LAST REVISED: Thursday 28th September, 2017 17:18
THE LATEST VERSION IS AVAILABLE AT http://www.mtholyoke.edu/~jjlee/Teaching/math_prerequisite.pdf

ESSENTIALS

OF

CALCULUS

AND

LINEAR ALGEBRA

FOR

STATISTICS

Contents

1 Basics of Discrete Mathematics **7**

1.1 Terminologies in Set Theory 7
1.2 Set Operations 11
1.3 Basics of Counting 16
1.4 Binomial Theorem and Its Consequences 19
1.5 Other Topics in Discrete Mathematics 23

2 Single Variable Calculus **29**

2.1 Limits and Continuity 29
2.2 Differentiation 37
2.3 Mean Value Theorem and Its Consequences 43
2.4 Finding Extreme Values 46
2.5 Integration 49
2.6 Techniques of Integration 57
2.7 Improper Integrals 62
2.8 Convergence of Sequences and Series 68
2.9 Power Series 74
2.10 Taylor Series 77

3 Linear Algebra **81**

3.1 Linear System and Matrices 81
3.2 Basics of Matrix Algebra 90
3.3 Vector Space and Matrix 99
3.4 Determinants 110
3.5 Orthogonal Projection 115
3.6 Eigenvectors and Eigenvalues 121
3.7 The Characteristic Equation 123
3.8 Diagonalization 125
3.9 Symmetric Matrices and Quadratic Forms 129

4 Multivariable Calculus **141**
4.1 Partial Derivatives and Interchange of Operations 141
4.2 Finding Extreme Values 143
4.3 Matrix Differentiation 145
4.4 Double Integrals 148

Preface

This monograph was designed as a review of standard results in mathematics for those who are planning to take courses in mathematical statistics. I tried to make it self-contained, but no proofs are given, since almost all results given in the monograph are well known and proofs are easily accessible on/off line.

Readers are assumed to have been introduced to calculus, linear algebra, and basic probability to some extent. Since the goal of this monograph is to provide the readers with necessary mathematical background that can help them understand and develop statistical tools, more emphasis was given to intuition and application rather than rigorous treatment of the topics. Whenever appropriate, each section contains examples in statistics that use the material in the section.

To enhance the understanding of the content, **Problem** and **Exercise** are included in addition to **Theorem**, **Remark**, and **Example**. The format of **Problem** is the same as **Exercise**, except that the former is followed by a completely worked out **Answer** ($\blacksquare$ indicates the end of **Answer**). The readers are strongly encouraged to work on **Exercise**.

J.-J. Lee

Chapter 1

Basics of Discrete Mathematics

1.1 Terminologies in Set Theory

Probability is a function defined on a certain collection of sets called events, and set theory lays the foundation for probability theory. In this section, we review basic terminologies of set theory.

Definition 1.1.1. A set is an unordered collection of objects. The objects in a set are called the elements, or members, of the set.

Example 1.1.2. The set consisting of all positive odd integers less than 10 can be expressed by $\{1, 3, 5, 7, 9\}$ or $\{1, 5, 7, 9, 3\}$. Let O denote this set. To indicate that 1 is an element of O, we use the notation $1 \in O$. To indicate that 2 is not an element of O, we write $2 \notin O$.

Example 1.1.3. The set consisting of all positive integers less than 10 can be denoted by $\{1, 2, 3, 4, 5, 6, 7, 8, 9\}$. $\{1, 2, 3, \ldots, 9\}$ is also used since the general pattern is obvious.

Another way to describe a set is to use *set builder* notation: elements in a given set are characterized by the property or properties they must have to be members of the set. For example, $O = \{1, 3, 5, 7, 9\}$ in Example 1.1.2 can also be written as

$$O = \{x : x \text{ is a positive odd integer less than } 10\}.$$

Problem 1.1.4. Describe the following sets by listing all the elements or giving general pattern.

(a) $A = \{x : x \text{ is a prime number less than } 20\}$.

(b) $B = \{n : n \text{ is a positive integer}\}$.

(c) $C = \{x : x^2 - 2x - 3 = 0\}$.

(d) $D = \{2n : n \text{ is an integer and } 1 \leq n \leq 4\}$.

Answer

(a) $A = \{2, 3, 5, 7, 11, 13, 17, 19\}$.

(b) $B = \{1, 2, 3, 4, \ldots\}$.

(c) The roots of the equation $x^2 - 2x - 3 = 0$ are -1 and 3, so $C = \{-1, 3\}$.

(d) Integers n such that $1 \leq n \leq 4$ are $1, 2, 3$, and 4. Since D is the collection of 2 times such n's, $D = \{2, 4, 6, 8\}$.

■

Some common notations for basic sets are given below.

Definition 1.1.5.

(a) $\emptyset$, The empty set (a set with *no* elements). $\{\quad\}$ is also used.

(b) $\mathbb{N} = \{1, 2, 3, \ldots\}$, the set of natural numbers.

(c) $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$, the set of integers.

(d) $\mathbb{Q} = \{p/q : p \in \mathbb{Z}, q \in \mathbb{Z}, q \neq 0\}$, the set of rational numbers.

(e) $\mathbb{R}$ or $(-\infty, \infty)$, the set of real numbers.

(f) $\mathbb{R}^+ = \{x : x \in \mathbb{R}$ and $x > 0\}$, the set of *positive* real numbers.

(g) $[a, b]$, the set of all real numbers x such that $a \leq x \leq b$.

(h) (a, b), the set of all real numbers x such that $a < x < b$.

(i) $(a, b]$, the set of all real numbers x such that $a < x \leq b$.

(j) $[a, b)$, the set of all real numbers x such that $a \leq x < b$.

(k) $(-\infty, a)$, the set of all real numbers x such that $x < a$.

(l) $[a, \infty)$, the set of all real numbers x such that $x \geq a$.

One of important notions in set theory is the equality of sets, which is defined in terms of a two-way containment.

Definition 1.1.6. A set A is said to be a subset of another set B if and only if every element of A is also an element of B. We use the notation $A \subseteq B$. Two sets A and B are said to be equal if they have the same elements. In other words, A and B are equal if $A \subseteq B$ and $B \subseteq A$.

Remark 1.1.7. $\emptyset$ is a subset of A for any set A. A is a subset of A itself, for any set A.

Example 1.1.8. By Definition 1.1.5, $\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R}$.

Example 1.1.9. $\{1,2,3,4\} \nsubseteq \{1,3,4,5,6,7\}$, because $2 \notin \{1,3,4,5,6,7\}$.

Problem 1.1.10. How many distinct subsets does the set $A = \{x, y\}$ have? How about $B = \{p, q, r\}$?

Answer Subsets of A are: $\emptyset, \{x\}, \{y\}$, and A itself, so A has 4 distinct subsets.
Subsets of B are: $\emptyset, \{p\}, \{q\}, \{r\}, \{p,q\}, \{p,r\}, \{q,r\}$, and B itself, so B has 8 distinct subsets. ■

In many cases, Definition 1.1.6 is the only tool available to identify inclusion relation of two sets.

Problem 1.1.11. Let

$$A = \{n : n = 2k + 1 \text{ for some } k \in \mathbb{Z}\}$$

and

$$B = \{m : m \text{ is the difference of two consecutive squares of integers}\}$$

Show that $A = B$.

Answer First we show that $A \subseteq B$. Let $n \in A$, then there is $k \in Z$ such that $n = 2k + 1$. Since $2k + 1 = (k + 1)^2 - k^2$, we conclude that $n \in B$. To show that $B \subseteq A$, let $m \in B$, then m can be written as $m = (\ell + 1)^2 - \ell^2$ for some $\ell \in \mathbb{Z}$. Since $(\ell + 1)^2 - \ell^2 = 2\ell + 1$, we see that $m \in A$. ■

We can measure the size of a given set by counting the number of elements in it.

Definition 1.1.12. By cardinality of a set S, we mean the number of elements in S. The cardinality of S is denoted by $|S|$.

Example 1.1.13. If $A = \{n : n \text{ is a prime number less than or equal to } 10\} = \{2, 3, 5, 7\}$, then $|A| = 4$. The cardinality of the set of natural numbers, $\mathbb{N}$, is infinite. We write $|\mathbb{N}| = \infty$.

Recall that in Problem 1.1.10, A has $2^2 = 4$ distinct subsets and B has $2^3 = 8$ distinct subsets. In fact, we have the following theorem (see Exercise 1.4.2 for a proof).

Theorem 1.1.14. *If S is a set with $|S| = n$, then S has 2^n distinct subsets.*

Now we consider the concept of the product of two sets. We start with ordered n-tuples.

Definition 1.1.15. The ordered n-tuple $(a_1, a_2, \ldots, a_n)$ is the ordered collection that has a_1 as its first element, a_2 as its second element,$\ldots$, and a_n as its n^{th} element. In particular, an ordered 2-tuple is called an ordered pair, and an ordered 3-tuple is called an ordered triple.

Remark 1.1.16. When $a \neq b$, we *distinguish* (a, b) from (b, a). In general, $(a_1, a_2, \ldots, a_n) = (b_1, b_2, \ldots, b_n)$ if and only if $a_1 = b_1$, $a_2 = b_2$, $\ldots$, and $a_n = b_n$.

Definition 1.1.17. Let A and B be sets. The Cartesian product of A and B, denoted by $A \times B$, is the set of all ordered pairs (a, b), where $a \in A$ and $b \in B$. In other words, $A \times B = \{(a, b) : a \in A \text{ and } b \in B\}$.

Example 1.1.18. Let $A = \{a, b\}$ and $B = \{1, 2, 3\}$. Then

$$A \times B = \{(a, 1), (a, 2), (a, 3), (b, 1), (b, 2), (b, 3)\}$$

and

$$B \times A = \{(1, a), (1, b), (2, a), (2, b), (3, a), (3, b)\}.$$

Remark 1.1.19. The Cartesian product can be defined for more than two sets. That is, if $A_1, A_2, \ldots, A_n$ are sets, then the Cartesian product of $A_1, A_2, \ldots, A_n$, denoted by $A_1 \times A_2 \times \cdots \times A_n$, is defined to be

$$A_1 \times A_2 \times \cdots \times A_n = \{(a_1, a_2, \ldots, a_n) : a_1 \in A_1, a_2 \in A_2, \ldots, a_n \in A_n\}.$$

Example 1.1.20. The Euclidean n-space $\mathbb{R}^n$ is defined to be

$$\mathbb{R}^n = \underbrace{\mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R}}_{n \text{ times}}.$$

In other words,

$$\mathbb{R}^n = \{(x_1, x_2, \ldots, x_n) : x_i \in \mathbb{R} \text{ for all } i\}.$$

Remark 1.1.21. Note that $6 = |A \times B| = |A||B| = 2 \times 3$ in Example 1.1.18. In general, if all A_i's have finitely many elements, then

$$|A_1 \times A_2 \times \cdots \times A_n| = |A_1||A_2| \cdots |A_n|.$$

See Example 1.3.4.

Exercise 1.1.

1. Describe the following sets by listing their elements:

(a) $A = \{n : n^2 - n - 12 \leq 0 \text{ and } n \in \mathbb{Z}\}$.

(b) $B = \{x : \log_2(x + 1) + \log_2(x - 1) = 3\}$.

2. Let

$$A = \{n : n = k^2 \text{ for some } k \in \mathbb{Z}\}$$

and

$$B = \{m : m = 3\ell \text{ or } m = 3\ell + 1 \text{ for some } \ell \in \mathbb{Z}\}.$$

Show that $A \subseteq B$ and $B \nsubseteq A$.
Hint: To show that $A \subseteq B$, note that if $k \in \mathbb{Z}$, then k must be of the form $k = 3p$, $k = 3p + 1$, or $k = 3p - 1$, where $p \in \mathbb{Z}$. To show that $B \nsubseteq A$, it is enough to find an element $b \in B$ such that $b \notin A$.

3. Choose all that are true:

(a) $0 \in \{0\}$

(b) $0 \in \emptyset$

(c) $0 \subseteq \{0\}$

(d) $\emptyset \subseteq \emptyset$

(e) $\emptyset \in \{\emptyset\}$

(f) $\emptyset \subseteq \{\emptyset\}$

(g) $\{1, 2, 3\} \subseteq \{1, 2, 3, 4, 5\}$

(h) $\{1\} \in \{1, 2, 3, 4, 5\}$

(i) $1 \subseteq \{1, 2, 3, 4, 5\}$

(j) $\{1\} \in \{\{1\}, 2, 3, \{2, 3\}\}$

(k) $\{2, 3\} \in \{\{1\}, 2, 3, \{2, 3\}\}$

(l) $\{2, 3\} \subseteq \{\{1\}, 2, 3, \{2, 3\}\}$

4. Find the Cartesian product of $A \times B \times C$, where $A = \{0, 1\}$, $B = \{1, 2\}$, and $C = \{0, 1, 2\}$.

1.2 Set Operations

In this section, we consider various set operations as a way of constructing new sets from old ones, and review some properties of these operations.

Definition 1.2.1. Let A and B be sets.

(a) The <u>union</u> of A and B, denoted by $A \cup B$, is the set that contains those elements that are either in A or B, or in both. In other words, $A \cup B = \{x : x \in A \text{ or } x \in B\}$.

(b) The <u>intersection</u> of A and B, denoted by $A \cap B$, is the set that contains those elements in both A and B. In other words, $A \cap B = \{x : x \in A \text{ and } x \in B\}$.

(c) Two sets are said to be <u>disjoint</u> if their intersection is the empty set.

Example 1.2.2. Let $A = \{1, 3, 4, 5\}$ and $B = \{1, 2, 3\}$, then $A \cup B = \{1, 2, 3, 4, 5\}$ and $A \cap B = \{1, 3\}$.

Example 1.2.3. If $S = (-\infty, 2)$ and $T = [-1, \infty)$, then $S \cap T = [-1, 2)$ and $S \cup T = (-\infty, \infty) = \mathbb{R}$.

Example 1.2.4. Let $O = \{2k + 1 : k \in \mathbb{Z}\}$ and $E = \{2k : k \in \mathbb{Z}\}$, then O and E are disjoint.

Obviously, $A \cup B$ contains both A and B, so $|A \cup B| \geq |A|$ and $|A \cup B| \geq |B|$. But exactly how many elements are there in $A \cup B$?

Theorem 1.2.5. *Let A, B be sets with finitely many elements, then*

$$|A \cup B| = |A| + |B| - |A \cap B|.$$

In particular, if A and B are disjoint, then $|A \cup B| = |A| + |B|$.

Note that in Theorem 1.2.5, $|A \cap B|$ is subtracted because elements in $A \cap B$ are double counted in $|A| + |B|$.

Problem 1.2.6. Let A and B be subsets of C. Suppose that $|C| = 50$, $|A| = 30$, and $|B| = 40$. Prove that $20 \leq |A \cap B| \leq 30$.

Answer Since $A \cup B \subseteq C$, $|A \cup B| \leq |C| = 50$. Since $|A \cap B| = |A| + |B| - |A \cup B| = 70 - |A \cup B|$, it follows that $|A \cap B| \geq 70 - 50 = 20$. On the other hand, since $A \cap B \subseteq A$, it follows that $|A \cap B| \leq |A| = 30$. ∎

Definition 1.2.7. Let A and B be sets.

(a) The <u>difference</u> of A and B, denoted by $A - B$ or $A \setminus B$, is the set containing those elements that are in A but not in B. In other words, $A - B = \{x : x \in A \text{ and } x \notin B\}$.

(b) The <u>universal set</u>, denoted by U, is the set of all objects under consideration in a given context.

(c) The <u>complement</u> of A, denoted by A^c, is $U - A$. In other words, $A^c = \{x : x \notin A\}$.

Example 1.2.8. Let $U = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$, $A = \{1, 2, 3, 4, 5\}$, and $B = \{3, 5, 7, 9\}$. Then $A - B = \{1, 2, 4\}$, $A^c = \{6, 7, 8, 9, 10\}$, and $B^c = \{1, 2, 4, 6, 8, 10\}$.

Example 1.2.9. Let O and E be as in Example 1.2.4. If $U = \mathbb{Z}$, then we see that $O^c = E$ and $E^c = O$.

Problem 1.2.10. Prove that $A - B = A \cap B^c$.

Answer $x \in A - B$ if and only if $x \in A$ and $x \notin B$. In other words, $x \in A - B$ if and only if $x \in A$ and $x \in B^c$. This shows that $A - B = A \cap B^c$. ∎

Remark 1.2.11. Some properties of set operations are summarized in the following table:

Identity	Name
$A \cap U = A$ $A \cup \emptyset = A$	Identity laws
$A \cup U = U$ $A \cap \emptyset = \emptyset$	Domination laws
$A \cup A = A$ $A \cap A = A$	Idempotent laws
$(A^c)^c = A$	Complementation law
$A \cup B = B \cup A$ $A \cap B = B \cap A$	Commutative laws
$A \cup (B \cup C) = (A \cup B) \cup C$ $A \cap (B \cap C) = (A \cap B) \cap C$	Associative laws
$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$	Distributive laws
$(A \cup B)^c = A^c \cap B^c$ $(A \cap B)^c = A^c \cup B^c$	De Morgan's laws
$A \cup (A \cap B) = A$ $A \cap (A \cup B) = A$	Absorption laws
$A \cup A^c = U$ $A \cap A^c = \emptyset$	Complement laws

Problem 1.2.12. Show that $(A - B) \cup B = A \cup B$.

Answer $(A - B) \cup B = (A \cap B^c) \cup B = (A \cup B) \cap (B^c \cup B) = (A \cup B) \cap U = A \cup B$. ∎

One can define union and intersection for more than two sets.

Definition 1.2.13. Let $A_1, A_2, \ldots, A_n$ be sets. Then

(a) the union $\bigcup_{i=1}^{n} A_i = A_1 \cup \cdots \cup A_n$ is the set containing those elements that are members of at least one of A_i, $1 \le i \le n$.

(b) the intersection $\bigcap_{i=1}^{n} A_i = A_1 \cap \cdots \cap A_n$ is the set containing those elements that are members of A_i for all i, $1 \le i \le n$.

(c) the union $\bigcup_{i=1}^{\infty} A_i = A_1 \cup A_2 \cup \cdots$ is the set containing those elements that are members of at least one of A_i, $i \ge 1$.

(d) the intersection $\bigcap_{i=1}^{\infty} A_i = A_1 \cap A_2 \cap \cdots$ is the set containing those elements that are members of A_i for all i, $i \ge 1$.

Problem 1.2.14. For $i = 1, 2, \ldots$, define $A_i = \{x : 0 \le x < \frac{1}{i}\}$. What is $\bigcap_{i=1}^{\infty} A_i$?

Answer We claim that $\bigcap_{i=1}^{\infty} A_i = \{0\}$. Clearly $0 \in A_i$ for every i, so it suffices to show that 0 is the only element in $\bigcap_{i=1}^{\infty} A_i$. Suppose $x \in \bigcap_{i=1}^{\infty} A_i$, then clearly $x \ge 0$. If $x > 0$, then there is sufficiently large $i_0 \in \mathbb{N}$ such that $x > \frac{1}{i_0}$ and hence $x \notin A_{i_0}$. This shows that if $x > 0$, then $x \notin \bigcap_{i=1}^{\infty} A_i$. ∎

We close this section by introducing the concept of the indicator function, which is useful in expressing piecewise defined functions.

Definition 1.2.15. Let A be a subset of $\mathbb{R}$. The <u>indicator function</u> of A, denoted by I_A, is a function defined by

$$I_A(x) = \begin{cases} 1, & x \in A, \\ 0, & x \notin A \end{cases}.$$

Example 1.2.16. Consider a piecewise defined function f given by

$$f(x) = \begin{cases} x^2, & x \geq 1, \\ 0, & x < 1. \end{cases}$$

Then f can be compactly expressed as $f(x) = x^2 I_A(x)$, where $A = \{x : x \geq 1\}$.

Example 1.2.17. Let X be a discrete random variable such that

$$P(X = n) = \frac{6}{\pi^2 n^2}, \quad n = 1, 2, 3, \ldots.$$

Then the probability mass function f of X can be expressed by $f(x) = \frac{6}{\pi^2 x^2} I_{\mathbb{N}}(x)$.

Exercise 1.2.

1. Let A, B, C be sets with finitely many elements. Prove that

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|.$$

Hint: Treat $B \cup C$ as a single set, then

$$|A \cup B \cup C| = |A| + |B \cup C| - |A \cap (B \cup C)| = |A| + |B \cup C| - |(A \cap B) \cup (A \cap C)|.$$

Use Theorem 1.2.5 to compute $|(A \cap B) \cup (A \cap C)|$.

2. Suppose $S \subseteq U$ and $U = \bigcup_{i=1}^{n} A_i$. Prove that $S = \bigcup_{i=1}^{n} (S \cap A_i)$.

3. Let A and B be sets.

(a) Prove that $(A - B) \cup (B - A) = (A \cup B) - (A \cap B)$.

(b) Prove that $|(A - B) \cup (B - A)| = |A| + |B| - 2|A \cap B|$.

4. For $i = 1, 2, \ldots$, define $A_i = \{x : -\frac{1}{i} < x < \frac{1}{i}\}$. Compute $\bigcap_{i=1}^{\infty} A_i$ and $\bigcup_{i=1}^{\infty} A_i$.

5. Let P be the set of all nonnegative numbers, that is, $P = \{x : x \geq 0\}$. Prove that $x(2I_P(x) - 1) = |x|$.

6. Let A and B be subsets of $\mathbb{R}$. Describe $I_{A \cap B}$ and $I_{A \cup B}$ in terms of I_A and I_B.

1.3 Basics of Counting

Many counting problems that appear in discrete probability can be solved by finding the number of ways to arrange a specified number of distinct elements of a set of a given size. We begin with what is called the *Product Rule*.

Theorem 1.3.1 (The Product Rule)**.** *Suppose that a procedure can be broken into a sequence of two tasks. If there are n_1 ways to do the first task and for each of these ways of doing the first task, there are n_2 ways to do the second task, then there are $n_1 n_2$ ways to do the procedure.*

Example 1.3.2. If chairs of an auditorium are to be labeled with a letter followed by a positive integer not exceeding 100, then $26 \times 100 = 2600$ chairs can be labeled differently.

The product rule extends to the case a procedure can be broken down into a sequence of more than just two tasks. Suppose that the procedure is carried out by performing the tasks $T_1, T_2, \ldots, T_k$ in sequence. If each task T_i can be done in n_i ways, regardless of how the previous tasks were done, then there are $n_1 n_2 \cdots n_k$ ways to carry out the procedure.

Problem 1.3.3. In how many ways can we select three students from a group of five students to stand in line for a picture?

Answer The first place can be filled with any of five students and the second place can take any of four remaining students. Similarly, the third place can be filled with any of remaining three students, so there are $5 \times 4 \times 3 = 60$ ways. ■

Example 1.3.4. Let A_i, $1 \leq i \leq n$, be sets with finitely many elements. All elements of $A_1 \times A_2 \times \cdots \times A_n$ are of the form $(a_1, a_2, \ldots, a_n)$, where $a_i \in A_i$. Note that a_1 can be any of $|A_1|$ elements of A_1 and for each choice of a_1, there are $|A_2|$ ways of selecting a_2. Proceeding in this fashion, we can verify that

$$|A_1 \times A_2 \times \cdots \times A_n| = |A_1||A_2| \cdots |A_n|.$$

Definition 1.3.5. A <u>permutation</u> of a set of distinct objects is an ordered arrangement of these objects. An ordered arrangement of r elements of a set is called an <u>r-permutation</u>. The number of r-permutations of a set with n elements is denoted by $P(n, r)$ or ${}_nP_r$.

Example 1.3.6. Let $S = \{1, 2, 3, 4, 5\}$. The ordered arrangement $(4, 2, 1, 5, 3)$ is a permutation of S. $(3, 1, 4)$ is a 3-permutation of S.

Example 1.3.7. Problem 1.3.3 shows that ${}_5P_3 = 60$.

Problem 1.3.8. How many ways are there to choose a chair and a vice chair from a group of four students?

Answer Let $\{A, B, C, D\}$ denote the group of students. We note that an ordered pair can be identified with a possible selection of a chair and a vice chair. For example, (A, D) corresponds to the selection of A as chair and D as vice chair. It follows that there are ${}_4P_2 = 12$ ways to choose a chair and a vice chair. Note that (A, B) should be distinguished from (B, A). ■

Recall that for $n \in \mathbb{N}$, the factorial $n!$ is defined to be

$$n! = 1 \cdot 2 \cdots (n-1) \cdot n$$

with the convention that $0! = 1$. In general, we get the following formula.

Theorem 1.3.9. *For $n \geq r \geq 0$,*

$$\begin{aligned} {}_nP_r &= \underbrace{n \cdot (n-1) \cdot (n-2) \cdots (n-r+1)}_{r \text{ factors}} \\ &= \frac{n \cdot (n-1) \cdot (n-2) \cdots (n-r+1)(n-r)(n-r-1) \cdots 2 \cdot 1}{(n-r)(n-r-1) \cdots 2 \cdot 1} \\ &= \frac{n!}{(n-r)!}. \end{aligned}$$

In particular,

$${}_nP_n = n \cdot (n-1) \cdot (n-2) \cdots 2 \cdot 1 = n!.$$

Problem 1.3.10. How many permutations of the letters $\{R, A, I, N, B, O, W\}$ start with R and end with W?

Answer The first and last spots are fixed (R and W, respectively), so the number of desired permutations is determined by remaining five spots in the middle, which can be filled in ${}_5P_5 = 5! = 120$ ways. ■

Example 1.3.11. Let n be a positive integer. The symmetric group on n letters, denoted by S_n, is the set of all bijections from $\{1, 2, \ldots, n\}$ onto itself. Note that each $\sigma \in S_n$ is in one-to-one correspondence with the n-tuple $(\sigma(1), \sigma(2), \ldots, \sigma(n))$. Since $\sigma(1)$ can be any of $\{1, 2, \ldots, n\}$, $\sigma(2)$ can be any of $\{1, 2, \ldots, n\}$ except for $\sigma(1)$, and $\sigma(3)$ can be any of $\{1, 2, \ldots, n\}$ except for $\sigma(1)$ and $\sigma(2)$, etc, it follows that $|S_n| = n!$.

Remark 1.3.12. For $\sigma \in S_n$, an ordered pair (i, j) is called an inversion of σ if $i < j$ and $\sigma(i) > \sigma(j)$. See Definition 3.4.1 for a use of this concept in the definition of the determinant of a square matrix.

We now turn our attention to counting *unordered* selection of objects.

Definition 1.3.13. An r-combination of elements of a set is an unordered selection of r elements from the set. The number of r-combinations of a set with n elements is denoted by $C(n, r)$, ${}_nC_r$, or $\binom{n}{r}$.

Example 1.3.14. Let $S = \{1, 2, 3, 4, 5\}$. Then $\{1, 3, 4\}$ is a 3-combination of S.

Problem 1.3.15. Find all 2-combinations of $\{1, 2, 3, 4, 5\}$.

Answer They are $\{1, 2\}, \{1, 3\}, \{1, 4\}, \{1, 5\}, \{2, 3\}, \{2, 4\}, \{2, 5\}, \{3, 4\}, \{3, 5\}, \{4, 5\}$. This shows that $\binom{5}{2} = 10$. ■

In general, we get the following formula.

Theorem 1.3.16. *The number of r-combinations of a set with n elements, where n is a nonnegative integer and r is an integer with $0 \le r \le n$, equals*

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}.$$

Remark 1.3.17. We explain Theorem 1.3.16 with $\binom{5}{3}$ as an example. Let $S = \{1, 2, 3, 4, 5\}$ and consider all 3-permutations of S consisting of $\{1, 2, 4\}$. They are:

$$(1, 2, 4), (1, 4, 2), (2, 1, 4), (2, 4, 1), (4, 1, 2), \text{ and } (4, 2, 1).$$

On the other hand, there is only one 3-combination of S consisting of $\{1, 2, 4\}$; just $\{1, 2, 4\}$. Hence there are 6 times as many permutations as combinations consisting of $\{1, 2, 4\}$. Here 6 comes from the total number of different permutations of $\{1, 2, 4\}$, that is, $6 = 3!$. Since this is true for all 3-combinations of S, we conclude that

$${}_5P_3 = 3! \times \binom{5}{3}.$$

In other words,

$$\binom{5}{3} = \frac{1}{3!} \times_5 P_3 = \frac{1}{3!} \times \frac{5!}{(5-2)!} = \frac{5!}{3!(5-2)!}.$$

Problem 1.3.18. Compute $\binom{n}{0}$, $\binom{n}{1}$, and $\binom{n}{n}$.

Answer $\binom{n}{0} = \frac{n!}{0!n!} = 1$, $\binom{n}{1} = \frac{n!}{1!(n-1)!} = n$, and $\binom{n}{n} = \frac{n!}{n!(n-n)!} = 1$. ■

Problem 1.3.19 (Pascal's Identity)**.** Let n and k be positive integers with $n \ge k$. Prove that

$$\binom{n+1}{k} = \binom{n}{k-1} + \binom{n}{k}.$$

Answer

$$\begin{aligned}
\binom{n}{k-1} + \binom{n}{k} &= \frac{n!}{(k-1)!(n-k+1)!} + \frac{n!}{k!(n-k)!} \\
&= \frac{n!k}{k!(n-k+1)!} + \frac{n!(n-k+1)}{k!(n-k+1)!} \\
&= \frac{(n+1)!}{k!(n-k+1)!} \\
&= \binom{n+1}{k}.
\end{aligned}$$

■

Exercise 1.3.

1. Consider permutations of the letters $ABCDEFG$.

(a) How many permutations contain the string ABC?

(b) How many permutations contain exactly two letters between A and B?

(c) In how many permutations does A come before B?

2. Prove that $\binom{n}{r} = \binom{n}{n-r}$.

3. What is the largest possible number of inversions of σ in S_n? When is it attained? *Hint*: It is clear that the number of inversions of $\sigma \in S_n$ cannot be larger than $\binom{n}{2} = \frac{n(n-1)}{2}$.

