Mathematical Reasoning I
Course Notes
Fall 2005

David Lyons
Mathematical Sciences
Lebanon Valley College

Mathematical Reasoning I
Course Notes

Fall 2005

David Lyons
Mathematical Sciences
Lebanon Valley College

Contents

0 Sets and Functions

The vocabulary of sets and functions is fundamental to all of mathematics, theoretical and applied. We present basic terminology and examples in this section.

0.1 Definitions for sets

A *set* is a collection of objects. The objects belonging to a set are called its *elements*, *members* or *points*. For example, the members of the set of odd digits are 1, 3, 5, 7 and 9. The word *points* for members of a set comes from geometry. For example, a line is a set of points in the plane or in space.

To indicate that a set A consists of elements called x, y and z, we write

$$A = \{x, y, z\}$$

listing the elements, separated by commas, inside curly brackets. The order in which the elements are listed does not matter, nor does redundancy in the listing. For example, we can say the following for the set A above.

$$A = \{x, y, z\} = \{y, z, x\} = \{x, x, y, z\}$$

(0.1.1) **Set builder notation.**

When a set has more than a few members, we usually describe the set rather than list all of its elements. For example, it is easier to say "the set of even whole numbers" than it is to list all of the even whole numbers. The mathematical way to write such verbal descriptions is called *set builder notation*, and has the following form.

$$\{x \mid \text{(a statement, in which } x \text{ is the subject)}\}$$

This denotes the set of all objects for which the statement is true. For example, the set of even whole numbers can be written like this.

$$\{x \mid x \text{ is an even whole number}\}$$

Sometimes a colon is used instead of the vertical bar. Here are some more examples. The set of positive real numbers can be written $\{x \colon x > 0\}$. The set of United States Citizens can be written $\{x \mid x \text{ is a US citizen}\}$. The set of points in the x, y-plane that lie above the x-axis ("above" means on the positive y-axis side of the x-axis) can be written $\{(x, y) \colon y > 0\}$.

A summary of vocabulary and special notation used for sets is given in (0.1.2).

(0.1.2) **Set vocabulary and notation.**

$\{\ldots\}$ (set brackets) indicates a set with a list or description of set members between the brackets

$\{x \mid p(x)\}$ (set builder notation, see (0.1.1)) denotes the set of all objects x for which the statement $p(x)$ holds true

$\{x\colon p(x)\}$ alternate form of set builder notation

$|A|$ the number of elements in a finite set A

$x \in A$ the object x is a member of the set A

$\emptyset$ the empty set (the set with no members)

$A \subseteq B$ (pronounced "A is contained in B," or "A is a subset of B") every member of the set A is also a member of the set B

$A \cap B$ (pronounced "A intersect B") the set of all objects which are members of *both* set A *and* set B

$A \cup B$ (pronounced "A union B") the set of all objects which are members of *either* set A *or* set B, *or both*

$A \setminus B$ (pronounced "A minus B") the set of all objects which are members of the set A *and are not* members of the set B

set difference refers to a set of the form $A \setminus B$

disjoint two sets are disjoint if their intersection is empty

ordered pair an ordered list (x, y) of two elements from a set, where we allow the possibility that x equals y

$A \times B$ (pronounced "A times B," or "(Cartesian) product of A and B," named after Rene Descartes) the set of all ordered pairs $(a, b) \in A \cup B$ such that $a \in A$ and $b \in B$

A^2 (pronounced "A squared") the product $A \times A$ of a set A with itself

(0.1.3) **Examples to illustrate set notation.** Let $A = \{1, 2, 3, 7\}$, $B = \{1, 2, 4, 5\}$ and $C = \{3, 5\}$ be subsets of the set of digits. Then we have the following.

$$
\begin{aligned}
\{x \in A \mid x > 1\} &= \{2, 3, 7\} \\
\{x\colon x \in A \text{ or } x \in C\} &= \{1, 2, 3, 5, 7\} \\
A \cap B &= \{1, 2\} \\
A \cup B &= \{1, 2, 3, 4, 5, 7\} \\
A \setminus B &= \{3, 7\} \\
B \setminus A &= \{4, 5\} \\
|A| &= 4 \\
|A \cup B| &= 6 \\
A \times C &= \{(1, 3), (1, 5), (2, 3), (2, 5), (3, 3), (3, 5), (7, 3), (7, 5)\} \\
C^2 &= C \times C = \{(3, 3), (3, 5), (5, 3), (5, 5)\}
\end{aligned}
$$

0.2 Venn diagrams

It is often helpful to use pictures to visualize the relationships between sets. *Venn diagrams* depict sets as 2-dimensional regions in the plane. Figure 1 shows a Venn diagram illustrating the relationships between the set R of all rectangles, the set S of all squares, the set T of all triangles and the set P of all polygons. Venn diagrams illustrating intersection, union and set difference involving sets A and B from example (0.1.3) are shown in Figures 2–4.

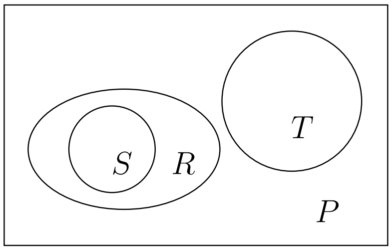


Figure 1
Venn diagram example

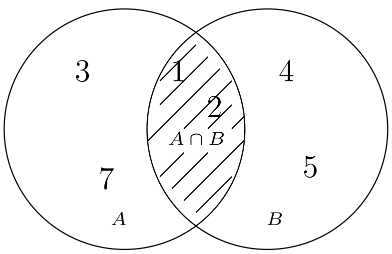


Figure 2
$A \cap B = \{1, 2\}$

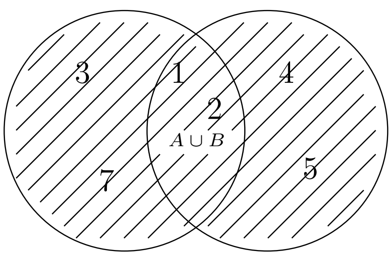


Figure 3
$A \cup B = \{1, 2, 3, 4, 5, 7\}$

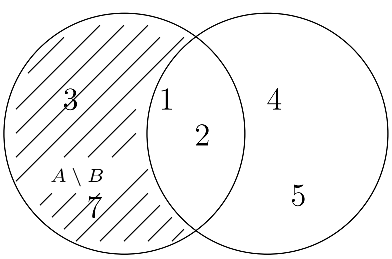


Figure 4
$A \setminus B = \{3, 7\}$

0.3 Some important sets

The Real Numbers

One of the most important sets in mathematics is the set $\mathbf{R}$ of real numbers, which is the set of points on a line. The name "real" indicates the notion that $\mathbf{R}$ is an appropriate set to represent quantities which can be measured in the "real" physical world, such as time, distance, temperature, etc. We represent $\mathbf{R}$ by drawings such as Figure 5. Two labeled points on the line indicate a scale and direction.

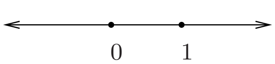


Figure 5
The real number line

Euclidean space

The set $\mathbf{R}^2 = \mathbf{R} \times \mathbf{R} = \{(x, y) \mid x, y \in \mathbf{R}\}$ is called the *x, y-coordinate plane* or the *Euclidean plane*, named after Euclid (ca. 300 BC) because it is the setting for classical plane geometry. We represent $\mathbf{R}^2$ with drawings such as Figure 6. We assume the reader is familiar with the identification of pairs (x, y) of real numbers with points in the plane.

The set $\mathbf{R}^3 = \mathbf{R} \times \mathbf{R} \times \mathbf{R} = \{(x, y, z) \mid x, y, z \in \mathbf{R}\}$ is called *real 3-dimensional space* or *Euclidean 3-space* and models the world in which we live. We represent $\mathbf{R}^3$ by drawings such as Figure 7, where we imagine the z-coordinate axis perpendicular to the flat plane in which the x and y-axes lie.

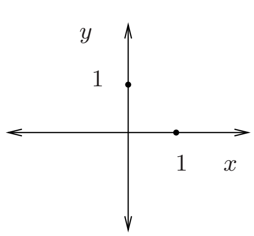


Figure 6
The Euclidean plane $\mathbf{R}^2$

Important Subsets of the Reals

The *integers* or *whole numbers*, is the set $\mathbf{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$. The *rational numbers* or *fractions*, denoted $\mathbf{Q}$, is the subset of all real numbers which can be written in the form m/n, where m and n are integers and $n \neq 0$. The *irrational numbers* is the set $\mathbf{R} \setminus \mathbf{Q}$ of points on the line, such as π and $\sqrt{2}$, which are not rational.

For two numbers a, b with $a < b$ we have the following subsets of $\mathbf{R}$, called *intervals*.

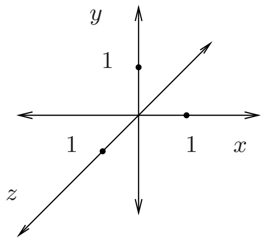
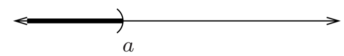
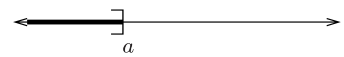
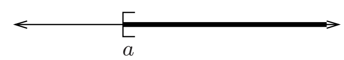


Figure 7
Euclidean space $\mathbf{R}^3$

Figure 8
Alternate sketch for $(a, b]$

Notation	Subset of $\mathbf{R}$	Type of interval	Sketch
(a, b)	$\{x \mid a < x < b\}$	open	
$[a, b]$	$\{x \mid a \le x \le b\}$	closed	
$(a, b]$	$\{x \mid a < x \le b\}$	half open, half closed	
$[a, b)$	$\{x \mid a \le x < b\}$	half open, half closed	
$(-\infty, a)$	$\{x \mid x < a\}$	open half line	
$(-\infty, a]$	$\{x \mid x \le a\}$	closed half line	
(a, ∞)	$\{x \mid a < x\}$	open half line	
$[a, \infty)$	$\{x \mid a \le x\}$	closed half line	

In the sketches for intervals, the round and square brackets marking the open and closed endpoints of the interval are sometimes shown by open and filled dots, respectively. For example, an alternate sketch of the interval $(a, b]$ is shown in Figure 8.

Notes on terminology

The symbol $\mathbf{Z}$ for the integers comes from the German "Zahlennummern," which means "counting numbers." The symbol $\mathbf{Q}$ for the rationals comes from the word "quotient." The word "rational" comes from the root for *ratio*, meaning proportion. Beware that the notation for the open interval $(a, b) = \{x \mid a < x < b\}$ is identical to the notation for the point (a, b) in the x, y-plane; if context does not make clear which is meant, then some additional comment is appropriate on the part of the user.

0.4 Definitions for functions

A function is a mathematical model for a process or machine that takes "inputs," does something to them, then produces "outputs." While the idea is not complicated, it is difficult to give a precise definition in everyday language, so some formality is required. The collections of inputs and outputs are modeled by sets. The function itself is modeled by pairs of the form (input value, output value). The machine is not allowed to be ambiguous; for an input value a, there must be exactly one output value b. Here is the formal definition, using the language of sets, that captures this idea.

A *function f from a set X to a set Y*, denoted $f\colon X \to Y$, is subset of $X \times Y$ in which each element $x \in X$ appears in exactly one ordered pair. We write $f(x) = y$ or $x \mapsto y$ to mean $(x, y) \in f$. The set X is called the *domain* of f, and the set Y is called the *codomain* or *range*. The arrows in the symbols $f\colon X \to Y$ and $x \mapsto y$ remind us that the machine takes input $x \in X$ and produces output $y = f(x) \in Y$. To illustrate that this is not an empty abstraction, here are some examples.

1. Eye color. Recording the eye color of each individual in a group of people can be expressed by a function where the domain is the set of people, the codomain is the set of colors, and the assignment $f(x)$ is the eye color for each individual x.

2. Temperature. A thermometer at a particular location leads naturally to the concept of temperature as a function of time. The domain is the set of time values, the codomain is the set of temperature values, and the assignment $T(t)$ is the temperature registered by the thermometer for each time t.

3. Position. A bug crawls around in the x, y-plane. The domain is the set of time values, the codomain is the set of points in the plane, and the assignment $P(t)$ is the location of the bug at each time t.

4. Projection. Line segments A and B are positioned as in Figure 9. Rays from a point-size light source positioned at point O cause each point on line segment A to cast a shadow on segment B. The domain is the segment A, the codomain is the segment B, and the assignment $f(p)$ is the projection p' for each point p.

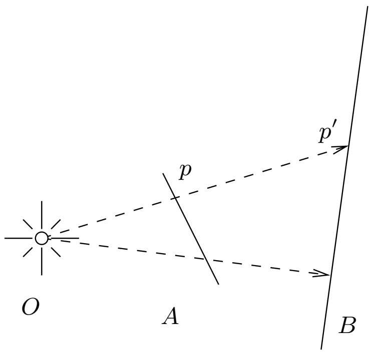


Figure 9
Projection

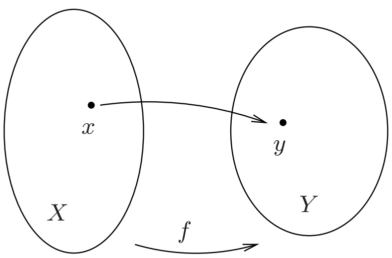


Figure 10
Schematic diagram of a function

Figure 10 shows a schematic representation of a function $f\colon X \to Y$. We use Venn diagrams for the domain and codomain sets with an arrow labeled f to indicate direction. An arrow from a point x in X to a point y in Y indicates that $f(x)$ is y. Figure 11 shows a version of a commonly used diagram that illustrates the conceptualization of a function as a machine.

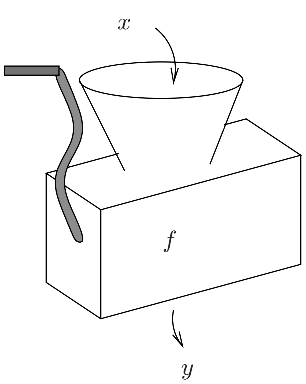


Figure 11
A function "machine"

Functions are often specified by equations. For example, we write $f(x) = x^2$ or $g(x) = 2x + 3$ to define functions f and g whose domains are sets of real numbers. We often refer to "the function $f(x) = x^2$," or simply, "the function x^2," to mean the function f defined by the equation $f(x) = x^2$. This language carries the potential to confuse the function f with its value $f(x)$; care must be taken when the distinction makes a difference.

Usually, when a function is specified by an equation, the domain is not explicitly given. For example, one might say "the function $h(x) = \sqrt{x-2}$." The convention in such cases is that the domain is all real x for which the equation specifying the function yields a meaningful real number. In this example, the domain for h is $\{x \mid 2 \leq x\}$.

A summary of vocabulary and notation for functions is given in (0.4.1).

