgraph (see Slide 15) isomorphic to K_5 or a subgraph isomorphic to $K_{3,3}$.

REMARK 2.13.1. *The important properties of a graph do not depend on how it is drawn. To see that two graphs, whose vertices have the same labels, are isomorphic, check that vertices are connected by an edge in one graph if and only if they are also connected by an edge in the other graph.*

3. Representing Graphs and Graph Isomorphism

3.1. Adjacency Matrix.

DEFINITION 3.1.1. *The* **adjacency matrix**, $A = [a_{ij}]$, *for a simple graph* $G = (V, E)$, *where* $V = \{v_1, v_2, ..., v_n\}$, *is defined by*

$$a_{ij} = \begin{cases} 1 & \text{if } \{v_i, v_j\} \text{ is an edge of } G, \\ 0 & \text{otherwise.} \end{cases}$$

Discussion

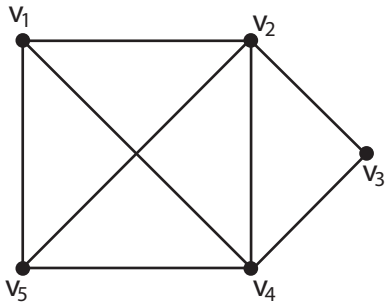
We introduce some alternate representations, which are extensions of connection matrices we have seen before, and learn to use them to help identify isomorphic graphs.

Remarks Here are some properties of the adjacency matrix of an undirected graph.

1. The adjacency matrix is always symmetric.
2. The vertices must be ordered: and the adjacency matrix depends on the order chosen.
3. An adjacency matrix can be defined for multigraphs by defining a_{ij} to be the **number** of edges between vertices i and j.
4. If there is a natural order on the set of vertices we will use that order unless otherwise indicated.

3.2. Example 3.2.1.

EXAMPLE 3.2.1. *An adjacency matrix for the graph*



is the matrix

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

Discussion

To find this matrix we may use a table as follows. First we set up a table labeling the rows and columns with the vertices.

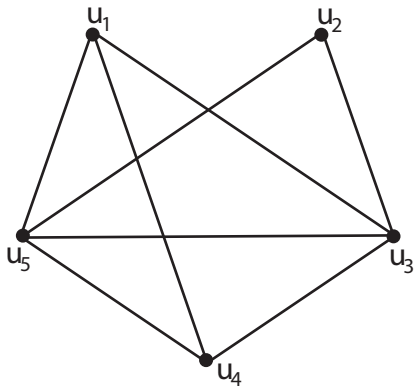
	v_1	v_2	v_3	v_4	v_5
v_1					
v_2					
v_3					
v_4					
v_5					

Since there are edges from v_1 to v_2, v_4, and v_5, but no edge between v_1 and itself or v_3, we fill in the first row and column as follows.

	v_1	v_2	v_3	v_4	v_5
v_1	0	1	0	1	1
v_2	1				
v_3	0				
v_4	1				
v_5	1				

We continue in this manner to fill the table with 0's and 1's. The matrix may then be read straight from the table.

EXAMPLE 3.2.2. *The adjacency matrix for the graph*



is the matrix

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

3.3. Incidence Matrices.

DEFINITION 3.3.1. *The* **incidence matrix**, $A = [a_{ij}]$, *for the undirected graph* $G = (V, E)$ *is defined by*

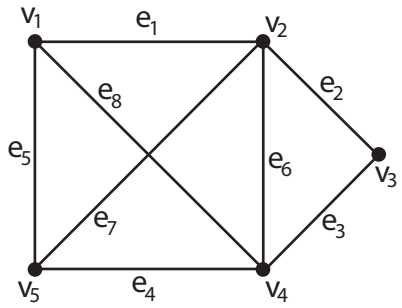
$$a_{ij} = \begin{cases} 1 & \text{if edge } j \text{ is incident with vertex } i \\ 0 & \text{otherwise.} \end{cases}$$

Discussion

Remarks:

(1) This method requires the edges and vertices to be labeled and depends on the order in which they are written.
(2) Every column will have exactly two 1's.
(3) As with adjacency matrices, if there is a natural order for the vertices and edges that order will be used unless otherwise specified.

EXAMPLE 3.3.1. *The incidence matrix for the graph*



is the matrix

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

Again you can use a table to get the matrix. List all the vertices as the labels for the rows and all the edges for the labels of the columns.

3.4. Degree Sequence.

DEFINITION 3.4.1. *The* **degree sequence** *a graph G with n vertices is the sequence $(d_1, d_2, ..., d_n)$, where $d_1, d_2, ..., d_n$ are the degrees of the vertices of G and $d_1 \geq d_2 \geq \cdots \geq d_n$.*

Note that a graph could conceivably have infinitely many vertices. If the vertices are *countable* then the degree sequence would be an infinite sequence. If the vertices are not countable, then this degree sequence would not be defined.