4. Suppose that there are seven faculty members in the mathematics department and six in the statistics department in a college. How many ways are there to select a committee to develop a mathematical statistics course at the school if the committee is to consist of three faculty members from the mathematics department and two from the statistics department?

5. Let m, n be integers greater than 1. Prove that

$$\binom{m+n}{n} = \binom{m+n-2}{n-2} + 2\binom{m+n-2}{n-1} + \binom{m+n-2}{n}.$$

1.4 Binomial Theorem and Its Consequences

When $(x+y)^5 = (x+y)(x+y)(x+y)(x+y)(x+y)$ is expanded, we obtain $2^5 = 32$ terms. Each term is determined by the choice of either x or y in each factor. For example, the choice of blue letters as in

$$(x+y)(x+y)(x+y)(x+y)(x+y)$$

gives a term $xxyxy = x^3y^2$. Note that each term is of the form x^5, x^4y, x^3y^2, x^2y^3, xy^4, or y^5. To obtain a term of the form x^5, x must be chosen from each of five factors and this can be done in only one way which implies that the coefficient of x^5 when $(x+y)^5$ is multiplied out is 1. To obtain a term of the form x^3y^2, x must be chosen from three of the five factors (and consequently y must be chosen from the remaining two factors). Since this can be done in $\binom{5}{3}$ ways, the coefficient of x^3y^2 when $(x+y)^5$ is expanded will be $\binom{5}{3}$. In general, we have the following theorem.

Theorem 1.4.1 (The Binomial Theorem)**.** *Let x and y be variables, and let n be a non-negative integer. Then*

$$\begin{aligned}
(x+y)^n &= \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k \\
&= \binom{n}{0} x^n + \binom{n}{1} x^{n-1} y + \binom{n}{2} x^{n-2} y^2 + \cdots + \binom{n}{n-1} x y^{n-1} + \binom{n}{n} y^n. \\
&= \binom{n}{n} x^n + \binom{n}{n-1} x^{n-1} y + \binom{n}{n-2} x^{n-2} y^2 + \cdots + \binom{n}{1} x y^{n-1} + \binom{n}{0} y^n. \\
&= \sum_{k=0}^{n} \binom{n}{n-k} x^{n-k} y^k.
\end{aligned}$$

Remark 1.4.2.

(a) The third equality in the Binomial Theorem above comes from Exercise 1.3.2.

(b) Since $x + y = y + x$, it follows that

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

Example 1.4.3. When $n = 5$, we get

$$\begin{aligned}
(x+y)^5 &= \binom{5}{0} x^5 + \binom{5}{1} x^4 y + \binom{5}{2} x^3 y^2 + \binom{5}{3} x^2 y^3 + \binom{5}{4} x y^4 + \binom{5}{5} y^5 \\
&= x^5 + 5x^4 y + 10x^3 y^2 + 10x^2 y^3 + 5xy^4 + y^5.
\end{aligned}$$

Problem 1.4.4. Find the coefficient of $x^{12}y^{13}$ in the expansion of $(x + y)^{25}$.

Answer The coefficient of $x^{12}y^{13}$ is given by $\binom{25}{12} = \frac{25!}{12!13!}$. ■

Problem 1.4.5. Determine whether we have a term of the form x^{14} in the expansion of $\left(x^2 + \frac{1}{x}\right)^{10}$. If we do, what is the coefficient?

Answer A general term of the expansion must be of the form $\binom{10}{k}(x^2)^{10-k}\left(\frac{1}{x}\right)^k = \binom{10}{k}x^{20-3k}$. To get the coefficient of x^{14}, we see that k must equal 2 and the corresponding coefficient is $\binom{10}{2} = 45$. ■

Problem 1.4.6. Show that $\binom{n}{0} + 2\binom{n}{1} + 2^2\binom{n}{2} + \cdots + 2^n\binom{n}{n} = 3^n$.

Answer $3^n = (2+1)^n = \binom{n}{0}2^0 \cdot 1^n + \binom{n}{1}2^1 \cdot 1^{n-1} + \binom{n}{2}2^2 \cdot 1^{n-2} + \cdots + \binom{n}{n}2^n \cdot 1^0$. ■

Example 1.4.7 (Binomial Distribution)**.** Let $X \sim Binom(n,p)$, that is, let X be a discrete random variable such that $P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$ for $k = 0, 1, 2, \ldots, n$. It follows from the Binomial Theorem that

$$\sum_{k=0}^{n} P(X=k) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p + (1-p))^n = 1.$$

Problem 1.4.8. Let $X \sim Binom(n,p)$. Compute the expectation $E(X)$ of X.

Answer

$$\begin{aligned}
E(X) &= \sum_{k=0}^{n} kP(X=k) \\
&= \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} \\
\text{(first term is zero)} &= \sum_{k=1}^{n} k \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} \frac{n!}{(n-k)!(k-1)!} p^k (1-p)^{n-k} \\
&= np \sum_{k=1}^{n} \frac{(n-1)!}{(n-k)!(k-1)!} p^{k-1} (1-p)^{n-k} \\
\text{(use } j = k-1\text{)} &= np \sum_{j=0}^{n-1} \frac{(n-1)!}{(n-1-j)!j!} p^j (1-p)^{n-1-j} \\
\text{(Binomial Theorem applied reversely)} &= np(p + (1-p))^{n-1} \\
&= np.
\end{aligned}$$

■

Problem 1.4.9. How big should a sample size be to detect an allele with a 0.5% frequency with 99% probability?

Answer Let N denote the sample size, then the observed number X of the allele with a 0.5% frequency will follow $Binom(2N, 0.005)$. Solving the inequality

$$0.01 > P(X=0) = \binom{2N}{0} (0.005)^0 (1-0.005)^{2N-0} = 0.995^{2N}$$

gives $N \geq 460$. ■

Problem 1.4.10 (Vandermonde's Identity)**.** For $m, n, r \in \mathbb{N}$, show that

$$\binom{m+n}{r} = \sum_{k=0}^{r} \binom{m}{k}\binom{n}{r-k}.$$

Answer We compute the coefficient of x^r in the expansion of $(1+x)^{m+n}$ in two ways. First, using the Binomial Theorem, we get

$$(1+x)^{m+n} = \sum_{r=0}^{m+n} \binom{m+n}{r} x^r.$$

On the other hand, since

$$(1+x)^m = \sum_{i=0}^{m} \binom{m}{i} x^i \quad \text{and} \quad (1+x)^n = \sum_{j=0}^{n} \binom{n}{j} x^j,$$

the x^r terms of $(1+x)^{m+n} = (1+x)^m(1+x)^n$ will be collected from the sum

$$\binom{m}{0}x^0\binom{n}{r}x^r + \binom{m}{1}x^1\binom{n}{r-1}x^{r-1} + \cdots + \binom{m}{r}x^r\binom{n}{0}x^0.$$

This leads to the desired identity. ■

Example 1.4.11. Consider an experiment of drawing n balls, without replacement, out of a box that contains N white or red balls, exactly $K(K \geq n)$ of which being red. Let X denote the number of red balls in this drawing, then

$$P(X=k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}, \quad k = 0, 1, 2, \ldots, n.$$

The random variable X is said to follow Hypergeometric Distribution and is denoted as $X \sim hypergeo(N, K, n)$. It follows from Vandermonde's Identity (see Problem 1.4.10) that

$$\sum_{k=0}^{n} P(X=k) = \frac{1}{\binom{N}{n}} \sum_{k=0}^{n} \binom{K}{k}\binom{N-K}{n-k} = \frac{1}{\binom{N}{n}} \binom{K+(N-K)}{n} = 1.$$

Exercise 1.4.

1. Find the coefficient of x^3y^7 in the expansion of $(2x-3y)^{10}$.

2. Let n be a positive integer.

(a) Show that $\binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{n} = 2^n$.

(b) Prove Theorem 1.1.14.
Hint: If $|S| = n$, the number of subsets of S having exactly k elements equals $\binom{n}{k}$.

(c) Consider a multiple linear regression with p covariates:

$$Y = \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \epsilon.$$

Show that there are $2^p - 1$ nested models within the given full model.

3. Let $X \sim Binom(n, p)$. Show that $E(X^2) = np(np + 1 - p)$ and $Var(X) = np(1 - p)$, where $Var(X)$ denotes the variance of X.
Hint: Note that

$$E(X^2) = \sum_{k=0}^{n} k^2 P(X = k) = \sum_{k=0}^{n} k(k-1)\binom{n}{k} p^k (1-p)^{n-k} + E(X).$$

4. Simplify $n\binom{n}{0} + (n-1)\binom{n}{1} + (n-2)\binom{n}{2} + \cdots + 2\binom{n}{n-2} + \binom{n}{n-1}$.
Hint: First, expand $(x+1)^n$ using the Binomial Theorem, then differentiate with respect to x.

5. Simplify each of the following summations:

(a) $\displaystyle\sum_{n=0}^{10} \binom{10}{n} 2^n$.

(b) $\displaystyle\sum_{n=0}^{10} n\binom{10}{n} 2^n$.

6. Let $X \sim hypergeo(N, K, n)$. Show that $E(X) = \frac{nK}{N}$.
Hint: You may want to use Vandermonde's Identity (see Problem 1.4.10)

$$\binom{N-1}{n-1} = \sum_{s=0}^{n-1} \binom{K-1}{s}\binom{N-K}{n-1-s}.$$

1.5 Other Topics in Discrete Mathematics

In this section, we cover miscellaneous topics in discrete mathematics. We begin with the formula for some special summations.

Theorem 1.5.1.

(a) $\displaystyle\sum_{k=1}^{n} ar^{k-1} = a + ar + ar^2 + \cdots + ar^{n-1} = \frac{ar^n - a}{r-1}$, *provided* $r \neq 1$. *When* $r = 1$, $\displaystyle\sum_{k=1}^{n} ar^{k-1} = a + a + \cdots + a = na$.

(b) $\displaystyle\sum_{k=1}^{n} k = 1 + 2 + 3 + \cdots + n = \frac{n(n+1)}{2}$.

(c) $\displaystyle\sum_{k=1}^{n} k^2 = 1^2 + 2^2 + 3^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6}$.

(d) $\displaystyle\sum_{k=1}^{n} k^3 = 1^3 + 2^3 + 3^3 + \cdots + n^3 = \frac{n^2(n+1)^2}{4}$.

Remark 1.5.2. Replacing n by $n-1$ in Theorem 1.5.1, we get

(a) $\displaystyle\sum_{k=1}^{n-1} k = \frac{n(n-1)}{2}$.

(b) $\displaystyle\sum_{k=1}^{n-1} k^2 = \frac{n(n-1)(2n-1)}{6}$.

(c) $\displaystyle\sum_{k=1}^{n-1} k^3 = \frac{n^2(n-1)^2}{4}$.

Example 1.5.3. $1 + 2 + 3 + \cdots + 99 + 100 = \frac{100(100+1)}{2} = 5050$.

Problem 1.5.4. The Spearman correlation coefficient ρ is defined as the Pearson correlation coefficient between the ranked variables. Let x_i, y_i $(1 \leq i \leq n)$ be ranked variables. Assuming that there are no ties, show that

$$\rho = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)},$$

where $d_i = x_i - y_i$.

Answer Note that by definition

$$\rho = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}, \tag{1.5.1}$$

where $\bar{x} = \bar{y} = \frac{\sum_{i=1}^n i}{n} = \frac{n+1}{2}$. Since $\sum_{i=1}^n x_i = \sum_{i=1}^n y_i = \sum_{i=1}^n i = \frac{n(n+1)}{2}$, the numerator of (1.5.1) becomes

$$\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n x_i y_i - \bar{y}\sum_{i=1}^n x_i - \bar{x}\sum_{i=1}^n y_i + n\bar{x}\bar{y} = \sum_{i=1}^n x_i y_i - \frac{n(n+1)^2}{4}.$$

In the denominator of (1.5.1),

$$\begin{aligned}\sum_{i=1}^n (x_i - \bar{x})^2 &= \sum_{i=1}^n (x_i^2 - 2\bar{x}x_i + \bar{x}^2) \\ &= \sum_{i=1}^n x_i^2 - n\bar{x}^2 \\ &= \frac{n(n+1)(2n+1)}{6} - n\left(\frac{n+1}{2}\right)^2 \\ &= \frac{n(n^2-1)}{12}.\end{aligned}$$

Similarly, $\sum_{i=1}^n (y_i - \bar{y})^2 = \frac{n(n^2-1)}{12}$ and hence the denominator of (1.5.1) reduces to $\frac{n(n^2-1)}{12}$ and it follows that

$$\rho = \frac{\sum_{i=1}^n x_i y_i - \frac{n(n+1)^2}{4}}{\frac{n(n^2-1)}{12}} = \frac{12\sum_{i=1}^n x_i y_i - 3n(n+1)^2}{n(n^2-1)}. \tag{1.5.2}$$

On the other hand,

$$\sum_{i=1}^n d_i^2 = \sum_{i=1}^n x_i^2 + \sum_{i=1}^n y_i^2 - 2\sum_{i=1}^n x_i y_i = \frac{n(n+1)(2n+1)}{3} - 2\sum_{i=1}^n x_i y_i,$$

so it follows that

$$\begin{aligned}1 - \frac{6\sum_{i=1}^n d_i^2}{n(n^2-1)} &= \frac{n(n^2-1) - 2n(n+1)(2n+1) + 12\sum_{i=1}^n x_i y_i}{n(n^2-1)} \\ &= \frac{12\sum_{i=1}^n x_i y_i - 3n(n+1)^2}{n(n^2-1)}.\end{aligned}$$

From this and (1.5.2) the result follows. ■

To motivate the Principle of Mathematical Induction, we consider an infinite row of dominoes labeled by $1, 2, 3, \ldots$. One way to knock over all dominoes is to knock off the first domino after making sure that neighboring dominoes are close enough so that the fall of the k^{th} domino guarantees the fall of the $k + 1^{\text{st}}$ domino, for all k. This is summarized in the following theorem.

Theorem 1.5.5 (Principle of Mathematical Induction)**.** *Let $P(n)$ be a statement involving a positive integer n. To prove that $P(n)$ is true for all positive integers n, it suffices to show the following two steps:*
- *Basis Step: verify that $P(1)$ is true.*
- *Inductive Step: show that the conditional statement $P(k) \to P(k+1)$ is true for all positive integers k.*

Problem 1.5.6. Use Mathematical Induction to prove that $1+2+3+\cdots+n = \frac{n(n+1)}{2}$ for all $n \in \mathbb{N}$.

Answer Let $P(n)$ denote the statement "$1+2+3+\cdots+n = \frac{n(n+1)}{2}$". We want to show that $P(n)$ is true for all positive integers n. We will do so by completing the Basis Step and the Inductive Step.
Basis Step: Is $P(1)$ true? Yes, because $P(1)$ means "$1 = \frac{1\cdot 2}{2}$", which is obviously true.
Inductive Step: Suppose $P(k)$ is true. To complete this step, we need to show that $P(k+1)$ is true. Since we are assuming that $P(k)$ is true, it means that

$$1+2+3+\cdots+k = \frac{k(k+1)}{2}. \tag{1.5.3}$$

Using this assumption, we want to show that

$$1+2+3+\cdots+k+(k+1) = \frac{(k+1)(k+2)}{2}. \tag{1.5.4}$$

Indeed, $1+2+3+\cdots+k+(k+1) = \frac{k(k+1)}{2}+(k+1)$ (since we assumed that (1.5.3) is true), and this in turn equals $(k+1)\left(\frac{k}{2}+1\right) = \frac{(k+1)(k+2)}{2}$, proving (1.5.4). This means that $P(k+1)$ is true. ■

Problem 1.5.7. Use Mathematical Induction to prove that $2^n < n!$ for every integer n with $n \geq 4$.

Answer Let $P(n)$ denote the statement "$2^n < n!$".
Basis Step: First note that this time the basis step is to show that $P(4)$ is true. Indeed $P(4)$ is true, because $2^4 = 16$ is less than $4! = 24$.
Inductive Step: Suppose that $P(k)$ is true, that is, suppose that $2^k < k!$ We need to show that $P(k+1)$ is true, that is, $2^{k+1} < (k+1)!$. Indeed,

$$\begin{aligned} (k+1)! - 2^{k+1} &= k!(k+1) - 2^{k+1} \\ (\text{since } 2^k < k!) &> 2^k(k+1) - 2^{k+1} \\ &= 2^k(k-1) \\ &> 0, \end{aligned}$$

as desired. ■

Exercise 1.5.

1. The goal of this exercise is to shows that $\displaystyle\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}$ (see Theorem 1.5.1).

(a) Verify that

$$(k+1)^3 - k^3 = 6k^2 + 6k + 1 \tag{1.5.5}$$

for all k.

(b) Taking summation over $k = 1, 2, \ldots, n$ on both sides of (1.5.5), show that

$$(n+1)^3 - 1^3 = 6\sum_{k=1}^{n} k^2 + 6\sum_{k=1}^{n} k + n. \tag{1.5.6}$$

(c) Using Problem 1.5.6 in (1.5.6), derive that $\displaystyle\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}$.

2. Show that $1 + 3 + 5 + \cdots + (2n-1) = n^2$ for all positive integers n using

(a) Theorem 1.5.1.

(b) Mathematical Induction.

3. Show that $1 + 2 + 2^2 + \cdots + 2^n = 2^{n+1} - 1$ for all *nonnegative* integers n using

(a) Theorem 1.5.1.

(b) Mathematical Induction.

4. Use Mathematical Induction to prove that $n < 2^n$ for all positive integers n.

Chapter 2

Single Variable Calculus

2.1 Limits and Continuity

Roughly speaking, A function is continuous if one can graph it without lifting a pen. In other words, a function is said to be continuous if it has neither jumps nor drops. The notion of continuity arises frequently in probability. For example, the cumulative distribution function (cdf) of a random variable is always right-continuous. In this section, we review basic properties and consequences of continuity such as the Intermediate Value Theorem and the Extreme Value Theorem. We begin with a rigorous definition of limits.

Definition 2.1.1. Let $f : \mathbb{R} \to \mathbb{R}$ and $a \in \mathbb{R}$.

(a) We say that the <u>limit</u> of f as x approaches a is ℓ if for every $\epsilon > 0$, there is $\delta > 0$ such that $0 < |x - a| < \delta$ implies $|f(x) - \ell| < \epsilon$. In this case, we write $\lim_{x \to a} f(x) = \ell$.

(b) f is said to be <u>continuous</u> at a if $\lim_{x \to a} f(x) = f(a)$.

(c) f is said to be continuous over an interval I if f is continuous at a for all $a \in I$.

Remark 2.1.2.

(a) The choice of δ depends on ϵ. In general, smaller ϵ requires smaller δ.

(b) f is continuous at a if and only if for every $\epsilon > 0$, there is $\delta > 0$ such that $|x - a| < \delta$ implies $|f(x) - f(a)| < \epsilon$. Note that the inequality $|x - a| > 0$ is not a required condition since $|x - a| = 0$ implies that $x = a$ and that $|f(x) - f(a)| = |f(a) - f(a)| = 0 < \epsilon$ for all $\epsilon > 0$.

Problem 2.1.3. Let $a \in \mathbb{R}$. Show that $\lim_{x \to a} 2x = 2a$.

Answer Let $\epsilon > 0$ be given. We need to find $\delta > 0$ such that $0 < |x - a| < \delta$ implies $|2x - 2a| < \epsilon$. Since $|2x - 2a| = 2|x - a|$, to make $|2x - 2a| < \epsilon$, it suffices to have $|x - a| < \frac{\epsilon}{2}$ and we can take $\delta = \frac{\epsilon}{2}$. In fact, any positive number less than $\frac{\epsilon}{2}$ would work as a δ. ∎

Definition 2.1.1 generalizes to *one-sided* versions.

Definition 2.1.4. Let $f : \mathbb{R} \to \mathbb{R}$ and $a \in \mathbb{R}$.

(a) We say that the limit from the right of f as x approaches a is ℓ if for every $\epsilon > 0$, there is $\delta > 0$ such that $a < x < a + \delta$ implies $|f(x) - \ell| < \epsilon$. In this case, we write $\lim\limits_{x \to a+} f(x) = \ell$ or $f(a+) = \ell$.

(b) We say that the limit from the left of f as x approaches a is ℓ if for every $\epsilon > 0$, there is $\delta > 0$ such that $a - \delta < x < a$ implies $|f(x) - \ell| < \epsilon$. In this case, we write $\lim\limits_{x \to a-} f(x) = \ell$ or $f(a-) = \ell$.

(c) f is said to be right-continuous at a if $\lim\limits_{x \to a+} f(x) = f(a)$.

(d) f is said to be left-continuous at a if $\lim\limits_{x \to a-} f(x) = f(a)$.

Remark 2.1.5.

(a) $\lim\limits_{x \to a} f(x) = \ell$ if and only if $\lim\limits_{x \to a+} f(x) = \lim\limits_{x \to a-} f(x) = \ell$. In particular, f is continuous at a if and only if f is both right- and left-continuous at a.

(b) For f to be continuous at a, the following three conditions must be met:

- $f(a)$ exists,
- $\lim\limits_{x \to a} f(x)$ also exists, and
- $\lim\limits_{x \to a} f(x) = f(a)$.

Example 2.1.6. In the graph of $y = f(x)$ in Figure 2.1, a solid bullet ($\bullet$) indicates the function value. For example, $f(0.5) = 1.8$. Note that $\lim\limits_{x \to 2} f(x) = -0.7$ and $\lim\limits_{x \to 3} f(x) = -0.9$. Since $f(2)$ does not exist, f is *not* continuous at 2. Although $f(3) = 1.5$ exists, since $f(3) = 1.5 \neq -0.9 = \lim\limits_{x \to 3} f(x)$, f is *not* continuous at 3 either.

Problem 2.1.7. Based on the graph of $y = f(x)$ in Figure 2.1, compute $\lim\limits_{x \to -1+} f(x)$ and $\lim\limits_{x \to -1-} f(x)$. Is f left-continuous at -1?

Answer $\lim\limits_{x \to -1+} f(x) = -1.3$ and $\lim\limits_{x \to -1-} f(x) = 1$. Since $f(-1) = 1$, f is left-continuous. ■

Example 2.1.8. It is easy to show that polynomial functions are continuous everywhere (see Example 2.1.14). In fact, it is known that all elementary functions such as trigonometric functions, exponential functions, and logarithmic functions are continuous on its domain.

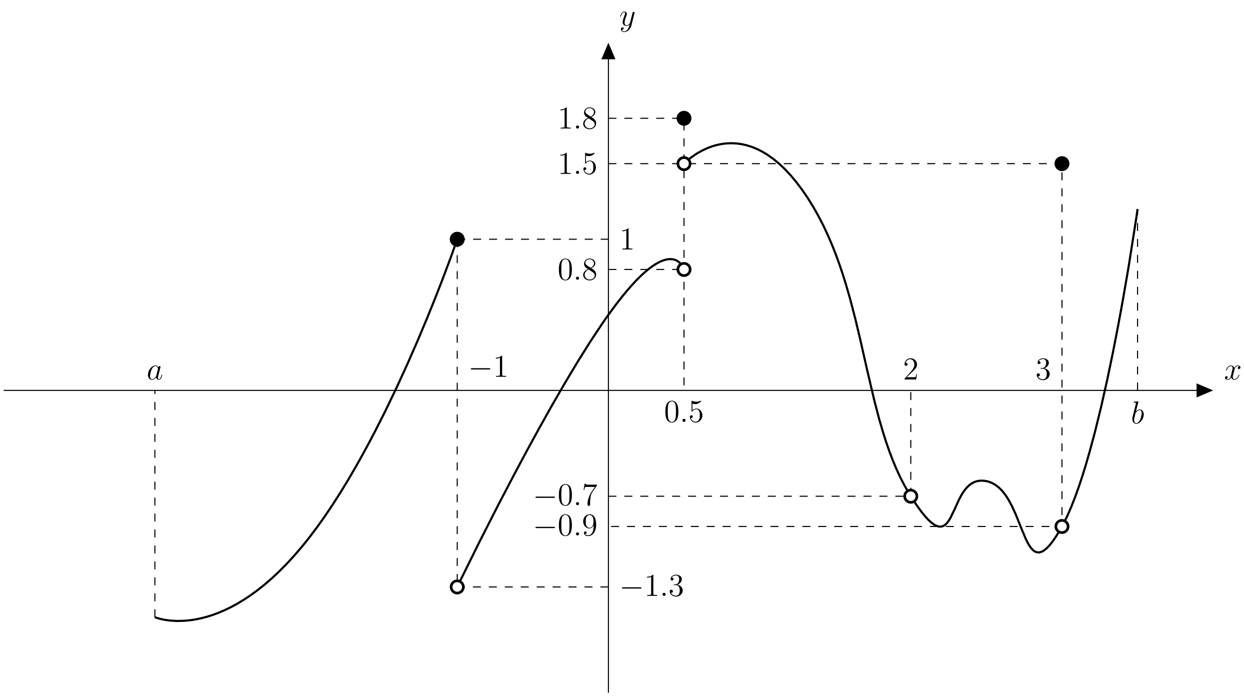

Figure 2.1: A function defined on $[a, b]$

The definition of the limit as x goes to $\pm\infty$ is given below.

Definition 2.1.9. Let $f : \mathbb{R} \to \mathbb{R}$.

(a) We say that the limit of f as x goes to ∞ is ℓ if for every $\epsilon > 0$, there is M such that $x > M$ implies $|f(x) - \ell| < \epsilon$. In this case, we write $\lim_{x\to\infty} f(x) = \ell$.

(b) We say that the limit of f as x goes to $-\infty$ is ℓ if for every $\epsilon > 0$, there is N such that $x < N$ implies $|f(x) - \ell| < \epsilon$. In this case, we write $\lim_{x\to-\infty} f(x) = \ell$.

Remark 2.1.10. Often we will use an expression such as $\lim_{x\to a} f(x) = \infty$ by which we mean that for any M, there is $\delta > 0$ such that $0 < |x - a| < \delta$ implies $f(x) > M$. Other expressions such as $\lim_{x\to a+} f(x) = \infty$ and $\lim_{x\to\infty} f(x) = -\infty$ can be defined analogously.

Example 2.1.11. By $\lim_{x\to a-} f(x) = -\infty$, we mean the following: for any N, there is $\delta > 0$ such that $a - \delta < x < a$ implies $f(x) < N$.

Some basic properties of limit are given below. Roughly speaking, limit is preserved by constant multiple, addition/subtraction, multiplication/division, and function composition.

Theorem 2.1.12. *Suppose that f, g, and h are functions. Let a be a real number or $a = \pm\infty$.*

(a) If $f(x) = c$ for all x, then $\lim\limits_{x\to a} f(x)$ exists and equals c.

(b) If both $\lim\limits_{x\to a} f(x)$ and $\lim\limits_{x\to a} g(x)$ exist (say $\lim\limits_{x\to a} f(x) = \ell$ and $\lim\limits_{x\to a} g(x) = k$), then

i. $\lim\limits_{x\to a}(cf(x))$ (c is a constant) also exists and equals $c\ell$.

ii. $\lim\limits_{x\to a}|f(x)|$ also exists and equals $|\ell|$.

iii. $\lim\limits_{x\to a}(f(x) \pm g(x))$ also exists and equals $\ell \pm k$.

iv. $\lim\limits_{x\to a}(f(x) \cdot g(x))$ also exists and equals $\ell \cdot k$.

v. If in addition $k \neq 0$, then $\lim\limits_{x\to a} \dfrac{f(x)}{g(x)}$ also exists and equals $\frac{\ell}{k}$.

(c) (Squeeze Theorem) If $f(x) \le h(x) \le g(x)$ for all $x \neq a$ and if $\lim\limits_{x\to a} f(x) = \lim\limits_{x\to a} g(x) = \ell$, then $\lim\limits_{x\to a} h(x)$ also exists and equals ℓ.

(d) If $\lim\limits_{x\to a} g(x) = b$ and f is continuous at b, then $\lim\limits_{x\to a} f(g(x)) = f(b)$.

Remark 2.1.13.

(a) Theorem 2.1.12 remains valid if the limit is replaced by one-sided limits.

(b) Theorem 2.1.12 implies that if f, g are continuous at a, then so are cf, $f \pm g$, fg, and f/g, provided that $g(a) \neq 0$.

Example 2.1.14. From Problem 2.1.3 and Theorem 2.1.12, it follows that

$$\lim_{x\to a} x = \lim_{x\to a} \frac{1}{2}(2x) = \frac{1}{2} \cdot 2a = a.$$

In general, if p is a polynomial and $a \in \mathbb{R}$, then $\lim\limits_{x\to a} p(x) = p(a)$. This shows that every polynomial is continuous on $\mathbb{R}$.

Example 2.1.15. Since $-1 \le \sin\frac{1}{x} \le 1$, it follows that $-x \le x\sin\frac{1}{x} \le x$ for all $x \neq 0$. Since $\lim\limits_{x\to 0}(-x) = \lim\limits_{x\to 0} x = 0$, we conclude that $\lim\limits_{x\to 0} x\sin\dfrac{1}{x} = 0$.

Example 2.1.16. It is known that $\lim\limits_{x\to 0}(1+x)^{\frac{1}{x}} = e$, where $e = 2.7182\cdots$, a number called <u>Euler's number</u> or the <u>base of natural logarithm</u>.

Problem 2.1.17. Compute $\lim\limits_{x\to 0} \dfrac{\ln(1+x)}{x}$.

Answer Theorem 2.1.12 and Example 2.1.16, $\lim_{x\to 0} \frac{\ln(1+x)}{x} = \lim_{x\to 0} \ln(1+x)^{\frac{1}{x}} = \ln e = 1$.
■

Now we study two main theorems regarding a continuous function defined over a closed interval. We start with the Intermediate Value Theorem.