(0.4.1) **Function vocabulary and notation.**

$f\colon X \to Y$	the function with domain X, codomain Y, and rule of assignment sending x to $f(x)$
$X \xrightarrow{f} Y$	ditto above
$x \mapsto y$	(pronounced "x goes to y," or "x maps to y") codomain element y is assigned to domain element x
map, mapping	synonyms for *function*
value of f at x	synonym for $f(x)$
image of x under f	ditto above
f evaluated at x	ditto above
image of a function	the set of images of all points in the domain
range of a function	(1) image or (2) codomain; since range has two meanings, care must be taken to avoid ambiguity
$f(X)$	the image of $f\colon X \to Y$
real-valued function	a function whose codomain is a subset of the real numbers
"plug in"	a colloquial term for the act of evaluating a function, as in, "we plug x into f and get $f(x)$"
identity function	a function whose domain and codomain are the same set, and which maps each element of the domain to itself
constant function	a function that has only one value (the *same* element in the codomain is assigned to every element in the domain)
$f \circ g$	the composition of functions f and g (see (0.5))
$kf, f+g, f\cdot g, f/g$	functions constructed from real-valued functions f, g and constant k (see (0.7))
invertible function	a function for which each element in the codomain is the image of *exactly one* element in the domain (see (0.6.1))
one-to-one correspondence	synonym for invertible function
f^{-1}	the inverse of an invertible function f (see (0.6.1))
$\displaystyle\sum_{i=m}^{n} x_i$	shorthand for the sum $x_m + x_{m+1} + x_{m+2} + \cdots + x_n$ (see (0.9))
$\displaystyle\sum_{i=m}^{n} f(i)$	shorthand for the sum $f(m) + f(m+1) + f(m+2) + \cdots + f(n)$ (see (0.9))

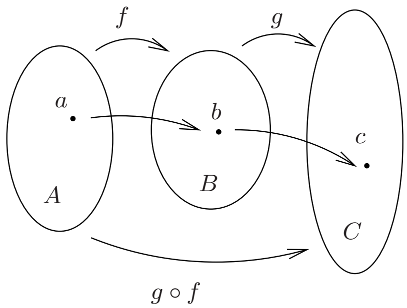


Figure 12
Composition of functions

0.5 Composition

Given two functions $f\colon A \to B$ and $g\colon B \to C$, the *composition* $g \circ f$ is the function $g \circ f\colon A \to C$ defined by $(g \circ f)(a) = g(f(a))$. Note that the order matters; $g \circ f$ is not the same as $f \circ g$. Figure 12 shows a schematic diagram.

Note: The order convention is sometimes reversed (especially in British usage). In other words, $g \circ f$ may mean $f \circ g$ as defined here. The order given here is the more standard.

0.6 Inverse functions

Functions operate on input values to produce output values. It is often worthwhile to reverse this process. For example, suppose we have a function f which tells us the dollar value $A = f(t)$ of an investment at time t. A practical problem would be to determine the time value needed to realize a given investment value. In other words, we are seeking a "reversing function," say g, that operates on dollar values and produces the corresponding time values. To say that g reverses the procedure f is to say $g(A) = t$ whenever $f(t) = A$. For any time value t or dollar value A, we would have the following.

$$g(f(t)) = t \qquad\qquad f(g(A)) = A$$

The above equations say that the composition $g \circ f$ is the identity function on the domain of time values of the function f, and $f \circ g$ is the identity function on the domain of dollar values of the function g. This example motivates the official definition of inverse functions.

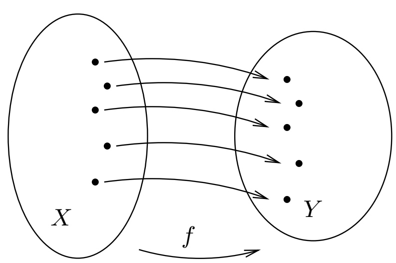


Figure 13
An invertible function

(0.6.1) **Definition of Inverse Function.** Suppose $f\colon X \to Y$ and $g\colon Y \to X$ satisfy the following equations

$$g \circ f = \mathrm{id}_X \qquad\qquad f \circ g = \mathrm{id}_Y$$

where id_X and id_Y denote the identity functions on X and Y, respectively. Then we say f and g are *inverses* of one another, and we write $g = f^{-1}$ and $f = g^{-1}$. The functions f and g are also called *invertible*. An invertible function is also called a *one-to-one correspondence*.

A visual representation of an invertible function $f\colon X \to Y$ (see Figure 13) shows the assignments made by f as arrows matching the elements of X and Y in a one-to-one manner. The picture of the inverse function f^{-1} is obtained by simply reversing the direction of all the arrows.

(0.6.2) **Examples of inverse functions.** Suppose X is a finite set, and $f\colon X \to Y$ is an invertible function. Since f matches the elements of X with the elements of Y in a one-to-one manner, Y must also be a finite set with the same number of elements as X.

Let $s\colon [0,\infty) \to [0,\infty)$ be the squaring function given by $x \mapsto x^2$. The inverse of s is the square root function $r\colon [0,\infty) \to [0,\infty)$ given by $x \mapsto \sqrt{x}$. Note that the domains are important here. Let $f\colon \mathbf{R} \to \mathbf{R}$ also be the squaring function $x \mapsto x^2$, but on the domain of all reals. The square root function is *not* an inverse for f because $r(f(-2)) = \sqrt{(-2)^2} = 2 \neq -2$. A lesson here is that a function given by an equation may be invertible with one domain, but not invertible with another domain.

0.7 Operations on real-valued functions

Given two functions $f\colon A \to \mathbf{R}$, $g\colon A \to \mathbf{R}$ and a constant real number k, we define the functions kf, $f+g$, $f-g$, $f\cdot g$ and f/g by the following formulas, for all a in A (note that the last equation is defined only when $g(a) \neq 0$, so the function f/g is defined to have domain $\{a \in A \mid g(a) \neq 0\}$).

$$\begin{aligned}
(kf)(a) &= kf(a) \\
(f+g)(a) &= f(a)+g(a) \\
(f-g)(a) &= f(a)-g(a) \\
(f\cdot g)(a) &= f(a)g(a) \\
(f/g)(a) &= f(a)/g(a)
\end{aligned}$$

Note: The notation fg is sometimes used to mean the product function $f\cdot g$, and sometimes to mean the composition $f \circ g$. Care should be taken when context does not make clear which is meant.

0.8 Sequences and functions

An ordered list

$$x_1, x_2, x_3, \ldots, x_n$$

of elements from a set X is called a *sequence* in X. The subscript i which marks the position in the list of the element x_i is called the *index* of x_i. For convenience, sequences sometimes begin with a whole number index different from 1.

A sequence $x_1, x_2, x_3, \ldots, x_n$ in a set X defines a function $f\colon \{1, 2, 3, \ldots, n\} \to X$ by

$$f(1) = x_1, f(2) = x_2, f(3) = x_3, \ldots, f(n) = x_n.$$

Conversely, a function $f\colon \{1, 2, 3, \ldots, n\} \to X$ determines a sequence in X by

$$x_1 = f(1), x_2 = f(2), x_3 = f(3), \ldots, x_n = f(n).$$

For example, the list of numbers $3, -1, 2, 3, 1, 5$ is a sequence

$$x_1 = 3, x_2 = -1, x_3 = 2, x_4 = 3, x_5 = 1, x_6 = 5$$

of elements of the set of integers, and also defines a function $f\colon \{1, 2, 3, 4, 5, 6\} \to \mathbf{Z}$ by

$$f(1) = 3, f(2) = -1, f(3) = 2, f(4) = 3, f(5) = 1, f(6) = 5.$$

0.9 Summation notation

A common operation in mathematics is to sum the values of a sequence of numbers. Let $x_1, x_2, x_3, \ldots, x_n$ be a sequence of numbers and let $f\{1, 2, 3, \ldots, n\} \to \mathbf{R}$ be the associated function given by $f(k) = x_k$ as described in the previous subsection. We write

$$\sum_{i=1}^{n} x_i \quad \text{or} \quad \sum_{i=1}^{n} f(i)$$

to denote the sum

$$x_1 + x_2 + \cdots + x_n = f(1) + f(2) + \cdots + f(n).$$

For example, for the sequence

$$1, 4, 9, 16, \ldots, 100$$

in which $x_k = k^2$ and $n = 10$, we write $\sum_{i=1}^{10} i^2$ to denote the sum $1 + 4 + 9 + \cdots + 100$. The symbol $\sum$ is the capital Greek letter sigma, and denotes a sum. The variable i is called the *index* of the sum. More generally, we write $\sum_{i=m}^{n} x_i$ to denote the sum

$$x_m + x_{m+1} + x_{m+2} + \cdots + x_n$$

where m, n are integers with $m \leq n$.

0.10 Exercises

1. List all subsets of the set $X = \{a, b, c\}$. (Note on convention: the empty set is considered to be a subset of any set. Also note that in the definition of subset, any set is a subset of itself.)

2. Let $A = \{1, 3, 5, 7, 9\}$, let $B = \{2, 3, 4, 5, 6\}$ and let $D = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ be the set of all ten digits. Find $A \cup B$, $A \cap B$, $A \setminus B$, and the $D \setminus A$. Draw a single Venn diagram showing the relationships between A, B, D, and the set $C = \{0, 2, 6\}$.

3. Sketch a Venn diagram for the *symmetric difference* $(A \setminus B) \cup (B \setminus A)$ of two sets A and B for the sets A and B in example (0.1.3).

4. Use a Venn diagram to illustrate the following property of set operations (one of *DeMorgan's Laws*).

$$A \setminus (B \cap C) = (A \setminus B) \cup (A \setminus C)$$

5. Let $A = \{a, b\}$ and $B = \{x, y, z\}$.

   (a) List all members of the set $A \times B$.

   (b) List all members of the set B^2.

6. Draw a single Venn diagram illustrating the relationships between the following sets: $\mathbf{R}$, $\mathbf{Q}$, $\mathbf{Z}$, and $[0, 1)$.

7. Let A and B be subsets of the real line $\mathbf{R}$ given by $A = \{x \mid -2 \leq x < 3\}$ and $B = \{x \mid 1 < x \leq 5\}$. Write in interval notation, set notation and sketch a picture of A, B, $A \cup B$, $A \cap B$, $A \setminus B$ and $\mathbf{R} \setminus A$. Example: in interval notation, set A is written $A = [-2, 3)$, in set notation $A = \{x \mid -2 \leq x < 3\}$, and the sketch of A is the shaded region of the real line marked with endpoint brackets at -2 on the left and 3 on the right that looks like the fourth sketch from the top in the chart of intervals in subsection (0.3).

8. Let $f(x) = x^2$ and $g(x) = x + 2$ define functions f and g from the reals to the reals.
   (a) Find $(f \circ g)(3)$
   (b) Find $(g \circ f)(3)$
   (c) Find $(g \cdot f)(3)$
   (d) Find $(f/g)(3)$
   (e) Find $(3f + g)(3)$
   (f) Write an equation for $(f \circ g)(x)$
   (g) Write an equation for $(g \circ f)(x)$

9. Let $X = \{a, b, c\}$ and $Y = \{1, 2, 3\}$. Describe all possible one-to-one correspondences between X and Y.

10. (a) Evaluate $\displaystyle\sum_{i=1}^{10}(2i + 1)$.
    (b) Write the sum
    $$(1^2+2\cdot1+3)+(2^2+2\cdot2+3)+(3^2+2\cdot3+3)+\cdots+(15^2+2\cdot15+3)$$
    using summation notation.

11. Let $A = \{x, y, z\}$ and let $f: A \to \mathbf{R}$ and $g: A \to \mathbf{R}$ be given by the following table of values.

element of A	value of f	value of g
x	2	5
y	0	1
z	-1	2

Let $s_1 = x, s_2 = y, s_3 = z$ be a sequence of elements in A. Find the following.

   (a) $\displaystyle\sum_{i=1}^{3} f(s_i)$
   (b) $\displaystyle\sum_{i=1}^{3}(f + g)(s_i)$
   (c) $\displaystyle\sum_{i=1}^{3}(f \cdot g)(s_i)$

1 Logic

Logic was originally developed to formalize common sense reasoning in order to analyze arguments and provide rules for clear reasoning and detecting flawed arguments. While logic is based on common sense, its demand for precision has led to certain usages which sometimes depart from everyday language (for example, see the discussion of the word "or" below). In this section we present the basics of logic that are routinely used in mathematics.

1.1 Statements

The fundamental object in logic is the *statement*. A statement is a sentence that is either true or false. In the examples below, 1 and 2 are statements. While 3 is a valid sentence, it is not considered a statement since it doesn't make sense to say whether it is true or false.

1. $2 + 2 = 5$
2. The moon is a satellite of the earth.
3. How I wish I could fly.

We often represent statements by letters p, q etc.

1.2 Negation

Every statement p has an opposite statement, denoted $\neg p$ or$\sim p$, called its *negation*. If p is true, its negation is false, and vice-versa. The negations of statements 1 and 2 above are the following.

1. $2 + 2 \neq 5$
2. The moon is not a satellite of the earth.

Negation can be confusing when a statement is made about some or all members of a set. Here are some examples.

1. Every person in this room works for the fire department.
2. No person in this room works for the fire department.
3. Someone in this room does not work for the fire department.
4. Some people like ice climbing.
5. Everyone does not like ice climbing.
6. No one likes ice climbing.
7. No bird has flown to the moon.
8. All birds have flown to the moon.
9. Some bird has flown to the moon.

Notice that 3, not 2, is the negation of 1, and that 6, not 5, is the negation of 4, and that 9, not 8 is the negation of 7.

1.3 Connectives

Statements can be combined into compound statements using *connectives*. The four essential connectives are *and*, *or*, *if-then*, and *if and only if*.

And

We write $p \wedge q$ to denote the statement "p and q." The statement $p \wedge q$ is true when both statements p and q are true, and false otherwise.

Or

In everyday English, we use the word "or" in two distinct ways. If we say to a child, "Do you want a cookie or a piece of cake for dessert?" we are implying that the child may choose one or the other, but not both. This usage is called the *exclusive or*. If we say, "Do you like movies or books?" we would accept "Both" as a reasonable answer. In logic and mathematics, we use the word "or" to mean "one or the other or both."

We write $p \vee q$ to denote the statement "p or q." The statement $p \vee q$ is true when either or both statements p and q are true, and false only when statements p and q are both false.

If-then

The statement "if it is raining, then I am carrying an umbrella" is a compound statement of the form "if p then q," where p is the statement "it is raining" and q is the statement "I am carrying an umbrella." If we observe that it is raining and I am not carrying an umbrella, we would declare the statement "if p then q" to be false. If it is not raining, common sense does not seem to demand that the statement "if p then q" be true or false. In logic and mathematics, however, every statement has a truth value. If p is false, the statement "if p then q" is considered to be true.

We write $p \Rightarrow q$ to denote the statement "if p then q" or "p implies q." The statement $p \Rightarrow q$ is true unless p is true and q is false. Statements of this form are called *if-then* statements or *implications*.

The statement $q \Rightarrow p$ is called the *converse* of the statement $p \Rightarrow q$. It is important to note that an if-then statement is *not* equivalent to its converse. This is perhaps the greatest source of poor logic in everyday usage. For example, suppose I am carrying an umbrella down the street on a sunny day. The statement "if it is raining, then I am carrying an umbrella" is *true* while the converse "if I am carrying an umbrella, then it is raining" is *false*.

The statement $(\neg q) \Rightarrow (\neg p)$ is called the *contrapositive* of the if-then statement $p \Rightarrow q$. The reader may check (see exercise 3 below) that an implication and its contrapositive have the same truth value for every possible combination of truth values of p and q.

If and only if

We write $p \Leftrightarrow q$ to denote the statement $(p \Rightarrow q) \wedge (q \Rightarrow p)$. In words, we say "$p$ if and only if q." The statement $p \Leftrightarrow q$ is true when p and q are either both true or both false. We also say that p and q are *logically equivalent* when $p \Leftrightarrow q$ is true. For example, the previous paragraph mentions that any implication is logically equivalent to its contrapositive.