3.5. Graph Invariants.

DEFINITION 3.5.1. *We say a property of graphs is a* **graph invariant** *(or, just invariant) if, whenever a graph G has the property, any graph isomorphic to G also has the property.*

THEOREM 3.5.1. *The following are invariants of a graph G:*

(1) G has r vertices.
(2) G has s edges.
(3) G has degree sequence $(d_1, d_2, ..., d_n)$.
(4) G is a bipartite graph.
(5) G contains r complete graphs K_n (as a subgraphs).
(6) G contains r complete bipartite graphs $K_{m,n}$.
(7) G contains r n-cycles.
(8) G contains r n-wheels.
(9) G contains r n-cubes.

Discussion

Recall from *Module 6.1 Introduction to Graphs* that two simple graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ are isomorphic if there is a bijection

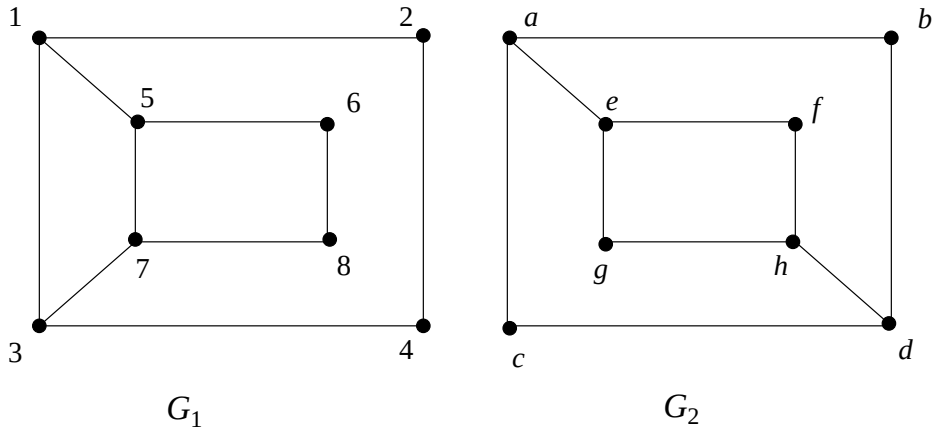
$$f : V_1 \to V_2$$

such that vertices u and v in V_1 are adjacent in G_1 if and only if $f(u)$ and $f(v)$ are adjacent in G_2. If there is such a function, we say f is an **isomorphism** and we write $G_1 \simeq G_2$.

It is often easier to determine when two graphs are *not* isomorphic. This is sometimes made possible by comparing *invariants* of the two graphs to see if they are different. The invariants in Theorem 3.5.1 may help us determine fairly quickly in some examples that two graphs are **not** isomorphic.

3.6. Example 3.6.1.

EXAMPLE 3.6.1. *Show that the following two graphs are not isomorphic.*



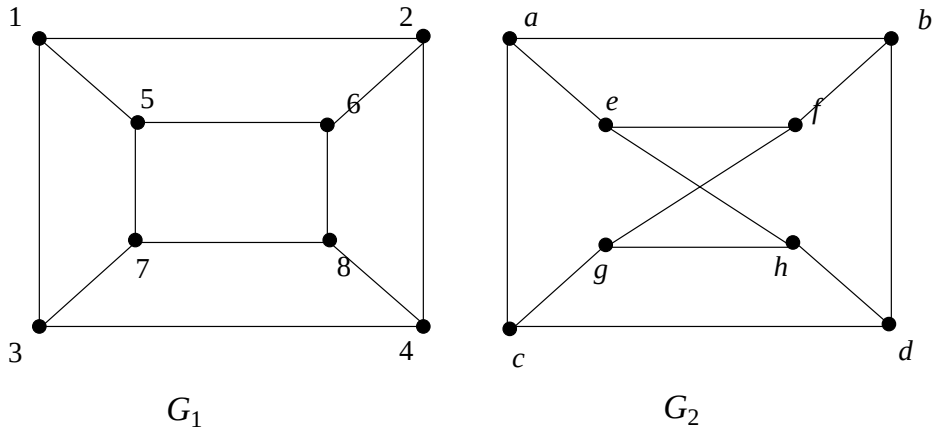
G_1 G_2

The two graphs have the same number of vertices, the same number of edges, and same degree sequences $(3, 3, 3, 3, 2, 2, 2, 2)$. *Perhaps the easiest way to see that they*

are not isomorphic is to observe that G_2 has only two 4-cycles, whereas G_1 has three 4-cycles. In fact, the four vertices of G_1 of degree 3 lie in a 4-cycle in G_1, but the four vertices of G_2 of degree 3 do not. Either of these two discrepancies is enough to show that the graphs are not isomorphic.

Another way we could recognize the graphs above are not isomorphic is to consider the adjacency relationships. Notice in G_1 all the vertices of degree 3 are adjacent to 2 vertices of degree 3 and 1 of degree 2. However, in graph G_2 all of the vertices of degree 3 are adjacent to 1 vertex of degree 3 and 2 vertices of degree 3. This discrepancy indicates the two graphs cannot be isomorphic.