Theorem 2.1.18 (Intermediate Value Theorem)**.** *Suppose f is continuous on a closed interval $[a,b]$. If k is any number between $f(a)$ and $f(b)$, then there is at least one number c in $[a,b]$ such that $f(c) = k$ (see Figure 2.2).*

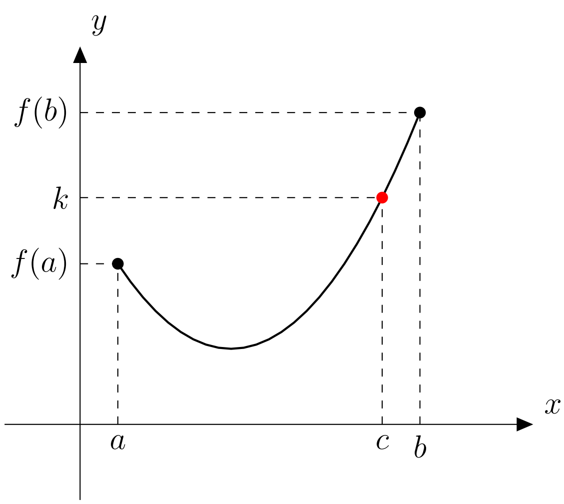

Figure 2.2: Intermediate Value Theorem

Example 2.1.19. Consider $f(x) = x^2 - 2x$. Note that $f(2) = 0$ and $f(6) = 24$. By the Intermediate Value Theorem, for any k between 0 and 24, there must be a c between 2 and 6 such that $f(c) = k$. In particular, if $k = 8$, then $c = 4$ works.

Problem 2.1.20. Show that there is $c \in [0,1]$ such that $ce^c = 1$.

Answer Let $f(x) = xe^x$, then clearly f is continuous on $[0,1]$. Since $f(0) = 0 < 1 < e = f(1)$, by Intermediate Value Theorem, there is $c \in [0,1]$ such that $f(c) = 1$, that is, $ce^c = 1$.
■

Remark 2.1.21 (Bisection Method)**.** The Intermediate Value Theorem can be used to find a numeric root of a continuous function. Suppose that f is a continuous function defined on a closed interval $[a,b]$ such that $f(a) \cdot f(b) < 0$. By the Intermediate Value Theorem, there exists x_0 such that $f(x_0) = 0$. Let $c = \frac{a+b}{2}$. If $f(c) = 0$, then c is a solution of the equation $f(x) = 0$ and we are done. If $f(c) \neq 0$, then either $f(a)f(c) < 0$ or $f(c)f(b) < 0$. Without loss of generality, assume that $f(c)f(b) < 0$, then the equation must have a solution in (c, b). Note that the length the interval in which a solution to $f(x) = 0$ decreased by half, and we can keep halving the intervals until a solution is found within desired accuracy. See Figure 2.3.

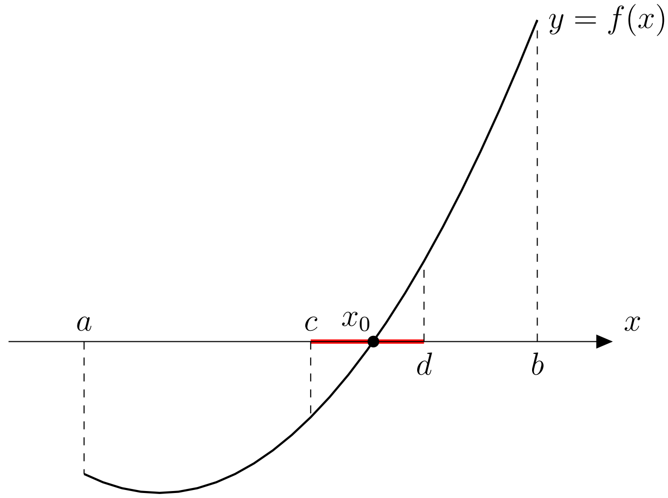

Figure 2.3: Bisection Method: a solution to $f(x) = 0$ is captured in the red line segment

Before we move on to the next result, some important notions regarding bounded subsets are in order.

Definition 2.1.22. Let S be a nonempty subset of $\mathbb{R}$. If there is m such that $m \leq s$ (respectively, $m \geq s$) for all $s \in S$, then we say that S is <u>bounded from below</u> (respectively, <u>bounded from above</u>) and m is called a <u>lower bound</u> (respectively, <u>upper bound</u>) of S. The largest lower bound of S is called the <u>infimum</u> and the smallest upper bound of S is called the <u>supremum</u>. The infimum (respectively, supremum) of S is denoted by $\inf S$ (respectively, $\sup S$).

Remark 2.1.23.

(a) Infimum and supremum generalize the notion of minimum and maximum, respectively. The main difference between infimum and minimum is in that $\inf S$ always exists whenever S is bounded from below. This is not the case when $\inf S$ is replaced by $\min S$: $\min S$ may not exist when S is bounded from below (see Example 2.1.24). Similar argument holds for $\sup S$ and $\max S$.

(b) $\inf S$ does not need be an element of S. In fact, $\inf S \in S$ if and only if S has the minimum element and in this case, we have $\inf S = \min S$. Likewise, $\sup S$ need not bean element of S and $\sup S \in S$ if and only if S has the maximum element and in this case, $\sup S = \max S$.

Example 2.1.24. Let $S = (1, 4] \cup \{5\}$, as described in Figure 2.4. Any number less than or equal to 1 is a lower bound of S and $\inf S = 1$. Any number greater than or equal to 5 is an upper bound of S and $\sup S = 5$. Note that $\max S = 5$, but $\min S$ does not exist.

Figure 2.4: A bounded subset of $\mathbb{R}$

The next theorem states that a continuous function defined on a finite closed interval achieves both maximum and minimum. See Figure 2.5.

Theorem 2.1.25 (Extreme Value Theorem). *Suppose that f is continuous on a closed interval $[a, b]$. Then there exist x_0 and x_1 in $[a, b]$ such that $f(x_0) \le f(x) \le f(x_1)$ for all $x \in [a, b]$.*

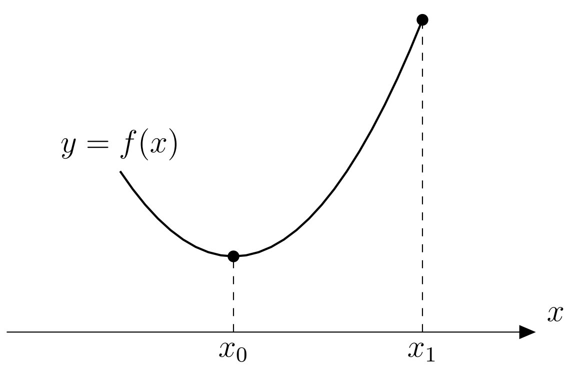

Figure 2.5: Extreme Value Theorem

Remark 2.1.26. The Extreme Value Theorem fails if the closed interval $[a, b]$ is replaced by a (half) open interval. Consider $f : (0, 5] \to \mathbb{R}$ defined by $f(x) = \frac{1}{x}$. It is easy to check that f fails to have the maximum. See Figure 2.6.

Exercise 2.1.

1. Based on the graph of $y = f(x)$ in Figure 2.1, compute $\lim\limits_{x \to 0.5+} f(x)$ and $\lim\limits_{x \to 0.5-} f(x)$. Is f left or right-continuous at 0.5?

2. Suppose that $f : \mathbb{R} \to \mathbb{R}$ satisfies $f(a + b) = f(a) + f(b)$ for all real numbers a and b. Suppose that f is continuous at $x = 0$. Show that f is continuous *everywhere*.
Hint: First, show that $f(0) = 0$. Let $x_0 \in \mathbb{R}$. For given $\epsilon > 0$, one needs to find $\delta > 0$

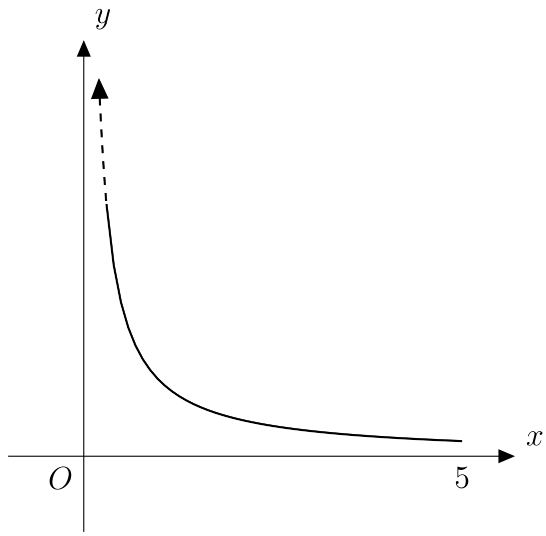

Figure 2.6: $f(x) = \frac{1}{x}$ on $(0, 5]$

such that $|x - x_0| < \delta$ implies that $|f(x) - f(x_0)| < \epsilon$. By assumption, there exists δ_0 such that $|y| < \delta_0$ implies $|f(y)| < \epsilon$. Show that taking $\delta = \delta_0$ works. You may want to use the fact that

$$f(x) - f(x_0) = f(x - x_0 + x_0) - f(x_0) = f(x - x_0) + f(x_0) - f(x_0) = f(x - x_0)$$

combined with the substitution $y = x - x_0$.

3. Prove that the equation

$$\frac{x^4 + x^2 + 5}{x - 1} + \frac{x^8 + 4x^5 + 1}{x - 7} = 0$$

has a solution between 1 and 7.
Hint: Consider a new equation $(x - 7)(x^4 + x^2 + 5) + (x - 1)(x^8 + 4x^5 + 1) = 0$.

4. Let $f : [0, 2] \to \mathbb{R}$ be a continuous function such that $f(0) = 0$ and $f(2) = 2$. Show that there is $x_0 \in [0, 1]$ such that $f(x_0 + 1) - f(x_0) = 1$.
Hint: Consider a function $g : [0, 1] \to \mathbb{R}$ given by $g(x) = f(x + 1) - f(x)$.

5. Let $f : (0, 5] \to \mathbb{R}$ defined by $f(x) = \frac{1}{x}$ (see Figure 2.6) and define $S = \{f(x) : x \in (0, 5]\}$. Is S bounded from below? from above? Does $\inf S$ exist? How about $\sup S$, $\min S$, and $\max S$?

2.2 Differentiation

Another important class of functions is that of differentiable functions. In probability, if the cdf of a random variable X is differentiable, then the probability density function (pdf) of X exists and equals the derivative of the cumulative distribution function. A formal definition of differentiability is in order.

Definition 2.2.1. For $f : \mathbb{R} \to \mathbb{R}$, consider the limit

$$\lim_{h\to 0} \frac{f(a+h)-f(a)}{h}.$$

If the limit exists, then we say that the function is differentiable at a and write

$$f'(a) = \lim_{h\to 0} \frac{f(a+h)-f(a)}{h}.$$

If the limit does not exist, then we say that the function is not differentiable at a.

Roughly speaking, a function is differentiable at $x = a$ if its graph is *smooth* there: a tangent line exists at the point $(a, f(a))$ (see Figure 2.7).

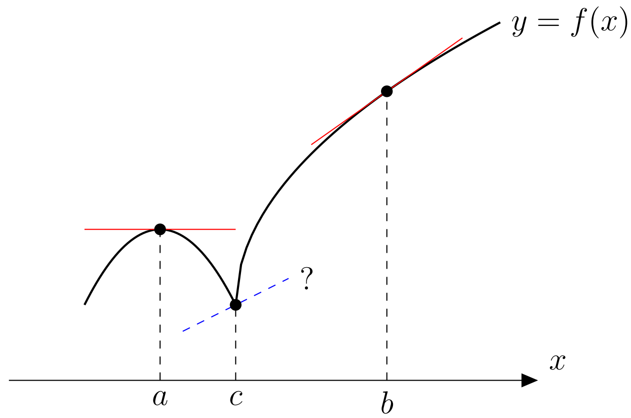

Figure 2.7: f is differentiable at $x = a$ and $x = b$, but not differentiable at $x = c$

Remark 2.2.2.

(a) $f'(a)$ is called the derivative of f at a.

(b) If f is differentiable at a, then f is continuous at a. The converse is not true.

(c) $f'(a)$ represents the *slope* of the tangent line to the graph of $y = f(x)$ at $x = a$. See Figure 2.8.

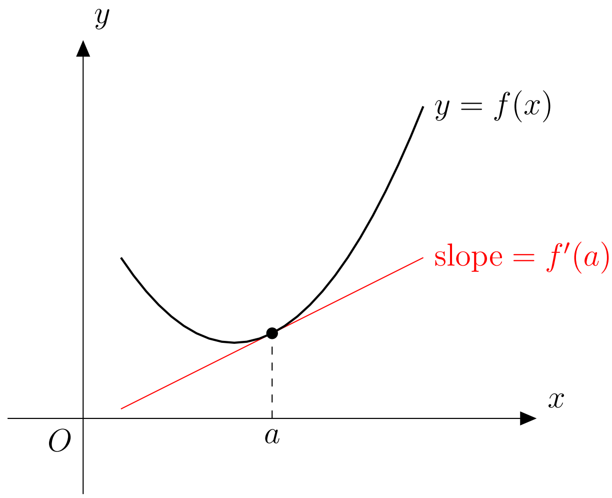

Figure 2.8: Geometric meaning of $f'(a)$

(d) If f is differentiable at every point in an interval, the function is said to be differentiable on that interval.

Example 2.2.3. Let $f(x) = |x|$. If $a > 0$, then for h close to 0, one gets

$$\frac{f(a+h) - f(a)}{h} = \frac{|a+h| - |a|}{h} = \frac{a+h-a}{h} = 1$$

and it follows that $f'(a) = \lim_{h \to 0} \dfrac{f(a+h) - f(a)}{h} = 1$. If $a < 0$, then for h close to 0,

$$\frac{f(a+h) - f(a)}{h} = \frac{|a+h| - |a|}{h} = \frac{-(a+h) - (-a)}{h} = -1,$$

so $f'(a) = -1$. We now claim that f is not differentiable at $x = 0$. Indeed,

$$\lim_{h \to 0+} \frac{f(0+h) - f(0)}{h} = \lim_{h \to 0+} \frac{|h|}{h} = \lim_{h \to 0+} \frac{h}{h} = 1,$$

while

$$\lim_{h \to 0-} \frac{f(0+h) - f(0)}{h} = \lim_{h \to 0-} \frac{|h|}{h} = \lim_{h \to 0+} \frac{-h}{h} = -1.$$

This shows that $\lim_{h \to 0} \dfrac{f(0+h) - f(0)}{h}$ does not exist.

Instead of computing $f'(a)$ for each point a, we now consider a generic derivative $f'(x)$. This *function* gives the slope of the tangent line to $y = f(x)$ at point x.

Definition 2.2.4. Let f be a function. The derivative function or simply derivative of f, denoted by f' or $\dfrac{df}{dx}$, is a function defined by

$$f'(x) = \lim_{h\to 0} \frac{f(x+h) - f(x)}{h}.$$

Example 2.2.5. Let $f(x) = \frac{1}{4}x^2 - 4$, then

$$f'(x) = \lim_{h\to 0} \frac{\frac{1}{4}(x+h)^2 - 4 - (\frac{1}{4}x^2 - 4)}{h} = \lim_{h\to 0} \frac{\frac{1}{2}xh + \frac{h^2}{4}}{h} = \frac{1}{2}x.$$

Problem 2.2.6. Let $f(x) = \frac{1}{x}$. Find $f'(x)$.

Answer

$$f'(x) = \lim_{h\to 0} \frac{\frac{1}{x+h} - \frac{1}{x}}{h} = \lim_{h\to 0} \frac{\frac{-h}{x(x+h)}}{h} = -\frac{1}{x^2}.$$

■

To understand the geometric meaning of the derivative function, we now compare the graphs of f and f' using $f(x) = \frac{1}{4}x^2 - 4$. Note that the *slope* of the tangent line to $y = f(x)$ at $x = 2$ (part of it is given in Figure 2.9) is 1.

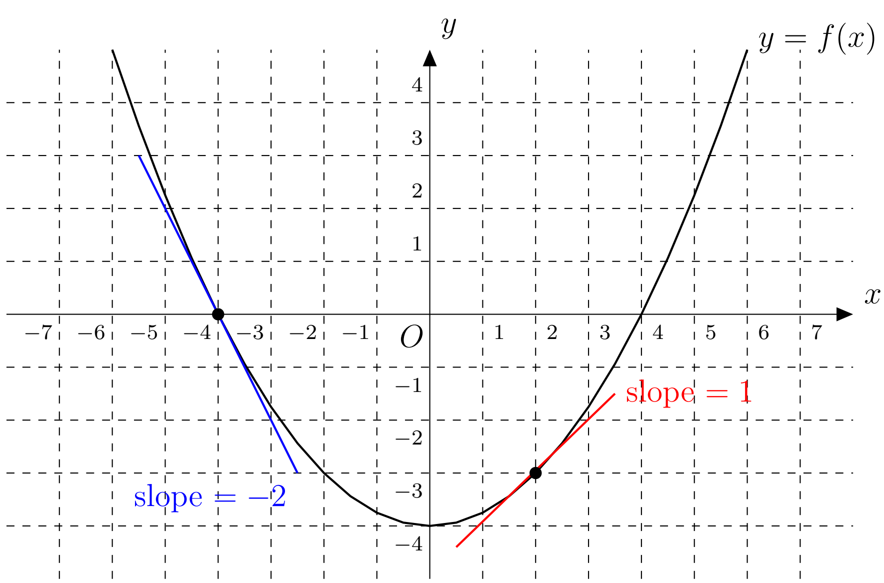

Figure 2.9: Graph of $y = f(x) = \frac{1}{4}x^2 - 4$

In Example 2.2.5, we observe that the derivative of f is given by $f'(x) = \frac{1}{2}x$, whose graph is given in Figure 2.10. Note that $f'(2) = 1$. Similarly, the *slope* of the tangent line

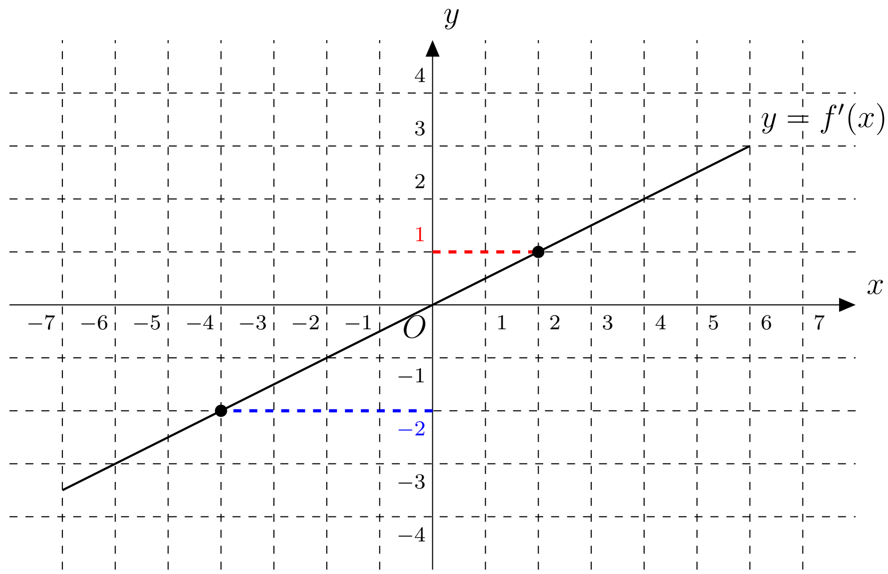

Figure 2.10: Graph of $y = f'(x) = \frac{1}{2}x$

to $y = f(x)$ at $x = -4$ is -2, so $f'(-4) = -2$.

It is known that all elementary functions are differentiable on its domain. Below is the collection of derivative functions of elementary functions.

Theorem 2.2.7. *Let a and c be constants, where $a > 0$, $a \neq 1$.*

(a) If $f(x) = c$, then $f'(x) = 0$.

(b) If $f(x) = x^n$, then $f'(x) = nx^{n-1}$.

(c) If $f(x) = e^x$, then $f'(x) = e^x$.

(d) If $f(x) = a^x$, then $f'(x) = a^x \ln a$.

(e) If $f(x) = \ln x$, then $f'(x) = \frac{1}{x}$.

(f) If $f(x) = \log_a x$, then $f'(x) = \frac{1}{x \ln a}$.

(g) If $f(x) = \sin x$, then $f'(x) = \cos x$.

(h) If $f(x) = \cos x$, then $f'(x) = -\sin x$.

(i) If $f(x) = \tan x$, then $f'(x) = \sec^2 x$.

(j) If $f(x) = \arctan x$, then $f'(x) = \frac{1}{1+x^2}$.

Some basic properties of derivatives, which are immediate consequences of Theorem 2.1.12, are recorded below.

Theorem 2.2.8. *Let f and g be differentiable functions and c be a constant.*

(a) (Constant Multiple Rule) $cf(x)$ is also differentiable and $(cf(x))' = cf'(x)$.

(b) (Sum Rule) $f(x) \pm g(x)$ is also differentiable and $(f(x) \pm g(x))' = f'(x) \pm g'(x)$.

(c) *(Product Rule)* $f(x)g(x)$ *is also differentiable and* $(f(x)g(x))' = f'(x)g(x)+f(x)g'(x)$.

(d) *(Quotient Rule)* $\frac{f(x)}{g(x)}$ *is also differentiable and* $\left(\frac{f(x)}{g(x)}\right)' = \frac{f'(x)g(x) - f(x)g'(x)}{(g(x))^2}$, *provided that* $g(x) \neq 0$.

(e) *(Chain Rule)* $(f(g(x)))' = f'(g(x))g'(x)$.

The second order derivative f'' of f is defined by the derivative function of f', provided that f' is also differentiable (In this case, we say that f is twice differentiable). Higher order derivatives are defined likewise.

Remark 2.2.9. The curvature κ of a plane curve $y = f(x)$ is given by

$$\kappa = \frac{|f''(x)|}{(1 + f'(x))^{3/2}}.$$

Therefore, the second derivative, together with the first derivative in the denominator, determines how rapidly a given curve bends.

The sign of the derivative function determines whether f is increasing, and the sign of the second derivative determines the concavity of f, as summarized in the following.

Theorem 2.2.10. *Let* f *be differentiable on an interval* I. *Then*

(a) $f' > 0$ *on* I *if and only if* f *is increasing over* I.

(b) $f' < 0$ *on* I *if and only if* f *is decreasing over* I.

If f *is twice differentiable, then*

(a) $f'' > 0$ *on* I *if and only if* f' *is increasing over* I *if and only if* f *is concave up* I.

(b) $f'' < 0$ *on* I *if and only if* f' *is increasing over* I *if and only if* f *is concave down* I.

We close this section with L'Hopital's Theorem, a powerful tool to compute limits.

Theorem 2.2.11 (L'Hopital's Theorem)**.** *Let* f *and* g *be differentiable functions. Let* c *be a real number or,* $c = \pm\infty$. *If* $\frac{f(c)}{g(c)}$ *is of the form* $\frac{0}{0}$ *or* $\frac{\infty}{\infty}$, *then*

$$\lim_{x\to c} \frac{f(x)}{g(x)} = \lim_{x\to c} \frac{f'(x)}{g'(x)},$$

provided that $\lim_{x\to c} \frac{f'(x)}{g'(x)}$ *exists.*

Problem 2.2.12. Compute $\lim_{x\to 2} \frac{2^x - 4}{x - 2}$.

Answer $\lim\limits_{x\to 2} \dfrac{2^x - 4}{x - 2} = \lim\limits_{x\to 2} \dfrac{2^x \ln 2}{1} = 4\ln 2.$ ■

Problem 2.2.13. Compute $\lim\limits_{x\to\infty} xe^{-2x}$.

Answer $\lim\limits_{x\to\infty} xe^{-2x} = \lim\limits_{x\to\infty} \dfrac{x}{e^{2x}} = \lim\limits_{x\to\infty} \dfrac{1}{2e^{2x}} = 0.$ ■

Exercise 2.2.

1. Show that $f(x) = x|x|$ is differentiable everywhere. Is f' continuous? differentiable?

2. Consider a piecewise defined function

$$f(x) = \begin{cases} ax + b, & x \le 1, \\ 2 - (x-2)^2, & x > 1. \end{cases}$$

Determine constants a and b so that f is continuous and differentiable everywhere.

3. The goal of this exercise is to prove the Quotient Rule using the Product Rule. Let f and g be differentiable function and $g(x) \neq 0$.

(a) Define $h(x) = \dfrac{f(x)}{g(x)}$. Compute $f'(x)$ in terms of $h(x), h'(x), g(x)$, and $g'(x)$.

(b) Express $h'(x)$ in terms of $f(x), f'(x), g(x)$, and $g'(x)$, and deduce the Quotient Rule.

4. Let $f : [0, \infty) \to \mathbb{R}$ be a differentiable function such that $f(0) = 0$ and $0 \le f'(x) \le f(x)$ for all $x \ge 0$. Show that $f(x) = 0$ for all $x \ge 0$.
Hint: Define $g : [0, \infty) \to \mathbb{R}$ by $g(x) = e^{-x} f(x)$, then $g(0) = 0$ and g is nonnegative for $x \ge 0$. Show that g is decreasing.

5. Compute the following limits:

(a) $\lim\limits_{x\to 3} \dfrac{x^2 - 9}{2x^2 - 5x - 3}$

(b) $\lim\limits_{x\to 0} \dfrac{\sin(3x)}{2x}$

(c) $\lim\limits_{t\to 0} \dfrac{t^2}{e^t - t - 1}$

(d) $\lim\limits_{x\to\infty} x^2 e^{-x}$

(e) $\lim\limits_{x\to\infty} \dfrac{\ln x}{\ln(x^2 + 1)}$

(f) $\lim\limits_{x\to 0} \left(\dfrac{1}{x} - \dfrac{1}{\sin x} \right)$

6. Let $a > 0$. For $x \neq 0$, define let $\psi(x) = \frac{a^x - 1}{x}$. Determine $\psi(0)$ (in terms of a) so that ψ is continuous at $x = 0$.

2.3 Mean Value Theorem and Its Consequences

One of the important properties of differentiable functions is depicted in the Mean Value Theorem. We begin with Rolle's Theorem, which can be thought as a special case of the Mean Value Theorem.

Theorem 2.3.1 (Rolle's Theorem)**.** *Let f be a continuous function defined on a closed interval $[a, b]$. Suppose that f is differentiable on (a, b) and $f(a) = f(b)$, then there is $c \in (a, b)$ such that $f'(c) = 0$.*

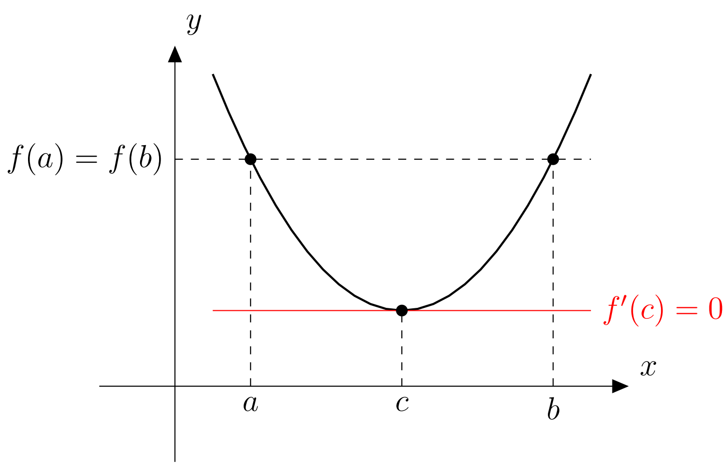

Figure 2.11: Geometric interpretation of Rolle's Theorem

The geometric meaning of Rolle's Theorem is described in Figure 2.11. Roughly speaking, if $f(a) = f(b)$, then there must be a point c between a and b at which the tangent line is horizontal. Note that there could be multiple such c's. What if $f(a) \neq f(b)$? Rolle's Theorem can be generalized to the following.

Theorem 2.3.2 (Mean Value Theorem)**.** *Let f be a continuous function defined on the closed interval $[a, b]$. Suppose that f is differentiable on (a, b), then there exists $c \in (a, b)$ such that*

$$\frac{f(b) - f(a)}{b - a} = f'(c). \tag{2.3.1}$$

The geometric interpretation of the Mean Value Theorem is given in Figure 2.12. Note that (2.3.1) can be written as $f(b) = f(a) + f'(c)(b - a)$ and this can be viewed as an approximation of $f(b)$ by $f(a)$ with an error $f'(c)(b - a)$.

Problem 2.3.3. Suppose that f is differentiable on (a, b) and $f'(x) = 0$ for all $x \in (a, b)$. Show that f is a constant function.

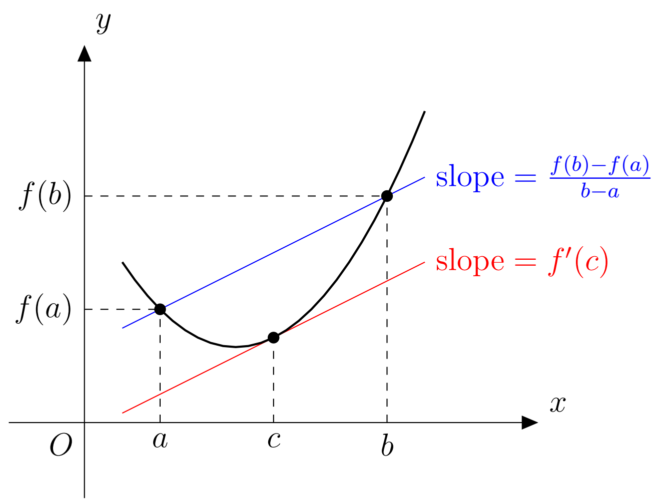

Figure 2.12: Geometric interpretation of Mean Value Theorem

Answer Pick any $s, t \in (a, b)$ with $s \neq t$. By the Mean Value Theorem, there is r between s and t such that

$$\frac{f(s) - f(t)}{s - t} = f'(r).$$

Since $f'(r) = 0$, it follows that $f(s) = f(t)$. ■

Problem 2.3.4. Let $f : \mathbb{R} \to \mathbb{R}$ with $f(0) \neq 0$. Suppose that $f(x + y) = f(x)f(y)$ for all $x, y \in \mathbb{R}$ and f is differentiable at 0 with $f'(0) = \lambda$. Show that $f(x) = e^{\lambda x}$.