1.4 Order of precedence in notation

When we calculate $4+3\cdot 2$, we follow the convention that multiplication happens before addition so the result is $4+6=10$ and not $7\cdot 2 = 14$. There are similar conventions for logic notation. When we write $\neg p \wedge q \Rightarrow r$, we mean $((\neg p)\wedge q) \Rightarrow r$. Negation has the highest precedence, and/or connectives have the next level of precedence, if-then and if-and-only-if have the lowest precedence. This is analogous to the minus sign (highest precedence), multiplication (next highest), and addition (lowest precedence) in arithmetic.

1.5 Truth tables

Truth tables provide a convenient method for analyzing compound statements. For example, to describe $p \wedge q$ we make a table of 3 columns, one column for each of the statements p and q involved in the compound statement and one column for the compound statement.

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

There is one row for each of the *logical possibilities*, or possible combinations of truth values, for p and q. The third column gives the truth value for the compound statement for each logical possibility.

Here is how to use a truth table to show that $p \Rightarrow q$ is logically equivalent to $\neg p \vee q$. (Note the order of precedence, as explained above: $\neg p \vee q$ means $(\neg p) \vee q$. If we want to negate $p \vee q$ we must use parentheses and write $\neg(p \vee q)$.)

p	q	$p \Rightarrow q$	$\neg p \vee q$
T	T	T	T
T	F	F	F
F	T	T	T
F	F	T	T

Since the truth values in the last two columns are equal in every row, we conclude that the compound statements are logically equivalent.

1.6 Proof

A *proof* is an argument which uses logical rules, called *rules of inference*, to establish the truth of a statement. For example, if we know that the statement p is true, and that the statement $p \Rightarrow q$ is true, then we may conclude that q is true. Most people would simply call this common sense; in the codification of logic, this rule of inference is called *modus ponens*. Slightly less trivial, if we know that $p \Rightarrow q$ is true and that q is false, we may conclude that p is false. This rule of inference is called *modus tollens*.

We will not attempt to give an exhaustive list of rules of inference here. We will, however, give examples of two special methods of proof that are widely used in mathematics.

Note on cultural practice: in mathematics, it is common to conclude a proof with a phrase such at "now we are done" or the Latin "Q.E.D." which stands for *quod erat demonstratum* and means "that which was to be shown."

Indirect Proof

Indirect proof, also called *proof by contradiction* or *reductio ad absurdum* is a surprisingly effective method which exploits *modus tollens* in a clever way. Here is the situation: we have a statement p that we believe to be true, and would like to construct a logical proof. However, direct attempts to conclude p from previously known facts don't seem to work. So we take an indirect approach and suppose that the opposite of p is true. If, from this assumption, we can use logical rules to prove that a false statement q is true, then we may conclude that p must be true! Here is a beautiful example.

(1.6.1) **The infinitude of primes.** *There are infinitely many prime numbers.*

Proof: Suppose that there are only finitely many primes $p_1, p_2, \ldots, p_n$. Let $q = p_1 p_2 \cdots p_n + 1$. Since q is larger than all the primes, q is a composite number, so it is divisible by some prime, say p_i. But long division of $q = p_1 p_2 \cdots p_n + 1$ by p_i leaves a remainder of 1, so q is *not* divisible by p_i. Clearly, this is impossible. We conclude that there must be infinitely many primes.

Mathematical induction

Suppose we have a collection of statements

$$p_1, p_2, p_3, \ldots$$

and suppose we can prove that statement p_1 is true, and also that the statement $p_k \Rightarrow p_{k+1}$ is true for all integers $k \geq 1$. Setting $k = 1$ we have $p_1 \Rightarrow p_2$. Using modus ponens on the pair of statements $p_1, p_1 \Rightarrow p_2$, we conclude that p_2 is true. Now let $k = 2$ and consider the pair $p_2, p_2 \Rightarrow p_3$. We conclude that p_3 is true. Intuitively, this goes on "forever," and all the statements $p_1, p_2, p_3, \ldots$ must be true. Mathematical induction is the formal statement that this intuition is correct. Formally, mathematical induction states that if p_1 is true, and $p_k \Rightarrow p_{k+1}$ is true for any integer $k \geq 1$, then p_n is true for all $n \geq 1$.

To use mathematical induction in a proof, we have to prove p_1 and $p_k \Rightarrow p_{k+1}$ for $k \geq 1$. Proving p_1 is called the *base case* and proving $p_k \Rightarrow p_{k+1}$ for $k \geq 1$

is called the *inductive step*. To prove the inductive step, we show that p_{k+1} follows from the assumption, called the *inductive hypothesis*, that p_k is true.

Here is an example.

(1.6.2) **Sum of the first n positive integers.** *The sum of the first n positive integers is*

$$1 + 2 + 3 + \cdots + n = (n)(n+1)/2.$$

Proof: This is really a collection of statements.

$$\begin{array}{rcl}
p_1 & : & 1 = (1)(2)/(2) \\
p_2 & : & 1 + 2 = (2)(3)/(2) \\
p_3 & : & 1 + 2 + 3 = (3)(4)/(2) \\
 & \vdots & \\
p_n & : & 1 + 2 + 3 + \cdots + n = (n)(n+1)/(2) \\
 & \vdots &
\end{array}$$

First, we establish the base case p_1. Simply note that the sum of the first 1 positive integer is indeed $1 = (1)(2)/2$, so the statement is true for $n = 1$.

Second, we do the inductive step. Suppose that the sum formula holds for $n = k$, that is, that p_k is true. We must show that p_{k+1} is true. We have

$$\begin{array}{rl}
 & 1 + 2 + 3 + \cdots + k + (k+1) \\
= & [1 + 2 + 3 + \cdots + k] + (k+1) \\
= & (k)(k+1)/2 + (k+1) \quad \text{(by inductive hypothesis)} \\
= & (k^2 + k)/2 + (2k+2)/2 \quad \text{(distributing, and getting a common denominator)} \\
= & (k^2 + 3k + 2)/2 \quad \text{(adding)} \\
= & (k+1)(k+2)/2 \quad \text{(factoring)}
\end{array}$$

which shows that the sum formula holds for $n = k + 1$. This completes the inductive step.

By mathematical induction, we conclude that formula (1.6.2) holds for all $n \geq 1$. Q.E.D.

1.7 Exercises

1. Form the negations of the following statements.
   (a) The cat is hungry.
   (b) All of the cats in the room are hungry.
   (c) Some of the cats in the room are hungry.
   (d) Some of the cats in the room are not hungry.
2. Use truth tables to establish *DeMorgan's Laws*.
   (a) $\neg(p \wedge q)$ is logically equivalent to $\neg p \vee \neg q$

(b) $\neg(p \vee q)$ is logically equivalent to $\neg p \wedge \neg q$

3. Use a truth table to show that an implication is logically equivalent to its contrapositive.

4. Let p, q and r represent logical statements. Three compound statements are given below. Are any of them logically equivalent? Explain why or why not.

   (i) $(p \wedge \neg q) \Rightarrow \neg r$

   (ii) $\neg r \Rightarrow (p \wedge \neg q)$

   (iii) $r \Rightarrow (q \vee \neg p)$

5. Give a proof by contradiction that there is no smallest positive number.

6. Use mathematical induction to prove that the number of subsets of a set with n elements is 2^n.

2 Counting

In this section we present some basic principles of counting which help us analyze problems such as the following.

> A particular state has automobile license plates with 3 capital letters followed by 4 digits. How many different license plates are possible in this format?

We solve this problem and others in the examples below.

2.1 How many in a row?

How many numbers are there in the list $6, 7, 8, \ldots, 12$? If you answer quickly, your first response might be 6, since 12 minus 6 equals 6. If you list the numbers and count more carefully, however, you find that there are 7 numbers in this list. This simple example illustrates that precision in counting should not be taken for granted, and leads to our first basic rule of counting.

(2.1.1) **Length of a string of consecutive numbers.** *Let m and n be integers with $m \leq n$. The count of numbers in the list $m, m+1, m+2, \ldots, n$ is $n - m + 1$.*

2.2 The multiplication principle

Here is a simpler version of the license plate problem: At a particular restaurant, you have a choice of 3 entrees and 2 salads. How many different dinners (entree with salad) could you possibly have? While you do not need fancy theory to solve this problem, we use it to demonstrate an organizational technique.

Choosing a dinner can be broken down into a sequence of choices: first choose the entree; then choose the salad. If we call the entrees A, B and C and call the salads S and T, we can draw a schematic diagram called a *tree* that captures the situation. See Figure 14.

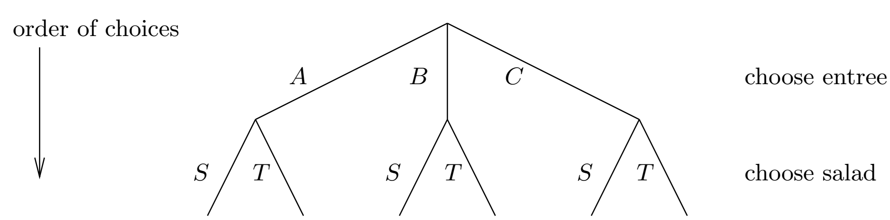


Figure 14 Tree of choices

The first thing to notice is that mathematicians tend to draw their trees upside-down. The line segments are called *branches* and the ends of the line segments are called *nodes* or *vertices*. The first choice (entree) is represented by the top level of branches and the second choice by the next level down.

There is a one-to-one correspondence between the nodes at the bottom of the tree and the set of possible dinners. For example, the second node from the left corresponds to entree A with salad T. By virtue of the one-to-one correspondence, we can count dinners by counting nodes at the bottom of the tree. But this is easy; since there are three branches at the entree choice level and two branches descending from each node at the top of the salad choice level, we clearly have $3 \cdot 2 = 6$ total nodes at the bottom of the tree. We can extend this reasoning to include further choices. Suppose that in addition to one of 3 entrees and 2 salads, a dinner includes one of 4 desserts d, e, f and g. The tree diagram of dinner choices now looks like Figure 15.

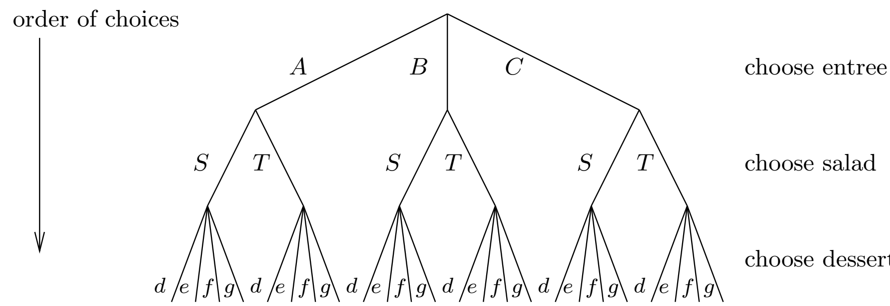


Figure 15 Tree of more choices

In this diagram, the 10th node from the left at the bottom of the diagram corresponds to the dinner with entree B, salad S and dessert e. The count of nodes at the bottom (and therefore the count of dinners) is $3 \cdot 2 \cdot 4 = 24$.

These examples lead to the following rule.

(2.2.1) **Multiplication Principle.** *Suppose that a sequence of n choices must be made and that there are m_1 ways to make the 1st choice, m_2 ways to make the 2nd choice, and so on, and m_n ways to make the final n-th choice. Then there are a total of $m_1 m_2 \cdots m_n$ different ways to make the sequence of choices.*

As an example, we solve the license plate problem stated at the beginning of this section. Making a license plate can be thought of as a sequence of 7 choices: the first letter, second letter, and so on, and finally the last digit. There are 26 ways to make each of the first three choices (the letters) and 10 ways to make each of the last 4 choices (the digits) so there are a total of $26^3 10^4 = 175,760,000$ possible license plates.

2.3 Permutations

There are 6 chairs numbered 1 through 6 on a stage. Six people are to be seated in the chairs, one in each chair. How many different seating arrangements are possible?

This is easy to solve using the multiplication principle. A seating arrangement can be thought of as a sequence of 6 choices: choose a person for chair 1, then for chair 2, and so on through chair 6. There are 6 different ways to make the first choice, then 5 ways to make the second choice (only 5 people remain after the first choice), then 4 ways to make the third choice, and so on, until there is only 1 way to make the last choice. Thus there are a total of $6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 720$ different seating arrangements.

The product $n(n-1)(n-2)\cdots 3\cdot 2\cdot 1$, where n is a positive integer, is so common in mathematics that it has a name and a shortcut notation. This product is called n *factorial* and is denoted using an exclamation mark as follows.

$$n! = n(n-1)(n-2)\cdots 3\cdot 2\cdot 1 \tag{2.3.1}$$

A *permutation* of a finite set of n elements is simply an ordering or labeling of the elements of the set with the numbers 1 through n. Thus, the number $n!$ is called the number of *permutations of n things*. By convention, $0!$ is defined to equal 1.

Now suppose that chairs 5 and 6 are removed from the stage but all 6 people remain. How many ways are possible to seat one person in each of the remaining chairs 1 through 4?

Again we use the multiplication principle. There are 6 ways to fill the first chair, then 5 ways to fill the second chair, and so on, and 3 ways to fill the 4th chair. Thus there are a total $6\cdot 5\cdot 4\cdot 3 = 360$ seating arrangements.

The product $6\cdot 5\cdot 4\cdot 3$ can be thought of as the entire factorial product $6\cdot 5\cdot 4\cdot 3\cdot 2\cdot 1$ with the "tail" $2\cdot 1$ chopped off. There is a simple way to express this "partial factorial."

Let n and k be positive integers with $k \le n$. The "partial factorial" of the first k factors of $n!$ (starting with n on the left and descending to the right) is the product $n(n-1)(n-2)\cdots(n-k+1)$. The part of $n!$ that is "chopped off" is $(n-k)! = (n-k)(n-k-1)\cdots 3\cdot 2\cdot 1$. Thus we can write this "partial factorial" as follows.

$$n(n-1)(n-2)\cdots(n-k+1) = \frac{n!}{(n-k)!}$$

The number $\dfrac{n!}{(n-k)!}$, where n and k are nonnegative integers with $k \le n$, is called the number of *permutations of n things taken k at a time* and is sometimes denoted $P(n,k)$ or nPk.

2.4 Combinations

A committee of 4 people is to be chosen from a group of 6 people. How many different committees are possible?

We have already counted the number ways to seat 4 of the 6 people in chairs labeled 1 through 4 ($P(6,4) = 6!/(6-4)! = 360$ ways). Each of these seatings can form a committee. However, some of these seatings yield the same committee. For example, Joe, Martha, Bob and Sue seated in chairs 1, 2, 3 and 4 is a *different* seating from Martha, Joe, Bob and Sue seated in order 1–4, but these two different seatings yield the *same* committee Joe-Martha-Bob-Sue or Martha-Joe-Bob-Sue. In fact, there are exactly $4! = 24$ different seatings (the number of ways to rearrange the 4 chairs) that yield the same committee. Each committee occurs $4!$ times in the list of seatings of 4 people, so there are $360/24 = 15$ possible committees.

The number $\dfrac{P(n,k)}{k!} = \dfrac{n!}{k!(n-k)!}$, where n and k are nonnegative integers with $k \le n$, is called the number of *combinations of n things chosen k at a time* or simply n *choose* k, and is denoted $\dbinom{n}{k}$, or sometimes $C(n,k)$ or nCk.