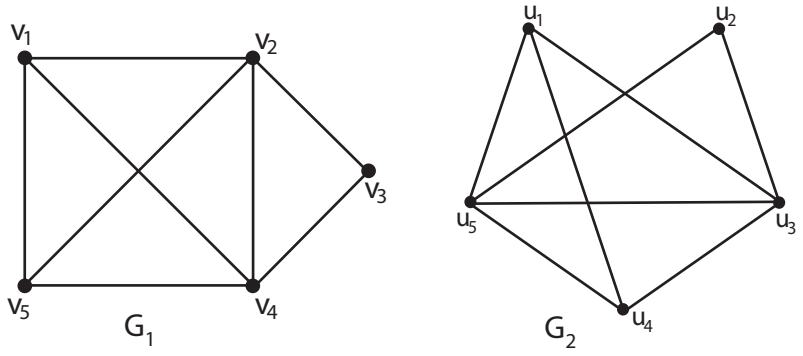
EXERCISE 3.6.1. *Show that the following two graphs are not isomorphic.*



G_1 G_2

3.7. Example .

EXAMPLE 3.7.1. *Determine whether the graphs G_1 and G_2 are isomorphic.*



Solution: We go through the following checklist that might tell us immediately if the two are not isomorphic.

- *They have the same number of vertices, 5.*
- *They have the same number of edges, 8.*
- *They have the same degree sequence* $(4, 4, 3, 3, 2)$.

Since there is no obvious reason to think they are not isomorphic, we try to construct an isomorphism, f. *(Note that the above does not tell us there is an isomorphism, only that there might be one.)*

The only vertex on each that have degree 2 are v_3 *and* u_2*, so we must have* $f(v_3) = u_2$.

Now, since $deg(v_1) = deg(v_5) = deg(u_1) = deg(u_4)$*, we must have either*

- $f(v_1) = u_1$ *and* $f(v_5) = u_4$*, or*
- $f(v_1) = u_4$ *and* $f(v_5) = u_1$.

It is possible only one choice would work or both choices may work (or neither choice may work, which would tell us there is no isomorphism).

We try $f(v_1) = u_1$ *and* $f(v_5) = u_4$.

Similarly we have two choices with the remaining vertices and try $f(v_2) = u_3$ *and* $f(v_4) = u_5$*. This defines a bijection from the vertices of* G_1 *to the vertices of* G_2*. We still need to check that adjacent vertices in* G_1 *are mapped to adjacent vertices in* G_2*. To check this we will look at the adjacency matrices.*

The adjacency matrix for G_1 *(when we list the vetices of* G_1 *by* v_1, v_2, v_3, v_4, v_5*) is*

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

We create an adjacency matrix for G_2*, using the bijection* f *as follows: since* $f(v_1) = u_1$*,* $f(v_2) = u_3$*,* $f(v_3) = u_2$*,* $f(v_4) = u_5$*, and* $f(v_5) = u_4$*, we rearrange the order of the vertices of* G_2 *to* u_1, u_3, u_2, u_5, u_4*. With this ordering, the adjacency*

matrix for G_2 is

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

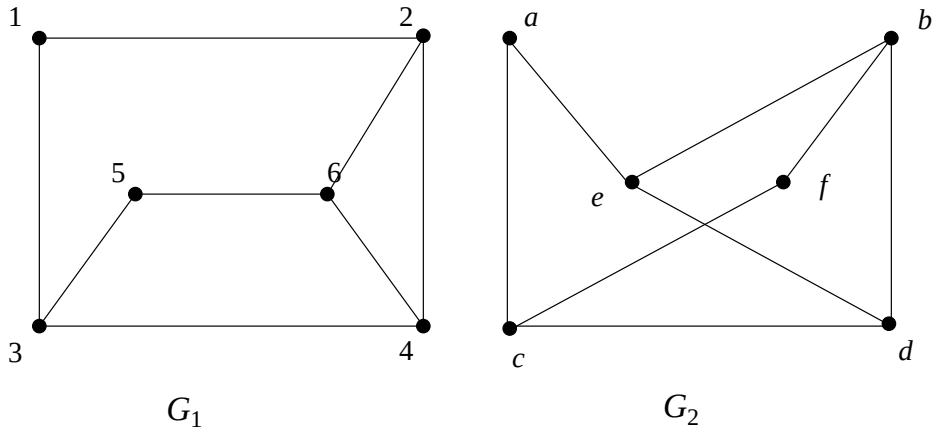
Since $A = B$, adjacency is preserved under this bijection. Hence the graphs are isomorphic.

Discussion

Notice that, trying to establish that the two graphs are isomorphic, it is *not* enough to show that they have the same number of vertices, edges, and degree sequence. In fact, if we knew they were isomorphic and we were asked to prove it, we would proceed to trying to find a bijection that preserves adjacency. That is, the check list is not necessary if you already know they are isomorphic. On the other hand, having found a bijection between two graphs that doesn't preserve adjacency doesn't tell us the graphs are not isomorphic, because some other bijection that would work. If we go down this path, we would have to show that *every* bijection fails to preserve adjacency.

The advantage of the checklist is that it will give you a quick and easy way to show two graphs are *not* isomorphic if some invariant of the graphs turn out to be different. If you examine the logic, however, you will see that if two graphs have all of the same invariants we have listed so far, we still wouldn't have a proof that they are isomorphic. Indeed, there is no known list of invariants that can be efficiently checked to determine when two graphs are isomorphic. The best algorithms known to date for determining graph isomorphism have exponential complexity (in the number n of vertices).