Answer First, we note that $f(0) = f(0+0) = f(0)f(0) = f(0)^2$. Since $f(0) \neq 0$, it follows that $f(0) = 1$. Since $f(x + h) = f(x)f(h)$, we get

$$\lim_{h \to 0} \frac{f(x+h) - f(x)}{h} = f(x) \lim_{h \to 0} \frac{f(h) - 1}{h} = f(x) \underbrace{\lim_{h \to 0} \frac{f(h+0) - f(0)}{h}}_{f'(0)} = \lambda f(x),$$

which implies that $f'(x) = \lambda f(x)$ for all x. Define $g : \mathbb{R} \to \mathbb{R}$ by $g(x) = f(x)e^{-\lambda x}$, then g is differentiable everywhere and moreover $g'(x) = e^{-\lambda x}(f'(x) - \lambda f(x)) = 0$ for all x. By Problem 2.3.3, we conclude that there is a constant c such that $f(x) = ce^{\lambda x}$. Since $f(0) = 1$, it follows that $f(x) = e^{\lambda x}$. ■

Exercise 2.3.

1. Let f and g be continuous on the closed interval $[a, b]$. Suppose that f and g are differentiable on (a, b) and $g(a) \neq g(b)$. Show that there exists $c \in (a, b)$ such that

$$\frac{f(b) - f(a)}{g(b) - g(a)} = \frac{f'(c)}{g'(c)}.$$

Hint: Define a function h on $[a,b]$ by $h(x) = g(x)(f(b)-f(a)) - f(x)(g(b)-g(a))$, then clearly h is continuous on $[a,b]$ and differentiable on (a,b). Show that $h(a) = h(b)$ and apply Rolle's Theorem. This result is often called Cauchy Mean Value Theorem.

2. Suppose that $f : \mathbb{R} \to \mathbb{R}$ satisfies

$$|f(x)| \leq 1000|x|^2$$

for all $x \in \mathbb{R}$. Show that f is differentiable at 0. What is $f'(0)$?

3. A function $f : \mathbb{R} \to \mathbb{R}$ is said to be *convex* if

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y) \tag{2.3.2}$$

for all $x, y \in \mathbb{R}$ and for all λ with $0 < \lambda < 1$. The purpose of this problem is to show that if $f : \mathbb{R} \to \mathbb{R}$ is twice differentiable with $f''(x) \geq 0$ for all x, then f is convex.

(a) Let $x < y$ be real numbers and $0 < \lambda < 1$. Let $z = \lambda x + (1-\lambda)y$. Show that $x < z < y$.

(b) Show that there exist $u, v \in \mathbb{R}$ with

$$x < u < z < v < y$$

such that

$$f'(u) = \frac{f(z)-f(x)}{z-x} \quad \text{and} \quad f'(v) = \frac{f(y)-f(z)}{y-z}.$$

(c) Suppose that f is twice differentiable with $f''(x) \geq 0$ for all x. Show that $f'(v) \geq f'(u)$ and from this conclude that (2.3.2) holds.

4. Let $f : \mathbb{R} \to \mathbb{R}$ be an infinitely differentiable function. Suppose that

$$f(1) = f(0) = f'(0) = f''(0) = f'''(0) = \cdots = f^{(99)}(0) = 0.$$

Prove that $f^{(100)}(x_0) = 0$ for some x_0 in $(0,1)$.

5. Suppose that $f : \mathbb{R} \to \mathbb{R}$ be twice differentiable with $f'(x) > 0$ and $f''(x) > 0$ for all $x \in \mathbb{R}$. Show that $\lim_{x\to\infty} f(x) = \infty$.
Hint: Suppose not, then f must be bounded from above and $\lim_{x\to\infty} f(x) = \sup_{x\in\mathbb{R}} f(x) = A$ exists (why?). Pick any $x_0 \in \mathbb{R}$ and let $h = f(x_0+1) - f(x_0)$, then $h > 0$. Choose M so that $M > \frac{A-f(x_0+1)}{h}$. Apply the Mean Value Theorem to $\frac{f(x_0+1+M)-f(x_0+1)}{M}$ to show that $f(x_0+1+M) > A$, a contradiction.

2.4 Finding Extreme Values

Maximum likelihood estimation is a method of estimating the parameter of a statistical model based. As the name suggests, the method involves finding the maximum of the likelihood, a function of the parameter determined by a given data set.

Definition 2.4.1. Let f be defined on an interval I and let $p \in I$.

(a) f is said to have a local minimum at p if $f(p)$ is less than or equal to the values of f for points near p.

(b) f is said to have a local maximum at p if $f(p)$ is greater than or equal to the values of f for points near p.

(c) f is said to have a local extremum at p if f has either a local minimum or a local maximum at p.

Example 2.4.2. In Figure 2.13, f has two local minima at b and d and two local maxima at a and c.

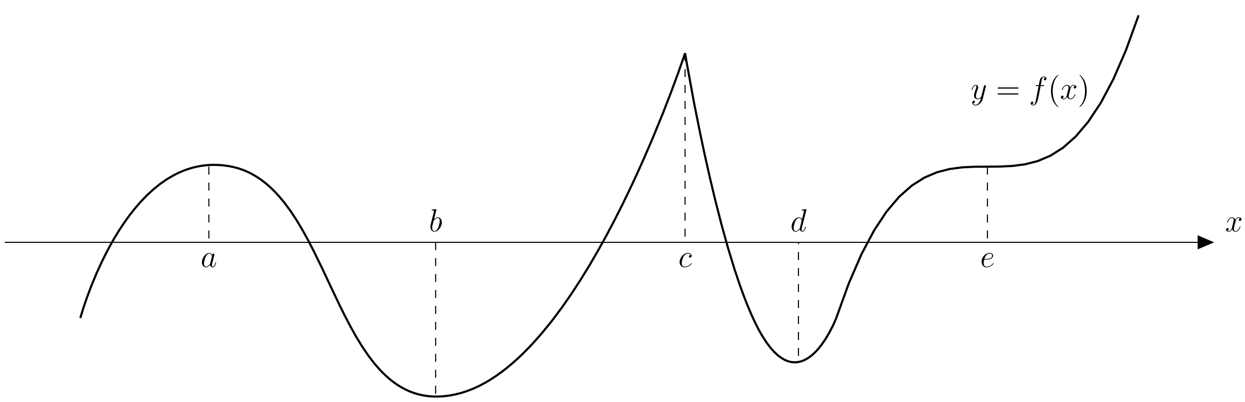

Figure 2.13: A function with local extrema

Definition 2.4.3. A point p in the domain of a function f where $f'(p) = 0$ or $f'(p)$ is undefined is called a critical point of the function.

Example 2.4.4. In Figure 2.13, f has five critical points: a, b, c, d, and e.

When the graph of a function is present, identifying local extrema is not that difficult. However, with only the formula of a function, finding local extrema might be challenging. The next theorems explain a systematic way to find local extrema without graphing the function.

Theorem 2.4.5 (Critical Point Theorem)**.** *Suppose f is defined on an interval and has a local extremum at p, which is not an endpoint of the interval. Then p must be a critical point of f.*

Remark 2.4.6.

(a) The converse of the Critical Point Theorem is not true, as the point e shows in Figure 2.13.

(b) If f is differentiable everywhere, the Critical Point Theorem means that local extrema occur only at points where f' equals zero.

Roughly speaking, critical points are the only candidates of local extrema. The next theorem, which is based on Theorem 2.2.10, explains how to classify critical points.

Theorem 2.4.7 (The Second Derivative Test)**.** *Let f be a function. Suppose $f'(p) = 0$ (so p is a critical point of f).*

(a) If $f''(p) > 0$, then f has a local minimum at p.

(b) If $f''(p) < 0$, then f has a local maximum at p.

(c) If $f''(p) = 0$, then the test fails. That is, f can have a local minimum, local maximum, or neither at p.

Problem 2.4.8. Classify the critical points of $f(x) = x^3 - 9x^2 - 48x + 52$ as local maxima or local minima.

Answer Note that $f'(x) = 3x^2 - 18x - 48 = 3(x-8)(x+2)$ and $f''(x) = 6x - 18$. It follows that $f'(x) = 0$ if $x = 8$ or $x = -2$, giving two critical points. Since $f''(8) > 0$ and $f''(-2) < 0$, we conclude that f has a local minimum -396 at 8 and a local maximum 104 at -2. ■

The single greatest (or least) value of a function f over an interval is called the <u>global maximum</u> (or <u>global minimum</u>) of f over the interval. It is apparent that the global extremum itself must be a local extremum.

Problem 2.4.9. Find the global maximum and global minimum of $g(x) = x + \frac{1}{x}$ defined on $(0, \infty)$.

Answer Since $g'(x) = 1 - \frac{1}{x^2}$ and $g''(x) = \frac{2}{x^3}$, we see that $g'(1) = 0$ and $g''(1) > 0$. It follows that g has a unique local minimum 2 at $x = 1$, which should be the global minimum. g does not have the global maximum since $\lim_{x\to\infty} g(x) = \infty$. ■

Problem 2.4.10. For $\sigma > 0$, define f by $f(\sigma) = \frac{1}{\sigma}e^{-\frac{1}{\sigma^2}}$. Find the global maximum of f over $(0, \infty)$.

Answer Since $f(\sigma) > 0$ for all $\sigma > 0$ and $\ln x$ is an increasing function, it suffices to maximize $\ell(\sigma) = \ln f(\sigma) = -\ln\sigma - \frac{1}{\sigma^2}$, which is easier to handle. From $\ell'(\sigma) = -\frac{1}{\sigma} + \frac{2}{\sigma^3} = 0$, we get $\ell'(\sqrt{2}) = 0$. Since $\ell''(\sigma) = \frac{1}{\sigma^2} - \frac{6}{\sigma^4}$, we have that $\ell''(\sqrt{2}) < 0$. It follows that ℓ, and hence f has the global maximum at $\sqrt{2}$. ■

For a continuous function defined on a bounded closed interval, there is an easy way to determine the global extremum.

Theorem 2.4.11 (The Global Extremum Theorem)**.** *Let f be a continuous function on a closed interval. Then the global maximum and global minimum are obtained either at critical points of the function or at the endpoints of the interval.*

Example 2.4.12. Consider $f(x) = x^3 - 9x^2 - 48x + 52$ on the interval $[-5, 15]$. From Problem 2.4.8, we see that f has a local minimum -396 at 8 and a local maximum 104 at -2. Since $f(-5) = -58$ and $f(15) = 682$, we conclude that the global maximum is 682 and the global minimum is -396.

Exercise 2.4.

1. Let $f : [0,1] \to \mathbb{R}$ be continuous on $[0,1]$ and differentiable on $(0,1)$. Suppose that $f(0) = 0$ and $|f'(x)| \leq |f(x)|$ for all $x \in (0,1)$. Prove that $f(x) = 0$ for all $x \in [0,1]$.
Hint: Since f is continuous on $[0,1]$, so is $|f|$. By the Extreme Value Theorem, there exists $y_0 \in [0,1]$ such that $|f(y_0)| = \max\{|f(x)| : x \in [0,1]\}$. If $y_0 = 0$, then the result follows obviously (why?). If $y_0 > 0$, then show that there is $z_0 \in (0, y_0)$ such that $f(y_0) = f'(z_0) \cdot y_0$ and hence $|f(y_0)| \leq |f'(z_0)||y_0| \leq |f(z_0)| \leq |f(y_0)|$. Conclude that f has either maximum or minimum at z_0. Use the Critical Point Theorem to show that $f'(z_0) = 0$.

2. Find the local extrema of $f(x) = \frac{2x}{x^2+1}$. Are they global extrema, too?

3. Find the global extrema of $g(x) = xe^{-x}$, if any.

4. For $p \in [0,1]$, define f by $f(p) = p^2(1-p)^6$. Find the global maximum of f.

5. Let $H_1, H_2, \ldots, H_m$ be null hypotheses of m independent tests with a common significance level α_0. The familywise error rate (FWER) is the probability of making even one type I error in the family: $1 - (1 - \alpha_0)^m$. Bonferroni correction states that if $\alpha_0 = \frac{\alpha}{m}$, then FWER is less than α. Prove this.
Hint: Let $f : [0,1] \to \mathbb{R}$ be defined to be

$$f(\alpha) = 1 - \left(1 - \frac{\alpha}{m}\right)^m.$$

Show that $f(\alpha) \leq \alpha$.

2.5 Integration

When X is a continuous random variable with the pdf f, the probability that X belongs to an interval $[a, b]$ is given by the definite integral of f over $[a, b]$. Moreover, moments of X, in particular the expectation of X, are all expressed by definite integrals. To define the definite integral of f over $[a, b]$, let $n \in \mathbb{N}$, and partition the interval $[a, b]$ into n subintervals of equal length, $\{a = t_0 < t_1 < t_2 < \cdots < t_n = b\}$, where $t_k = a + k\frac{b-a}{n}$. Consider the upper sum

$$U(f, n) = \sum_{k=1}^{n} \sup\{f(x) : t \in [t_{k-1}, t_k]\} \cdot \frac{b-a}{n}$$

and the lower sum

$$L(f, n) = \sum_{k=1}^{n} \inf\{f(x) : t \in [t_{k-1}, t_k]\} \cdot \frac{b-a}{n}.$$

Note that both upper and lower sums represent the sum of rectangular areas. See the first three plots in Figure 2.14 for an illustration when $n = 6, 12, 30$.

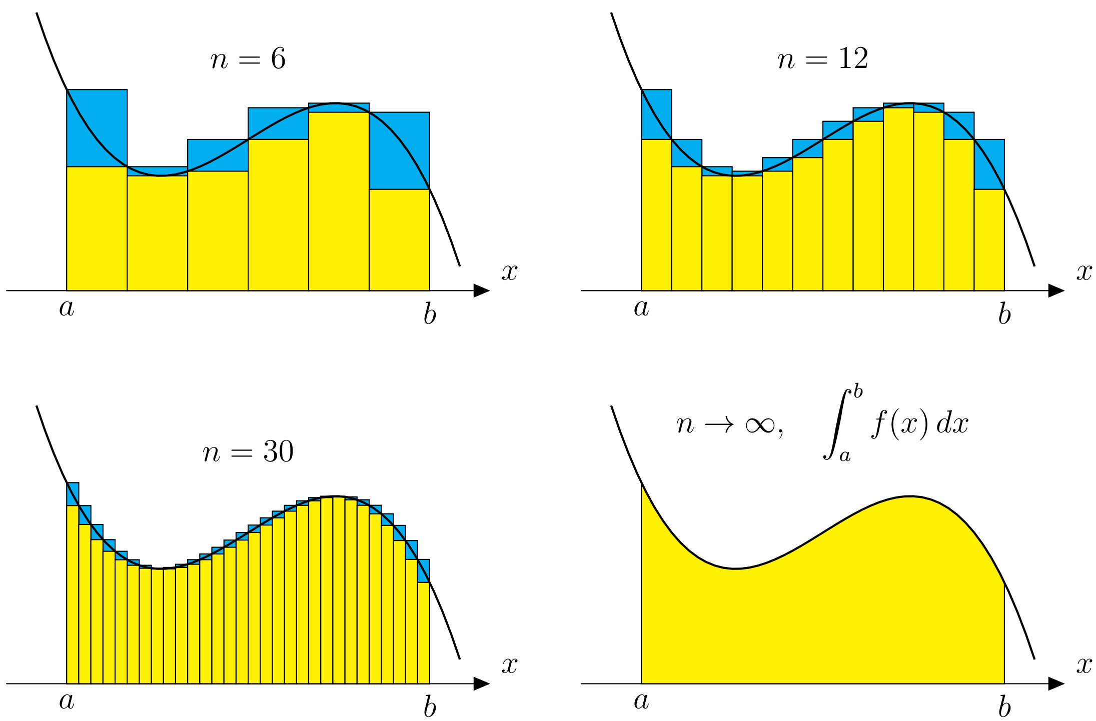

Figure 2.14: Upper and lower sums

As can be seen in Figure 2.14, as n gets bigger and bigger, the difference between $U(f, n)$ and $L(f, n)$ gets smaller and smaller. In fact, it is known that if f has at most finitely many discontinuities in $[a, b]$, then both limits $\lim_{n\to\infty} U(f, n)$ and $\lim_{n\to\infty} L(f, n)$ exist and equal each other. When this happens, we say that f is Riemann integrable or simply integrable, and define the definite integral $\int_a^b f(x)\,dx$ of f between a and b by

$$\int_a^b f(x)\,dx = \lim_{n\to\infty} U(f, n),$$

which also equals $\lim_{n\to\infty} L(f, n)$.

As the last plot in Figure 2.14 shows, when f is nonnegative over $[a, b]$, $\int_a^b f(x)\,dx$ can be interpreted as the area under the graph of f and above the x-axis between vertical lines $x = a$ and $x = b$.

Example 2.5.1. The graph of $y = \sin x$ is given in Figure 2.15. The areas A and B can be described by $A = \int_{-\frac{3\pi}{2}}^{-\frac{5\pi}{4}} \sin x\,dx$ and $B = \int_{\frac{\pi}{2}}^{\pi} \sin x\,dx$.

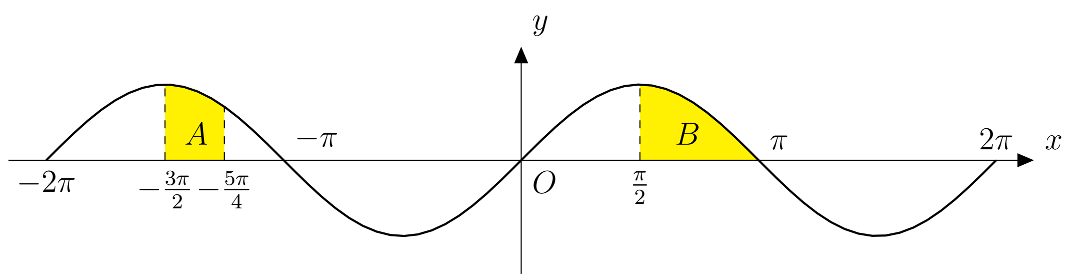

Figure 2.15: Graph of $f(x) = \sin x$

Problem 2.5.2. Compute $\int_{-3}^{3} \sqrt{9 - x^2}\,dx$.

Answer Note that the graph of $y = \sqrt{9 - x^2}$ is the upper semicircle with radius 3 centered at the origin (see Figure 2.16), so $\int_{-3}^{3} \sqrt{9 - x^2}\,dx = \frac{1}{2} \cdot \pi \cdot 3^2 = \frac{9\pi}{2}$. ■

How do we interpret the definite integral of f over $[a, b]$ when f is not always nonnegative? If f is positive for some x values and negative for others, then $\int_a^b f(x)\,dx$ is interpreted as the *signed* area, that is, it is the sum of areas above the x-axis, counted positively, and areas below the x-axis, counted negatively. For example, in Figure 2.17,

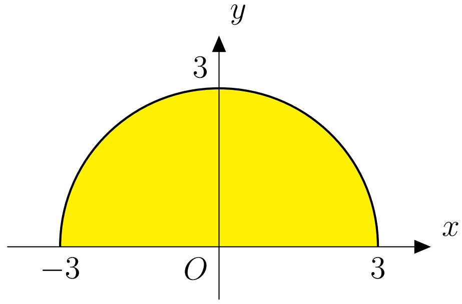

Figure 2.16: Graph of $y = \sqrt{9 - x^2}$

suppose the shaded area above the x-axis is 2 and the shaded area below the x-axis is 3. Then

$$\int_a^b f(x)\, dx = 2 + (-3) = -1.$$

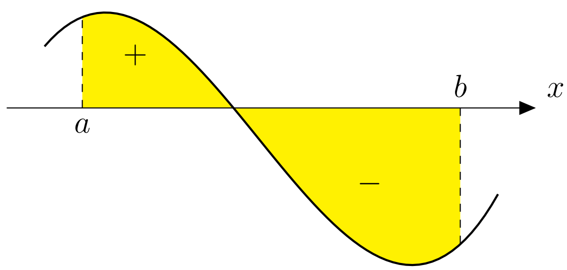

Figure 2.17: A function that has both positive and negative values over $[a, b]$

Problem 2.5.3. It is known that $\displaystyle\int_{\frac{\pi}{2}}^{\pi} \sin x\, dx = 1$ (see Problem 2.5.9). Using this, compute

(a) $\displaystyle\int_0^{\frac{3\pi}{2}} \sin x\, dx$

(b) $\displaystyle\int_0^1 \arcsin x\, dx$

Answer

(a) By symmetry of sine, we see that the area under the sine curve between $x = 0$ and $x = \frac{\pi}{2}$ equals 1. Similarly, the area enclosed by the sine curve, the x-axis, and vertical lines $x = \pi$ and $x = \frac{3\pi}{2}$ also equals 1. It follows that $\displaystyle\int_0^{\frac{3\pi}{2}} \sin x\, dx = 1 + 1 + (-1) = 1.$ See Figure 2.18.

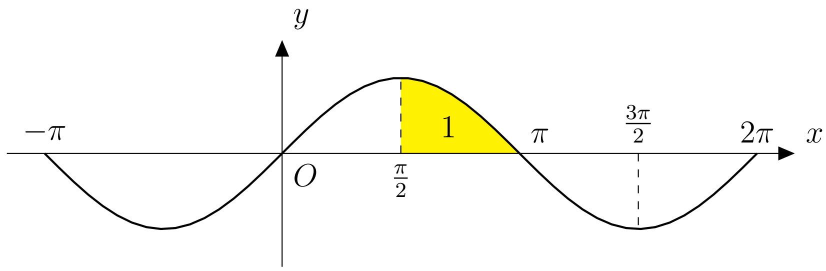

Figure 2.18: Graph of sine

(b) It follows from Figure 2.19 that

$$\begin{aligned}\int_{\frac{1}{2}}^{1} \arcsin x\,dx &= \frac{\pi}{2}\cdot 1 - \int_{0}^{\frac{\pi}{2}} \sin x\,dx \\ &= \frac{\pi}{2} - \int_{\frac{\pi}{2}}^{\pi} \sin x\,dx \\ &= \frac{\pi}{2} - 1.\end{aligned}$$

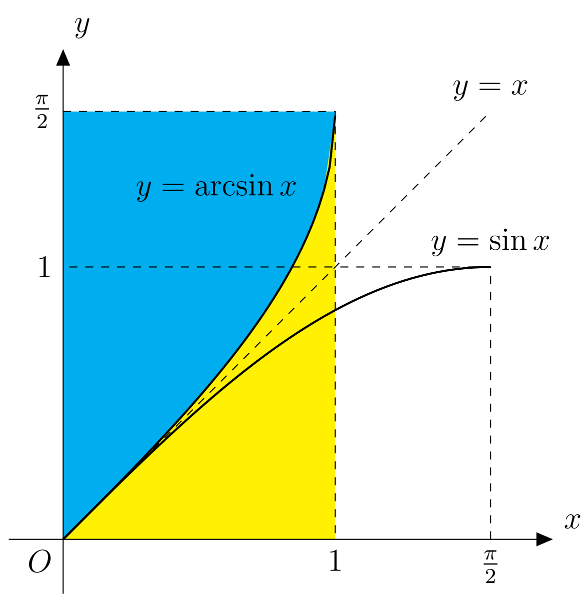

Figure 2.19: Graph of arcsine

■

Definition 2.5.4. Let f be a function. If $F'(x) = f(x)$ for all x, then we say that F is an antiderivative of f.

Example 2.5.5. $F(x) = \frac{1}{3}x^3$ is an antiderivative of $f(x) = x^2$. Note that $G(x) = \frac{1}{3}x^3 + 2$ is also an antiderivative of f.

Remark 2.5.6. In general, if $F(x)$ is an antiderivative of $f(x)$, then so is $F(x)+C$ for any constant C. Moreover, if F and G are both antiderivatives of f, then F and G differ by only a constant by the Mean Value Theorem (apply Problem 2.3.3 to $F(x) - G(x)$). The collection of all antiderivatives of f is called the indefinite integral of f and denoted by

$$\int f(x)\,dx.$$

Therefore, if F is an antiderivative of f, then we have

$$\int f(x)\,dx = F(x) + C,$$

where C is an arbitrary constant.

Theorem 2.5.7 (The Fundamental Theorem of Calculus)**.** *Let f be a continuous function on the interval $[a,b]$ and $f(x) = F'(x)$ (so F is an antiderivative of f). Then*

$$\int_a^b f(x)\,dx = [F(x)]_a^b = F(b) - F(a).$$

Problem 2.5.8. Compute $\displaystyle\int_{-1}^{3} (2x+3)\,dx$ using the Fundamental Theorem of Calculus.

Answer Let $F(x) = x^2 + 3x$, then it is easy to check that $F'(x) = 2x + 3$. It follows that $\displaystyle\int_{-1}^{3} (2x+3)\,dx = F(3) - F(-1) = 18 - (-2) = 20.$ ■

Problem 2.5.9. (see Problem 2.5.3) Find $\displaystyle\int_{\frac{\pi}{2}}^{\pi} \sin x\,dx$.

Answer Since $(-\cos x)' = \sin x$, by the Fundamental Theorem of Calculus,

$$\int_{\frac{\pi}{2}}^{\pi} \sin x\,dx = [-\cos x]_{\frac{\pi}{2}}^{\pi} = -\cos\pi + \cos\frac{\pi}{2} = 1.$$

■

By the Fundamental Theorem of Calculus, finding definite integral of a function f boils down to finding its indefinite integral. From Theorem 2.2.7 and 2.2.8, we have the following theorems.

Theorem 2.5.10. *For $n \neq -1$ and $a \neq -1$, $a > 0$,* *(a)* $\displaystyle\int x^n\,dx = \frac{x^{n+1}}{n+1} + C.$

(b) $\displaystyle\int \frac{1}{x}\,dx = \ln|x| + C.$

(c) $\displaystyle\int e^x\,dx = e^x + C.$

(d) $\displaystyle\int a^x\,dx = \frac{a^x}{\ln a} + C.$

(e) $\displaystyle\int \sin x\,dx = -\cos x + C.$

(f) $\displaystyle\int \cos x\,dx = \sin x + C.$

(g) $\displaystyle\int \sec^2 x\,dx = \tan x + C.$

(h) $\displaystyle\int \frac{1}{1+x^2}\,dx = \arctan x + C.$

Theorem 2.5.11. *Let f and g be integrable functions and let a, b, and c be constants. Then*

(a) $\displaystyle\int_a^c f(x)\,dx = \int_a^b f(x)\,dx + \int_b^c f(x)\,dx.$

(b) $\displaystyle\int_a^b (f(x) \pm g(x))\,dx = \int_a^b f(x)\,dx \pm \int_a^b g(x)\,dx.$

(c) $\displaystyle\int_a^b cf(x)\,dx = c\int_a^b f(x)\,dx.$

Recall that $f : \mathbb{R} \to \mathbb{R}$ is called odd (respectively, even) if $f(-x) = -f(x)$ (respectively, $f(-x) = f(x)$) for all $x \in \mathbb{R}$. In other words, f is an odd (respectively, even) function if f is symmetric with respect to the origin (respectively, the y-axis). Integral of a symmetric function over a symmetric region has the following convenient properties.

Theorem 2.5.12. *Let f be a continuous function (see Figure 2.20).*

(a) If f is odd, then $\displaystyle\int_{-a}^a f(x)\,dx = 0.$

(b) If f is even, then $\displaystyle\int_{-a}^a f(x)\,dx = 2\int_0^a f(x)\,dx.$

Problem 2.5.13. Compute $\displaystyle\int_{-2}^2 \left(\frac{1}{\sqrt{2}}x^{101} - 100x^{97} + 23x^5 + 3x^2 + 4\right) dx.$

Answer Note that the function $\frac{1}{\sqrt{2}}x^{101} - 100x^{97} + 23x^5 + 3x^2 + 4$ can be written as the sum of f and g, where $f(x) = \frac{1}{\sqrt{2}}x^{101} - 100x^{97} + 23x^5$ and $g(x) = 3x^2 + 4$, Since f is odd and g is even, it follows that

$$\int_{-2}^2 \left(\frac{1}{\sqrt{2}}x^{101} - 100x^{97} + 23x^5 + 3x^2 + 4\right) dx = 2\int_0^2 g(x)\,dx = 2[x^3 + 4x]_0^2 = 32.$$

■

Theorem 2.5.11 gives a way to integrate piecewise defined functions.