The numbers $\binom{n}{k}$ are also called the *binomial coefficients* since they give the coefficients of monomials in the expansion of powers of a binomial. To be precise, there is the formula

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}.$$

The thoughtful reader will enjoy proving this formula as a puzzle. Readers who fancy this topic will find further reference material in any introductory combinatorics text.

2.5 Summary of counting formulas

Here we summarize the five basic counting formulas of this section.

$$\begin{array}{rcl}
\text{\# items in list } m, m+1, \ldots, n & = & n-m+1 \\
\begin{array}{c}\text{\# ways to make sequence of } n \text{ choices} \\ \text{with } m_i \text{ ways to make } i\text{th choice}\end{array} & = & m_1 m_2 \cdots m_n \\
\text{\# permutations of } n \text{ objects} & = & n! \\
\begin{array}{c}\text{\# permutations of } n \text{ things} \\ \text{taken } k \text{ at a time}\end{array} & = & \dfrac{n!}{(n-k)!} \\
\begin{array}{c}\text{\# combinations of } n \text{ things} \\ \text{taken } k \text{ at a time}\end{array} = \dbinom{n}{k} & = & \dfrac{n!}{k!(n-k)!}
\end{array}$$

2.6 Exercises

1. How many integers are in the open interval of the real line $\{x\colon 53 < x < 186\}$?

2. How many integers are in the closed interval of the real line $\{x\colon 53 \le x \le 186\}$?

3. How many distinct telephone numbers are there under the following rules: a telephone number has 10 digits (3 digit area code plus 7 digit local number), the first digit (first digit of the area code) must not be a zero or a one, and the fourth digit (first digit of the local number) must not be a zero or a one?

4. How many different ways are there to make an ordered list of 20 people?

5. A DJ has recordings of 50 songs. How many different playlists (where order of the songs is part of the playlist) of 20 songs can she possibly make, with no songs repeated?

6. In the problem above, how many playlists are there if the DJ can repeat any number of songs any number of times?

7. Twenty players sit on a bench. How many different teams of 11 players are there?

8. (a) Verify the formula $C(n,k)P(k,k) = P(n,k)$.
   (b) Describe a sequence of two choices so that the multiplication principle applies to the formula in part (a) with $m_1 = C(n,k)$ and $m_2 = P(k,k)$.

(c) Suppose that you do not yet know the formula for $C(n, k)$, but that you know $P(n, k)$ for all values of n and k. Show how to use part (b) to derive $C(n, k)$.

3 Arithmetic and Geometric Sequences

In this section we present the basic theory of two fundamental types of number patterns.

3.1 Arithmetic sequences

An *arithmetic sequence* is a list of numbers of the form

$$a, a + d, a + 2d, a + 3d, a + 4d, \ldots.$$

The number a is called the *initial term*, and the number d is called the *common difference* because it is the difference between any two consecutive terms in the sequence. For example, the sequence

$$2, 5, 8, 11, 14, \ldots$$

is the arithmetic sequence with initial term $a = 2$ and common difference $d = 3$.

The adjective "arithmetic" carries the stress on the third syllable, rather than the second as in the noun form.

3.2 Geometric sequences

A *geometric sequence* is a list of numbers of the form

$$a, ar, ar^2, ar^3, ar^4 \ldots.$$

The number a is called the *initial term*, and the number r is called the *common ratio* because it is the ratio of any two consecutive terms in the sequence. For example, the sequence

$$2, 6, 18, 54, 162, \ldots$$

is the geometric sequence with initial term $a = 2$ and common ratio $r = 3$.

3.3 Explicit formulas

If we set $a_n = a + nd$ for $n = 0, 1, 2, 3 \ldots$, then the sequence

$$a_0, a_1, a_2, a_3, \ldots$$

is the arithmetic sequence with initial term a and common difference d.

Similarly, if we define $g_n = ar^n$ for $n = 0, 1, 2, 3 \ldots$, then the sequence

$$g_0, g_1, g_2, g_3, \ldots$$

is the geometric sequence with initial term a and common ratio r.

The formulas $a_n = a + nd$ and $g_n = ar^n$ are called *explicit* or *closed form* formulas for the given sequences.

3.4 Recursive formulas

The pair of equations

$$\begin{aligned} a_0 &= a \\ a_n &= a_{n-1} + d, \quad n = 1, 2, 3 \ldots \end{aligned}$$

is called a *recursive* description of the arithmetic sequence with initial term a and common difference d. The first equation tells us that the sequence starts with a, and the second equation tells us that the number in the sequence at index n, that is, the nth term of the sequence, is obtained by adding d to the $(n-1)$st term of the sequence.

Similarly, the pair of equations

$$\begin{aligned} g_0 &= a \\ g_n &= r g_{n-1}, \quad n = 1, 2, 3 \ldots \end{aligned}$$

is a recursive description of the geometric sequence with initial term a and common ratio r. The first equation tells us that the sequence starts with a, and the second equation tells us that the nth term of the sequence is obtained by multiplying the $(n-1)$st term of the sequence times r.

In general, a recursive description tells us how to generate the next term in a sequence by applying a regular process to the previous terms in the sequence.

3.5 Sums of terms in a sequence

One of the most common operations performed on sequences is to sum a number of consecutive terms. For example, if $I_1, I_2, I_3, \ldots$ is a sequence of dollar values, where I_k is the amount of interest earned by a certain investment in its kth year, then $I_1 + I_2 + \cdots + I_{10}$ is the total amount of interest earned in years 1 through 10. For an arbitrary sequence of numbers

$$a_0, a_1, a_2, \ldots$$

finding the sum $a_0 + a_1 + \ldots + a_n$ of the first $n+1$ terms requires n addition operations (first add a_1 to a_0, then add a_2 to that result, and so on). For arithmetic and geometric sequences, there are alternative formulas for sums that involve only a few operations. First, we demonstrate a method for sums of terms of arithmetic sequences.

Let s_n be the sum

$$s_n = a + (a+d) + (a+2d) + \cdots + (a+nd) \tag{3.5.1}$$

of the first $n+1$ terms of the arithmetic sequence with initial term a and common difference d. Listing the summands in reverse order we have

$$s_n = (a+nd) + (a+(n-1)d) + (a+(n-2)d) + \cdots + a. \tag{3.5.2}$$

The sum of the first term on the right hand side of (3.5.1) with the first term on the right hand side of (3.5.2) is $a + (a+nd) = 2a + nd$. Notice that the sum of the second terms on the right hand sides of these two equations is also $2a+nd$, and so on. Thus we have

$$2s_n = (2a+nd) + (2a+nd) + \cdots + (2a+nd) = (n+1)(2a+nd).$$

Solving for s_n we have

$$s_n = (n+1)(2a+nd)/2. \tag{3.5.3}$$

Next we demonstrate a method for sums of terms of geometric sequences. Let

$$s_n = a + ar + ar^2 + \cdots + ar^{n-1} + ar^n \tag{3.5.4}$$

be the sum of the first $n+1$ terms of the geometric sequence with initial term a and common ratio r. Multiplying both sides by r, we obtain

$$rs_n = ar + ar^2 + \cdots + ar^n + ar^{n+1} \tag{3.5.5}$$

Subtracting (3.5.5) from (3.5.4), we get

$$s_n - rs_n = a - ar^{n+1}.$$

Factoring both sides yields

$$s_n(1-r) = a(1-r^{n+1}).$$

Finally, dividing both sides by $1-r$, we obtain the following formula for s_n.

$$s_n = a\left(\frac{1-r^{n+1}}{1-r}\right) \tag{3.5.6}$$

3.6 Geometric Series and Convergence

In this subsection we make some observations and a definition about the sum of terms of a geometric sequence (3.5.6).

Let r be a number such that $|r| < 1$. It is clear that

$$|r| > |r|^2 > |r|^3 > \cdots$$

and that $|r|^n$ is "very small" when n is "very large". We express this observation with the symbols

$$\lim_{n\to\infty} |r|^n = 0$$

or $|r|^n \to 0$ as $n \to \infty$ and say "the limit as n goes to infinity of $|r|^n$ is zero" or "$|r|^n$ goes to zero as n goes to infinity" (because the intuitive meaning is clear, it is customary to use this limit language whether or not the reader understands the formal definition of limit from first year calculus).

If $|r| < 1$, the term $|r|^{n+1}$ in the expression for s_n in (3.5.6) vanishes to zero as $n \to \infty$, so it makes sense to say $s_n \to a/(1-r)$ as $n \to \infty$. We call expression

$$\lim_{n\to\infty} s_n$$

a *geometric series*, which we also write as an "infinite sum"

$$a + ar + ar^2 + ar^3 + \cdots.$$

When $|r| < 1$, we say that the geometric series *converges* to $s = a/(1-r)$. We also call s the *sum* of the geometric series and write

$$s = a + ar + ar^2 + ar^3 + \cdots.$$

3.7 Linear and exponential growth

For constants a and d, let $L\colon \mathbf{R} \to \mathbf{R}$ be the linear function defined by $L(x) = a + dx$. Observe that the sequence of values

$$L(0), L(1), L(2), L(3), \ldots$$

is the arithmetic sequence with initial term a and common difference d.

Similarly, given constants a and r with $a \neq 0$, $r > 0$ and $r \neq 1$, let $E\colon \mathbf{R} \to \mathbf{R}$ be the exponential function given by $E(x) = ar^x$. Then the sequence

$$E(0), E(1), E(2), E(3), \ldots$$

is the geometric sequence with initial term a and common ratio r.

Because of these natural associations between sequences and functions, arithmetic sequences are sometimes called *linear* and geometric sequences are sometimes called *exponential*. When the function L or E is increasing (that is, when the graph of the function rises from left to right) we say that values of an arithmetic sequence follow *linear growth* or *grow linearly* and that values of a geometric sequence follow *exponential growth* or *grow exponentially*. When L or E decreases (when the graph drops from left to right) we say that the associated arithmetic sequence follows *linear decay* or *decays linearly*, and that the associated geometric sequence follows *exponential decay* or *decays exponentially*.

3.8 Exercises

1. Fill in the missing terms of the following arithmetic and geometric sequences. Identify the initial term and the common difference or common ratio for each.

   (a) $5, 2, -1, \_\_, \_\_, \_\_, \ldots$

   (b) $5, 2, 0.8, \_\_, \_\_, \_\_, \ldots$

   (c) $\_\_, 2, \_\_, 5, \_\_, 8, \ldots$

   (d) $\_\_, 2, \_\_, 4, \_\_, 8, \ldots$

2. Find the sum of the first 100 positive integers.

3. Find the given sums of terms of arithmetic and geometric sequences.

   (a) $2 + 5 + 8 + 11 + \cdots + 302$

   (b) $2 + 5 + 8 + 11 + \cdots + 1571$

   (c) $2 + 6 + 18 + 54 + \cdots + 2(3^{100})$

   (d) $2 + 6 + 18 + 54 + \cdots + 9565938$

4. In the formula (3.5.6), what happens if $r = 1$? Is there any way to find the sum s_n?

5. The following questions are about the definitions of linear and exponential growth and decay.

   (a) For the exponential function associated to a geometric sequence, why do we make the restrictions $a \neq 0$, $r > 0$ and $r \neq 1$ for the exponential function?

(b) What values of the common difference d correspond to linear growth? To linear decay?

(c) What values of the common ratio r correspond to exponential growth? To exponential decay?

6. Consider the following true life situation.

> "I want to rent this room for the month of July," he said. The clerk wheezed. He peered through the narrow slits of his bloodshot eyes, glaring through the murk of the humid dusty darkness of the fleabag lobby, and said, "for you—a deal." "How much?" said the big guy, sweat trickling down his face, staining the collar of his dingy shirt which didn't appear to have been washed in weeks. Noticing the telltale bulge of a revolver under the stranger's dirt stained jacket, the clerk replied, "First day—one cent. Second day—two cents. Third day—four. Every day it doubles." The stranger's face drew into a knot as he scrutinized the greasy poker faced clerk. He said, "That's nothin'. What's the hitch?"

(a) How much would the stranger pay on July 31st?

(b) What would the bill be for the month of July?

7. You get a letter in the mail that says, "Send a dollar to each of the five people on this list. Add your name to the bottom, take the top name off, and send a copy of the new list plus these instructions to five new people. P.S. If you break the chain you will have to sit in a math lecture every day for the rest of your life."

(a) Assuming nobody broke the chain, and every letter was passed on in one day, how much money would you have after 10 days? 20 days? One hundred days?

(b) Assuming no person ever received the letter twice, and each letter was passed on in one day (and nobody broke the chain) how long would it take for everyone on the planet to get a letter?

8. Gumby and Pokey decide to go on a diet together. Gumby and Pokey both weigh 10 ounces. Starting their diets on the same day, Gumby loses 1/10 of an ounce each day, while Pokey loses half his body weight each day.

(a) Who wins the race to the body weight of 1 ounce?

(b) Explain how you know, without calculating, that Gumby will win the race to zero body weight.

(c) How much of Pokey is left when Gumby vanishes?

9. The Greek philosopher Zeno (ca. 450 BC) considered the following motion problem. A rabbit and a turtle agree to run a race. Displaying good sportsmanship, the rabbit, who can run faster, gives the turtle a head start. Zeno argued that the rabbit will never catch the turtle, as follows. To catch the turtle, the rabbit must first travel from the starting point to the turtle's starting point. During this time, the turtle will advance. Let's call the turtle's new location point 2. Now the rabbit must travel to point 2, but during that time, the turtle advances to point 3. And so on. This process defines an infinite sequence of distinct points. Traveling from each one to the next requires a positive amount of time. Since an infinite sum of positive numbers must be infinite, Zeno concludes that the rabbit will never catch the turtle. Of course, Zeno knew there must be some flaw in this argument, but was unable to resolve the paradox satisfactorily.

   Analyze Zeno's paradox under the following assumptions: the rabbit travels at a constant speed of 5 feet per second; the turtle travels at a constant speed of 2 feet per second; and the head start distance is 10 feet. Place coordinates on the race track with the rabbit beginning at 0 and the turtle beginning at 10, with both running in the positive direction.

   (a) Using high school algebra and the formula

   $$(\text{distance}) = (\text{rate})(\text{time})$$

   find the location where the rabbit catches the turtle, and the time elapsed from the beginning of the race to the point where the rabbit catches the turtle.

   (b) Find the sequence of distances traveled by the rabbit from the rabbit's starting point to the turtle's starting point, from there to point 2, from point 2 to point 3, etc.

   (c) Find the sequence of times that elapse while the rabbit traveled the distance intervals in the previous part.

   (d) Use the theory of geometric sequences to resolve Zeno's paradox by reconciling your findings in the previous steps of this problem.

10. The following problem is a modern version of Zeno's paradox. It is taken from an article by George Andrews in the January 1998 issue of the *American Mathematical Monthly*, Vol. 105 No. 1 and is attributed to a Prof. Sleator.

    Two trees are one mile apart. A drib flies from one tree to the other and back, making the first trip at 10 miles per hour, the return at 20 miles per hour, the next at 40 and so on, each successive mile at twice the speed of the preceding.

    (a) Write the first five terms of the sequence of velocities for trip numbers 1, 2, 3, etc. Write an explicit formula for this sequence.

    (b) Write the first five terms of the sequence times taken for trips 1, 2, 3, etc. Write an explicit formula for this sequence.