EXERCISE 3.7.1. *Determine whether the following two graphs are isomorphic.*



EXERCISE 3.7.2. *How many different isomorphism (that is, bijections that preserve adjacencies) are possible between the graphs in Example 3.7.1?*

EXERCISE 3.7.3. *There are 14 nonisomorphic pseudographs with 3 vertices and 3 edges. Draw all of them.*

EXERCISE 3.7.4. *How many nonisomorphic simple graphs with 6 vertices, 5 edges, and no cycles are there. In other words, how many different simple graphs satisfying the criteria that it have 6 vertices, 5 edges, and no cycles can be drawn so that no two of the graphs are isomorphic?*

3.8. Proof of Theorem 3.5.1 Part 3 for finite simple graphs.

PROOF. Let G_1 and G_2 be isomorphic finite simple graphs having degree sequences. By part 1 of Theorem 3.5.1 the degree sequences of G_1 and G_2 have the same number of elements. Let $f : V(G_1) \to V(G_2)$ be an isomorphism and let $v \in V(G_1)$. We claim $deg_{G_1}(v) = deg_{G_2}(f(v))$. If we show this, then f defines a bijection between the vertices of G_1 and G_2 that maps vertices to vertices of the same degree. This will imply the degree sequences are the same.

Proof of claim: Suppose $deg_{G_1}(v) = k$. Then there are k vertices in G_1 adjacent to v, say $u_1, u_2, \ldots, u_k$. The isomorphism maps each of the vertices to k distinct vertices adjacent to $f(u)$ in G_2 since the isomorphism is a bijection and preserves adjacency. Thus $deg_{G_2}(f(u)) \geq k$. Suppose $deg_{G_2}(f(u)) > k$. Then there would be a vertex, $w_{k+1} \in V(G_2)$, not equal to any of the vertices $f(u_1), \ldots, f(u_k)$, and adjacent to $f(u)$. Since f is a bijection there is a vertex u_{k+1} in G_1 that is not equal to any of $u_1, \ldots, u_k$ such that $f(u_{k+1}) = w_{k+1}$. Since f preserves adjacency we would have u_{k+1} and v are adjacent. But this contradicts that $deg_{G_1}(v) = k$. Thus $deg_{G_2}(f(u)) = k = deg_{G_1}(u)$ □

EXERCISE 3.8.1. *Prove the first 2 properties listed in Theorem 3.5.1 for finite simple graphs using only the properties listed before each and the definition of isomorphism.*

CHAPTER 7

Introduction to Relations

1. Relations and Their Properties

1.1. Definition of a Relation. Definition: A **binary relation from a set** A **to a set** B is a subset

$$R \subseteq A \times B.$$

If $(a, b) \in R$ we say a is **related** to b by R.

A is the **domain** of R, and

B is the **codomain** of R.

If $A = B$, R is called a **binary relation on the set** A.

Notation:

- If $(a, b) \in R$, then we write aRb.
- If $(a, b) \notin R$, then we write $a \not R b$.

Discussion

Notice that a relation is simply a subset of $A \times B$. If $(a, b) \in R$, where R is some relation from A to B, we think of a as being assigned to b. In these senses students often associate relations with functions. In fact, a function is a special case of a relation as you will see in Example 1.2.4. Be warned, however, that a relation may differ from a function in two possible ways. If R is an arbitrary relation from A to B, then

- it is possible to have both $(a, b) \in R$ and $(a, b') \in R$, where $b' \neq b$; that is, an element in A could be related to any number of elements of B; or
- it is possible to have some element a in A that is not related to any element in B at all.

Often the relations in our examples do have special properties, but be careful not to assume that a given relation must have any of these properties.

1.2. Examples.

EXAMPLE 1.2.1. *Let $A = \{a, b, c\}$ and $B = \{1, 2, 3, 4\}$, and let $R_1 = \{(a, 1), (a, 2), (c, 4)\}$.*

EXAMPLE 1.2.2. *Let $R_2 \subset \mathbb{N} \times \mathbb{N}$ be defined by $(m, n) \in R_2$ if and only if $m|n$.*

EXAMPLE 1.2.3. *Let A be the set of all FSU students, and B the set of all courses offered at FSU. Define R_3 as a relation from A to B by $(s, c) \in R_3$ if and only if s is enrolled in c this term.*

Discussion

There are many different types of examples of relations. The previous examples give three very different types of examples. Let's look a little more closely at these examples.

Example 1.2.1. This is a completely abstract relation. There is no obvious reason for a to be related to 1 and 2. It just is. This kind of relation, while not having any obvious application, is often useful to demonstrate properties of relations.

Example 1.2.2. This relation is one you will see more frequently. The set R_2 is an infinite set, so it is impossible to list all the elements of R_2, but here are some elements of R_2:

$$(2, 6), (4, 8), (5, 5), (5, 0), (6, 0), (6, 18), (2, 18).$$

Equivalently, we could also write

$$2R_26, 4R_28, 5R_25, 5R_20, 6R_20, 6R_218, 2R_218.$$

Here are some elements of $\mathbb{N} \times \mathbb{N}$ that are **not** elements of R_2:

$$(6, 2), (8, 4), (2, 5), (0, 5), (0, 6), (18, 6), (6, 8), (8, 6).$$

Example 1.2.3. Here is an element of R_3: (you, MAD2104).