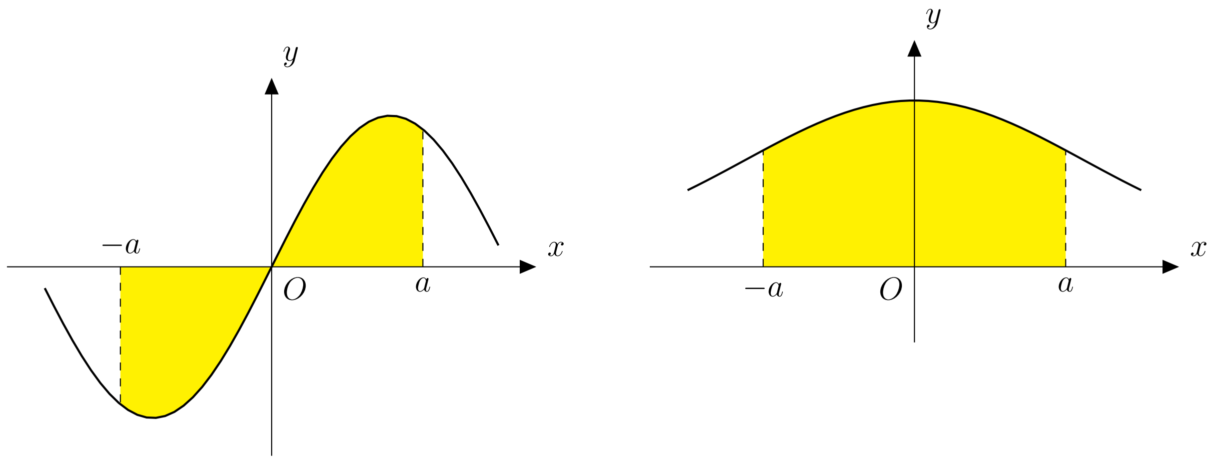

Figure 2.20: Integral of symmetric functions

Example 2.5.14. Let $f : [-2, 5] \to \mathbb{R}$ be defined by

$$f(x) = \begin{cases} 1, & -2 \leq x < 1, \\ \frac{x^2}{2}, & 1 \leq x < 3, \\ 4 - x, & 3 \leq x \leq 5. \end{cases}$$

Consider a function $g : [-2, 5] \to \mathbb{R}$ defined by $g(x) = \displaystyle\int_{-2}^{x} f(t)\,dt$ (see Figure 2.21). It is clear that we need to treat cases $x < 1$, $1 \leq x < 3$, and $x \geq 3$ separately. For $x < 1$, $g(x) = \displaystyle\int_{-2}^{x} dt = x + 2$. If $1 \leq x < 3$, then $g(x) = \displaystyle\int_{-2}^{1} dt + \int_{1}^{x} \frac{t^2}{2}\,dt = \frac{x^3 + 17}{6}$. If $x \geq 3$, then $g(x) = \displaystyle\int_{-2}^{1} dt + \int_{1}^{3} \frac{t^2}{2}\,dt + \int_{3}^{x} (4 - t)\,dt = 4x - \frac{x^2}{2} - \frac{1}{6}$.

Problem 2.5.15. Determine a constant c so that the function $\displaystyle\int_{0}^{2} (c - 2|x - 1|)\,dx = 1$.

Answer Since

$$\int_{0}^{2} (c - 2|x - 1|)\,dx = \int_{0}^{1} (c - 2(1 - x))\,dx + \int_{1}^{2} (c - 2(x - 1))\,dx = 2c - 2,$$

it follows that $c = \frac{3}{2}$. ■

Exercise 2.5.

1. Determine a constant a that minimizes the value of $\displaystyle\int_{a}^{a+1} |x|\,dx$.

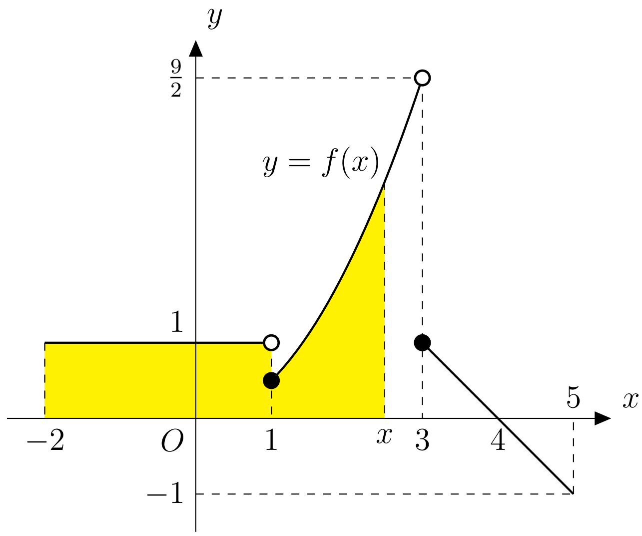

Figure 2.21: A piecewise defined function

2. Let $p, q > 1$ be such that $\frac{1}{p} + \frac{1}{q} = 1$. Prove Young's Inequality:

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q} \tag{2.5.1}$$

for all nonnegative numbers a and b, with equality when $a^p = b^q$.
Hint: Note that $\frac{1}{p-1} = q - 1$ and $\frac{1}{q-1} = p - 1$. Consider the geometric interpretation of both sides of (2.5.1) using Figure 2.22.

3. Determine a constant c so that $\displaystyle\int_0^1 \frac{c}{1+x^2}\,dx = 1$.

4. Compute $\displaystyle\lim_{x\to 0} \frac{1}{e^x - 1} \int_0^x e^{t^2}\,dt$.
Hint: Combine L'Hopital's Theorem and the Fundamental Theorem of Calculus.

5. Suppose $f(x) \geq g(x)$ for all x in $[a, b]$. Show that $\displaystyle\int_a^b f(x)\,dx \geq \int_a^b g(x)\,dx$.
Hint: Consider $\displaystyle\int_a^b h(x)\,dx$, where $h(x) = f(x) - g(x)$.

6. Consider a piecewise defined function

$$f(x) = \begin{cases} x^2 + 2, & x \leq 0, \\ e^{-x}, & x > 0. \end{cases}$$

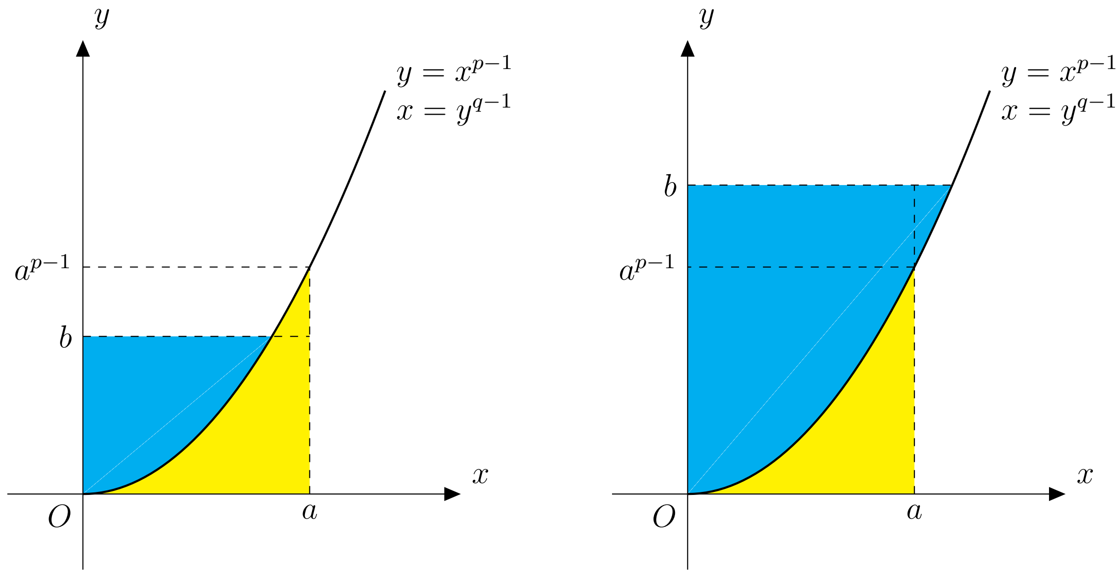

Figure 2.22: Young's inequality

For $y \in \mathbb{R}$, define $g(y) = \displaystyle\int_{-2}^{y} f(x)\,dx$.

(a) Compute $g(1)$.

(b) Is g continuous at 0?

(c) Is g differentiable at 0?

2.6 Techniques of Integration

In this section we discuss several techniques of finding indefinite integrals. We begin with Integration by Substitution. Consider the composition $f(g(x))$ of two differentiable functions f and g. By the Chain Rule, we know that

$$(f(g(x)))' = f'(g(x))g'(x),$$

and this gives

$$\int f'(g(x))g'(x)\,dx = f(g(x)) + C.$$

Example 2.6.1. Consider $\displaystyle\int 2x\cos(x^2)\,dx$. Note that $2x\cos(x^2) = f'(g(x))g'(x)$, where $f(x) = \sin x$ and $g(x) = x^2$. Therefore, $\displaystyle\int 2x\cos(x^2)\,dx = \sin(x^2) + C$.

In practice, it is hard to pick out f and g. Integration by Substitution is a technique with which we can resolve this problem.

Integration by Substitution

- Step 1: Pick a function inside a function and call it u.
- Step 2: Find $\frac{du}{dx}$, and express dx in terms of du, treating $\frac{du}{dx}$ as a fraction.
- Step 3: Replace dx and all x's using du and u's. Now you have an integral of a function of u.
- Step 4: Integrate and replace all u's with functions of x.
- Step 5: Check your answer by differentiating it.

Example 2.6.2. Following the steps described above, we redo Example 2.6.1. One can easily note that x^2 is a function inside another function, say $\cos x$, so we let $u = x^2$. Now $\frac{du}{dx} = 2x$ and $dx = \frac{du}{2x}$, so we get

$$\int 2x\cos(x^2)\,dx = \int 2x\cos u\,\frac{du}{2x} = \int \cos u\,du.$$

The last integral contains only u's and can be easily integrated: $\displaystyle\int \cos u\,du = \sin u + C$. Replacing u in terms of a function of x, we get that $\displaystyle\int 2x\cos(x^2)\,dx = \sin(x^2) + C$.

Example 2.6.3. Consider $\displaystyle\int_2^3 (3x-7)^4\,dx$. Note that $3x-7$ is a function inside another function, so we let $u = 3x-7$ and get $\frac{du}{dx} = 3$, which gives $dx = \frac{du}{3}$. It follows that $\displaystyle\int (3x-7)^4\,dx = \int u^4\,\frac{du}{3} = \frac{u^5}{15} + C = \frac{(3x-7)^5}{15} + C$. In particular, it follows that $\displaystyle\int_2^3 (3x-7)^4\,dx = \left[\frac{(3x-7)^5}{15}\right]_2^3 = \frac{33}{15}$.

Example 2.6.4. Consider $\displaystyle\int xe^{-x^2}\,dx$. Let $u = -x^2$, then $\frac{du}{dx} = -2x$ and it follows that $\displaystyle\int xe^{-x^2}\,dx = \int xe^u\left(-\frac{du}{2x}\right) = -\frac{1}{2}\int e^u\,du = -\frac{1}{2}e^u + C = -\frac{1}{2}e^{-x^2} + C$.

There is also a definite integral version of Integration by Substitution. It is often called the Change of Variables Formula.

Theorem 2.6.5. *Let $\varphi : [a,b] \to \mathbb{R}$ be a differentiable function. Then*

$$\int_{\varphi(a)}^{\varphi(b)} f(u)\,du = \int_a^b f(\varphi(x))\varphi'(x)\,dx.$$

Example 2.6.6. Consider $\displaystyle\int_2^3 (3x-7)^4\,dx$ (see Example 2.6.3). Let $\varphi(x) = 3x-7$, then $\varphi(2) = -1$, $\varphi(3) = 2$, and $\varphi'(x) = 3$. By Theorem 2.6.5, it follows that

$$\int_2^3 (3x-7)^4\,dx = \frac{1}{3}\int_2^3 (3x-7)^4 \cdot 3\,dx = \frac{1}{3}\int_{-1}^2 u^4\,du = \left[\frac{u^5}{15}\right]_{-1}^2 = \frac{33}{15}.$$

We now introduce another technique of integration. By the Product Rule, we get

$$(uv)' = u'v + uv' \quad \text{or} \quad uv' = (uv)' - u'v.$$

Integrating both sides, we get

$$\int uv'\,dx = uv - \int u'v\,dx.$$

Example 2.6.7. Consider $\displaystyle\int xe^x\,dx$. Letting $u = x$ and $v = e^x$, we get that the given integral equals $\displaystyle\int uv'\,dx$, which in turn equals $uv - \displaystyle\int u'v\,dx = xe^x - \int e^x\,dx = xe^x - e^x + C$.

Problem 2.6.8. Integrate $\displaystyle\int x\cos x\,dx$.

Answer Let $u = x$ and $v = \sin x$, then

$$\int x\cos x\,dx = \int uv'\,dx = uv - \int u'v\,dx = x\sin x - \int \sin x\,dx = x\sin x + \cos x + C.$$

■

The procedure used in the previous examples is called the Integration by Parts and can be summarized as in the following:

Integration by Parts

- Step 1: Break up the function into the product of u and v'. The following chart helps in determining functions to be uses as u or v':

u	$\ln x$	$\ldots, x^{-2}, x^{-1}, 1, x, x^2, \ldots$	$\sin x, \ \cos x$	e^x	v'

  In general, the function that is easier to integrate is selected as v'.

- Step 2: Find u' by differentiating u. Find v by integrating v'.
- Step 3: Integrate $u'v$.
- Step 4: Compute $\displaystyle\int uv'\,dx = uv - \int u'v\,dx$.

Problem 2.6.9. Integrate $\displaystyle\int x\ln x\,dx$.

Answer The integrand is the product of two functions, say x and $\ln x$. According to the chart, x is closer to v' and $\ln x$ is closer to u, so we let $u = \ln x$ and $v' = x$, which gives $u' = \frac{1}{x}$ and $v = \frac{x^2}{2}$. It therefore follows that

$$\int x\ln x\,dx = (\ln x)\left(\frac{x^2}{2}\right) - \int\left(\frac{1}{x}\right)\left(\frac{x^2}{2}\right)dx = \frac{x^2}{2}\ln x - \frac{x^2}{4} + C.$$

■

Some integrals require both Integration by Substitution and Integration by Parts.

Example 2.6.10. Consider $\displaystyle\int \frac{x}{e^{3x}}\,dx$. We begin with Integration by Substitution. Letting $w = 3x$, we get that $\frac{dw}{dx} = 3$, or $dx = \frac{dw}{3}$, and the given integral becomes

$$\int \frac{x}{e^{3x}}\,dx = \int \frac{\frac{w}{3}}{e^w}\frac{dw}{3} = \frac{1}{9}\int we^{-w}\,dw.$$

Now using Integration by Parts with $u = w$ and $v' = e^{-w}$, we obtain

$$\int we^{-w}\,dw = -we^{-w} + \int e^{-w}\,dw = -we^{-w} - e^{-w} + C.$$

Finally, it follows that

$$\int \frac{x}{e^{3x}}\,dx = -\frac{1}{9}we^{-w} - \frac{1}{9}e^{-w} + C = -\frac{1}{9}(3x+1)e^{-3x} + C.$$

Problem 2.6.11. Let $\alpha > 0$. Show that $\displaystyle\int x^\alpha e^{-x}\,dx = -x^\alpha e^{-x} + \alpha \int x^{\alpha-1}e^{-x}\,dx$.

Answer Let $u = x^\alpha$ and $v' = e^{-x}$, then $u' = \alpha x^{\alpha-1}$ and $v = -e^{-x}$, so it follows that

$$\int x^\alpha e^{-x}\,dx = -x^\alpha e^{-x} + \alpha \int x^{\alpha-1}e^{-x}\,dx.$$

■

There is also a definite integral version of Integration by Parts. It is straightforward to check

$$\int_a^b uv'\,dx = [uv]_a^b - \int_a^b u'v\,dx = u(b)v(b) - u(a)v(a) - \int_a^b u'v\,dx.$$

We end this section with some special shortcuts in Integration by Substitution.

Theorem 2.6.12. *We have the following integrals:*

(a) $\displaystyle\int \frac{f'(x)}{f(x)}\,dx = \ln|f(x)| + C.$

(b) If $F'(x) = f(x)$, *then* $\displaystyle\int f(ax+b)\,dx = \frac{1}{a}F(ax+b) + C.$

Example 2.6.13. Consider $\displaystyle\int \frac{2x}{x^2+1}\,dx$. Since $(x^2+1)' = 2x$, it follows that

$$\int \frac{2x}{x^2+1}\,dx = \ln|x^2+1| + C = \ln(x^2+1) + C.$$

Example 2.6.14. Recall that the function inside a function in Example 2.6.3 was $3x-7$, a linear function. Therefore, $\displaystyle\int (3x-7)^4\,dx = \frac{1}{3}(3x-7)^5 + C.$

Exercise 2.6.

1. Integrate:

(a) $\displaystyle\int x^2 e^{x^3+4}\,dx.$

(b) $\displaystyle\int x\sqrt{3-x^2}\,dx.$

(c) $\displaystyle\int \frac{1}{x\ln x}\,dx.$

2. Integrate:

(a) $\displaystyle\int x^2 \ln x \, dx$

(b) $\displaystyle\int \ln x \, dx$.
Hint: $\ln x$ can be viewed as $1 \cdot \ln x$.

3. Integrate:

(a) $\displaystyle\int \frac{1}{x \ln x \ln(\ln x)} \, dx$

(b) $\displaystyle\int (5x+11)^{\frac{1}{3}} \, dx$.

4. For $x \in \mathbb{R}$, define f by $f(x) = \displaystyle\int_0^x \ln(t^2+1) \, dt$.

(a) Compute $f(1)$.

(b) Compute $\displaystyle\lim_{x\to 0} \frac{f(x)}{x^3}$.

2.7 Improper Integrals

The definition of integrals extends to infinite intervals and they are called improper integrals. There are various types of improper integrals, but we focus on one type, say, the improper integral of the form $\displaystyle\int_a^\infty f(x)\,dx$.

Definition 2.7.1. Suppose that f is a function defined on $[a,\infty)$. If $\displaystyle\lim_{b\to\infty} \int_a^b f(x)\,dx$ exists, then we say that $\displaystyle\int_a^\infty f(x)\,dx$ is convergent and write

$$\int_a^\infty f(x)\,dx = \lim_{b\to\infty} \int_a^b f(x)\,dx.$$

If the limit does not exist, then we say that the integral is divergent.

Remark 2.7.2. $\displaystyle\int_{-\infty}^a f(x)\,dx$ is defined in a similar way: if $\displaystyle\lim_{b\to-\infty} \int_b^a f(x)\,dx$ exists, we say that $\displaystyle\int_{-\infty}^a f(x)\,dx$ is convergent and write

$$\int_{-\infty}^a f(x)\,dx = \lim_{b\to-\infty} \int_b^a f(x)\,dx.$$

Otherwise, $\int_{-\infty}^{a} f(x)\,dx$ is divergent. If both $\int_{-\infty}^{a} f(x)\,dx$ and $\int_{a}^{\infty} f(x)\,dx$ are convergent for some a, then we say that $\int_{-\infty}^{\infty} f(x)\,dx$ is convergent and write

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{a} f(x)\,dx + \int_{a}^{\infty} f(x)\,dx.$$

We also write $\int_{\mathbb{R}} f(x)\,dx$ for $\int_{-\infty}^{\infty} f(x)\,dx$.

Example 2.7.3. Consider a binary classfication in which prediction is based on *score*, a continuous random variable X. To be more precise, for a fixed quantity T, let an instance be classified as *positive* if $X > T$, and *negative* if $X \leq T$. Suppose that X follows a pdf f_+ (respectively, f_-) if the instance actually belongs to *positive* (respectively, *negative*). Then the sensitivity $sens(T)$ of the test is given by

$$sens(T) = P(X > T|+) = \int_{T}^{\infty} f_+(x)\,dx.$$

On the other hand, the specificity $spec(T)$ of the test is given by

$$spec(T) = P(X \leq T|-) = \int_{-\infty}^{T} f_-(x)\,dx.$$

As T decreases from ∞ to $-\infty$, $sens(T)$ increases from 0 to 1 and $spec(T)$ decreases from 1 to 0. The parametric curve given by coordinates $(1 - spec(T), sens(T))$ is called the receiver operating characteristic (ROC) curve. Typical ROC curves are given in Figure 2.23. Note that the ROC curves reflect a trade off between sensitivity and specificity. The area under the ROC curve has a special meaning. See Example 4.4.9.

Example 2.7.4. Consider $\int_{1}^{\infty} \frac{1}{x^2}\,dx$. Since $\int_{1}^{b} \frac{1}{x^2}\,dx = 1 - \frac{1}{b} \to 0$ as $b \to \infty$, the given improper integral exists and $\int_{1}^{\infty} \frac{1}{x^2}\,dx = 1$. However, since $\int_{2}^{b} \frac{1}{x}\,dx = \ln b - \ln 2 \to \infty$ as $b \to \infty$, $\int_{2}^{\infty} \frac{1}{x}\,dx$ is divergent.

Problem 2.7.5. Compute $\int_{0}^{\infty} xe^{-x}\,dx$, if it is convergent.

Answer Using Integration by Parts, one can easily see that $\int xe^{-x}\,dx = -(x+1)e^{-x} + C$. Therefore, it follows that $\int_{0}^{b} xe^{-x}\,dx = \left[-(x+1)e^{-x}\right]_{0}^{b} = -(b+1)e^{-b} + 1$. Finally,

$$\int_{0}^{\infty} xe^{-x}\,dx = \lim_{b\to\infty} (-(b+1)e^{-b} + 1) = 1$$

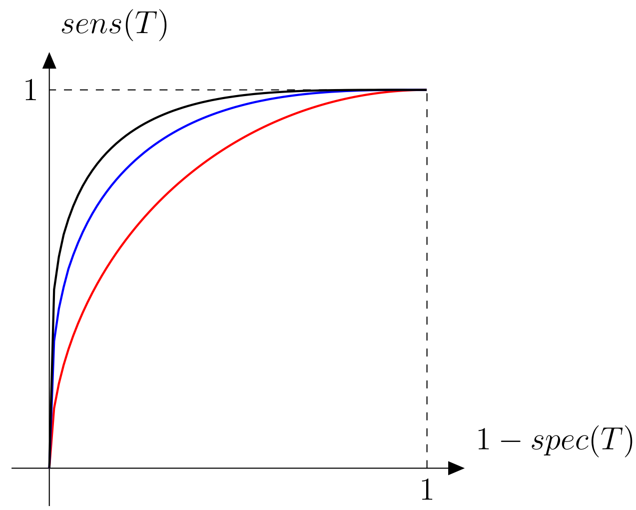

Figure 2.23: Typical ROC curves. For any *cut off* value T, the test represented by black curve performs better than blue, and blue performs better than red.

by L'Hopital's Theorem. ■

Problem 2.7.6. For a fixed positive constant λ, define

$$f(x) = \begin{cases} \frac{\lambda}{2}e^{-\lambda x}, & x \geq 0, \\ \frac{\lambda}{2}e^{\lambda x}, & x < 0. \end{cases}$$

(a) Show that $\displaystyle\int_{-\infty}^{\infty} f(x)\,dx = 1$.

(b) Compute $\displaystyle\int_{-\infty}^{x} f(t)\,dt$.

Answer

(a)

$$\begin{aligned} \int_{-\infty}^{\infty} f(x)\,dx &= \frac{1}{2}\int_{-\infty}^{0} \lambda e^{\lambda x}\,dx + \frac{1}{2}\int_{0}^{\infty} \lambda e^{-\lambda x}\,dx \\ &= \frac{1}{2}\left[e^{\lambda x}\right]_{-\infty}^{0} - \frac{1}{2}\left[e^{-\lambda x}\right]_{0}^{\infty} \\ &= 1. \end{aligned}$$

(b) Let $F(x) = \displaystyle\int_{-\infty}^{x} f(t)\,dt$. For $x < 0$,

$$F(x) = \int_{-\infty}^{x} \frac{1}{2}\lambda e^{\lambda t}\,dt = \frac{1}{2}e^{\lambda x}.$$

For $x \geq 0$,

$$F(x) = \int_{-\infty}^{0} \frac{1}{2}\lambda e^{\lambda t}\,dt + \int_{0}^{x} \frac{1}{2}\lambda e^{-\lambda t}\,dt = \frac{1}{2} + \left(-\frac{1}{2}e^{-\lambda x} + \frac{1}{2}\right) = 1 - \frac{1}{2}e^{-\lambda x}.$$

■

Problem 2.7.7. Suppose that f is an even function such that $\int_{-\infty}^{\infty} f(x)\,dx = 1$.

(a) Compute $\int_{0}^{\infty} f(x)\,dx$.

(b) If $\int_{-\infty}^{2} f(x)\,dx = 0.8$, then what is $\int_{-2}^{0} f(x)\,dx$?

Answer

(a) Since f is even, $\int_{0}^{\infty} f(x)\,dx = \int_{-\infty}^{0} f(x)\,dx$. From

$$1 = \int_{-\infty}^{\infty} f(x)\,dx = \int_{0}^{\infty} f(x)\,dx + \int_{-\infty}^{0} f(x)\,dx,$$

it follows that $\int_{0}^{\infty} f(x)\,dx = \int_{-\infty}^{0} f(x)\,dx = 0.5$.

(b) Note that $\int_{0}^{2} f(x)\,dx = \int_{-\infty}^{2} f(x)\,dx - \int_{-\infty}^{0} f(x)\,dx = 0.8 - 0.5 = 0.3$. Since f is even, we get $\int_{-2}^{0} f(x)\,dx = \int_{0}^{2} f(x)\,dx = 0.3$. See Figure 2.24.

■

Example 2.7.8. For $\alpha > 0$, we define the gamma function Γ by $\Gamma(\alpha) = \int_{0}^{\infty} x^{\alpha-1}e^{-x}\,dx$. By Problem 2.6.11,

$$\begin{aligned}
\Gamma(\alpha+1) &= \lim_{b\to\infty} \int_{0}^{b} x^{\alpha}e^{-x}\,dx \\
&= \lim_{b\to\infty} \left(\left[-x^{\alpha}e^{-x}\right]_0^b + \alpha \int_{0}^{b} x^{\alpha-1}e^{-x}\,dx \right) \\
&= -\lim_{b\to\infty} \frac{b^{\alpha}}{e^{b}} + \alpha \lim_{b\to\infty} \int_{0}^{b} x^{\alpha-1}e^{-x}\,dx.
\end{aligned}$$

Here $\lim_{b\to\infty} \frac{b^{\alpha}}{e^{b}} = 0$ by L'Hopital's Theorem, and $\lim_{b\to\infty} \int_{0}^{b} x^{\alpha-1}e^{-x}\,dx = \Gamma(\alpha)$, so it follows that

$$\Gamma(\alpha+1) = \alpha\Gamma(\alpha). \tag{2.7.1}$$

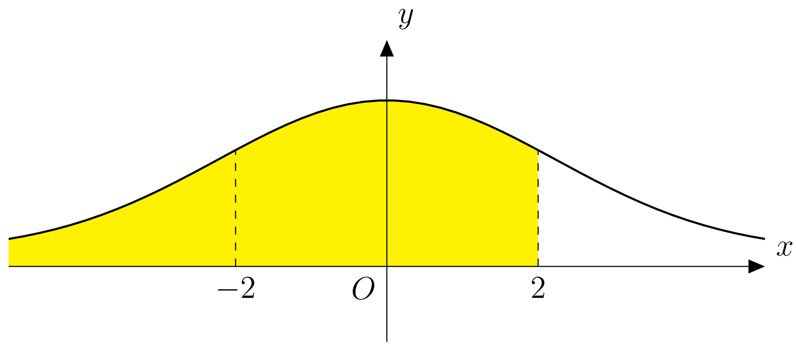

Figure 2.24: Improper integral of an even function

One can easily show that $\Gamma(1) = \int_0^\infty e^{-x}\,dx = 1$, so by (2.7.1) and mathematical induction, we obtain

$$\Gamma(n) = (n-1)!$$

for a positive integer n.

Example 2.7.9. A continuous random variable X is said to follow the Gamma Distribution with the shape parameter α and the scale parameter β, denoted $X \sim Gamma(\alpha, \beta)$, if its pdf is given by

$$f(x|\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}}, \quad 0 < x < \infty, \quad \alpha > 0, \quad \beta > 0.$$

We show that f is indeed a density, that is, we show that $\int_0^\infty \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}}\,dx = 1$. Starting with Integration by Substitution $y = \frac{x}{\beta}$, we get $\int_0^b x^{\alpha-1}e^{-\frac{x}{\beta}}\,dx = \int_0^{\frac{b}{\beta}} \beta^\alpha y^{\alpha-1} e^{-y}\,dy$, so it follows that

$$\int_0^\infty \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}}\,dx = \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty \beta^\alpha y^{\alpha-1} e^{-y}\,dy = \frac{1}{\Gamma(\alpha)\beta^\alpha} \cdot \beta^\alpha \Gamma(\alpha) = 1.$$

Example 2.7.10. The beta function B is defined for $\alpha > 0$, $\beta > 0$ by

$$B(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,dx.$$

It is easy to show that B is symmetric: $B(\alpha,\beta) = B(\beta,\alpha)$. In fact, it can be shown that

$$B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

A continuous random variable X is said to follow Beta Distribution, denoted $X \sim Beta(\alpha, \beta)$, if its pdf is given by

$$f(x|\alpha,\beta) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad 0 < x < 1, \quad \alpha > 0, \quad \beta > 0.$$

Example 2.7.11. A continuous random variable X is said to follow Exponential Distribution with parameter λ, denoted $X \sim Exp(\lambda)$, if its pdf is given by

$$f(x|\lambda) = \frac{1}{\lambda} e^{-\frac{x}{\lambda}}, \quad x > 0.$$

It can be viewed as a special case of the Gamma Distribution. To be precise, $Exp(\lambda)$ distribution is the same as $Gamma(1, \lambda)$ distribution. It therefore follows that $E(X) = \lambda$ and $Var(X) = \lambda^2$. See Exercise 2.7.5.

Example 2.7.12. A continuous random variable X is said to follow Chi-squared Distribution with k degrees of freedom, denoted $X \sim \chi^2(k)$, if its pdf is given by

$$f(x|k) = \frac{x^{\frac{k}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{k}{2}} \Gamma(\frac{k}{2})}, \quad x > 0.$$

Note that $\chi^2(k)$ distribution is the same as $\Gamma(\frac{k}{2}, 2)$.

Example 2.7.13. A continuous random variable X is said to follow Normal Distribution with mean μ and variance σ^2, $\sigma > 0$, denoted $X \sim N(\mu, \sigma^2)$, if its pdf is given by

$$f(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty.$$

One can show that this is indeed a pdf (see Exercise 4.4.3). In particular, if $\mu = 0$ and $\sigma = 1$, then X is said to follow the Standard Normal Distribution.

Problem 2.7.14. Suppose that $Y = \frac{1}{X}$, where $X \sim Gamma(\alpha, \beta)$ with $\alpha > 1$ and $\beta > 0$. Compute $E(Y)$.