    (c) Write the first five terms of the sequence of how much time it takes the drib to travel 1 mile, 2 miles, 3 miles, etc. Write an explicit formula for this sequence.

    (d) Where is the drib 12 minutes after the first trip begins?

(e) What limitation of the physical world prevents this paradox?

11. *Koch's snowflake* is a geometric figure built recursively, as follows. Stage 1 is an equilateral triangle whose sides are 1 unit in length. To produce Stage n from Stage $n-1$, perform the replacement of edges illustrated in Figure 16.

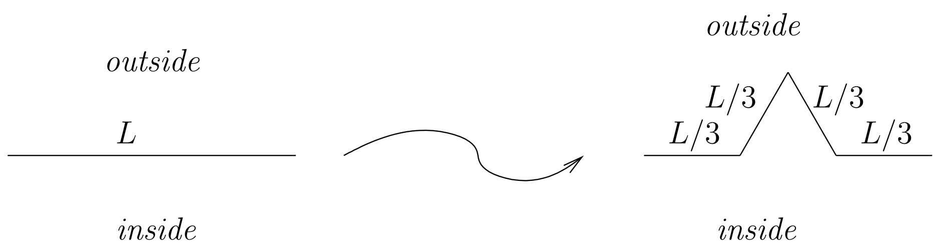


Figure 16 Each edge at stage $n-1$ is replaced by four edges at stage n

The snowflake is the figure obtained after infinitely many stages. Stages 1, 2, 3 and the finished snowflake are shown in Figure 17 (this image is from the website of the Geometry Center of the University of Minnesota).

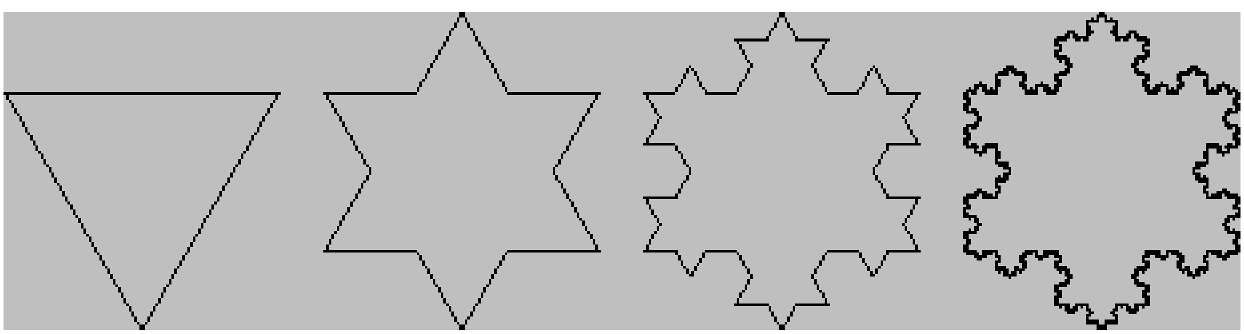


Figure 17 The first three stages and the finished snowflake

(a) Write the first 5 terms of the sequence of the number of edges in Stages 1, 2, 3, etc. Write an explicit formula for this sequence.

(b) Write the first 5 terms of the sequence of the lengths of each edge in Stages 1, 2, 3, etc. Write an explicit formula for this sequence.

(c) Write the first 5 terms of the sequence of the total perimeter of the figure for Stages 1, 2, 3, etc. Write an explicit formula for this sequence. Hint: Multiply your results from the previous two parts.

(d) Write the first 5 terms of the sequence of the number of new equilateral triangle "bumps" added at Stages 1, 2, 3, etc. Write an explicit formula for this sequence. Hint: Adapt your result from part (a).

(e) Write the first 5 terms of the sequence of areas of each new equilateral triangle "bump" added at Stages 1, 2, 3, etc. Write an explicit formula for this sequence. Hint: The area of an equilateral triangle whose side measures s units of length is $s^2\sqrt{3}/4$. Hint: Use part (b).

(f) Write the first 5 terms of the sequence of the new area added (total area of all the new "bumps") at Stages 1, 2, 3, etc. Write an explicit formula for this sequence. Hint: Multiply your results from the previous two parts.

(g) Write the first 5 terms of the sequence of total area for Stages 1, 2, 3, etc. Write an explicit formula for this sequence. Hint: Sum the terms of the geometric sequence from the previous part. Watch out! The first couple of terms may not fit the pattern of the sequence.

(h) What is the perimeter of the snowflake?

(i) What is the area of the snowflake?

12. Here is a recursive description for a sequence that is neither arithmetic nor geometric, made by combining addition and multiplication in the recursive process.

$$\begin{aligned} a_0 &= 2 \\ a_n &= 3a_{n-1} + 1 \end{aligned}$$

(a) Find the terms $a_0, a_1, a_2, a_3, a_4, a_5$ of the sequence.

(b) Find a closed form for the sequence.

(c) Find the terms a_{10}, a_{50}, a_{100}.

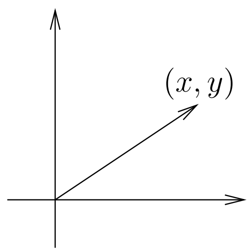


Figure 18
The vector (x, y)

4 Complex Numbers

Motivation

The Euclidean line is the set used to measure a great many "real world" phenomena such as time, distance, mass, temperature, and so on. It is for this reason that we call the line *real*. Analysis of mathematical models involving functions on the real line is made rich and powerful by the algebraic structure of the real numbers. Key features of this algebra are the operations of addition and multiplication, together with the distributive law which governs their interaction.

The Euclidean plane is the set used to measure any kind of data that consists of *pairs* of real numbers. Examples include graphs of functions on the real line and two-dimensional geometric figures. A natural question is whether the algebra of the line extends in any useful ways to the plane. The answer is yes. The extension of the algebra of the reals to the plane forms the *complex* numbers.

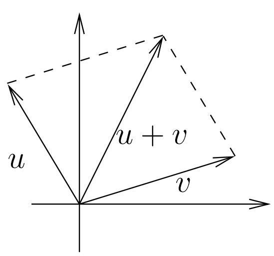


Figure 19
The parallelogram law
of vector addition

Vectors and vector operations

The complex numbers, denoted $\mathbf{C}$, is defined to be the set of points in the Euclidean plane, that is, the set

$$\mathbf{R}^2 = \mathbf{R} \times \mathbf{R} = \{(x, y) \mid x, y \in \mathbf{R}\}$$

where $\mathbf{R}$ denotes the set of real numbers. We can think of a point (x, y) as a *vector*, or directed line segment, which begins at the origin and ends at (x, y). In drawings, we represent vectors by arrows as in Figure 18. This comes from physics where vectors model physical quantities such as displacement, velocity or force.

The way in which physical vector quantities combine leads to the operation of *vector addition*. The sum of two vectors (a, b), (c, d) is defined to be

$$(a, b) + (c, d) = (a + c, b + d). \tag{4.0.1}$$

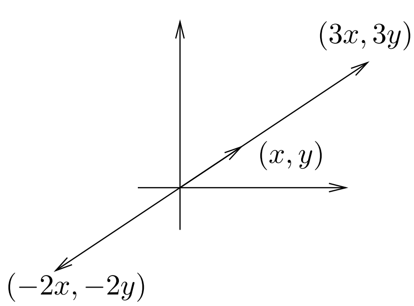


Figure 20
Scalar multiples of
a vector

Geometrically, two vectors v, w and their sum $v+w$ form two sides and a diagonal of a parallelogram; for this reason, equation (4.0.1) is called the *parallelogram law*. See Figure 19.

Given a vector $v = (x, y)$ and a real number k, the *scalar product* of k times v, denoted kv, is defined to be (kx, ky). The vector kv is $|k|$ times as long as v and points in the same direction as v if k is positive and in the opposite direction from v if k is negative. See Figure 20.

Vector addition and scalar multiplication obey the *distributive law*. For any two vectors v, w and any real number k, we have

$$k(v + w) = kv + kw. \tag{4.0.2}$$

Notice that vector operations say how to add two points in the plane, and how to multiply a point in the plane by a real number, but *not* how to multiply two vectors. We explain how to multiply complex numbers in §4 below.

Points, vectors and scalars as complex numbers

While points and vectors may seem to be different types of objects, the set which represents them both is the same, namely $\mathbf{R}^2$. We refer to an ordered pair (x, y) alternately as a point or as a vector, depending on convenience. This takes some getting used to, but it turns out to be useful to think of complex numbers both ways.

Similarly, a scalar $k \in \mathbf{R}$ and a vector $v \in \mathbf{R}^2 = \mathbf{C}$ seem to be different kinds of things. It is also natural, however, to think of the real number line as a *subset* of the plane. The x-axis, that is, the set of points of the form $\{(x, 0) \mid x \in \mathbf{R}\}$, is in one-to-one correspondence with the real numbers via $(x, 0) \leftrightarrow x$. In this way, we think of the real number x as the complex number $(x, 0)$. This is what we mean when we say that that the Euclidean plane *extends* the Euclidean line.

To summarize this subsection: Complex numbers may be thought of as either points or vectors in the plane; real numbers are also complex numbers since points on the real line (the x-axis) are also points in the plane.

Polar coordinates

The *norm* of a point $p = (x, y)$, denoted $|p|$, is defined to be

$$|p| = \sqrt{x^2 + y^2}.$$

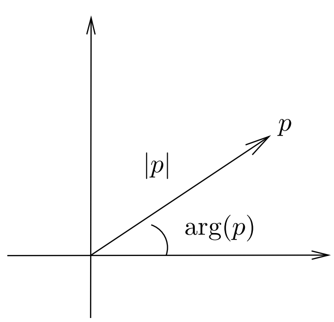


Figure 21
Norm and argument

Geometrically, the norm of a point is its length as a vector, which is the same as its distance from the origin $(0, 0)$. The *argument* of p, denoted $\arg(p)$, is the measure in radians of the directed angle with vertex at the origin, whose initial ray is the positive x-axis and whose terminal ray passes through p. See Figure 21. We agree that two arguments are the same if they differ by an integer multiple of 2π.

The norm and argument of a point are called the *polar coordinates* of that point. By contrast, the numbers x, y in the ordered pair $(x, y) \in \mathbf{R}^2$ are called the *rectangular* or *Cartesian* coordinates of the point. In these notes, we shall use the notation $P_{(r,\theta)}$ to denote the point whose polar coordinates are (r, θ).

Multiplication

The "obvious" way to attempt multiplication of vectors is to mimic vector addition by defining $(a, b) \cdot (c, d) = (ac, bd)$. However, this leads to an unacceptable conflict, as we shall describe.

Let k be a real number and let $v = (x, y)$ be a vector. As a point in the plane, the real number k is $(k, 0)$, so we would have $kv = (k, 0) \cdot (x, y) = (kx, 0)$. On the other hand, we have already defined (scalar multiplication) kv to be $k \cdot (x, y) = (kx, ky)$. But $(kx, 0)$ does not equal (kx, ky) unless one or both of k, y are zero. This example says that if we want scalar multiplication of vectors to be compatible with multiplying pairs of complex numbers, we must look for a different way to multiply points than component-wise.

To see another way, we examine some multiplications of real numbers using polar coordinates.

$$\begin{array}{rcccccl} P_{(2,0)} \cdot P_{(3,0)} & = & 2 \cdot 3 & = & 6 & = & P_{(6,0)} \\ P_{(2,0)} \cdot P_{(3,\pi)} & = & 2 \cdot -3 & = & -6 & = & P_{(6,\pi)} \\ P_{(2,\pi)} \cdot P_{(3,\pi)} & = & -2 \cdot -3 & = & 6 & = & P_{(6,2\pi) = P(6,0)} \end{array}$$

These examples suggest that in polar coordinates, norms multiply and arguments add. That is the inspiration for the following definition of multiplication.

$$P_{(r,\theta)} \cdot P_{(s,\varphi)} = P_{(rs,\theta+\phi)} \tag{4.0.3}$$

This multiplication does indeed extend the ordinary multiplication of real numbers, and resolves the conflict from the previous paragraph. It is not immediately obvious, but it turns out that other desirable properties of the algebra of the real numbers also hold in the plane. In particular, extended addition and multiplication satisfy the distributive law

$$z(u+v) = zu + zv \tag{4.0.4}$$

for all z, u, v in the plane. Note that we normally omit the dot between two points and write zw to denote the product of two points z and w, just as we do for real numbers.

Real and imaginary parts, rectangular form

The complex number $(0,1) = P_{(1,\pi/2)}$ is denoted by the symbol i. It has the curious property that its square is negative one. Note that the number $z = (x, y)$ is equal to $x(1,0) + y(0,1) = x + yi$. The real number x is called the *real part* and the real number y is called the *imaginary part* of $z = x + yi$, and we write $\mathrm{Re}(z) = x$ and $\mathrm{Im}(z) = y$. A number of the form $(0, y) = yi$ is called *pure imaginary*. The coordinates (x, y) of z are called the *rectangular* or *Cartesian* coordinates, and the expression $x + yi$ is called the *rectangular form* of z.

Next we demonstrate how to multiply two complex numbers in rectangular form without converting to polar coordinates. Let $(x, y) = x+yi$ and let $(u, v) = u+vi$ be two complex numbers. Using the distributive law, we have

$$\begin{array}{rcll} (x,y) \cdot (u,v) & = & (x+yi)(u+vi) & \\ & = & xu + yui + xvi + yvi^2 & \text{(distributing)} \\ & = & xu + yui + xvi - yv & (i^2 = -1) \\ & = & (xu - yv) + (yu + xv)i & \text{(collecting real and imaginary parts).} \end{array}$$

Exponential notation and polar form

The elementary functions of a single real variable, including polynomials, rational functions, sine, cosine, and the natural exponential function can be extended to functions of a complex variable. This is done in the subject of *complex analysis*. It turns out that the natural exponential function (that is, the function that sends the complex number z to the complex number e^z) has the following formula for pure imaginary exponents.

(4.0.5) $$e^{it} = \cos t + i \sin t \qquad \text{for } t \in \mathbf{R}$$

Equation (4.0.5) is not supposed to be obvious; it takes a bit of work to even define what e^z means for a complex number z. However, it is not necessary to know the theory of the complex exponential function to use (4.0.5). For now, you may think of (4.0.5) as a definition of the symbols e^{it}.

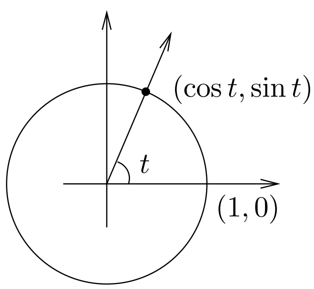


Figure 22
Point on
the unit circle

Recall that $p = (\cos t, \sin t)$ is the point on the unit circle intersected by the terminal ray of a directed angle of t radians in standard position. See Figure 22. It follows that $e^{it} = P_{(1,t)}$ has norm 1 and argument t. If z is an arbitrary complex number with norm r and argument θ, then we have

(4.0.6) $$z = P_{(r,\theta)} = rP_{(1,\theta)} = re^{i\theta}.$$

The expression on the right-hand side of the above equation is called the *polar form* of the complex number z.

Conjugation

Let $z = (x, y) = P_{(r,\theta)} = x + iy = re^{i\theta}$ be a complex number. The *conjugate* of z, denoted $\overline{z}$ is defined to be

(4.0.7) $$\overline{z} = (x, -y) = P_{(r,-\theta)} = x - iy = re^{-i\theta}.$$

Geometrically, the conjugate of a point is the reflection of that point across the x-axis. Here is a useful relation that involves the conjugate.