EXAMPLE 1.2.4. *Let A and B be sets and let $f\colon A \to B$ be a function. The graph of f, defined by graph$(f) = \{(x, f(x))|x \in A\}$, is a relation from A to B.*

Notice the previous example illustrates that any function has a relation that is associated with it. However, not all relations have functions associated with them.

EXERCISE 1.2.1. *Suppose $f : \mathbb{R} \to \mathbb{R}$ is defined by $f(x) = \lfloor x/2 \rfloor$.*

(1) Find 5 elements of the relation graph(f).

(2) Find 5 elements of $\mathbb{R} \times \mathbb{R}$ *that are not in graph*(f).

EXERCISE 1.2.2. *Find a relation from* $\mathbb{R}$ *to* $\mathbb{R}$ *that cannot be represented as the graph of a functions.*

EXERCISE 1.2.3. *Let* n *be a positive integer. How many binary relations are there on a set* A *if* $|A| = n$*? [Hint: How many elements are there in* $|A \times A|$*?]*

1.3. Directed Graphs.

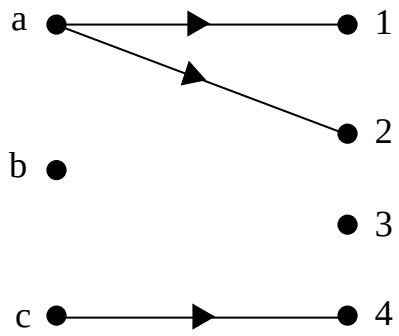
DEFINITIONS 1.3.1.

- *A* **directed graph** *or a* **digraph** D *from* A *to* B *is a collection of* **vertices** $V \subseteq A \cup B$ *and a collection of* **edges** $R \subseteq A \times B$.
- *If there is an ordered pair* $e = (x, y)$ *in* R *then there is an* **arc** *or* **edge** *from* x *to* y *in* D.
- *The elements* x *and* y *are called the* **initial** *and* **terminal** *vertices of the edge* $e = (x, y)$*, respectively.*

Discussion

A digraph can be a useful device for representing a relation, especially if the relation isn't "too large" or complicated.

The digraph that represents R_1 in Example 1.2.1 is:

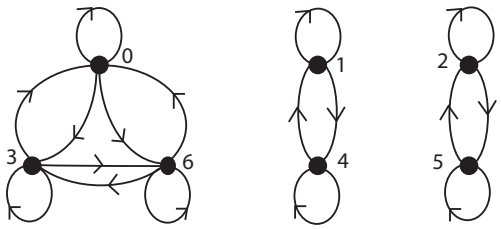


Discussion

If R is a relation on a set A, we simplify the digraph D representing R by having only one vertex for each $a \in A$. This results, however, in the possibility of having **loops**, that is, edges from a vertex to itself, and having more than one edge joining distinct vertices (but with opposite orientations).

A digraph for R_2 in Example 1.2.2 would be difficult to illustrate (and impossible to draw completely), since it would require infinitely many vertices and edges. We could draw a digraph for some finite subset of R_2. It is possible to indicate what the graph of some infinite relations might look like, but this one would be particularly difficult.

EXAMPLE 1.3.1. *Let R_5 be the relation from $\{0,1,2,3,4,5,6\}$ defined by mR_5n if and only if $m \equiv n (mod\ 3)$. The digraph that represents R_5 is*



1.4. Inverse Relation.

DEFINITION 1.4.1. *Let R be a relation from A to B. Then $R^{-1} = \{(b,a)|(a,b) \in R\}$ is a relation from B to A.*

R^{-1} is called the **inverse of the relation** *R.*

Discussion

The inverse of a relation R is simply the relation obtained by reversing the ordered pairs of R. The inverse relation is also called the **converse relation**.

EXAMPLE 1.4.1. *Recall Example 1.2.1 $A = \{a,b,c\}$ and $B = \{1,2,3,4\}$ and $R_1 = \{(a,1),(a,2),(c,4)\}$. Then $R^{-1} = \{(1,a),(2,a),(4,a)\}$.*

EXERCISE 1.4.1. *Recall Example 1.2.4. A and B are sets and $f\colon A \to B$ is a function. The graph of f, $graph(f) = \{(x,f(x))|x \in A\}$ is a relation from A to B.*

(1) What is the inverse of this relation?
(2) Does f have to be invertible for the inverse of this relation to exist?
(3) If f is invertible, find the inverse of the relation $graph(f)$ in terms of the inverse function f^{-1}.

1.5. Special Properties of Binary Relations.

DEFINITIONS 1.5.1. *Let A be a set, and let R be a binary relation on A.*

(1) R *is* **reflexive** *if*
$\forall x[(x \in A) \to ((x, x) \in R)]$.
(2) R *is* **irreflexive** *if*
$\forall x[(x \in A) \to ((x, x) \notin R)]$.
(3) R *is* **symmetric** *if*
$\forall x \forall y[((x, y) \in R) \to ((y, x) \in R)]$.
(4) R *is* **antisymmetric** *if*
$\forall x \forall y[([(x, y) \in R] \wedge [(y, x) \in R]) \to (x = y)]$.
(5) R *is* **asymmetric** *if*
$\forall x \forall y[((x, y) \in R) \to ((y, x) \notin R)]$.
(6) R *is* **transitive** *if*
$\forall x \forall y \forall z[([(x, y) \in R] \wedge [(y, z) \in R]) \to ((x, z) \in R)]$.