Answer We need to compute $\int_0^\infty \frac{1}{x} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}}\, dx = \int_0^\infty \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-2} e^{-\frac{x}{\beta}}\, dx$. Note that the latter integrand is a constant multiple of a $Gamma(\alpha - 1, \beta)$ density, so it is natural to transform $\frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-2} e^{-\frac{x}{\beta}}$ into $\frac{1}{\beta(\alpha-1)} \cdot \frac{1}{\Gamma(\alpha-1)\beta^{\alpha-1}} x^{\alpha-2} e^{-\frac{x}{\beta}}$ and it follows that

$$\int_0^\infty \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-2} e^{-\frac{x}{\beta}}\, dx = \frac{1}{\beta(\alpha-1)} \int_0^\infty \underbrace{\frac{1}{\Gamma(\alpha-1)\beta^{\alpha-1}} x^{\alpha-2} e^{-\frac{x}{\beta}}}_{\text{pdf of } Gamma(\alpha-1,\beta)}\, dx = \frac{1}{\beta(\alpha-1)}.$$

■

Exercise 2.7.

1. A continuous random variable X is said to follow the Cauchy Distribution if its pdf is given by $f(x) = \frac{1}{\pi(1+x^2)}$, $-\infty < x < \infty$. Show that f is indeed a pdf, that is, show that $\displaystyle\int_{-\infty}^{\infty} \frac{dx}{\pi(1+x^2)} = 1$.

2. Determine the value of a constant c so that

$$\int_0^{\infty} \frac{c\ln(2x+1)}{4x^2+4x+1}\,dx = 1.$$

3. Compute $\displaystyle\int_0^{\infty} xe^{-x^2}\,dx$ and $\displaystyle\int_{-\infty}^{\infty} xe^{-x^2}\,dx$.
Hint: See Example 2.6.4.

4. Compute $\displaystyle\int_0^{\infty} \frac{e^x}{(1+e^x)^2}\,dx$.

5. Suppose that X follows $Gamma(\alpha, \beta)$ distribution. Show that $E(X) = \alpha\beta$ and $Var(X) = \alpha\beta^2$.

6. Let $M(t) = \displaystyle\int_0^{\infty} \frac{1}{4} xe^{(t-\frac{1}{2})x}\,dx$. Show that $M(t)$ converges if and only if $t < \frac{1}{2}$ and when $t < \frac{1}{2}$, show that $M(t) = \frac{1}{(2t-1)^2}$.

7. Suppose that X follows $B(\alpha, \beta)$ distribution. Show that

$$E(X) = \frac{\alpha}{\alpha+\beta} \quad \text{and} \quad Var(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

2.8 Convergence of Sequences and Series

In probability, there are several modes of convergence that describe the limit behavior of a sequence of random variables, and some of them can be defined in terms of the limit of a sequence of real numbers. For example, a sequence (X_n) of random variables is said to converge to X in probability if for any $\epsilon > 0$,

$$\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0.$$

(X_n) is said to converge to X in distribution or in law if

$$\lim_{n\to\infty} F_n(x) = F(x)$$

for all x at which F is continuous, where F_n and F are cdf's of X_n and X, respectively. The convergence of a sequence is defined in a way similar to Definition 2.1.9 and Remark 2.1.10.

Definition 2.8.1. Let (a_n) be a sequence of real numbers. We say that the limit of (a_n) is ℓ if for every $\epsilon > 0$, there is a corresponding N such that $n > N$ implies $|a_n - \ell| < \epsilon$. In this case, we write $\lim_{n\to\infty} a_n = \ell$ or $a_n \to \ell$. We write $\lim_{n\to\infty} a_n = \infty$ (respectively, $\lim_{n\to\infty} a_n = -\infty$) or $a_n \to \infty$ (respectively, $a_n \to -\infty$) if for every M, there exists N such that $n > N$ implies $a_n > M$ (respectively, $a_n < M$).

Remark 2.8.2. The same properties as in Theorem 2.1.12 hold for limit of a sequence.

Example 2.8.3. We claim that $\frac{1}{n} \to 0$. Let $\epsilon > 0$, then there is N such that $\frac{1}{N} < \epsilon$. It then follows that for any $n > N$,

$$\left|\frac{1}{n} - 0\right| = \frac{1}{n} < \frac{1}{N} < \epsilon.$$

Example 2.8.4. Let $x \in \mathbb{R}$, then there exists a sequence (q_n) such that $q_n \in \mathbb{Q}$ and $\lim_{n\to\infty} q_n = x$.

Example 2.8.5. Some more examples of limits are in order.

(a) $\lim_{n\to\infty} \frac{n+1}{2n} = \frac{1}{2}$.

(b) $\lim_{n\to\infty} \frac{5n^2 - 100}{2n^2 - 3n + 5} = \frac{5}{2}$.

(c) $\lim_{n\to\infty} n^2 = \infty$.

(d) $\lim_{n\to\infty} \frac{(-1)^n}{n} = 0$.

(e) $\lim_{n\to\infty} \frac{\sin n}{n} = 0$.

(f) $\lim_{n\to\infty} (1 + (-1)^n)$ does not exist.

(g) $\lim_{n\to\infty} \left(\frac{1}{4}\right)^n = 0$.

(h) $\lim_{n\to\infty} \left(-\frac{1}{2}\right)^n = 0$.

(i) $\lim_{n\to\infty} \left(\frac{4}{3}\right)^n = \infty$.

(j) $\lim_{n\to\infty} (-2)^n$ does not exist.

Remark 2.8.6. In summary we get

$$\lim_{n\to\infty} \frac{a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0}{b_n x^n + b_{n-1}x^{n-1} + \cdots + b_1 x + b_0} = \frac{a_n}{b_n} \quad \text{(provided } b_n \neq 0\text{)}$$

and

$$\lim_{n\to\infty} r^n = \begin{cases} \infty & \text{if } r > 1 \\ 1 & \text{if } r = 1 \\ 0 & \text{if } -1 < r < 1 \\ \text{does not exist} & \text{if } r \leq -1 \end{cases}$$

Problem 2.8.7. Examine the limit of each of the following sequences:

(a) $a_n = \frac{1+3^n}{2-5^n}$

(b) $b_n = \frac{2+4^n}{3^n}$

(c) $c_n = \frac{6^n+3^{2n}}{3^n+5\cdot 9^n}$

(d) $d_n = \frac{2^n+4^n}{3^n+2^n}$

Answer

(a) $$\lim_{n\to\infty} \frac{1+3^n}{2-5^n} = \lim_{n\to\infty} \frac{\frac{1}{5^n}+\left(\frac{3}{5}\right)^n}{\frac{2}{5^n}-1} = \frac{0+0}{0-1} = 0.$$

(b) $$\lim_{n\to\infty} \frac{2+4^n}{3^n} = \lim_{n\to\infty} \frac{\frac{2}{3^n}+\left(\frac{4}{3}\right)^n}{1} = \infty.$$

(c) $$\lim_{n\to\infty} \frac{6^n+3^{2n}}{3^n+5\cdot 9^n} = \lim_{n\to\infty} \frac{\left(\frac{6}{9}\right)^n+1}{\left(\frac{3}{9}\right)^n+5} = \frac{1}{5}.$$

(d) $$\lim_{n\to\infty} \frac{2^n+4^n}{3^n+2^n} = \lim_{n\to\infty} \frac{\left(\frac{2}{3}\right)^n+\left(\frac{4}{3}\right)^n}{1+\left(\frac{2}{3}\right)^n} = \infty.$$

■

Let (a_n) be a sequence of real numbers. If $a_n \leq a_{n+1}$ (respectively, $a_n \geq a_{n+1}$) for all $n \geq 1$, we say that (a_n) is nondecreasing (respectively, nonincreasing). Next theorem gives a condition on a sequence that guarantees the existence of the limit.

Theorem 2.8.8. *Suppose that (a_n) is a nondecreasing sequence that is bounded from above, that is, there is M such that $a_n \leq M$ for all $n \geq 1$. Then $\lim_{n\to\infty} a_n$ exists. Similarly, if (a_n) is a nonincreasing sequence that is bounded from below, that is, there is M such that $a_n \geq M$ for all $n \geq 1$, then $\lim_{n\to\infty} a_n$ exists.*

Sequences can be used to check continuity of a function.

Theorem 2.8.9. *Let f be a function and $x_0 \in \mathbb{R}$. Then f is continuous at x_0 if and only if $\lim_{n\to\infty} f(x_n) = f(x_0)$ for all sequences (x_n) such that $\lim_{n\to\infty} x_n = x_0$.*

Problem 2.8.10. Suppose that $f : \mathbb{R} \to \mathbb{R}$ satisfies $f(a+b) = f(a) + f(b)$ for all real numbers a and b. Suppose that f is continuous at $x = 0$. Show that $f(x) = \lambda x$ for some $\lambda \in \mathbb{R}$.

Answer By Exercise 2.1.2, f is continuous *everywhere*. Let $\lambda = f(1)$. Using Mathematical Induction, one can easily show that $f(nx) = nf(x)$ for all positive integer n and for all $x \in \mathbb{R}$. Using this and the fact that $f(0) = 0$, one can also get that $f(nx) = nf(x)$ for *all* $n \in \mathbb{Z}$. Since $f(1) = f(n \cdot \frac{1}{n})$, we obtain that $f(\frac{1}{n}) = \frac{\lambda}{n}$, $n \neq 0$, and that $f(\frac{m}{n}) = \frac{m}{n}\lambda$, $m, n \in \mathbb{Z}$, $n \neq 0$. This shows that $f(q) = \lambda q$ for all $q \in \mathbb{Q}$. Finally, let $x \in \mathbb{R}$. By Example

2.8.4, there is a sequence $(q_n) \subseteq \mathbb{Q}$ such that $\lim\limits_{n\to\infty} q_n = x$ and by Theorem 2.8.9, it follows that

$$f(x) = \lim_{n\to\infty} f(q_n) = \lim_{n\to\infty} \lambda q_n = \lambda x.$$

■

Let X be a continuous random variable. If

$$P(X > s + t | X > s) = P(X > t) \tag{2.8.1}$$

for all $s, t \geq 0$, then we say that X has the memoryless property. The argument in the answer to Problem 2.8.10 can be used to show that the only class of random variables that has the memoryless property is the exponential distribution (see Example 2.7.11). To see this, let F denote the cdf of X, then (2.8.1) implies that

$$\frac{1 - F(s+t)}{1 - F(s)} = 1 - F(t)$$

for all $s, t \geq 0$. In particular, $1 - F(0) = 1$. Define $G : [0, \infty) \to \mathbb{R}$ by $G(x) = 1 - F(x)$, then $G(0) = 1$ and

$$G(s + t) = G(s)G(t) \tag{2.8.2}$$

for all $s, t \geq 0$. For any $x \geq 0$, substitution $s = t = \frac{x}{2}$ in (2.8.2) gives $G(x) = (G(\frac{x}{2}))^2 \geq 0$. If there exists $x_0 \geq 0$ such that $G(x_0) = 0$ (necessarily $x_0 \neq 0$), then by (2.8.2), we would have $G(\frac{x_0}{2^n}) = 0$ for $n = 1, 2, \ldots$ Since G is right-continuous at 0, it would imply that (a slight modification of Theorem 2.8.9)

$$G(0) = \lim_{n\to\infty} G\left(\frac{x_0}{2^n}\right) = 0,$$

and this contradiction shows that $G(x) > 0$ for all $x \geq 0$. Consequently $H : [0, \infty) \to \mathbb{R}$, $H(x) = \ln G(x)$ is well-defined. Note that $H(s + t) = H(s) + H(t)$ for all $s, t \geq 0$. Since H is right-continuous at 0, a slight modification of the answer to Problem 2.8.10 gives $H(x) = \mu x$ for some μ. In other words, $F(x) = 1 - G(x) = 1 - e^{H(x)} = 1 - e^{-\frac{1}{\lambda}x}$, where $\lambda = -\frac{1}{\mu}$.

Next, we consider the sum of each term of a given sequence.

Definition 2.8.11. Let (a_n) be a sequence. An infinite series $\sum\limits_{n=1}^{\infty} a_n$ is defined to be

$$\sum_{n=1}^{\infty} a_n = \lim_{k\to\infty} S_k,$$

where $S_k = \sum_{n=1}^{k} a_n = a_1 + a_2 + \cdots + a_{k-1} + a_k$, the k^{th} partial sum. In other words,

$$\sum_{n=1}^{\infty} a_n = \lim_{k\to\infty} S_k = \lim_{k\to\infty} \sum_{n=1}^{k} a_n.$$

Example 2.8.12. It is well known that the infinite series $\sum_{n=1}^{\infty} \frac{1}{n^p}$ converges if and only if $p > 1$. In particular, the harmonic series $\sum_{n=1}^{\infty} \frac{1}{n}$ diverges.

The next theorem is an infinite series version of Theorem 1.5.1.

Theorem 2.8.13. *A geometric series with the initial term a and the common ratio r*

$$\sum_{n=1}^{\infty} ar^{n-1} = a + ar + ar^2 + ar^3 + \cdots$$

converges if and only if $-1 < r < 1$. When $-1 < r < 1$,

$$\sum_{n=1}^{\infty} ar^{n-1} = \frac{a}{1-r}.$$

Example 2.8.14. The geometric series $1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots$ converges and equals $\frac{1}{1-\frac{1}{2}} = 2$.

Problem 2.8.15. Let $0 < p < 1$. Prove that the infinite sum

$$\sum_{n=1}^{\infty} p(1-p)^{n-1} = p + p(1-p) + p(1-p)^2 + p(1-p)^3 + \cdots$$

converges. What is the sum?

Answer The given sum is a geometric series with the initial term p and the common ratio $1 - p$. Since $0 < 1 - p < 1$, the infinite sum converges. In fact, by Theorem 2.8.13, we get that

$$p + p(1-p) + p(1-p)^2 + p(1-p)^3 + \cdots = \frac{p}{1-(1-p)} = 1.$$

■

Remark 2.8.16. Problem 2.8.15 shows that $P(X = n) = p(1-p)^{n-1}$, $n = 1, 2, \ldots$ defines a discrete random variable X. Such a random variable is said to have a Geometric Distribution with parameter p.

Exercise 2.8.

1. Examine the limit of each of the following sequences:

(a) $a_n = \frac{8n^2}{9n^2-5n+1}$

(b) $b_n = \frac{\ln n}{\sqrt{n}}$

(c) $c_n = \frac{\sin n+(-1)^n}{n^3}$

(d) $d_n = \frac{(-2)^n+5^n}{6^n+(-1)^n}$

2. The purpose of this exercise is to prove that $\lim\limits_{n\to\infty}\left(1+\dfrac{1}{n}\right)^n$ exists. Let $s_n = \left(1+\dfrac{1}{n}\right)^n$.

(a) For a fixed integer $n > 0$, let $f(x) = (n+1)x^n - nx^{n+1}$.

i. Compute $f(1)$.

ii. Show that $f'(x) < 0$ for $x > 1$.

iii. Using i and ii, explain why the inequality

$$x^n(n+1-nx) < 1 \tag{2.8.3}$$

holds for all $x > 1$.

(b) Let $x = \dfrac{1+\frac{1}{n}}{1+\frac{1}{n+1}}$.

i. Show that $x > 1$.

ii. Substitute x into (2.8.3) above and show that

$$\left(\frac{1+\frac{1}{n}}{1+\frac{1}{n+1}}\right)^n\left(\frac{n+1}{n+2}\right) < 1. \tag{2.8.4}$$

iii. Using (2.8.4), show that $s_n < s_{n+1}$ for all $n \in \mathbb{N}$. This proves that (s_n) is a nondecreasing sequence.

(c) Let $x = 1 + \dfrac{1}{2n}$.

i. Substitute x into (2.8.3) to show that

$$\left(1+\frac{1}{2n}\right)^n < 2. \tag{2.8.5}$$

ii. Use (2.8.5) to show that $s_{2n} < 4$ for all $n \in \mathbb{N}$ $n > 0$.

iii. Show that (s_n) is bounded from above.

(d) Use Theorem 2.8.8 show that s_n converges. The limit, denoted by e, is called Euler's Number and is the base of natural logarithm.

3. When does the infinite series $1 + 2x + 4x^2 + 8x^3 + 16x^4 + \cdots$ converge? Simplify the sum.

4. Compute $\dfrac{1}{2!} + \dfrac{2}{3!} + \dfrac{3}{4!} + \cdots = \displaystyle\sum_{k=1}^{\infty} \frac{k}{(k+1)!}$.

Hint: Note that $\dfrac{k}{(k+1)!} = \dfrac{k+1-1}{(k+1)!} = \dfrac{1}{k!} - \dfrac{1}{(k+1)!}$, so $\displaystyle\sum_{k=1}^{n} \frac{k}{(k+1)!} = \frac{1}{1!} - \frac{1}{(n+1)!}$.

2.9 Power Series

Just as 2 can be written as an infinite sum, say $2 = 1 + \frac{1}{2} + \frac{1}{4} + \cdots$, some functions can be described in terms of infinite series. For example, some periodic functions can be written as its Fourier expansion, an infinite series involving trigonometric functions. In this section, we consider power series as a way of describing a certain class of functions.

Definition 2.9.1. A power series about $x = c$ is a series of the form

$$\sum_{n=0}^{\infty} a_n(x-c)^n = a_0 + a_1(x-c) + a_2(x-c)^2 + \cdots + a_n(x-c)^n + \cdots .$$

Here c is called the center of the power series.

Example 2.9.2. $2 + (x-1) + \frac{(x-1)^2}{2} + \frac{(x-1)^3}{3} + \frac{(x-1)^4}{4} + \cdots$ is a power series about $x = 1$. Here $a_0 = 2$ and $a_n = \frac{1}{n}$ for $n \geq 1$.

Example 2.9.3. $1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24} + \frac{x^5}{120} + \cdots$ is a power series about $x = 0$. Here $a_n = \frac{1}{n!}$ for all n.

Remark 2.9.4. A power series can be regarded as a polynomial in $(x-c)$ with an *infinite* degree.

A power series is an infinite series with a variable x in it, so its convergence depends on the value of x. Our first interest is in determining the values of x for which a given power series converges.

Example 2.9.5. The power series $\displaystyle\sum_{n=0}^{\infty} \frac{x^n}{2^n}$ converges when $x = -1$ because it is then a geometric series with the common ratio $-\frac{1}{2}$. However, if $x = 3$, the series diverges because the common ratio $\frac{3}{2}$ is bigger than 1.

Problem 2.9.6. Find all x such that the power series $\displaystyle\sum_{n=0}^{\infty} \frac{x^n}{2^n}$ converges.

Answer The first few terms of the series are $1 + \frac{x}{2} + (\frac{x}{2})^2 + (\frac{x}{2})^3 + \cdots$, so the given sum is a geometric series with the initial term 1 and the common ratio $\frac{x}{2}$. By Theorem 2.8.13, it converges if and only if $-2 < x < 2$, and when $-2 < x < 2$, the series converges to $\frac{1}{1-\frac{x}{2}} = \frac{2}{2-x}$. ■

It is easy to observe that a power series about $x = c$ converges for at least one x, say $x = c$. In fact, the following is known:

Theorem 2.9.7. *Let a power series* $\displaystyle\sum_{n=0}^{\infty} a_n(x-c)^n$ *be given. Then exactly one of the following is true for the power series:*

(a) *There is a positive number R, called the* <u>*radius of convergence*</u>*, such that the series converges for $|x-c|<R$ and diverges for $|x-c|>R$.*

(b) *The series converges only for $x=c$. In this case we say that the radius of convergence R equals 0.*

(c) *The series converges for all values of x. In this case we say that the radius of convergence R equals ∞.*

Example 2.9.8. The radius of convergence for the power series in Problem 2.9.6 is 2.

Remark 2.9.9. Let R be the radius of convergence of a power series $\sum_{n=0}^{\infty} a_n(x-c)^n$. By definition, $\sum_{n=0}^{\infty} a_n(x-c)^n$ converges for all $x \in (c-R, c+R)$ and diverges for all $x \in (-\infty, c-R) \cup (c+R, \infty)$. The convergence at the boundary points $\{c-R, c+R\}$ depends on the power series. See Figure 2.25.

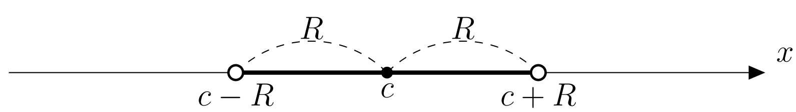

Figure 2.25: Radius of convergence

There is an easy method for computing the radius of convergence.

Theorem 2.9.10. *Let a power series $\sum_{n=0}^{\infty} a_n(x-c)^n$ be given.*

(a) *If $\lim_{n\to\infty} \dfrac{|a_{n+1}|}{|a_n|} = \infty$, then $R=0$.*

(b) *If $\lim_{n\to\infty} \dfrac{|a_{n+1}|}{|a_n|} = 0$, then $R=\infty$.*

(c) *If $\lim_{n\to\infty} \dfrac{|a_{n+1}|}{|a_n|} = K$, then $R=\dfrac{1}{K}$.*

Example 2.9.11. Consider the power series

$$\sum_{n=1}^{\infty} \frac{(-1)^{n+1}x^n}{n} = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots .$$

Since

$$\lim_{n\to\infty} \frac{\left|\frac{(-1)^{n+2}}{n+1}\right|}{\left|\frac{(-1)^{n+1}}{n}\right|} = \lim_{n\to\infty} \frac{n}{n+1} = 1,$$

the radius of convergence of the power series is equal to 1.

Let R be the radius of convergence of a power series $\sum_{n=0}^{\infty} a_n(x-c)^n$, then $f(x) = \sum_{n=0}^{\infty} a_n(x-c)^n$ is a well-defined function on the interval $(c-R, c+R)$. In fact, more is true.

Theorem 2.9.12. *Let R be the radius of convergence of a power series $\sum_{n=0}^{\infty} a_n(x-c)^n$. For $x \in (c-R, c+R)$, define $f(x) = \sum_{n=0}^{\infty} a_n(x-c)^n$, then*

(a) f is differentiable on $(c-R, c+R)$ and $f'(x) = \sum_{n=1}^{\infty} na_n(x-c)^{n-1}$.

(b) For any $y \in (c-R, c+R)$, $\int_c^y f(x)\,dx = \sum_{n=0}^{\infty} \frac{a_n}{n+1}(y-c)^{n+1}$.

Remark 2.9.13. Theorem 2.9.12 means that power series can be termwise differentiable and integrable.

Example 2.9.14. Let

$$f(x) = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}x^n}{n} = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots.$$

In Example 2.9.11, we observed that its radius of convergence is 1. By Theorem 2.9.12, f is differentiable on $(-1, 1)$ and $f'(x) = 1 - x + x^2 - x^3 + \cdots$. Moreover, by Theorem 2.8.13, it follows that

$$f'(x) = 1 - x + x^2 - x^3 + \cdots = \frac{1}{1-(-x)} = \frac{1}{1+x}.$$

Since $f(0) = 0$, we see that f must be a solution to a differential equation $f'(x) = \frac{1}{1+x}$, $f(0) = 0$, so we conclude that $f(x) = \ln(1+x)$.

Exercise 2.9.

1. When does $\sum_{n=0}^{\infty} x^n = 1 + x + x^2 + x^3 + \cdots$ converge? Find the sum.

2. Find the radius of convergence of $\sum_{n=1}^{\infty} nx^{n-1} = 1 + 2x + 3x^2 + \cdots$. Compute the sum.
Hint: Use the result from Exercise 2.9.1 and Theorem 2.9.12.

3. Prove that $\sum_{n=1}^{\infty} \frac{n^3}{2^n} = 26$.
Hint: Let $f(x) = \sum_{n=1}^{\infty} x^n$. First, consider the derivative of $xf'(x)$. You may want to use Exercise 2.9.2.

4. Let X have a Geometric Distribution with parameter p so that $P(X = n) = p(1-p)^{n-1}$, $n = 1, 2, \ldots$. Show that $E(X) = \frac{1}{p}$.
Hint: Use Exercise 2.9.2.

2.10 Taylor Series

The Taylor series expansion of a function gives a way to approximate the function with a polynomial, which is in most cases easier to handle. In statistics, the so called Delta Method, which can be viewed as a generalized Central Limit Theorem applied to a function of asymptotically normal random variables, relies heavily on the Taylor expansion of the function. To motivate, consider a problem of approximating the graph of $y = f(x)$ by a linear function $y = P_1(x) = c_0 + c_1(x - a)$ near $x = a$. In order for $P_1(x)$ to be the best approximation of $f(x)$ near $x = a$, we impose conditions $f(a) = P_1(a)$ (same function values at $x = a$) and $f'(a) = P_1'(a)$ (same derivative at $x = a$) and we get that $c_0 = f(a)$ and $c_1 = f'(a)$. In other words, the best linear approximation $P_1(x)$ of $f(x)$ near $x = a$ is given by $P_1(x) = f(a) + f'(a)(x - a)$. Now consider a problem of approximating $f(x)$ by a quadratic function $y = P_2(x) = d_0 + d_1(x - a) + d_2(x - a)^2$ near $x = a$. By imposing conditions $f(a) = P_2(a)$, $f'(a) = P_2'(a)$, and $f''(a) = P_2''(a)$, we get $d_0 = f(a)$, $d_1 = f'(a)$, and $d_2 = \frac{f''(a)}{2}$. Similarly, the best cubic approximation is obtained (see Figure 2.26). Continuing in this fashion, we can consider a problem of approximating $f(x)$ by a degree n polynomial $P_n(x)$ and it turns out that the best degree n polynomial $P_n(x)$ approximating $f(x)$ near $x = a$ is given by

$$
\begin{aligned}
P_n(x) &= f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n \\
&= \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(x-a)^k.
\end{aligned}
$$

The polynomial $P_n(x)$ is called the Taylor polynomial of order n about $x = a$. Note that the degree of $P_n(x)$ may not be equal to n, because $f^{(n)}(a)$ could equal 0.

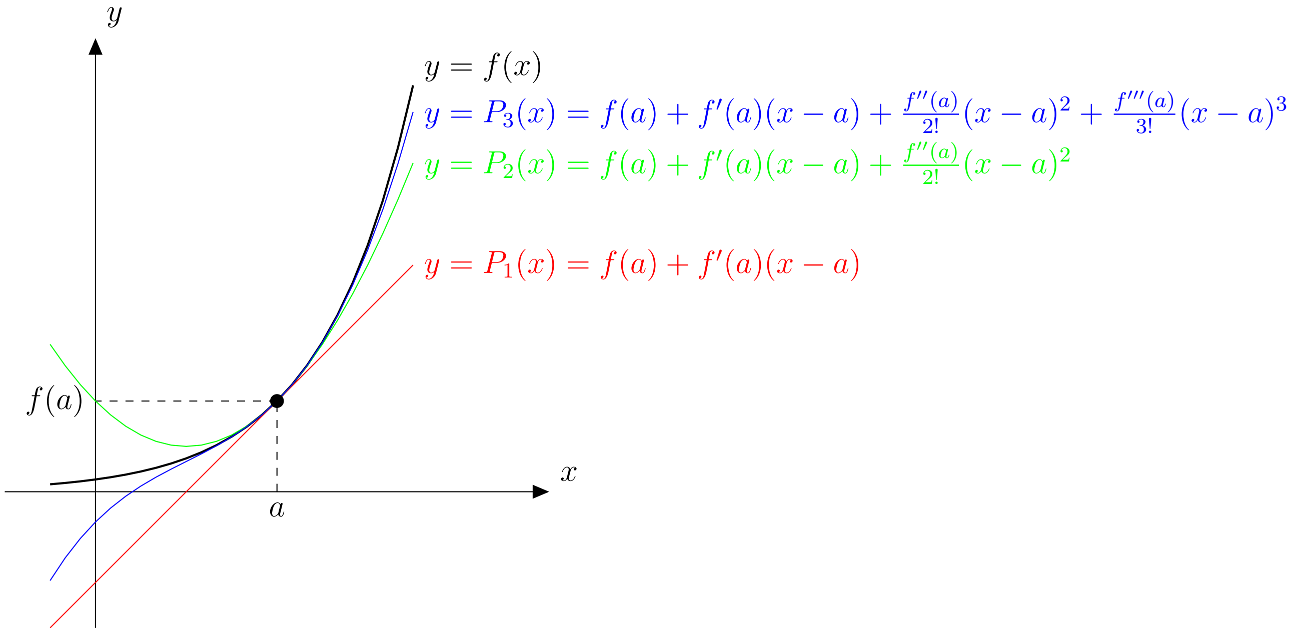

Figure 2.26: Taylor polynomials

Taylor polynomial $P_n(x)$ is a good approximation of $f(x)$ near $x = a$, but $f(x)$ is not exactly equal to $P_n(x)$ unless f itself is a polynomial. The next result, however, gives a way to get an equality at the cost of modifying Taylor polynomial: adding a remainder term or having infinitely many terms.

Theorem 2.10.1 (Taylor Series Expansion)**.** *Let f be a continuous function defined on the closed interval $[a, b]$. Suppose that f is infinitely differentiable on (a, b) and let $x_0 \in (a, b)$. Then for every $x \in (a, b)$, $x \neq x_0$ and for every n, there is an x_1 between x_0 and x such that*

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k + \frac{f^{(n+1)}(x_1)}{n!}(x - x_0)^{n+1}. \tag{2.10.1}$$

Moreover, if there exists a constant M such that $|f^{(n)}(x)| \leq M$ for all n and for all $x \in [a, b]$, then

$$f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k. \tag{2.10.2}$$

(2.10.1) is often called the Taylor expansion with remainder about $x = a$ and (2.10.2) is called the Taylor series expansion of f about $x = a$.

The following is the collection of Taylor series expansion of some elementary functions about $x = 0$ and values at which the expansion is valid.