(4.0.8) $$z\overline{z} = |z|^2$$

Exercises

1. What is the difference between polar coordinates and polar form? What is the difference between rectangular coordinates and rectangular form? Write formulas for converting from polar to rectangular coordinates and vice-versa.

2. Express each of the following in rectangular and polar form.

   (a) $3(2 - i) + 6(1 + i)$
   (b) $\left(2e^{i\pi/6}\right)\left(3e^{-i\pi/3}\right)$
   (c) $(2 + 3i)(4 - i)$
   (d) $(1 + i)^3$

3. Prove the following property of norm, for all complex numbers z, w.

   $$|zw| = |z||w|$$

   Do the proof using rectangular and polar forms. Which is easier?

4. Prove the following property of norm, called the *triangle inequality*. For any two complex numbers z, w, we have

   $$|z + w| \leq |z| + |w|.$$

5. Prove (4.0.8).

6. Let p and q be complex numbers. Prove that the distance (ordinary distance between points in the plane) between p and q is $|p - q|$. Hint: Use rectangular form.

7. Verify the distributive law (4.0.4). Suggestion: First prove case (i) where z is a real number. Next prove case (ii) where z is has norm 1 (use the fact that the diagonal of a rotated parallelogram is the rotation of the diagonal of the original parallelogram. Finally prove the general case where $z = re^{i\theta}$.

8. This exercise outlines the definitions and properties of complex division.

   (a) Notice that the real number 1, considered as the complex number (1,0), has the property that $1z = z$ for any complex number z. For this reason, 1 is called a *multiplicative identity* for **C**. Are there any other multiplicative identities? That is, does there exist any other complex number u with the property that $uz = z$ for every complex number z? If so, find one. If not, explain why none exists.

   (b) Given two complex numbers u and v with $v \neq 0$, the quotient of u divided by v, denoted u/v, is defined to be the complex number z with the property that $u = vz$. Write expressions for $z = u/v$ and $w = 1/v$ in polar form if $u = re^{i\theta}$ and $v = se^{i\varphi}$.

       Note a new definition: we call $1/v$ the *reciprocal* or *multiplicative inverse* of v, and also write it as v^{-1}. Notice that multiplicative identity, division and multiplicative inverse are defined the same as for real numbers.

   (c) Suppose that z, w are nonzero complex numbers. Prove that

   $$(1/z)(1/w) = 1/(zw).$$

   (d) Find the multiplicative inverse of $z = x + iy$ in rectangular form (assume $z \neq 0$). Hint: Multiply $1/z$ by $\overline{z}/\overline{z}$ and use the previous exercise. This is called *rationalizing the denominator.*

9. Express each of the following in rectangular and polar form.

   (a) $\dfrac{2+i}{3-i}$

   (b) $\dfrac{1+2i}{1-2i}$

   (c) $\dfrac{2e^{i\pi/4}}{3e^{-i\pi/2}}$

10. Verify the following formulas. For any complex number z, we have

    (a) $\mathrm{Re}(z) = \dfrac{z+\overline{z}}{2}$, and

    (b) $\mathrm{Im}(z) = \dfrac{z-\overline{z}}{2i}$.

11. Given a nonzero complex number z, explain why z has exactly two square roots, and explain how to find them.

12. Find all complex solutions of the following equations.
    (a) $z^2 + 3z + 5 = 0$
    (b) $(z - i)(z + i) = 1$
    (c) $\dfrac{2z + i}{-z + 3i} = z$
13. Derive the double angle identities for $\cos 2\theta$ and $\sin 2\theta$ by computing $(e^{i\theta})^2$ two ways: in polar form and in rectangular form. Then compare real and imaginary parts.
14. Graph the solutions to the following complex equations.
    (a) $|z - 2| = 3$
    (b) $|4z - 2i| = 3$
    (c) $\mathrm{Im}(z) = 3$
    (d) $\mathrm{Im}(2e^{i\pi/4}z - 2 + 3i) = 0$

5 Vectors, Linear Maps, and Matrices

Euclidean Space

Let $\mathbf{R}$ denote the set of real numbers. For a given positive integer n, the set

$$\mathbf{R}^n = \underbrace{\mathbf{R} \times \mathbf{R} \times \cdots \times \mathbf{R}}_{n \text{ factors}} = \{(x_1, x_2, x_3, \ldots, x_n)\}$$

of ordered n-tuples of real numbers is called n-dimensional Euclidean space or *Euclidean n-space*.

The numbers $\{x_i\}$ are called the coordinates of the point $x = (x_1, \ldots, x_n)$. The space $\mathbf{R}^1 = \mathbf{R}$ is the real number line, $\mathbf{R}^2$ is the plane of high school geometry and algebra, and $\mathbf{R}^3$ is the mathematical abstraction for the familiar 3-space in which we live. The space $\mathbf{R}^0$ is defined to be the one point set $\mathbf{R}^0 = \{0\}$.

Vector Operations

Points in Euclidean space are sometimes called vectors, and real numbers are sometimes called scalars. In multivariate calculus and physics courses, vectors are often denoted using an arrow decoration like "$\vec{x}$", but it is also common to omit any decorations, as we choose to do in these notes. Given two vectors $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ and a scalar α, we define two operations called scaling and vector addition. The vector αx is defined to be

(5.0.9)
$$\alpha x = (\alpha x_1, \alpha x_2, \ldots, \alpha x_n).$$

We say that αx is the vector x *scaled by the factor* α. The vector $x + y$, called the vector sum of x and y, is defined to be

(5.0.10)
$$x + y = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n).$$

Scalar multiplication and vector addition obey the following distributive law, which is easy to verify.

(5.0.11) **Distributive law for vector operations.** For any scalar α and vectors x, y of the same dimension, we have

$$\alpha(x + y) = \alpha x + \alpha y.$$

The vector $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ with a 1 in the ith coordinate and zeroes in all other coordinates is called the ith standard basis vector in $\mathbf{R}^n$. In $\mathbf{R}^2$, the standard basis vectors $e_1 = (1, 0)$ and $e_2 = (0, 1)$ are also called $\mathbf{i}$ and $\mathbf{j}$, respectively. In $\mathbf{R}^3$, the standard basis vectors $e_1 = (1, 0, 0)$, $e_2 = (0, 1, 0)$ and $e_3 = (0, 0, 1)$ are also called $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$, respectively. Given a vector $x = (x_1, \ldots, x_n)$, we have the following representation of x as a sum of scalar multiples of the standard basis vectors (note that the summation sign indicates vector addition).

(5.0.12)
$$x = \sum_{i=1}^{n} x_i e_i$$

The inner product or dot product[1] of two vectors $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ is defined to be the scalar quantity

$$\langle x, y\rangle = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n.$$

In terms of inner product, the ith coordinate x_i of the vector $x = (x_1, \ldots, x_n)$ is given by

$$x_i = \langle e_i, x\rangle \tag{5.0.13}$$

and (5.0.12) becomes

$$x = \sum_{i=1}^{n} \langle e_i, x\rangle e_i. \tag{5.0.14}$$

Linear Maps and Matrices

Because vector operations are useful, it is natural to consider functions or maps that respect vector operations. We call these maps *linear*.

(5.0.15) **Definition of Linear Map.** A function or map $L\colon \mathbf{R}^n \to \mathbf{R}^m$ is called linear if

(i) $L(\alpha x) = \alpha L(x)$, and

(ii) $L(x + y) = L(x) + L(y)$

for all vectors x, y in $\mathbf{R}^n$ and scalars α in $\mathbf{R}$. We describe properties (i) and (ii) by saying that L preserves or respects vector operations of scaling and addition.

Given a vector $x = (x_1, \ldots, x_n)$ and a linear map $L\colon \mathbf{R}^n \to \mathbf{R}^m$, we have

$$\begin{aligned} L(x) &= L(x_1, x_2, \ldots, x_n) \\ &= L\left(\sum_{j=1}^{n} x_j e_j\right) \\ &= \sum_{j=1}^{n} L(x_j e_j) \\ &= \sum_{j=1}^{n} x_j L(e_j) \end{aligned} \tag{5.0.16}$$

A consequence of this equation is that a linear map is determined by its values on the standard basis vectors $e_1, e_2, \ldots, e_n$. We can write an explicit formula for the coordinates $(y_1, y_2, \ldots, y_m)$ of the value $y = L(x)$. Let $f_1, f_2, \ldots, f_m$ denote the standard basis vectors for $\mathbf{R}^m$. Then we have

$$y_i = \langle f_i, L(x)\rangle \tag{5.0.17}$$

[1] In multivariable calculus, the dot product of vectors x and y is denoted by $x \cdot y$, whence the name "dot" product. In order to avoid confusion with matrix operations soon to be defined, we do not use this notation in linear algebra (see the comment preceding 3A on p.143 of the text).There are many variations in use for inner product notation. Strang uses (x, y), but we avoid this because it looks like an ordered pair. Physicists use the symbols $\langle x|y\rangle$, called *Dirac* notation, and refer to the "bra" $\langle x|$ and the "ket" $|y\rangle$ of the "bracket" $\langle x|y\rangle$.

$$= \langle f_i, \sum_{j=1}^{n} x_j L(e_j) \rangle$$
$$= \sum_{j=1}^{n} x_j \, \langle f_i, L(e_j) \rangle.$$

This last expression shows that the values of a linear function are completely determined by the numbers

$$a_{ij} = \langle f_i, L(e_j) \rangle \tag{5.0.18}$$

where i is in the range $1 \le i \le m$ and j is in the range $1 \le j \le n$. We call the rectangular array of numbers

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \tag{5.0.19}$$

the matrix for L, and denote it by $[L]$ or $[a_{ij}]$. The numbers a_{ij} are called the entries of the matrix. Rows of the matrix are numbered top to bottom, and columns are numbered left to right. A matrix with m rows and n columns is called an $m \times n$ matrix.

Written out fully, the equations for the value $y = L(x)$ are the following.

$$\begin{array}{rcccccccc} y_1 & = & a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n \\ y_2 & = & a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n \\ & \vdots & & & & & & & \\ y_i & = & a_{i1}x_1 & + & a_{i2}x_2 & + & \cdots & + & a_{in}x_n \\ & \vdots & & & & & & & \\ y_m & = & a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n \end{array} \tag{5.0.20}$$

The expressions on the right sides are inner products.

$$y_i = \langle (\text{the } i\text{th row of } [L]), x \rangle \tag{5.0.21}$$

For the special case when the input vector x is a standard basis vector e_j, we see that

$$y = L(e_j) = (a_{1j}, a_{2j}, \ldots, a_{mj}) \tag{5.0.22}$$

or, in words,

$$L(e_j) \text{ is the } j\text{th column of } [L]. \tag{5.0.23}$$

Exercises

1. Write each of the following vectors x in the form $(x_1, x_2, \ldots, x_n)$ and $\sum x_i e_i$. For $n = 2, 3$, also write x using $\mathbf{i}, \mathbf{j}, \mathbf{k}$ notation.
   Example: Given $x = 3(2, 4, -1)$, write $x = (6, 12, -3) = 6e_1 + 12e_2 - 3e_3 = 6\mathbf{i} + 12\mathbf{j} - 3\mathbf{k}$.

(a) $x = (3, 2) - (5, -2)$
(b) $x = 2(-1, 2, 1) + 3(2, -2, 0)$
(c) $x = 2e_1 - 3e_2 + 4e_4 - (e_1 - e_2 + e_3)$

2. Prove the distributive law (5.0.11).
3. Verify equation (5.0.12).
4. Show that for $x = (x_1, \ldots, x_n)$, we have $x_i = e_i \cdot x$.
5. Show that
$$e_i \cdot e_j = \delta_{ij}$$
where δ_{ij}, called the Kronecker delta, is given by
$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$
6. Let $L: \mathbf{R}^2 \to \mathbf{R}^2$ be a linear map such that $L(e_1) = (2, 3)$ and $L(e_2) = (-1, -2)$. Find $L(1, 2)$.
7. Let $L: \mathbf{R}^3 \to \mathbf{R}$ be a linear map. Find $L(\mathbf{k})$ if $L(\mathbf{i}) = 2$, $L(\mathbf{j}) = -1$, and $L(1, 2, 3) = 0$.
8. Let $L: \mathbf{R}^3 \to \mathbf{R}^2$ be a linear map with $L(e_1) = (1, 2)$, $L(e_2) = (-1, 1)$, and $L(e_3) = (0, 1)$.
   (a) Write the matrix for L.
   (b) Evaluate $L(2, 1, 3)$.
   (c) Evaluate $L(0, 1, 1)$.
9. Let $L: \mathbf{R}^2 \to \mathbf{R}^3$ be a linear map with the following matrix.
$$\begin{bmatrix} 1 & 2 \\ 3 & -1 \\ 0 & 1 \end{bmatrix}$$
   (a) Evaluate $L(1, 2)$.
   (b) Evaluate $L(-2, 1)$.

[2T p.126]

10. Show that the two linearity properties in the definition (5.0.15) of linear map are equivalent to the single property
$$L(\alpha x + \beta y) = \alpha L(x) + \beta L(y) \tag{5.0.24}$$
for all vectors x, y and scalars α, β.
11. Justify each equality in (5.0.16).
12. In formulas (5.0.17) and (5.0.18), what is the difference between e_i and f_i? Aren't both of these vectors with 1 in the ith coordinate and 0 elsewhere?
13. The dot product has the following properties that look like the properties in the definition of linear map.
$$\begin{aligned} u \cdot (\alpha v) &= \alpha u \cdot v \\ u \cdot (v + w) &= u \cdot v + u \cdot w \end{aligned}$$
for all u, v, w in $\mathbf{R}^n$ and scalars α.

(a) Show that these properties hold.

(b) Exactly where in this section did we use these properties?

14. Justify the steps of the derivation (5.0.17).

15. Verify (5.0.23).

16. Prove that the map $r\colon \mathbf{R}^2 \to \mathbf{R}^2$ which rotates the plane $1/4$ turn counterclockwise about the origin is a linear map. Find the matrix for this linear map.