Discussion

Study the definitions of the definitions of the properties given above. You must know these properties, be able to recognize whether or not a relation has a particular property, and be able to prove that a relation has or does not have a particular property. Notice that the definitions of reflexive and irreflexive relations are not complementary. That is, a relation on a set may be both reflexive and irreflexive or it may be neither. The same is true for the symmetric and antisymmetric properties, as well as the symmetric and asymmetric properties. Some texts use the term antireflexive for irreflexive.

EXERCISE 1.5.1. *Before reading further, find a relation on the set $\{a, b, c\}$ that is neither*

(a) reflexive nor irreflexive.
(b) symmetric nor antisymmetric.
(c) symmetric nor asymmetric.

1.6. Examples of Relations and their Properties.

EXAMPLE 1.6.1. *Suppose A is the set of FSU students and R is the relation given by aRb if students a and b have the same last name. This relation is...*

- *reflexive*
- *not irreflexive*
- *symmetric*
- *not antisymmetric*

- *not asymmetric*
- *transitive*

EXAMPLE 1.6.2. *Suppose T is the relation on the set of integers given by xTy if $2x - y = 1$. This relation is...*

- *not reflexive*
- *not irreflexive*
- *not symmetric*
- *antisymmetric*
- *not asymmetric*
- *not transitive*

EXAMPLE 1.6.3. *Suppose $A = \{a, b, c, d\}$ and R is the relation $\{(a, a)\}$. This relation is...*

- *not reflexive*
- *not irreflexive*
- *symmetric*
- *antisymmetric*
- *not asymmetric*
- *transitive*

Discussion

The examples above illustrate three rather different relations. Some of the relations have many of the properties defined on Section 1.5, whereas one has only one of the property. It is entirely possible to create a relation with none of the properties given in Section 1.5.

EXERCISE 1.6.1. *Give an example of a relation that does not satisfy any property given in Section 1.5.*

1.7. Proving or disproving relations have a property.

EXAMPLE 1.7.1. *Suppose T is the relation on the set of integers given by xTy if $2x - y = 1$. This relation is*

- *not reflexive*

  PROOF. 2 is an integer and $2 \cdot 2 - 2 = 2 \neq 1$. This shows that $\forall x[x \in \mathbb{Z} \to (x, x) \in T]$ is **not true**. □

- *not irreflexive*

  PROOF. 1 is an integer and $2 \cdot 1 - 1 = 1$. This shows that $\forall x[x \in \mathbb{Z} \to (x, x) \notin T]$ is **not true**. □

- *not symmetric*

  PROOF. Both 2 and 3 are integers, $2 \cdot 2 - 3 = 1$, and $2 \cdot 3 - 2 = 4 \neq 1$. This shows $2R3$, but $3 \not R 2$; that is, $\forall x \forall y[(x, y) \in \mathbf{Z} \rightarrow (y, x) \in T]$ is **not true**. □

- *antisymmetric*

  PROOF. Let $m, n \in \mathbb{Z}$ be such that $(m, n) \in T$ and $(n, m) \in T$. By the definition of T, this implies both equations $2m - n = 1$ and $2n - m = 1$ must hold. We may use the first equation to solve for n, $n = 2m - 1$, and substitute this in for n in the second equation to get $2(2m-1) - m = 1$. We may use this equation to solve for m and we find $m = 1$. Now solve for n and we get $n = 1$.

  This shows that the only integers, m and n, such that both equations $2m - n = 1$ and $2n - m = 1$ hold are $m = n = 1$. This shows that $\forall m \forall n[((m, n) \in T \wedge (n, m) \in T) \rightarrow m = n]$. □

- *not asymmetric*

  PROOF. 1 is an integer such that $(1, 1) \in T$. Thus $\forall x \forall y[((x, y) \in T \rightarrow (b, a) \notin T]$ is **not true** (counterexample is $a = b = 1$). □

- *not transitive*

  PROOF. 2, 3, and 5 are integers such that $(2, 3) \in T$, $(3, 5) \in T$, but $(2, 5) \notin T$. This shows $\forall x \forall y \forall z[(x, y) \in T \wedge (y, z) \in T \rightarrow (x, z) \in T]$ is **not true**. □

EXAMPLE 1.7.2. *Recall Example 1.2.2: $R_2 \subset \mathbb{N} \times \mathbb{N}$ was defined by $(m, n) \in R_2$ if and only if $m|n$.*

- *reflexive*

  PROOF. Since $n|n$ for all integers, n, we have nR_2n for every integer. This shows R_2 is reflexive. □

- *not irreflexive*

  PROOF. 1 is an integer and clearly $1R_21$. This shows R_2 is not irreflexive. (you could use any natural number to show R_2 is not irreflexive). □