Taylor Series Expansion of Some Elementary Functions

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!} + \cdots \quad \text{for } -\infty < x < \infty.$$

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots + (-1)^n \frac{x^{2n+1}}{(2n+1)!} + \cdots \quad \text{for } -\infty < x < \infty.$$

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots + (-1)^n \frac{x^{2n}}{(2n)!} + \cdots \quad \text{for } -\infty < x < \infty.$$

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \cdots + x^n + \cdots \quad \text{for } -1 < x < 1.$$

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots + (-1)^{n+1}\frac{x^n}{n} + \cdots \quad \text{for } -1 < x \leq 1.$$

$$(1+x)^p = 1 + px + \frac{p(p-1)}{2!}x^2 + \frac{p(p-1)(p-2)}{3!}x^3 + \cdots \quad \text{for } -1 < x < 1.$$

Example 2.10.2. Using the Taylor series expansion of e^x, we get

$$e = e^1 = 1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \cdots = 2.7182818284\cdots.$$

Example 2.10.3. For a positive constant λ, let X be a discrete random variable such that

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}, \quad k = 0, 1, 2, 3, \ldots.$$

Using the Taylor series expansion of e^x, we get

$$\sum_{k=0}^{\infty} P(X = k) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda}e^{\lambda} = 1.$$

Such a random variable is said to follow Poisson Distribution with parameter λ and is denoted $X \sim Poisson(\lambda)$.

Problem 2.10.4. Let $X \sim Poisson(\lambda)$. Compute $E(X)$.

Answer Note that

$$E(X) = \sum_{k=0}^{\infty} k\frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda}\sum_{k=0}^{\infty} k\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} = \lambda e^{-\lambda}\sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!}.$$

Using the change of index $j = k - 1$, we get

$$E(X) = \lambda e^{-\lambda}\sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = \lambda e^{-\lambda}e^{\lambda} = \lambda.$$

■

Exercise 2.10.

1. Let $X \sim Poisson(\lambda)$. Compute $Var(X)$.
   Hint: Consider $E(X^2 - X) = E(X(X-1))$ first.

2. Suppose that a continuous function $f : [-1, 1] \to \mathbb{R}$ is thrice differentiable on $(-1, 1)$ with
   $$f(-1) = 0, \quad f(0) = 0, \quad f(1) = 1, \quad \text{and} \quad f'(0) = 0.$$
   Show that there is $x_0 \in (-1, 1)$ such that $f'''(x_0) \geq 3$.
   Hint: Suppose that $f'''(x) < 3$ for all $x \in (-1, 1)$. By Taylor Expansion, for any $x \in [-1, 1]$, there is x_0 between 0 and x such that
   $$f(x) = \frac{f''(0)}{2!}x^2 + \frac{f'''(x_0)}{3!}x^3.$$
   Substitute 1 and -1 for x and derive a contradiction.

3. Compute $\displaystyle\sum_{n=0}^{\infty} \frac{1}{(2n)!} = 1 + \frac{1}{2!} + \frac{1}{4!} + \frac{1}{6!} + \cdots$.
   Hint: Consider the Taylor expansion of e^x and e^{-x}.

4. Find the Taylor expansion of $f(x) = \ln x$ about $x = 1$.
   Hint: Note that $\ln x = \ln(1 + (x - 1))$.

allele fre
percent
sample s
to detec
one allel
99 perce
lele freq
0.05 per
required
ple size
4600. (1
$1 - 2Np$
Problem

Chapter 3

Linear Algebra

3.1 Linear System and Matrices

Linear algebra is an essential tool in statistics, especially in statistical modelling. The concept of the column space of a matrix in particular plays an important role interpreting linear models in geometric context. In this section, we consider linear systems and their matrix representation.

Definition 3.1.1. A linear equation in the variables $x_1, x_2, \ldots, x_n$ is an equation of the form

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b,$$

where $a_1, a_2, \ldots, a_n$ and b are real numbers, called coefficients.

Example 3.1.2. $3x_1 + 2x_2 - x_3 = 5$ is a linear equation having 3 variables. $2x - y + z^2 = -1$ is *not* a linear equation, because it contains z^2.

Definition 3.1.3. A linear system in the variables $x_1, x_2, \ldots, x_n$ is a collection of linear equations in the variables $x_1, x_2, \ldots, x_n$. A solution of a linear system is a list $(s_1, s_2, \ldots, s_n)$ that makes each equation in the system hold when variables $x_1, x_2, \cdots, x_n$ are replaced by $s_1, s_2, \ldots, s_n$. In this case, we say that $x_1 = s_1, x_2 = s_2, \ldots, x_n = s_n$ is a solution to the system. The solution set is the collection of all possible solutions. Two systems are said to be equivalent if they have the same solution set.

Example 3.1.4.

$$\left\{\begin{array}{rrrr} 2x_1 & -x_2 & +\frac{3}{2}x_3 = & 8 \\ x_1 & & -4x_3 = & -7 \end{array}\right. \tag{3.1.1}$$

is a linear system having 2 equations and 3 variables. $x_1 = -7, x_2 = -22, x_3 = 0$ is a solution to the system (3.1.1). $x_1 = 5, x_2 = 6.5, x_3 = 3$ is another solution to (3.1.1).

Example 3.1.5. Consider systems

$$\text{I.}\left\{\begin{array}{rrr} x_1 & -2x_2 = & -1 \\ -x_1 & +3x_2 = & 3 \end{array}\right. \qquad \text{II.}\left\{\begin{array}{rrr} 2x_1 & -4x_2 = & -2 \\ x_1 & -3x_2 = & -3 \end{array}\right. \qquad \text{III.}\left\{\begin{array}{rrr} x_1 & -2x_2 = & -1 \\ 2x_1 & -4x_2 = & 1 \end{array}\right.$$

It is easy to check that I and II are equivalent linear systems: $x_1 = 3$, $x_2 = 2$ is a common unique solution to both I and II. One can easily check that III has *no* solutions.

There are linear systems that have infinitely many solutions. In fact, the following is known (see Theorem 3.1.39).

Theorem 3.1.6. *A linear system has either*

(a) no solutions, or

(b) exactly one solution, or

(c) infinitely many solutions.

Definition 3.1.7. A system is said to be consistent if it has either one solution or infinitely many solutions. A system is inconsistent if it has no solutions.

Example 3.1.8. Linear systems I and II in Example 3.1.5 are consistent, while III is inconsistent.

A linear system can be conveniently represented by its augmented matrix. First we recall the definition of a matrix.

Definition 3.1.9. An $m \times n$ matrix is a rectangular array of numbers having m rows and n columns. In particular, an $n \times 1$ matrix is called an n-dimensional column vector (or simply n-vector). Each number in a matrix is called an entry or element or component. The entry at the i^{th} row and j^{th} column is called the (i, j)-entry. If the (i, j)-entry of an $m \times n$ matrix A is given by a_{ij}, then we use notation $A = [a_{ij}]_{i,j=1}^{m,n}$ (or simply $A = [a_{ij}]_{i,j=1}^{n}$ if $m = n$).

Remark 3.1.10.

(a) The set of all n-vectors can be identified with $\mathbb{R}^n$ and from now on, we will use $\mathbb{R}^n$ for the notation of the set of all n-vectors.

(b) In this monograph, column vectors will be denoted using bold face, for example, $\boldsymbol{u}, \boldsymbol{v}$.

(c) An $1 \times n$ matrix is called an n-dimensional row vector. We often write $(x_1, \ldots, x_n)$ to denote a row vector $\begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}$.

Example 3.1.11. Let $A = \begin{bmatrix} 2 & 3 & 1 \\ -1 & 0 & 7 \end{bmatrix}$ and $\boldsymbol{u} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$. Then A is a 2×3 matrix and $\boldsymbol{u}$ is a 2-vector.

We use an example to define matrices that are associated with a given linear system. Consider a system

$$\left\{\begin{array}{rrrr} x_1 & -2x_2 & +x_3 = & 0 \\ & 2x_2 & -8x_3 = & 8 \\ -4x_1 & +5x_2 & +9x_3 = & -9 \end{array}\right. . \tag{3.1.2}$$

The coefficient matrix of the linear system (3.1.2) is $\begin{bmatrix} 1 & -2 & 1 \\ 0 & 2 & -8 \\ -4 & 5 & 9 \end{bmatrix}$.

The augmented matrix of the linear system (3.1.2) is $\left[\begin{array}{rrr|r} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ -4 & 5 & 9 & -9 \end{array}\right]$. Note that the augmented matrix contains all information about the corresponding linear system. In other words, if a matrix A is given, one can recover a unique linear system that has A as the augmented matrix.

Problem 3.1.12. Let $A = \left[\begin{array}{rrrr|r} -2 & 0 & 3 & 1 & 7 \\ 1 & 0 & -2 & 0 & 4 \\ 5 & -4 & 2 & 2 & 1 \end{array}\right]$ be the augmented matrix of a linear system. What is the system?

Answer From A, we recover $\left\{\begin{array}{rrrrr} -2x_1 & & 3x_3 & +x_4 = & 7 \\ x_1 & & -2x_3 & = & 4 \\ 5x_1 & -4x_2 & +2x_3 & +2x_4 = & 1 \end{array}\right.$. ■

We are mainly interested in solving a given system, that is, finding the solution set of the system. We will develop several methods to do so through examples.

Example 3.1.13. Consider systems

$$\text{I.} \left\{\begin{array}{rrrr} x_1 & -2x_2 & +x_3 = & 0 \\ & 2x_2 & -8x_3 = & 8 \\ -4x_1 & +5x_2 & +9x_3 = & -9 \end{array}\right. \quad \text{and} \quad \text{II.} \left\{\begin{array}{rrrr} x_1 & -2x_2 & -x_3 = & 7 \\ & x_2 & -2x_3 = & -10 \\ & & x_3 = & 4 \end{array}\right. .$$

We can easily solve II using backward substitution:

$$x_3 = 4, \quad x_2 = 2x_3 - 10 = 2 \cdot 4 - 10 = -2, \quad x_1 = 2x_2 + x_3 + 7 = 2 \cdot (-2) + 4 + 7 = 7.$$

Note that system II in Example 3.1.13 is easy to solve because it is *steplike*, by which we mean that the leading variable in the n^{th} equation is x_n with coefficient equal to 1. Therefore, one idea to solve a general and more complicated system like system I is to transform it into a system like II. When we transform the original system, however, of course we need to be careful in order *not* to change the solution set of the system. We introduce three important operations that can be used to transform a general linear system into an easier one without altering the solution set.

Theorem 3.1.14. *Let a linear system be given.*

(a) *scaling operation: if we replace an equation in the linear system by a nonzero multiple of the equation, then the solution set remains unchanged.*

(b) *interchange operation: if we interchange two equations in the system, the solution set remains unchanged.*

(c) *replacement operation: if we replace one equation in the system by the sum of itself and a multiple of another equation, the solution set remains unchanged.*

Definition 3.1.15. The operations in Theorem 3.1.14 are called elementary row operations.

Example 3.1.16. Let

$$\mathrm{I}'.\begin{cases} 2x_1 & -4x_2 & +2x_3 = & 0 \\ & 2x_2 & -8x_3 = & 8 \\ -4x_1 & +5x_2 & +9x_3 = & -9 \end{cases} \quad \text{and} \quad \mathrm{I}''.\begin{cases} x_1 & -2x_2 & +x_3 = & 0 \\ & x_2 & -4x_3 = & 4 \\ -4x_1 & +5x_2 & +9x_3 = & -9 \end{cases},$$

then both I' and I'' are equivalent to I in Example 3.1.13. In other words, systems I, I', and I'' are essentially the same. Note that I' is obtained from I by multiplying row 1 by 2. Similarly, I'' is obtained from I by multiplying row 2 by $\frac{1}{2}$. To indicate that I' (respectively, I'') is obtained from I using the scaling operation described above, we use notation $\mathrm{I} \xrightarrow{R_1 \mapsto 2R_1} \mathrm{I}'$ (respectively, $\mathrm{I} \xrightarrow{R_2 \mapsto \frac{1}{2}R_2} \mathrm{I}''$).

Example 3.1.17. Consider

$$\mathrm{I}'''.\begin{cases} x_1 & -2x_2 & +x_3 = & 0 \\ -4x_1 & +5x_2 & +9x_3 = & -9 \\ & 2x_2 & -8x_3 = & 8 \end{cases},$$

then I''' is equivalent to I in Example 3.1.13. Note that I''' is obtained from I by interchanging row 2 and row 3. To indicate the use of interchange operation, we use notation $\mathrm{I} \xrightarrow{R_2 \leftrightarrow R_3} \mathrm{I}'''$.

Example 3.1.18. The system

$$\mathrm{I}''''.\begin{cases} x_1 & -2x_2 & +x_3 = & 0 \\ & 2x_2 & -8x_3 = & 8 \\ & -3x_2 & +13x_3 = & -9 \end{cases}$$

is obtained by replacing row 3 of system I in Example 3.1.13 by the sum of row 3 and 4 times row 1 of I. It is easy to check that I and I'''' are equivalent. To indicate the use of replacement operation, we use notation $\mathrm{I} \xrightarrow{R_3 \mapsto R_3 + 4R_1} \mathrm{I}''''$.

Example 3.1.19. In this example, we explain how to transform a general system into a steplike one. Consider system I in Example 3.1.13 and we apply elementary row operations as in the following:

$$\left\{\begin{array}{rrrr} x_1 & -2x_2 & +x_3 = & 0 \\ & 2x_2 & -8x_3 = & 8 \\ -4x_1 & +5x_2 & +9x_3 = & -9 \end{array}\right. \xrightarrow{R_3 \mapsto R_3+4R_1} \left\{\begin{array}{rrrr} x_1 & -2x_2 & +x_3 = & 0 \\ & 2x_2 & -8x_3 = & 8 \\ & -3x_2 & +13x_3 = & -9 \end{array}\right.$$

$$\xrightarrow{R_2 \mapsto \frac{1}{2}R_2} \left\{\begin{array}{rrrr} x_1 & -2x_2 & +x_3 = & 0 \\ & x_2 & -4x_3 = & 4 \\ & -3x_2 & +13x_3 = & -9 \end{array}\right. \xrightarrow{R_3 \mapsto R_3+3R_2} \left\{\begin{array}{rrrr} x_1 & -2x_2 & +x_3 = & 0 \\ & x_2 & -4x_3 = & 4 \\ & & x_3 = & 3 \end{array}\right. .$$

It is easy to find a (unique) solution to the last system: $x_3 = 3$, $x_2 = 16$, $x_1 = 29$. Since elementary row operations do not change the solution set, we see that $x_1 = 29$, $x_2 = 16$, $x_3 = 3$ is a unique solution to I, too. Note that the procedure above can be described using the augmented matrices

$$\left[\begin{array}{rrr|r} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ -4 & 5 & 9 & -9 \end{array}\right] \xrightarrow{R_3 \mapsto R_3+4R_1} \left[\begin{array}{rrr|r} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ 0 & -3 & 13 & -9 \end{array}\right]$$

$$\xrightarrow{R_2 \mapsto \frac{1}{2}R_2} \left[\begin{array}{rrr|r} 1 & -2 & 1 & 0 \\ 0 & 1 & -4 & 4 \\ 0 & -3 & 13 & -9 \end{array}\right] \xrightarrow{R_3 \mapsto R_3+3R_2} \left[\begin{array}{rrr|r} 1 & -2 & 1 & 0 \\ 0 & 1 & -4 & 4 \\ 0 & 0 & 1 & 3 \end{array}\right],$$

where the last matrix can be decoded to produce the corresponding linear system

$$\left\{\begin{array}{rrrr} x_1 & -2x_2 & +x_3 = & 0 \\ & x_2 & -4x_3 = & 4 \\ & & x_3 = & 3 \end{array}\right. ,$$

yielding the same solution $x_1 = 29$, $x_2 = 16$, $x_3 = 3$ as expected. Note that instead of stopping at the last matrix, we can perform additional elementary row operations to get

$$\left[\begin{array}{rrr|r} 1 & -2 & 1 & 0 \\ 0 & 1 & -4 & 4 \\ 0 & 0 & 1 & 3 \end{array}\right] \xrightarrow{R_1 \mapsto R_1+2R_2} \left[\begin{array}{rrr|r} 1 & 0 & -7 & 8 \\ 0 & 1 & -4 & 4 \\ 0 & 0 & 1 & 3 \end{array}\right] \xrightarrow[R_2 \mapsto R_2+4R_3]{R_1 \mapsto R_1+7R_3} \left[\begin{array}{rrr|r} 1 & 0 & 0 & 29 \\ 0 & 1 & 0 & 16 \\ 0 & 0 & 1 & 3 \end{array}\right],$$

which, after recovering the corresponding linear system, yields the solution $x_1 = 29$, $x_2 = 16$, $x_3 = 3$ right away.

Remark 3.1.20. The procedure described in Example 3.1.19 is called the <u>Gauss-Jordan elimination</u>.

Example 3.1.21. Consider the system

$$\begin{cases} \quad\quad\; x_2 \; -4x_3 = 8 \\ 2x_1 \; -3x_2 \; +2x_3 = 1 \\ 5x_1 \; -8x_2 \; +7x_3 = 1 \end{cases}.$$

Note that the corresponding augmented matrix can be transformed as

$$\left[\begin{array}{rrr|r} 0 & 1 & -4 & 8 \\ 2 & -3 & 2 & 1 \\ 5 & -8 & 7 & 1 \end{array}\right] \xrightarrow{R_1 \leftrightarrow R_2} \left[\begin{array}{rrr|r} 2 & -3 & 2 & 1 \\ 0 & 1 & -4 & 8 \\ 5 & -8 & 7 & 1 \end{array}\right] \xrightarrow{R_1 \mapsto \frac{1}{2}R_1} \left[\begin{array}{rrr|r} 1 & -\frac{3}{2} & 1 & \frac{1}{2} \\ 0 & 1 & -4 & 8 \\ 5 & -8 & 7 & 1 \end{array}\right]$$

$$\xrightarrow{R_3 \mapsto R_3 - 5R_1} \left[\begin{array}{rrr|r} 1 & -\frac{3}{2} & 1 & \frac{1}{2} \\ 0 & 1 & -4 & 8 \\ 0 & -\frac{1}{2} & 2 & -\frac{3}{2} \end{array}\right] \xrightarrow{R_3 \mapsto R_3 + \frac{1}{2}R_2} \left[\begin{array}{rrr|r} 1 & -\frac{3}{2} & 1 & \frac{1}{2} \\ 0 & 1 & -4 & 8 \\ 0 & 0 & 0 & \frac{5}{2} \end{array}\right]$$

from which we recover a system

$$\begin{cases} x_1 \; -\frac{3}{2}x_2 \; +x_3 = \frac{1}{2} \\ \quad\quad\; x_2 \; -4x_3 = 8 \\ \quad\quad\quad\quad\quad\; 0 = \frac{5}{2} \end{cases}.$$

Therefore, the system is inconsistent.

Example 3.1.22. Consider a linear system

$$\begin{cases} x_1 \; +6x_2 \; +3x_4 = 0 \\ \quad\quad x_3 \; -4x_4 = 5 \\ \quad\quad\quad\quad x_5 = 7 \end{cases}$$

which has 5 variables and 3 equations. Obviously $x_5 = 7$. How about x_4? As one can see, x_4 need not be a fixed number. In other words, x_4 is <u>free</u>. However, once x_4 is chosen, x_3 is completely determined because of equation 2. That is, $x_3 = 4x_4 + 5$. Looking at the top equation, we see that x_2 is another free variable, independent of x_4. Note, however, that the values of x_4 and x_2 completely determine x_1, say, $x_1 = -6x_2 - 3x_4$. In summary, the solution set of the system is

$$\begin{cases} x_1 & = -6x_2 - 3x_4 \\ x_2 & \text{is free} \\ x_3 & = 4x_4 + 5 \\ x_4 & \text{is free} \\ x_5 & = 7 \end{cases}.$$

If we take $x_2 = 3$ and $x_4 = 1$, then we have a solution

$$x_1 = -21, \quad x_2 = 3, \quad x_3 = 9, \quad x_4 = 1, \quad x_5 = 7.$$

Setting $x_2 = -1$ and $x_4 = 0$ gives

$$x_1 = 6, \quad x_2 = -1, \quad x_3 = 5, \quad x_4 = 0, \quad x_5 = 7.$$

In this way, we can supply infinitely many solutions of the given linear system. Note that presence of a free variable makes the system have infinitely many solutions.

Definition 3.1.23. A variable which is not free is called basic.

Example 3.1.24. In Example 3.1.22, x_1, x_3, x_5 are basic variables and x_2, x_4 are free variables.

So far we have observed that if a system is obtained from another using elementary row operations, then the systems are equivalent to each other. Since the augmented matrix of a linear system completely describes the system, this observation leads to the following definition.

Definition 3.1.25. Two matrices are row equivalent if there is a sequence of elementary row operations that transforms one matrix into the other.

Remark 3.1.26. It is easy to check that two systems are equivalent if and only if the corresponding augmented matrices are row equivalent.

By Remark 3.1.26, solving a linear system boils down to transforming its augmented matrix into a steplike one. We make precise what it means by a steplike matrix.

Definition 3.1.27. A matrix is said to be in a reduced row echelon form or a reduced row echelon matrix if it has the following properties:

(a) All nonzero rows are above any rows of all zeros.

(b) Each leading entry of a row (i.e., the leftmost nonzero entry in the row) is 1 and it is in a column to the right of the leading entry of the row above it. In other words, the leading 1's move to the right.

(c) The leading entry is the only nonzero number in its column.

Example 3.1.28. $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ is in a reduced row echelon form.

Example 3.1.29. $\begin{bmatrix} 1 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$ is *not* in a reduced row echelon form, because the row of all zeros is above the last row, which contains a nonzero number. If row 2 and 3 are switched, then it would be in the reduced row echelon form.

Example 3.1.30. $\begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ is *not* in a reduced raw echelon form, because the leading entries of rows 2 and 3 are not the only nonzero number in their respective columns.

Example 3.1.31. $\begin{bmatrix} 1 & 0 & 3 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ is in a reduced row echelon form. Note that the $(2,3)$-entry is *not* the leading entry in row 2, so having 3 above it does *not* violate the conditions in Definition 3.1.27.

As seen in Example 3.1.19, the importance of a reduced row echelon form is in that the linear system associated with a matrix in a reduced row echelon form is easy to solve. The next theorem states that it is always possible to transform a general matrix into a unique reduced echelon matrix.

Theorem 3.1.32. *Each matrix A is row equivalent to one and only one reduced echelon matrix. This unique matrix is called the reduced row echelon form of A.*

Example 3.1.33. Consider the reduced row echelon form of $A = \begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ -4 & 5 & 9 & -9 \end{bmatrix}$.

In Example 3.1.19, we observed that A is row equivalent to $\begin{bmatrix} 1 & 0 & 0 & 29 \\ 0 & 1 & 0 & 16 \\ 0 & 0 & 1 & 3 \end{bmatrix}$, which is in a reduced row echelon form. By Theorem 3.1.33, we see that $\begin{bmatrix} 1 & 0 & 0 & 29 \\ 0 & 1 & 0 & 16 \\ 0 & 0 & 1 & 3 \end{bmatrix}$ is the reduced row echelon form of A.

Problem 3.1.34. Find the reduced row echelon form of $\begin{bmatrix} 0 & 1 & 2 & -4 \\ 1 & 0 & 2 & -2 \\ -2 & 4 & 6 & -4 \end{bmatrix}$.

Answer

$$\begin{bmatrix} 0 & 1 & 2 & -4 \\ 1 & 0 & 2 & -2 \\ -2 & 4 & 6 & -4 \end{bmatrix} \xrightarrow[R_3 \mapsto -\frac{1}{2}R_3]{R_1 \leftrightarrow R_2} \begin{bmatrix} 1 & 0 & 2 & -2 \\ 0 & 1 & 2 & -4 \\ 1 & -2 & -3 & 2 \end{bmatrix} \xrightarrow{R_3 \mapsto R_3 - R_1} \begin{bmatrix} 1 & 0 & 2 & -2 \\ 0 & 1 & 2 & -4 \\ 0 & -2 & -5 & 4 \end{bmatrix}$$

$$\xrightarrow{R_3 \mapsto R_3 + 2R_2} \begin{bmatrix} 1 & 0 & 2 & -2 \\ 0 & 1 & 2 & -4 \\ 0 & 0 & -1 & -4 \end{bmatrix} \xrightarrow{R_3 \mapsto -R_3} \begin{bmatrix} 1 & 0 & 2 & -2 \\ 0 & 1 & 2 & -4 \\ 0 & 0 & 1 & 4 \end{bmatrix}$$

$$\xrightarrow[R_2 \mapsto R_2 - 2R_3]{R_1 \mapsto R_1 - 2R_3} \begin{bmatrix} 1 & 0 & 0 & -10 \\ 0 & 1 & 0 & -12 \\ 0 & 0 & 1 & 4 \end{bmatrix}.$$

■

Problem 3.1.35. Solve a linear system

$$\begin{cases} \quad\quad x_2 \ +2x_3 = -4 \\ x_1 \quad\quad +2x_3 = -2 \\ -2x_1 \ +4x_2 \ +6x_3 = -4 \end{cases}.$$

Answer Note that the augmented matrix of the system is given in Problem 3.1.34. Since it is row equivalent to $\begin{bmatrix} 1 & 0 & 0 & -10 \\ 0 & 1 & 0 & -12 \\ 0 & 0 & 1 & 4 \end{bmatrix}$, we conclude that $x_1 = -10$, $x_2 = -12$, and $x_3 = 4$. ■

We are now ready to define the rank of a matrix.

Definition 3.1.36. A pivot position in a matrix A is a location in A that corresponds to a leading 1 in the reduced row echelon form of A. A pivot column is a column of A that contains a pivot position. The number of pivot positions of a matrix A is called the rank of A and denoted by $\text{rank}(A)$.

Example 3.1.37. In Example 3.1.33, we had three pivot columns: columns 1, 2, and 3. In particular, its rank equals 3.

Remark 3.1.38. The rank of a matrix has an important meaning regarding the matrix. See Theorem 3.3.39 for example.

Next theorem gives a way to determine whether a given system has no solutions, a unique solution, or infinitely many solution (see Theorem 3.1.6).

Theorem 3.1.39. *A linear system is inconsistent if and only if the rightmost column of the augmented matrix is a pivot column. When a system is consistent (i.e., the rightmost column of its augmented matrix is not pivotal), it contains either (i) a unique solution, when there are no free variables, or (ii) infinitely many solutions, when there is at least one free variable.*

Remark 3.1.40. Let a consistent system be given, so it has either a unique solution or infinitely many solutions. Note that pivot columns of the coefficient matrix of the system correspond to basic variables, while non-pivot columns correspond to free variables (see Example 3.1.22). We conclude that a consistent system has a unique solution if and only if all columns of the coefficient matrix of the system are pivotal.

Exercise 3.1.

1. Solve the following system:

$$\begin{cases} x_1 \quad\quad -2x_3 \ -5x_4 \quad\quad = 8 \\ \quad\quad x_2 \quad\quad +2x_4 \quad\quad = 2 \\ \quad\quad\quad\quad\quad\quad\quad\quad x_5 = 1 \end{cases}.$$

2. Solve the following system:

$$\begin{cases} -x_1 & +x_2 & -x_3 = & 2 \\ x_1 & -x_2 & +2x_3 = & 0 \\ 2x_1 & -3x_2 & +2x_3 = & -2 \end{cases}.$$

3. Solve the following system:

$$\begin{cases} x_1 & +6x_2 & -x_3 & +3x_4 & +2x_5 & = -2 \\ -2x_1 & +x_2 & x_3 & -2x_4 & +x_5 & = 3 \\ & x_2 & & -2x_4 & x_5 & = -10 \end{cases}.$$

4. Find the reduced row echelon form of the matrix $\begin{bmatrix} 2 & 4 & 6 & -2 \\ 1 & 0 & -3 & 1 \\ 1 & -1 & 3 & 0 \end{bmatrix}$ and using it, solve

$$\begin{cases} 2x & +4y & +6z = & -2 \\ x & & -3z = & 1 \\ x & -y & +3z = & 0 \end{cases}.$$

5. Find the rank of each of the following matrices.

(a) $\begin{bmatrix} -1 & 0 & 0 & 2 \\ 5 & 0 & 2 & 7 \\ 1 & -3 & 1 & -2 \end{bmatrix}$ (b) $\begin{bmatrix} 0 & -5 & 2 & -2 & 5 & 1 \\ 1 & 3 & -3 & 0 & -3 & 1 \\ -5 & 4 & 1 & -1 & 2 & 0 \end{bmatrix}$

3.2 Basics of Matrix Algebra

In this section, we discuss basic operations and analytic properties of matrices that will play a key role in handling multidimensional data.

Definition 3.2.1. Two $m \times n$ matrices are said to be equal if their corresponding entries are equal. The sum $A + B$ of two $m \times n$ matrices A, B is an $m \times n$ matrix obtained by adding corresponding entries. The scalar multiple cA of an $m \times n$ matrix A by c is the $m \times n$ matrix obtained by multiplying each entry in A by c. The difference $A - B$ of two $m \times n$ matrices A, B is defined by $A + (-1)B$. In other words, $A - B$ is obtained by computing entrywise difference. The transpose A^t of an $m \times n$ matrix A is the $n \times m$ matrix whose columns are formed from the corresponding rows of A.

Example 3.2.2. If $\begin{bmatrix} x-2 & 2 \\ 5 & y+2 \end{bmatrix} = \begin{bmatrix} 3 & 2 \\ x & -1 \end{bmatrix}$, then $x = 5$ and $y = -3$.

Example 3.2.3. For $\boldsymbol{u} = \begin{bmatrix} 3 \\ 4 \\ -2 \end{bmatrix}$, $7\boldsymbol{u} = \begin{bmatrix} 21 \\ 28 \\ -14 \end{bmatrix}$ and $\frac{1}{2}\boldsymbol{u} = \begin{bmatrix} \frac{3}{2} \\ 2 \\ -1 \end{bmatrix}$.

Problem 3.2.4. Compute $3\boldsymbol{u} - 2\boldsymbol{v}$, where $\boldsymbol{u} = \begin{bmatrix} 3 \\ 4 \\ -2 \end{bmatrix}$ and $\boldsymbol{v} = \begin{bmatrix} 1 \\ -2 \\ 5 \end{bmatrix}$.