6 Operations on Linear Maps and Matrices

The set of linear maps is equipped with several natural operations. The first is scalar multiplication. Given a scalar α and a linear map $L\colon \mathbf{R}^n \to \mathbf{R}^m$, the map $\alpha L\colon \mathbf{R}^n \to \mathbf{R}^m$ is defined by

$$(\alpha L)(x) = \alpha L(x). \tag{6.0.1}$$

The next is addition of linear maps. Given maps $L, L'\colon \mathbf{R}^n \to \mathbf{R}^m$ the map $L + L'\colon \mathbf{R}^n \to \mathbf{R}^m$ is given by

$$(L + L')(x) = L(x) + L'(x). \tag{6.0.2}$$

The last is composition. Given $L\colon \mathbf{R}^n \to \mathbf{R}^m$ and $M\colon \mathbf{R}^p \to \mathbf{R}^n$, the composition $L \circ M\colon \mathbf{R}^p \to \mathbf{R}^m$ is given by

$$(L \circ M)(x) = L(M(x)). \tag{6.0.3}$$

By virtue of the association $L \leftrightarrow [L]$ that pairs a linear map $L\colon \mathbf{R}^n \to \mathbf{R}^m$ with its associated $m \times n$ matrix $[L]$, we can impose these operations on matrices. Given a scalar α, two $m \times n$ matrices A, A', and an $n \times p$ matrix B, let L, L' be the associated linear maps for A, A' (that is, $A = [L]$ and $A' = [L']$) and let M be the associated linear map for B (so $B = [M]$). We define new matrices αA, $A + A'$, and AB in terms of the associated linear maps as follows.

$$\alpha A = [\alpha L], \tag{6.0.4}$$

$$A + A' = [L + L'], \text{ and} \tag{6.0.5}$$

$$AB = [L \circ M]. \tag{6.0.6}$$

These operations are called scalar multiplication, matrix addition, and matrix multiplication, respectively. Here are formulas for these three basic matrix operations in terms of matrix entries of A, A' and B.

$$i, j \text{ entry of } \alpha A = \alpha(i, j \text{ entry of } A) \tag{6.0.7}$$

$$i, j \text{ entry of } A + A' = (i, j \text{ entry of } A) + (i, j \text{ entry of } A') \tag{6.0.8}$$

$$i, j \text{ entry of } AB = \langle (i\text{th row of } A), (j\text{th column of } B) \rangle \tag{6.0.9}$$

In most texts, the above formulas are given as *definitions.* In these notes, these formulas are *consequences* of the definitions (6.0.4), (6.0.5), and (6.0.6). We make this choice to emphasize that matrix algebra operations are natural because they come from the corresponding natural operations on linear maps. There is a fourth basic operation, called transposition, whose corresponding operation on linear maps is less easy to describe. We define the transpose of matrix A, denoted A^T, by

$$i, j \text{ entry of } A^T = j, i \text{ entry of } A. \tag{6.0.10}$$

Formulas (6.0.7) and (6.0.8) are easy to see, but (6.0.9) is less straightforward. To see why (6.0.9) is true, let f_i be the ith standard basis vector in $\mathbf{R}^p$ and let e_j be the jth standard basis vector in $\mathbf{R}^n$. We have

$$\begin{aligned}
i, j \text{ entry of } AB &= i, j \text{ entry of matrix for } L \circ M \qquad \text{(definition (6.0.6)} \\
&= \langle f_i, L(M(e_j)) \rangle \qquad \text{(applying (5.0.18))} \\
&= \langle f_i, L(j\text{th column of } M) \rangle \qquad \text{(applying (5.0.23))} \\
&= i\text{th entry of } L(j\text{th column of } M) \qquad \text{applying (5.0.13)} \\
&= \langle (i\text{th row of } [L]), (j\text{th column of } M) \rangle \qquad \text{(by (5.0.21).}
\end{aligned}$$

Vectors and Column Matrices

Given a vector $x = (x_1, x_2, \ldots, x_n)$, we define the matrix for x, denoted $[x]$, to the the $n \times 1$ matrix

$$[x] = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

A matrix with one column is called a column matrix. The identification of vectors with column matrices is standard and nearly universal. It is normal usage to use the symbol x to refer to the vector and the column matrix without bothering to use square brackets or other notation to make a distinction. With this convention, we have a way to compute $L(x)$ using matrix multiplication. The proof is an exercise.

(6.0.11) **Evaluation and Matrix Multiplication.** The matrix product $[L][x]$ is the column matrix for $L(x)$. In other words,

$$[L(x)] = [L][x].$$

Handy Matrix Multiplication Formulas

In this section, for a vector $x = (x_1, \ldots, x_n)$ in $\mathbf{R}^n$, we will follow standard practice and use the symbol x to denote both the vector and its associated column matrix $[x]$.

Let x, y be vectors in $\mathbf{R}^n$. The matrix product $x^T y$ is the 1×1 matrix

$$x^T y = [x_1 y_1 + x_2 y_2 + \cdots + x_n y_n]$$

whose single entry is the inner product $\langle x, y \rangle$. We identify 1×1 matrices with numbers, and so we write

(6.0.12) $$x^T y = \langle x, y \rangle.$$

Let $e_1, e_2, \ldots, e_n$ be the standard basis vectors for $\mathbf{R}^n$, and let $f_1, f_2, \ldots, f_m$ be the standard basis vectors for $\mathbf{R}^m$. Given an $m \times n$ matrix A, we have

(6.0.13) $$Ae_j = j\text{th column of } A,$$

(6.0.14) $$f_i^T A = i\text{th row of } A, \textit{and}$$

(6.0.15) $$f_i^T A e_j = i, j \text{ entry of } A.$$

Block Multiplication

Suppose A is an $m \times n$ matrix and B is an $n \times p$ matrix, so that the product AB is defined. If we subdivide A and B into blocks (submatrices) of compatible sizes, we can perform matrix multiplication using the blocks. Here is a picture of the situation.

$$\underbrace{\left[\begin{array}{c|c} A_{11} & A_{12} \\ m_1 \times n_1 & m_1 \times n_2 \\ \hline A_{21} & A_{22} \\ m_2 \times n_1 & m_2 \times n_2 \end{array}\right]}_{A} \underbrace{\left[\begin{array}{c|c} B_{11} & B_{12} \\ n_1 \times p_1 & n_1 \times p_2 \\ \hline B_{21} & B_{22} \\ n_2 \times p_1 & n_2 \times p_2 \end{array}\right]}_{B}$$

The result is the following.

$$AB = \left[\begin{array}{c|c} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ \hline A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{array}\right] \tag{6.0.16}$$

Exercises

1. Perform the matrix multiplications below.

   (a)
   $$\begin{bmatrix} 1 & -2 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} 0 & 2 \\ 1 & -1 \end{bmatrix}$$

   (b)
   $$\begin{bmatrix} 1 & -2 & 0 \\ -1 & 3 & 1 \end{bmatrix} \begin{bmatrix} 0 & 2 \\ 1 & -1 \\ 2 & 0 \end{bmatrix}$$

2. Let $L: \mathbf{R}^2 \to \mathbf{R}$ and $K: \mathbf{R} \to \mathbf{R}^3$ be linear maps such that $L(\mathbf{i}) = 1$, $L(\mathbf{j}) = -2$ and $K(1) = (2, 1, 3)$.
   (a) Write the matrices for L and K.
   (b) Find the matrix for $K \circ L$.
   (c) Find $(K \circ L)(2, -1)$.

3. Prove that the composition of two linear maps is a linear map.

4. Prove (6.0.11).

5. Let r_θ and r_φ be rotations of the plane about the origin by θ and φ radians, respectively. Use matrix multiplication to derive the formulas for cosines and sines of a sum of two angles.

6. Let z_0 be a fixed complex number, and let $f: \mathbf{C} \to \mathbf{C}$ be the map given by $f(w) = z_0 w$. Let $I: \mathbf{R}^2 \to \mathbf{C}$ be the identification of $\mathbf{R}^2$ with $\mathbf{C}$ given by $I(x, y) = x + iy$. Let $L = I^{-1} \circ f \circ I$.
   (a) Verify that the map $L: \mathbf{R}^2 \to \mathbf{R}^2$ is linear.
   (b) Find the matrix for L.
   (c) Let $z_0 = e^{i\theta}$, where θ is a real number. Verify that L is the same thing as r_θ from the previous problem.

7. Let $e_1, e_2, \ldots, e_n$ denote the standard basis vectors in $\mathbf{R}^n$.
   (a) Show that $e_k e_k^T$ is an $n \times n$ matrix with a 1 in the k, k entry and zeroes elsewhere.
   (b) Show that $\displaystyle\sum_{k=1}^{n} e_k e_k^T$ is the $n \times n$ identity matrix.
   [1.4 ex.19 p.28] (c) Given an $m \times n$ matrix A and an $n \times p$ matrix B, show that
   $$AB = \sum_{k=1}^{n} ((\text{column } k \text{ of } A)(\text{row } k \text{ of } B).$$
   Hint: Use (6.0.13) and (6.0.14).

8. Let A be an $m \times n$ matrix, and let $L: \mathbf{R}^n \to \mathbf{R}^m$ be the associated linear map. Define $L^T: \mathbf{R}^m \to \mathbf{R}^n$ to be the linear map associated to the matrix A^T. Let $\langle , \rangle_n$, $\langle , \rangle_m$, denote the inner products in $\mathbf{R}^n, \mathbf{R}^m$, respectively. Show that
$$\langle Lx, y \rangle_m = \langle x, L^T y \rangle_n$$
for all x in $\mathbf{R}^n$, y in $\mathbf{R}^m$.

9. Prove (6.0.16).

Solutions to Exercises

Note: Most of the "solutions" posted here are not solutions at all, but are merely final answer keys, although some are complete. These are posted so that you can check your work; reading the answer keys is not a substitute for working the problems yourself. For homework, quizzes and exams, you need to show the steps of whatever procedure you are using—not just the final result. Sometimes you will be asked to explain your thinking in complete sentences.

0.10 Solutions

1. $\emptyset, \{a\}, \{b\}, \{c\}, \{a,b\}, \{a,c\}, \{b,c\}, \{a,b,c\}$

2. $$\begin{aligned} A \cup B &= \{1,2,3,4,5,6,7,9\} \\ A \cap B &= \{3,5\} \\ A \setminus B &= \{1,7,9\} \\ D \setminus A &= \{0,2,4,6,8\} \end{aligned}$$

   (also show Venn diagram)

3. (Venn diagram)

4. (Venn diagram)

5. (a) $A \times B = \{(a,x),(a,y),(a,z),(b,x),(b,y),(b,z)\}$
   (b) $B^2 = \{(x,x),(x,y),(x,z),(y,x),(y,y),(y,z),(z,x),(z,y),(z,z)\}$

6. (Venn diagram)

7. $$\begin{aligned} A &= \{x\colon -2 \le x < 3\} = [-2,3) \\ B &= \{x\colon 1 < x \le 5\} = (1,5] \\ A \cup B &= \{x\colon -2 \le x \le 5\} = [-2,5] \\ A \cap B &= \{x\colon 1 < x < 3\} = (1,3) \\ A \setminus B &= \{x\colon -2 \le x \le 1\} = (-2,1) \\ R \setminus A &= \{x\colon x < -2 \text{ or } 3 \le x\} = (-\infty,-2) \cup [3,\infty) \end{aligned}$$

   (also sketch intervals)

8. Let $f(x) = x^2$ and $g(x) = x + 2$ define functions f and g from the reals to the reals.

   (a) $f(g(3)) = 25$
   (b) $g(f(3)) = 11$
   (c) $(g \cdot f)(3) = 45$
   (d) $(f/g)(3) = 9/5$
   (e) $(3f + g)(3) = 32$
   (f) $f(g(x)) = (x+2)^2$

(g) $g(f(x)) = x^2 + 2$

9. There are six 1-1 correspondences, indicated below.

```
abc abc abc abc abc abc
123 132 213 231 312 321
```

10. (a) 120

    (b) $\sum_{i=1}^{15}(i^2 + 2i + 3)$

11. (a) 1

    (b) 9

    (c) 8

1.7 Solutions

1. (a) The cat is not hungry.

   (b) Some of the cats in the room are not hungry.
   or
   There is at least one cat in the room that is not hungry.

   (c) None of the cats in the room are hungry.

   (d) All of the cats in the room are hungry.

2. (a) Here is the truth table.

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

   Since the two compound statements have the same truth values for all four logical possibilities, we are done.

   (b) (similar to (a))

3. (similar to the previous problem)

4. (complete the steps)

5. First observe that the number of subsets of a zero element set is $2^0 = 1$ (only the empty set) so the statement holds for $n = 0$. Now suppose the statement is true for $n = k$. Let X be a set with $k + 1$ elements. Label the elements of X with indices 1 through $k + 1$. Let X' be the k element set consisting of elements from X with indices 1 through k. For each subset S of X', let $T(S)$ be the set S union with the single element set containing the $(k+1)$st element of X. Note that the collection $\{S : S \subseteq X'\} \cup \{T(S) : S \subseteq X'\}$ is in fact the collection of all subsets of X. Since there are 2^k different subsets S, it follows that there are $2(2^k) = 2^{k+1}$ subsets of X. Q.E.D.

2.6 Solutions

1. 132
2. 134
3. $8^2 \times 10^8 = 6.4$ billion
4. $20! \approx 2.4 \times 10^{18}$
5. $P(50, 20) \approx 1.15 \times 10^{32}$
6. $50^{20} \approx 9.54 \times 10^{33}$
7. $\binom{20}{11} = 167,960$

3.8 Solutions

1. (a) $-4, -7, -10$, $a = 5$, $d = -3$
   (b) $5(2/5)^3, 5(2/5)^4, 5(2/5)^5$, $a = 5$, $r = 2/5$
   (c) $1/2, 7/2, 13, 2$, $a = 1/2$, $d = 3/2$
   (d) $2^{1/2}, 2^{3/2}, 2^{5/2}$, $a = 2^{1/2}$, $r = 2^{1/2}$
2. 5050
3. (a) 15,352
   (b) 412,126
   (c) $3^{101} - 1 \approx 1.55 \times 10^{48}$
   (d) $3^{15} - 1 = 14,348,906$
4. The expression on the right hand side of (3.5.6) is undefined if $r = 1$, but in this case we simply have $s_n = a + a + \cdots + a = a(n + 1)$.
5. (a) If $a = 0$ or if $r = 1$ or $r = 0$, the function and its associated sequence are constant (at least after the initial term), and so hardly deserve any attention. If r is negative there is a problem of definition for r^x. For example, when $r = -2$ and $x = 1/2$, r^x is not defined.
   (b) Growth corresponds to $d > 0$, decay to $d < 0$.
   (c) Growth corresponds to $r > 1$, decay to $0 < r < 1$.
6. (a) 2^{30} cents
   (b) $2^{31} - 1$ cents
7. (a) $-5 + 5 + 5^2 + 5^3 + 5^4 + 5^5 = \$3900$
   (b) Solve $5\left(\dfrac{5^{n+1} - 1}{5 - 1}\right) > 7 \times 10^9$ (or pick your favorite world population) to get $n = 15$ days
8. (a) Body weight functions are $G(t) = 10 - t/10$ and $P(t) = 10(1/2)^t$. Solve $G(t) = 1$ and $P(t) = 1$ to see that Pokey wins (about 3.3 days versus 90 days).
   (b) There is no t for which $P(t) = 0$.

(c) It takes 100 days for $G(t)$ to hit zero. At that time, Pokey weights $10(1/2)^{100} \approx 7.9 \times 10^{-30}$ ounces.

9. (a) 16 2/3 feet from the rabbit's start, 3 1/3 seconds
   (b) $10, 10(2/5), 10(2/5)^2, \ldots$
   (c) $2, 2(2/5), 2(2/5)^2, \ldots$
   (d) Using the formula for the sum of a geometric series, we see that as $n \to \infty$, the total time approaches $\lim_{n\to\infty} 2\left(\frac{1-(2/5)^{n+1}}{1-2/5}\right) = 2\left(\frac{1}{1-2/5}\right) = 10/3$ and the total distance approaches $\lim_{n\to\infty} 10\left(\frac{1-(2/5)^{n+1}}{1-2/5}\right) = 10\left(\frac{1}{1-2/5}\right) = 50/3$.