- *not symmetric*

  PROOF. 2 and 4 are natural numbers with $2|4$, but $4 \not| 2$, so $2R_24$, but $4 \not R_2 2$. This shows R_2 is not reflexive. □

- *antisymmetric*

  PROOF. Let $n, m \in \mathbb{N}$ be such that nR_2m and mR_2n. By the definition of R_2 this implies $n|m$ and $m|n$. Hence we must have $m = n$. This shows R_2 is antisymmetric. □

- *not asymmetric*

  PROOF. Let $m = n$ be any natural number. Then nR_2m and mR_2n, which shows R_2 is not asymetric. (You may use a particular number to show R_2 is not asymmetric. □

- *transitive*

  PROOF. Let $p, q, r \in \mathbb{N}$ and assume pR_2q and qR_2r. By the definition of R_2 this means $p|q$ and $q|r$. We have proven in *Integers and Division* that this implies $p|r$, thus pR_2r. This shows R_2 is transitive. □

Discussion

When proving a relation, R, on a set A **has** a particular property, the property must be shown to hold for all possible members of the set. For example, if you wish to prove that a given relation, R, on A is reflexive, you must take an arbitrary element x from A and show that xRx. Some properties, such as the symmetric property, are defined using implications. For example, if you are asked to show that a relation, R, on A is symmetric, you would suppose that x and y are arbitrary elements of A such that xRy, and then try to prove that yRx. It is possible that a property defined by an implication holds **vacuously** or **trivially**.

EXERCISE 1.7.1. *Let R be the relation on the set of real numbers given by xRy if and only if $x < y$. Prove R is antisymmetric.*

When proving R does **not** have a property, it is enough to give a counterexample. Recall $\neg[\forall x \forall y P(x, y)] \Leftrightarrow \exists x \exists y \neg P(x, y)$.

EXERCISE 1.7.2. *Prove whether or not each of the properties in Section 1.5 holds for the relation in Example 1.6.1.*

EXERCISE 1.7.3. *Prove whether or not each of the properties in Section 1.5 holds for the relation in Example 1.6.3.*

1.8. Combining Relations. Important Question: Suppose property P is one of the properties listed in Section 1.5, and suppose R and S are relations on a set A, each having property P. Then the following questions naturally arise.

(1) Does $\overline{R}$ (necessarily) have property P?
(2) Does $R \cup S$ have property P?
(3) Does $R \cap S$ have property P?
(4) Does $R - S$ have property P?

1.9. Example of Combining Relations.

EXAMPLE 1.9.1. *Let R_1 and R_2 be transitive relations on a set A. Does it follow that $R_1 \cup R_2$ is transitive?*

Solution: No. Here is a counterexample:

$$A = \{1, 2\}, \quad R_1 = \{(1, 2)\}, \quad R_2 = \{(2, 1)\}$$

$$\textit{Therefore,} \qquad R_1 \cup R_2 = \{(1, 2), (2, 1)\}$$

Notice that R_1 and R_2 are both transitive (vacuously, since there are no two elements satisfying the conditions of the property). However $R_1 \cup R_2$ is not transitive. If it were it would have to have $(1, 1)$ and $(2, 2)$ in $R_1 \cup R_2$.

Discussion

Example 1.9.1 gives a counterexample to show that the union of two transitive relations is not necessarily transitive. Note that you could find an example of two transitive relations whose union *is* transitive. However, the question asks if the given property holds for two relations must it hold for the binary operation of the two relations. This is a general question and to give the answer "yes" we must know it is true for *every* possible pair of relations satisfying the property.

Here is another example:

EXAMPLE 1.9.2. *Suppose R and S are transitive relations on the set A. Is $R \cap S$ transitive?*

Solution: Yes.

PROOF. Assume R and S are both transitive and let $(a, b), (b, c) \in R \cap S$. Then $(a, b), (b, c) \in R$ and $(a, b), (b, c) \in S$. It is given that both R and S are transitive, so $(a, c) \in R$ and $(a, c) \in S$. Therefore $(a, c) \in R \cap S$. This shows that for arbitrary $(a, b), (b, c) \in R \cap S$ we have $(a, c) \in R \cap S$. Thus $R \cap S$ is transitive. □

1.10. Composition.

DEFINITIONS 1.10.1.

(1) Let

- *R_1 be a relation from A to B, and*
- *R_2 be a relation from B to C.*

Then the **composition** *of R_1 with R_2, denoted $R_2 \circ R_1$, is the relation from A to C defined by the following property:*

$(x, z) \in R_2 \circ R_1$ *if and only if there is a $y \in B$ such that $(x, y) \in R_1$ and $(y, z) \in R_2$.*

(2) Let R be a binary relation on A. Then R^n is defined recursively as follows:

Basis: $R^1 = R$

Recurrence: $R^{n+1} = R^n \circ R$, *if* $n \geq 1$.

Discussion

The composition of two relations can be thought of as a generalization of the composition of two functions, as the following exercise shows.

EXERCISE 1.10.1. *Prove: If $f\colon A \to B$ and $g\colon B \to C$ are functions, then $graph(g \circ f) = graph(g) \circ graph(f)$.*

EXERCISE 1.10.2. *Prove the composition of relations is an associative operation.*

EXERCISE 1.10.3. *Let R be a relation on A. Prove $R^n \circ R = R \circ R^n$ using the previous exercise and induction.*

EXERCISE 1.10.4. *Prove an ordered pair $(x, y) \in R^n$ if and only if, in the digraph D of R, there is a directed path of length n from x to y.*

Notice that if there is no element of B such that $(a, b) \in R_1$ and $(b, c) \in R_2$ for some $a \in A$ and $c \in C$, then the composition is empty.

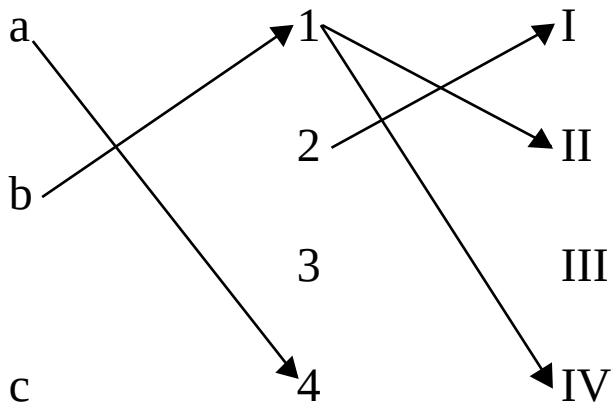
1.11. Example of Composition.

EXAMPLE 1.11.1. *Let $A = \{a, b, c\}$, $B = \{1, 2, 3, 4\}$, and $C = \{I, II, III, IV\}$.*

- $R_1 = \{(a, 4), (b, 1)\}$
- $R_2 = \{(1, II), (1, IV), (2, I)\}$
- *Then* $R_2 \circ R_1 = \{(b, II), (b, IV)\}$

Discussion

It can help to consider the following type of diagram when discussing composition of relations, such as the ones in Example 1.11.1 as shown here.



EXAMPLE 1.11.2. *If R and S are transitive binary relations on A, is $R \circ S$ transitive?*

Solution: No. Here is a counterexample: Let

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

Then both R and S are transitive (vacuously). However,

$$R \circ S = \{(2,4),(4,2)\}$$

is not transitive. (Why?)

EXAMPLE 1.11.3. *Suppose R is the relation on $\mathbf{Z}$ defined by aRb if and only if $a|b$. Then $R^2 = R$.*

EXERCISE 1.11.1. *Let R be the relation on the set of real numbers given by xRy if and only if $\frac{x}{y} = 2$.*

(1) Describe the relation R^2.
(2) Describe the relation R^n.

EXERCISE 1.11.2. *Let P be a property given below and let R and S be relations on A satisfying property P. When does the relation obtained by combining R and S using the operation given satisfy property P?*

(1) P is the reflexive property.
  (a) $R \cup S$
  (b) $R \cap S$
  (c) $R \oplus S$
  (d) $R - S$
  (e) $R \circ S$
  (f) R^{-1}
  (g) R^n
(2) P is the symmetric property.

(a) $R \cup S$
(b) $R \cap S$
(c) $R \oplus S$
(d) $R - S$
(e) $R \circ S$
(f) R^{-1}
(g) R^n

(3) *P is the transitive property.*
(a) $R \cup S$
(b) $R \cap S$
(c) $R \oplus S$
(d) $R - S$
(e) $R \circ S$
(f) R^{-1}
(g) R^n

1.12. Characterization of Transitive Relations.

THEOREM 1.12.1. *Let R be a binary relation on a set A. R is transitive if and only if $R^n \subseteq R$, for $n \geq 1$.*

PROOF. To prove (R transitive) $\rightarrow$ ($R^n \subseteq R$) we assume R is transitive and prove $R^n \subseteq R$ for $n \geq 1$ by induction.

Basis Step, $n = 1$. $R^1 = R$, so this is obviously true.

Induction Step. Prove $R^n \subseteq R \rightarrow R^{n+1} \subseteq R$.

Assume $R^n \subseteq R$ for some $n \geq 1$. Suppose $(x, y) \in R^{n+1}$. By definition $R^{n+1} = R^n \circ R$, so there must be some $a \in A$ such that $(x, a) \in R$ and $(a, y) \in R^n$. However, by the induction hypothesis $R^n \subseteq R$, so $(a, y) \in R$. R is transitive, though, so $(x, a), (a, y) \in R$ implies $(x, y) \in R$. Since (x, y) was an arbitrary element of R^{n+1}, this shows $R^{n+1} \subseteq R$.

Now we must show the other direction : $R^n \subseteq R$, for $n \geq 1$, implies R is transitive. We prove this directly.

Assume $(x, y), (y, z) \in R$. But by the definition of composition, this implies $(x, z) \in R^2$. But $R^2 \subseteq R$, so $(x, z) \in R$. This shows R is transitive.

□

Discussion

Theorem 1.12.1 gives an important theorem characterizing the transitivity relation. Notice that, since the statement of the theorem was a property that was to be proven for all positive integers, induction was a natural choice for the proof.

EXERCISE 1.12.1. *Prove that a relation R on a set A is transitive if and only if $R^2 \subseteq R$. [Hint: Examine not only the statement, but the proof of Theorem 1.12.1.]*