10. (a) $10, 10\cdot 2, 10\cdot 2^2, 10\cdot 2^3, 10\cdot 2^4$, $10\cdot 2^{n-1}$, $n \geq 1$
   (b) $1/10 \cdot (1/2)^{n-1}$, $n \geq 1$
   (c) $\frac{1}{10}\left(\frac{1-(1/2)^n}{1-1/2}\right) = \frac{1}{5}\left(1-(1/2)^n\right)$
   (d) Since the sequence in the previous part goes to 1/5 hour or 12 minutes, this means that in 12 minutes the drip has completed an infinite number of 1-mile trips. Perhaps the drib is in both trees at once?
   (e) Since the drib cannot exceed the speed of light, this troubling experiment could not take place.

11. (a) $3\cdot 4^{n-1}$, $n \geq 1$
   (b) $(1/3)^{n-1}$, $n \geq 1$
   (c) $3(4/3)^{n-1}$, $n \geq 1$
   (d) $3\cdot 4^{n-2}$, $n \geq 2$. Note that the first term in this sequence is 1.
   (e) $\frac{\sqrt{3}}{4}(\frac{1}{9})^{n-1}$, $n \geq 1$
   (f) $\frac{3\sqrt{3}}{16}(\frac{4}{9})^{n-1}$, $n \geq 2$. Note that the first term in this sequence is $\sqrt{3}/4$.
   (g) $\frac{\sqrt{3}}{4} + \frac{\sqrt{3}}{12}\left(\frac{1-(4/9)^n}{1-4/9}\right)$
   (h) The perimeter is the limit as $n \to \infty$ of the sequence in part (c), which is infinity.
   (i) The area is the limit as $n \to \infty$ of the sequence in part (g), which is $\frac{2\sqrt{3}}{5}$.

12. (a) $2, 7, 22, 67, 202, 607$
   (b) $2\cdot 3^n + \frac{3^n-1}{2}$
   (c) $147,622$, $\approx 1.79 \times 10^{24}$, $\approx 1.29 \times 1048$

4.9 Solutions

1. Let z be a complex number, let $x = \text{Re}(z)$, $y = \text{Im}(z)$, $r = |z|$ and $\theta = \arg(z)$. The pair (r, θ) is called the polar coordinates for z, while the expression $re^{i\theta}$ is called the polar form for z. The pair (x, y) is called the rectangular coordinates for z, while the expression $x + iy$ is called the rectangular form for z.

   To convert from polar to rectangular, use the equations $x = r\cos\theta, y = r\sin\theta$ (show sketches to explain these formulas). To convert from rectangular to polar, use $r = \sqrt{x^2 + y^2}$ and $\tan\theta = y/x$. For the last equation, you must use judgment when $x = 0$ to decide whether θ should be $\pi/2$ or $-\pi/2$. You must also use judgment to interpret any machine computation of the arctangent function to decide which quadrant θ should be in.

2. (a) $12 + 3i = \sqrt{153}\, e^{i \arctan(1/4)}$
   (b) $6e^{-i\pi/6} = 3\sqrt{3} - 3i$
   (c) $11 + 10i = \sqrt{221}\, e^{i \arctan(10/11)}$
   (d) $-2 + 2i = 2\sqrt{2}\, e^{i3\pi/4}$

3. Let $z = x + iy = re^{i\theta}$ and $w = u + iv = se^{i\varphi}$. In rectangular form, we have

$$\begin{aligned}
|zw| &= |(x+iy)(u+iv)| \\
&= |(xu - yv) + i(xv + yu)| \quad \text{(multiply)} \\
&= \sqrt{(xu - yv)^2 + (xv + yu)^2} \quad \text{(definition of norm)} \\
&= \sqrt{x^2u^2 - 2xyuv + y^2v^2 + x^2v^2 + 2xyuv + y^2u^2} \quad \text{(multiply)} \\
&= \sqrt{(x^2 + y^2)(u^2 + v^2)} \quad \text{(simplify and factor)} \\
&= |z||w|.
\end{aligned}$$

   Alternatively, and more simply, in polar coordinates we have

$$\begin{aligned}
|zw| &= |re^{i\theta}se^{i\varphi}| \\
&= |rse^{i(\theta+\varphi)}| \quad \text{(definition(!) of multiplication)} \\
&= rs \quad \text{(definition of norm)} \\
&= |z||w|.
\end{aligned}$$

   Q.E.D.

4. (This is a challenge problem. The simplest approach is geometric: sketch the parallelogram for vector subtraction and use the fact from elementary geometry that the total length for any two sides of a triangle is more than the length of the remaining side. It is harder, but a fun exercise, to do an algebraic proof.)

5. Let $z = re^{i\theta}$. Then $\overline{z} = re^{-i\theta}$, and we have

$$\begin{aligned}
z\overline{z} &= re^{i\theta}re^{-i\theta} \\
&= r^2e^0 \\
&= r^2 = |z|^2.
\end{aligned}$$

6. Let $p = a + ib$ and $q = c + id$. We have

$$\begin{aligned} |p - q| &= |(a + ib) - (c + id)| \\ &= |(a - c) + i(b - d)| \\ &= \sqrt{(a - c)^2 + (b - d)^2}. \end{aligned}$$

The latter expression is the distance from p to q, so we are done.

7. We prove (4.5.2) in two stages. First we prove the special case where z is a real number. Second we prove the case where $z = e^{i\theta}$ is a complex number with norm 1. Then we put the two results together to get the distributive law.

First, if $z = k$ is a real number, then (4.5.2) becomes (4.2.2). Here is the proof. Let $u = (a, b)$ and $v = (c, d)$ be complex numbers. We have

$$\begin{aligned} z(u + v) &= k((a, b) + (c, d)) \\ &= k((a + c, b + d)) && \text{(definition of addition)} \\ &= (k(a + c), k(b + d)) && \text{(definition of scalar multiplication)} \\ &= (ka + kc, kb + kd) && \text{(distributive law for real numbers)} \\ &= (ka, kb) + (kc, kd) && \text{(definition of addition)} \\ &= k(a, b) + k(c, d) && \text{(definition of scalar multiplication)} \\ &= ku + kv. \end{aligned}$$

This proves the first case.

Second, let $z = e^{i\theta}$. Geometrically, u, v and $u + v$ are three corners of a parallelogram whose fourth corner is the origin. If we rotate this parallelogram θ radians counterclockwise, we have a new parallelogram with corners zu, zv, $z(u + v)$ and the origin. Since the diagonal of this new parallelogram is the sum of the two edge vectors, we have $z(u + v) = zu + zv$.

Finally, let $z = re^{i\theta}, u, v$ be any complex numbers. Then

$$\begin{aligned} z(u + v) &= re^{i\theta}(u + v) \\ &= r(e^{i\theta}u + e^{i\theta}v) && \text{(by the second case above)} \\ &= re^{i\theta}u + re^{i\theta}v && \text{(by the first case above)} \\ &= zu + zv. \end{aligned}$$

Q.E.D.

8. (a) Suppose that the complex number u has the property that $uz = z$ for all complex numbers z. Then we must have $|u||z| = |uz| = |z|$. Dividing by $|z|$ yields $|u| = 1$. We also have $\arg(u) + \arg(z) = \arg(uz) = \arg(z)$ (where it is understood that two arguments are considered equal if they differ by an integer multiple of 2π), so subtracting $\arg(z)$ yields $\arg(u) = 0$. The only complex number with norm 1 and argument 0 is the real number 1, so the multiplicative identity is indeed unique.

(b) Let $z = te^{i\psi}$. Since $re^{i\theta} = u = vz = ste^{i(\varphi+\psi)}$, we must have $r = st$ and $\theta = \varphi + \psi$ (plus possibly some multiple of 2π). Clearly then $z = (r/s)e^{i(\theta-\varphi)}$ satisfies the equation $u = vz$. It is also clear that this z is unique. Setting $u = 1$ gives $w = 1/v = (1/s)e^{-i\varphi}$.

(c) Let $z = re^{i\theta}, w = se^{i\varphi}$. Then $zw = rse^{i(\theta+\varphi)}$, $1/z = (1/r)e^{-i\theta}$, $1/w = (1/s)e^{-i\varphi}$, and $1/(zw) = (1/(rs))e^{-i(\theta+\varphi)}$. The identity $(1/z)(1/w) = 1/(zw)$ clearly holds.

(d) Let $z = x + iy$. Then

$$1/z = \frac{1}{z}\frac{\overline{z}}{\overline{z}} = \frac{\overline{z}}{|z|^2} = \frac{x - iy}{x^2 + y^2} = \frac{x}{x^2 + y^2} - i\frac{y}{x^2 + y^2}.$$

9. (a) $\frac{1}{2} + \frac{1}{2}i = \frac{\sqrt{2}}{2}e^{i\pi/4}$

(b) $-\frac{3}{5} + \frac{4}{5}i = e^{i(\arctan(-4/3)+\pi)}$

(c) $\frac{2}{3}e^{i3\pi/4} = -\frac{\sqrt{2}}{3} + \frac{\sqrt{2}}{3}i$

10. Let $z = x + iy$. Then we have

(a) $\frac{z + \overline{z}}{2} = \frac{2x}{2} = x = \text{Re}(z)$, and

(b) $\frac{z - \overline{z}}{2i} = \frac{2iy}{2i} = y = \text{Im}(z)$.

11. Since squaring a number squares the norm and doubles the argument, a square root can be found by taking the square root of the norm and dividing the argument by two. That is, for $z = re^{i\theta}$, a square root of z is $\sqrt{r}e^{i\theta/2}$. Another square root of z is the negative of that expression. Any other square root of z would have to have norm $\sqrt{|z|}$ and argument $\theta/2$ plus or minus an integer multiple of π, so these must be all the square roots of z.

12. (a) $-\frac{3}{2} \pm i\frac{\sqrt{11}}{2}$

(b) 0

(c) $(1/2)[(-2\pm 281^{1/4}\cos\varphi)+i(3\pm 281^{1/4}\sin\varphi)]$, where $\varphi = (\arctan(16/5)+\pi)/2$

13. On the one hand, we have $(e^{i\theta})^2 = e^{i2\theta} = \cos 2\theta + i\sin 2\theta$. On the other hand, we have $(e^{i\theta})^2 = (\cos\theta + i\sin\theta)^2 = (\cos^2\theta - \sin^2\theta) + i(2\cos\theta\sin\theta)$. Comparing real and imaginary parts yields the double angle identities.

14. (a) (circle of radius 3 centered at 2)

(b) (circle of radius 3/4 centered at $(1/2)i$)

(c) (line $y = 3$)

(d) (line $y = -x - \frac{3\sqrt{2}}{2}$)

5.3 Solutions

1. (a) $x = (-2, 4) = -2e_1 + 4e_2 = -2\mathbf{i} + 4\mathbf{j}$
   (b) $x = (4, -2, 2) = 4e_1 - 2e_2 + 2e_3 = 4\mathbf{i} - 2\mathbf{j} + 2\mathbf{k}$
   (c) $x = (1, -2, -1, 4) = e_1 - 2e_2 - e_3 + 4e_4$

2. $$\begin{aligned} k(x+y) &= k(x_1+y_1, \ldots, x_n+y_n) \quad \text{(definition of vector addition)} \\ &= (k(x_1+y_1), \ldots, k(x_n+y_n)) \quad \text{(definition of scalar multiplication)} \\ &= (kx_1+ky_1), \ldots, kx_n+ky_n) \quad \text{(distributive law for reals)} \\ &= (kx_1, \ldots, kx_n) + (ky_1, \ldots, ky_n) \quad \text{(definition of vect addition)} \\ &= kx + ky \quad \text{(definition of scalar multiplication)} \end{aligned}$$

3. $x = (x_1, x_2, \ldots, x_n) = x_1e_1 + x_2e_2 + \cdots + x_ne_n$

5.5 Solutions

1. $(0, -1)$
2. $L(\mathbf{k}) = 0$
3. 1st equality: definition
   2nd equality: equation (5.2.2)
   3rd equality: first of linearity properties (5.4.1)
   4th equality: second of linearity properties (5.4.2)
4. Let $L(1) = m$. For any x, we have
   $$L(x) = L(x \cdot 1) = xL(1) \text{ (by linearity)} = mx.$$
5. An intuitive geometric proof is simplest. The sum of two vectors is the diagonal of the parallelogram formed by the vectors. If we add two vectors, then rotate their sum, we get the same thing if we rotate the two vectors, then add them. Similarly, if we scale a vector, then rotate the result, we get the same result if we rotate the vector first, then scale the result.

5.7 Solutions

1. (a) $\begin{bmatrix} 1 & -1 & 0 \\ 2 & 1 & 1 \end{bmatrix}$
   (b) $(1, 8)$
   (c) $(-1, 2)$
2. (a) $(5, 1, 2)$
   (b) $(0, -7, 1)$
3. $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$
4. $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$

5.9 Solutions

1. (a) $\begin{bmatrix} -2 & 4 \\ 3 & -5 \end{bmatrix}$

   (b) $\begin{bmatrix} -2 & 4 \\ 5 & -5 \end{bmatrix}$

2. (a) $L = \begin{bmatrix} 1 & -2 \end{bmatrix}$

   $K = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}$

   (b) $\begin{bmatrix} 2 & -4 \\ 1 & -2 \\ 3 & -6 \end{bmatrix}$

   (c) $(8, 4, 12)$

3. (a) This is the definition of the coefficients in the matrix for the linear map $S \circ T$.

   (b) The first equality is true because $(S \circ T)(e_j) = S(T(e_j))$. To obtain the right hand side, replace $T(e_j)$ by a summation expression involving the matrix coefficients for T.

   The second equality is true because S is linear. We can pass the linear map S by the summation sign and then factor out the constants t_{kj}.

   The third equality is simply replacing $S(f_k)$ by a summation involving the matrix coefficients for S.

   The fourth equality is the distributive law; we can pass the constant t_{kj} across the summation.

   The fifth equality simply reorders the summands in the double summation.

   The final summation is the distributive law; we factor out the vectors g_i.

   (c) The fact that the left hand side equals the right hand side in the previous part means (by definition) that the matrix coefficients for $S \circ T$ are given by $c_{ij} = \sum_{k=1}^{v} s_{ik} t_{kj}$.

4. Clearly, $r_{(\varphi+\theta)} = r_\varphi \circ r_\theta$. In matrices, we have

$$\begin{aligned} &\begin{bmatrix} \cos(\varphi+\theta) & -\sin(\varphi+\theta) \\ \sin(\varphi+\theta) & \cos(\varphi+\theta) \end{bmatrix} \\ = &\begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \\ = &\begin{bmatrix} (\cos\varphi\cos\theta - \sin\varphi\sin\theta) & (-\cos\varphi\sin\theta - \sin\varphi\cos\theta) \\ (\sin\varphi\cos\theta + \cos\varphi\sin\theta) & (-\sin\varphi\sin\theta + \cos\varphi\cos\theta) \end{bmatrix}. \end{aligned}$$

   Equating the top left and bottom left entries of the first and last matrices, we have the formulas for cosine and sine of the sum of two angles.

5. (a) Let $u = (a, b)$, $v = (c, d)$ be two vectors in $\mathbf{R}^2$, let k be a real number, and let $z_0 = x_0 + iy_0$. We need to show that $L(u + v) = L(u) + L(v)$ and $L(ku) = kL(u)$. First, show that

$$L(a, b) = (x_0 a - y_0 b, x_0 b + y_0 a).$$

Then verify the two properties of linearity.

(b) $\begin{bmatrix} x_0 & -y_0 \\ y_0 & x_0 \end{bmatrix}$

(c) Since $e^{i\theta} = \cos\theta + i\sin\theta$, by part (b), the matrix for L is the same as the matrix for r_θ.