Answer

$$3\boldsymbol{u} - 2\boldsymbol{v} = 3\begin{bmatrix} 3 \\ 4 \\ -2 \end{bmatrix} - 2\begin{bmatrix} 1 \\ -2 \\ 5 \end{bmatrix} = \begin{bmatrix} 7 \\ 16 \\ -16 \end{bmatrix}.$$

■

Example 3.2.5. If $A = \begin{bmatrix} 1 & 0 \\ -2 & 5 \\ 5 & 1 \end{bmatrix}$, then $A^t = \begin{bmatrix} 1 & -2 & 5 \\ 0 & 5 & 1 \end{bmatrix}$.

Remark 3.2.6. It is straightforward to check that $(A^t)^t = A$.

Definition 3.2.7. The zero matrix, denoted by O, is the matrix whose entries are all equal to 0. The $n \times 1$ zero matrix is called the n-dimensional zero vector and denoted by $\mathbf{0}$. A matrix is called a square matrix if it has the same number of rows and columns. A square matrix is said to be upper triangular (respectively, lower triangular), if its (i, j)-entry is zero for all i, j with $i > j$ (respectively, $i < j$). A square matrix is called a diagonal matrix if its (i, j)-entry is zero for all i, j with $i \neq j$, that is, if all off diagonal entries are zero. The $n \times n$ identity matrix, denoted by I_n, is the $n \times n$ diagonal matrix with 1 along the diagonal.

Example 3.2.8. $\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$ is the 3-dimensional zero vector and $I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ is the 2×2 identity matrix.

The following properties follow immediately.

Theorem 3.2.9. *For any $m \times n$ matrices A, B, C and for any scalars $c, d \in \mathbb{R}$, the following are true:*

(a) $A + B = B + A$

(b) $(A + B) + C = A + (B + C)$

(c) $A + O = O + A = A$

(d) $A + (-A) = -A + A = O$

(e) $c(A + B) = cA + cB$

(f) $(c + d)A = cA + dA$

(g) $c(dA) = (cd)A$

(h) $1A = A$

Theorem 3.2.9 has nice geometric explanations when the matrices are replaced by n-vectors. First note that an n-vector may be identified with an arrow starting from the origin. See Figure 3.1.

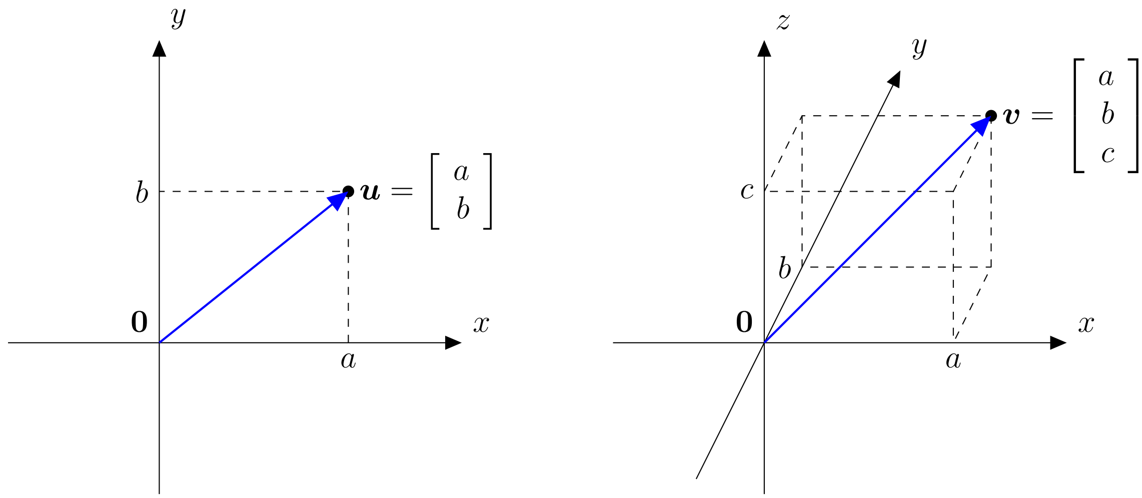

Figure 3.1: Geometric interpretation of a 2-vector and a 3-vector

The sum, scaling, and additive inverse of n-vectors are described in Figure 3.2.

Next, we consider matrix multiplication. To define the product of two matrices, we begin with the dot product of two column vectors.

Definition 3.2.10. Let $\boldsymbol{u} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$ and $\boldsymbol{v} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$. The dot product (or inner product or scalar product) of $\boldsymbol{u}$ and $\boldsymbol{v}$, denoted by $\langle \boldsymbol{u}, \boldsymbol{v} \rangle$ or $\boldsymbol{u} \cdot \boldsymbol{v}$, is the number given by $\sum_{i=1}^n a_i b_i = a_1 b_1 + \cdots + a_n b_n$.

Example 3.2.11. $\begin{bmatrix} 1 \\ -2 \\ 3 \\ -1 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 2 \\ 5 \\ -2 \end{bmatrix} = 1 \cdot 0 + (-2) \cdot 2 + 3 \cdot 5 + (-1) \cdot (-2) = 13.$

It is easy to verify the following properties of inner product.

Theorem 3.2.12. *For all* $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^n$,

(a) $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \langle \boldsymbol{v}, \boldsymbol{u} \rangle$.

(b) $\langle \boldsymbol{u} + \boldsymbol{v}, \boldsymbol{w} \rangle = \langle \boldsymbol{u}, \boldsymbol{w} \rangle + \langle \boldsymbol{v}, \boldsymbol{w} \rangle$.

(c) $\langle c\boldsymbol{u}, \boldsymbol{v} \rangle = c\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \langle \boldsymbol{u}, c\boldsymbol{v} \rangle$.

(d) $\langle \boldsymbol{u}, \boldsymbol{u} \rangle \geq 0$, *and* $\langle \boldsymbol{u}, \boldsymbol{u} \rangle = 0$ *if and only if* $\boldsymbol{u} = \boldsymbol{0}$.

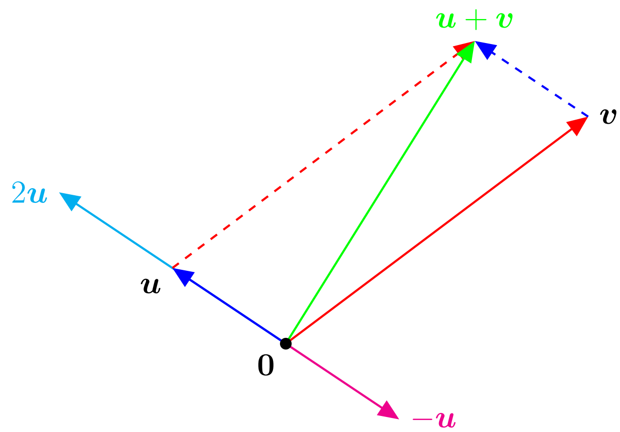

Figure 3.2: Geometric interpretation of vector addition, scaling, and additive inverse

Note that, by Pythagorean Theorem, the length of the arrow $\boldsymbol{v}$ in Figure 3.1 is given by $\sqrt{a^2+b^2+c^2} = \sqrt{\langle \boldsymbol{v}, \boldsymbol{v} \rangle}$. This justifies the following definitions.

Definition 3.2.13. Let $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$.

(a) The length of $\boldsymbol{u}$, denoted by $\|\boldsymbol{u}\|$, is defined by $\|\boldsymbol{u}\| = \sqrt{\langle \boldsymbol{u}, \boldsymbol{u} \rangle}$.

(b) The distance between $\boldsymbol{u}$ and $\boldsymbol{v}$ is defined to be $\|\boldsymbol{u} - \boldsymbol{v}\|$.

(c) $\boldsymbol{u}$ is called a unit vector if $\|\boldsymbol{u}\| = 1$.

Before we proceed, one of the fundamental inequalities in linear algebra is in order.

Theorem 3.2.14 (Cauchy-Schwarz Inequality)**.** *For all* $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$,

$$|\langle \boldsymbol{u}, \boldsymbol{v} \rangle| \leq \|\boldsymbol{u}\| \|\boldsymbol{v}\|.$$

Cauchy-Schwarz Inequality opens a way to define the concept of an angle between $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\mathbb{R}^n$, even when $n > 3$. Indeed, for nonzero vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\mathbb{R}^n$, the angle $\angle(\boldsymbol{u}, \boldsymbol{v})$ between $\boldsymbol{u}$ and $\boldsymbol{v}$ is defined to be

$$\angle(\boldsymbol{u}, \boldsymbol{v}) = \arccos\left(\frac{\langle \boldsymbol{u}, \boldsymbol{v} \rangle}{\|\boldsymbol{u}\| \|\boldsymbol{v}\|}\right).$$

Note that $\frac{\langle \boldsymbol{u}, \boldsymbol{v} \rangle}{\|\boldsymbol{u}\| \|\boldsymbol{v}\|}$ is between -1 and 1 by Cauchy-Schwarz Inequality, so $\angle(\boldsymbol{u}, \boldsymbol{v})$ ranges from 0 to π. In particular, if $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = 0$, then $\angle(\boldsymbol{u}, \boldsymbol{v}) = \frac{\pi}{2}$. In this case, $\boldsymbol{u}$ and $\boldsymbol{v}$ are said to be orthogonal or perpendicular.

Example 3.2.15 (sample correlation coefficient)**.** For paired data points (X_i, Y_i), $1 \le i \le N$, the sample correlation coefficient r is defined to be

$$r = \frac{\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{N}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{N}(Y_i - \bar{Y})^2}},$$

where $\bar{X}$ (respectively, $\bar{Y}$) is the sample mean of $(X_i)_{i=1}^{N}$ (respectively, $(Y_i)_{i=1}^{N}$). Define $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^N$ by

$$\boldsymbol{u} = \begin{bmatrix} X_1 - \bar{X} \\ \vdots \\ X_N - \bar{X} \end{bmatrix} \quad \text{and} \quad \boldsymbol{v} = \begin{bmatrix} Y_1 - \bar{Y} \\ \vdots \\ Y_N - \bar{Y} \end{bmatrix},$$

then it is easy to verify that $r = \frac{\langle \boldsymbol{u}, \boldsymbol{v} \rangle}{\|\boldsymbol{u}\|\|\boldsymbol{v}\|}$ and therefore $-1 \le r \le 1$.

Now we define the product of matrices.

Definition 3.2.16. Let A be an $\ell \times m$ matrix, B an $m \times n$ matrix. Then their product AB is an $\ell \times n$ matrix whose (i, j)-entry is given by $\sum_{k=1}^{m} a_{ik} b_{kj}$, where a_{ik} is the (i, k)-entry of A and b_{kj} is the (k, j)-entry of B. In other words, (i, j)-entry of AB equals the inner product of the i^{th} row of A and j^{th} column of B.

Remark 3.2.17.

(a) For AB to be defined, the number of columns of A must equal the number of rows of B.

(b) The inner product $\langle \boldsymbol{u}, \boldsymbol{v} \rangle$ between two n-vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ is the same as $\boldsymbol{u}^t \boldsymbol{v}$.

Example 3.2.18. The product of a 2×3 matrix and a 3×2 matrix is a 2×2 matrix. For example,

$$\begin{aligned} \begin{bmatrix} 1 & 2 & -1 \\ 0 & -5 & 3 \end{bmatrix} \begin{bmatrix} 4 & 2 \\ 3 & 1 \\ 7 & 0 \end{bmatrix} &= \begin{bmatrix} 1 \cdot 4 + 2 \cdot 3 + (-1) \cdot 7 & 1 \cdot 2 + 2 \cdot 1 + (-1) \cdot 0 \\ 0 \cdot 4 + (-5) \cdot 3 + 3 \cdot 7 & 0 \cdot 2 + (-5) \cdot 1 + 3 \cdot 0 \end{bmatrix} \\ &= \begin{bmatrix} 3 & 4 \\ 6 & -5 \end{bmatrix}. \end{aligned}$$

Example 3.2.19. The product of a 2×3 matrix and a 3-vector is a 2-vector. For example,

$$\begin{aligned} \begin{bmatrix} 5 & 4 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 4 \\ 3 \end{bmatrix} &= \begin{bmatrix} 5 \cdot 2 + 4 \cdot 4 + 1 \cdot 3 \\ 1 \cdot 2 + 1 \cdot 4 + 0 \cdot 3 \end{bmatrix} \\ &= \begin{bmatrix} 29 \\ 6 \end{bmatrix}. \end{aligned}$$

Example 3.2.20. $\begin{bmatrix} 1 & 2 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 4 \\ -1 \end{bmatrix}$ is *not* defined, because the number of columns of the former, 2, is different from the number of rows of the latter, 3.

Example 3.2.21. $\begin{bmatrix} -2 & 0 & 1 \\ 1 & 2 & 3 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -2x_1 + x_3 \\ x_1 + 2x_2 + 3x_3 \\ x_3 \end{bmatrix} = x_1\boldsymbol{v}_1 + x_2\boldsymbol{v}_2 + x_3\boldsymbol{v}_3$, where $\boldsymbol{v}_i$ denotes the i^{th} column of the former matrix. See Example 3.3.8.

Example 3.2.22. Matrix multiplication can be used to represent a linear system. Consider the system

$$\begin{cases} \begin{array}{rrrr} x_1 & -2x_2 & +x_3 = & 0 \\ & 2x_2 & -8x_3 = & 8 \\ -4x_1 & +5x_2 & +9x_3 = & -9 \end{array} \end{cases},$$

then it can be described as $A\boldsymbol{x} = \boldsymbol{b}$, where $A = \begin{bmatrix} 1 & -2 & 1 \\ 0 & 2 & -8 \\ -4 & 5 & 9 \end{bmatrix}$, $\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$, and $\boldsymbol{b} = \begin{bmatrix} 0 \\ 8 \\ -9 \end{bmatrix}$.

Problem 3.2.23. A system of the form $A\boldsymbol{x} = \mathbf{0}$ is said to be <u>homogeneous</u>. Prove that every homogeneous system is consistent.

Answer It suffices to provide a solution to $A\boldsymbol{x} = \mathbf{0}$. Indeed $\boldsymbol{x} = \mathbf{0}$ works. For another approach, note that the rightmost column of the augmented matrix of the system $A\boldsymbol{x} = \mathbf{0}$, being a column of all zeros, cannot be pivotal, so the result follows from Theorem 3.1.39. ■

Example 3.2.24. Matrix multiplication can be used to extract a column or a row in a matrix. Let $A = [a_{ij}]_{i,j=1}^n$ and $\boldsymbol{e}_i = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$, a unit column vector with 1 at the i^{th} position and 0 elsewhere. Then $A\boldsymbol{e}_i$ equals the i^{th} column of A, while $\boldsymbol{e}_i^t A$ gives the i^{th} row of A. In particular, $\boldsymbol{e}_i^t A\boldsymbol{e}_j = \langle \boldsymbol{e}_i, A\boldsymbol{e}_j \rangle$ gives the (i, j)-element a_{ij}.

Problem 3.2.25. Let $A = [a_{ij}]_{i,j=1}^{n}$, $\boldsymbol{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}$, and $\boldsymbol{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$. Prove that $\langle A\boldsymbol{u}, \boldsymbol{v}\rangle = \langle \boldsymbol{u}, A^t\boldsymbol{v}\rangle$.

Answer Since the k^{th} component of $A\boldsymbol{u}$ is given by $\sum_{i=1}^{n} a_{ki}u_i$, we get

$$\langle A\boldsymbol{u}, \boldsymbol{v}\rangle = \sum_{k=1}^{n} \left(\sum_{i=1}^{n} a_{ki}u_i \right) v_k.$$

Note that the k^{th} component of $A^t\boldsymbol{v}$ is given by $\sum_{i=1}^{n} a_{ik}v_i$, so it follows that

$$\langle \boldsymbol{u}, A^t\boldsymbol{v}\rangle = \sum_{k=1}^{n} u_k \left(\sum_{i=1}^{n} a_{ik}v_i \right).$$

The assertion follows from simple index change. In fact, the result holds even when A is a non-square matrix. See Exercise 3.2.7. ■

Remark 3.2.26.

(a) When AB is defined, BA is not necessarily defined. For example, if A is a 3×2 matrix and B is a 2×4 matrix, then AB is a well-defined 3×4 matrix while BA is not defined.

(b) Even when both AB and BA are defined, they might not have the same size. For example, if A is a 3×2 matrix and B is a 2×3 matrix, then AB is a 3×3 matrix while BA is a 2×2 matrix.

(c) Even if A, B are both $n \times n$ matrices so that AB and BA have the same size, AB is not necessarily equal to BA. Let $A = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$, then $AB = \begin{bmatrix} 4 & 5 \\ 1 & 2 \end{bmatrix}$, while $BA = \begin{bmatrix} 2 & 5 \\ 1 & 4 \end{bmatrix}$.

(d) For any $m \times n$ matrix A, $I_m A = A = A I_n$.

Theorem 3.2.27. *If A is an $\ell \times m$ matrix and B, C are $m \times n$ matrices and c is a real number, then*

(a) $A(B + C) = AB + AC$

(b) $A(cB) = c(AB) = (cA)B$

Now we define the concept of a norm of a matrix. Roughly speaking, a norm measures the *size* of a matrix. Since an $m \times n$ matrix can be viewed as a collection of (n many m dimensional) data points, a norm of a matrix can be viewed as a measure of the size of data.

Definition 3.2.28. Let A be an $m \times n$ matrix. A nonnegative function $\|\cdot\|$ defined on the set of all $m \times n$ matrices is called a norm if it has the following properties for all $m \times n$ matrices A, B and all constant c.

(a) $\|A + B\| \le \|A\| + \|B\|$,

(b) $\|cA\| = |c|\|A\|$, and

(c) $\|A\| = 0$ if and only if $A = O$, the zero matrix.

Remark 3.2.29.

(a) For an $m \times n$ matrix $A = [a_{ij}]_{i,j=1}^{m,n}$, define

$$\|A\|_{HS} = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2 \right)^{1/2}.$$

It is well known that $\|\cdot\|_{HS}$ defines a norm. $\|\cdot\|_{HS}$ is called the Hilbert-Schmidt norm or Frobenius norm). Note that $\|A\|_{HS}^2$ is the sum of the squares of the norm of columns of A.

(b) Note that an $m \times n$ matrix A can be viewed as a function from $\mathbb{R}^n$ to $\mathbb{R}^m$, via multiplication: $\boldsymbol{x} \mapsto A\boldsymbol{x}$. Define $\|\cdot\|_{op}$ on A by

$$\|A\|_{op} = \max\{\|A\boldsymbol{x}\| : \|\boldsymbol{x}\| = 1\}.$$

It is also well known that $\|\cdot\|_{op}$ defines a norm. $\|\cdot\|_{op}$ is called the operator norm.

Definition 3.2.30. The trace of an $n \times n$ square matrix $M = [m_{ij}]_{i,j=1}^{n}$, denoted by $tr(M)$, is defined to be the sum of all diagonal entries of M. In other words,

$$tr(M) = \sum_{i=1}^{n} m_{ii}.$$

Example 3.2.31. Let $A = [a_{ij}]_{i,j=1}^{m,n}$ be an $m \times n$ matrix, then both AA^t and A^tA are well-defined and it is easy to verify that

$$tr(AA^t) = tr(A^tA) = \|A\|_{HS}^2.$$

Remark 3.2.32. The notion of the Hilbert-Schmidt norm and trace can be interpreted in statistical context. For $1 \leq i \leq N$, let $\boldsymbol{X}_i$ be a multivariate data point in $\mathbb{R}^p$ and $X = \begin{bmatrix} | & | & & | \\ \boldsymbol{X}_1 & \boldsymbol{X}_2 & \cdots & \boldsymbol{X}_N \\ | & | & & | \end{bmatrix}$ be the corresponding $p \times N$ matrix of observations. Let $\bar{\boldsymbol{X}} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{X}_i$ denote the sample mean of the observations and let B be the <u>mean-deviation form</u> of X defined by

$$B = \begin{bmatrix} | & | & & | \\ \boldsymbol{X}_1 - \bar{\boldsymbol{X}} & \boldsymbol{X}_2 - \bar{\boldsymbol{X}} & \cdots & \boldsymbol{X}_N - \bar{\boldsymbol{X}} \\ | & | & & | \end{bmatrix},$$

then the <u>total variance</u> of $\{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N\}$ is defined to be $\frac{1}{N-1} tr(B^t B) = \frac{1}{N-1}\|B\|_{HS}^2$. See Figure 3.3 for a geometric interpretation of the Hilbert-Schmidt norm.

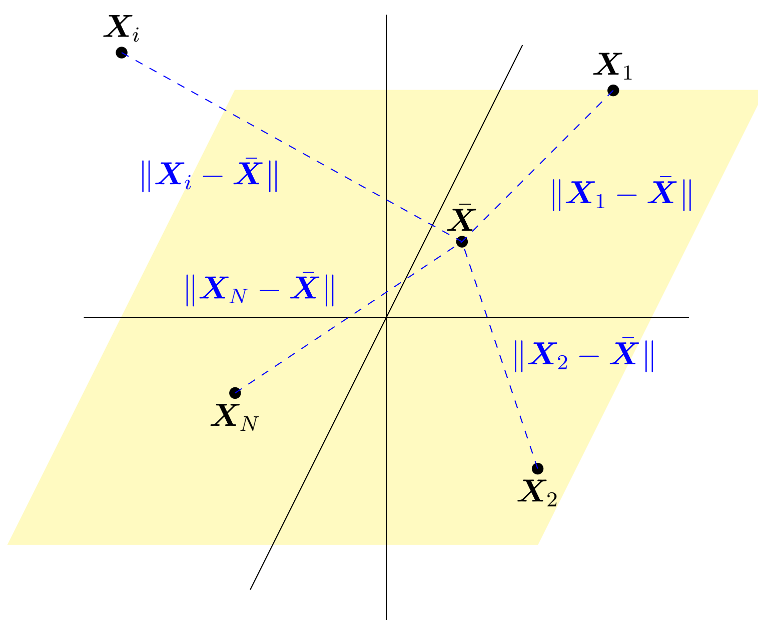

Figure 3.3: Geometric meaning of $\|B\|_{HS}$. Note that $\|B\|_{HS}^2 = \sum_{i=1}^{N} \|\boldsymbol{X}_i - \bar{\boldsymbol{X}}\|^2$.

We close this section with the inverse of a matrix.

Definition 3.2.33. Let A be an $n \times n$ square matrix. A is said to be <u>invertible</u> if there exists an $n \times n$ matrix B such that $AB = I_n = BA$, where I_n is the $n \times n$ identity matrix.

Remark 3.2.34. If A is invertible, then there is only one B such that $AB = I_n = BA$ (see Exercise 3.2.3). In this case, B is called the <u>inverse</u> of A and denoted by A^{-1}.

Problem 3.2.35. Let A be an invertible $n \times n$ matrix. Show that A^t is also invertible and $(A^t)^{-1} = (A^{-1})^t$.

Answer It is sufficient to show that $(A^{-1})^t A^t = A^t (A^{-1})^t = I_n$. Indeed, by Exercise 3.2.5,

$$(A^{-1})^t A^t = (AA^{-1})^t = I_n^t = I_n$$

and

$$A^t (A^{-1})^t = (A^{-1}A)^t = I_n^t = I_n.$$

■

Exercise 3.2.

1. Let $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$. Suppose that $\langle \boldsymbol{x}, \boldsymbol{z} \rangle = \langle \boldsymbol{y}, \boldsymbol{z} \rangle$ for all $\boldsymbol{z} \in \mathbb{R}^n$. Show that $\boldsymbol{x} = \boldsymbol{y}$.
Hint: It is sufficient to show that $\langle \boldsymbol{x} - \boldsymbol{y}, \boldsymbol{x} - \boldsymbol{y} \rangle = 0$. Take $\boldsymbol{z} = \boldsymbol{x} - \boldsymbol{y}$.

2. The goal of this exercise is to prove Cauchy-Schwarz Inequality. Let $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$.

(a) Show that Cauchy-Schwarz Inequality holds when $\boldsymbol{u} = \boldsymbol{0}$, so it remains to show the inequality for $\boldsymbol{u} \neq \boldsymbol{0}$.

(b) Let a, b, c be constants with $a > 0$. Show that $ax^2 - 2bx + c \geq 0$ for all $x \in \mathbb{R}$ if and only if $b^2 \leq ac$.
Hint: Consider the vertex of the parabola $y = ax^2 - 2bx + c = a(x - \frac{b}{a})^2 + \frac{ac-b^2}{a}$.

(c) Show that $\|\boldsymbol{u}\|^2 x^2 - 2\langle \boldsymbol{u}, \boldsymbol{v} \rangle x + \|\boldsymbol{v}\|^2 \geq 0$ for all $x \in \mathbb{R}$.
Hint: Expand $\langle x\boldsymbol{u} - \boldsymbol{v}, x\boldsymbol{u} - \boldsymbol{v} \rangle$.

(d) Use (b) and (c) to conclude that $(\langle \boldsymbol{u}, \boldsymbol{v} \rangle)^2 \leq \|\boldsymbol{u}\|^2 \|\boldsymbol{v}\|^2$ from which Cauchy-Schwarz Inequality follows.

3. Suppose that A, B, C are $n \times n$ matrices such that $AB = BA = AC = CA = I_n$. Show that $B = C$.

4. Suppose that A is an $n \times n$ square matrix such that $A^3 = O$, the zero matrix.

(a) Show that A is not invertible.

(b) Show that $I_n - A$ is invertible. Here I_n denotes the $n \times n$ identity matrix.
Hint: Consider $(I_n - A)(I_n + A + A^2)$.

5. Show that if AB is defined, then so is $B^t A^t$ and $(AB)^t = B^t A^t$. In general, if $A_1 A_2 \cdots A_k$ is defined, then show that $(A_1 A_2 \cdots A_k)^t = (A_k^t \cdots A_2^t A_1^t)$.

6. Let $tr(M)$ denote the trace of a square matrix M.

(a) Show that $tr(M) = tr(M^t)$.

(b) Show that $tr(A+B) = tr(A) + tr(B)$, provided that A and B are square matrices of the same dimensions.

(c) Show that $tr(AB) = tr(BA)$, provided both AB and BA are square matrices, not necessarily of same size.

(d) Show that $tr(A^tA) = 0$ if and only if $A = O$, the zero matrix.
Hint: See Example 3.2.31.

7. This exercise generalizes Problem 3.2.25. Let A be an $m \times n$ matrix, $\boldsymbol{u} \in \mathbb{R}^n$, and $\boldsymbol{v} \in \mathbb{R}^m$. Prove that $\langle A\boldsymbol{u}, \boldsymbol{v}\rangle = \langle \boldsymbol{u}, A^t\boldsymbol{u}\rangle$.

3.3 Vector Space and Matrix

Recall that the set of all n-vectors is denoted by $\mathbb{R}^n$.

Definition 3.3.1. Let $H \subseteq \mathbb{R}^n$. H is called a subspace of $\mathbb{R}^n$ if it has the following properties:

(a) $\mathbf{0} \in H$,

(b) If $\boldsymbol{u}, \boldsymbol{v} \in H$, then $\boldsymbol{u} + \boldsymbol{v} \in H$, and

(c) If $\boldsymbol{u} \in H$ and $c \in \mathbb{R}$, then $c\boldsymbol{u} \in H$.

Example 3.3.2. Let $H = \left\{ \begin{bmatrix} x \\ 0 \end{bmatrix} : x \in \mathbb{R} \right\}$. Clearly $\mathbf{0} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \in H$. Let $\boldsymbol{u} = \begin{bmatrix} x \\ 0 \end{bmatrix}$ and $\boldsymbol{v} = \begin{bmatrix} y \\ 0 \end{bmatrix}$, then $\boldsymbol{u} + \boldsymbol{v} = \begin{bmatrix} x+y \\ 0 \end{bmatrix}$ belongs to H. Finally, for any constant c and for any $\boldsymbol{u} = \begin{bmatrix} x \\ 0 \end{bmatrix}$ in H, $c\boldsymbol{u} = \begin{bmatrix} cx \\ 0 \end{bmatrix}$ belongs to H, too. This shows that H is a subspace of $\mathbb{R}^2$.

Example 3.3.3. $\{\mathbf{0}\}$ is a subspace of $\mathbb{R}^n$ and called the trivial subspace of $\mathbb{R}^n$. $\mathbb{R}^n$ itself is a subspace of $\mathbb{R}^n$.

Problem 3.3.4. Again consider $(\mathbb{R}^2, +, \cdot)$. Let $K = \left\{ \begin{bmatrix} x \\ x^2 \end{bmatrix} : x \in \mathbb{R} \right\}$. Is K a subspace of $\mathbb{R}^2$?

Answer No. $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 2 \\ 4 \end{bmatrix}$ both belong to K, but their sum $\begin{bmatrix} 3 \\ 5 \end{bmatrix}$ does not. ■

Definition 3.3.5. Given $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_p \in \mathbb{R}^n$ and $c_1, c_2, \ldots, c_p \in \mathbb{R}$, the vector $\boldsymbol{u} = c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_p\boldsymbol{v}_p$ is called a linear combination of $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_p$ with weights (or coefficients) $c_1, c_2, \ldots, c_p$. The span of $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_p\}$, denoted by $\text{Span}\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_p\}$, is the collection of all linear combinations of vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_p$.

Remark 3.3.6.

(a) What is the difference between $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_p\}$ and $\mathrm{Span}\{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_p\}$? The former contains only finitely many vectors but the second has infinitely many vectors in general. In fact, the former is a subset of the latter.

(b) Let S and T be finite subsets of $\mathbb{R}^n$. Suppose that S is a subset of T. Since T contains more vectors than S, the set of linear combinations of vectors in T contains that of S, that is, $\mathrm{Span}\, S \subseteq \mathrm{Span}\, T$. Exercise 3.3.2 is to show this rigorously.

Geometrically, $\mathrm{Span}\, S$ is the collection of all points that can be obtained by scaling and adding vectors in S. Figure 3.4 illustrates geometric interpretation of the span.

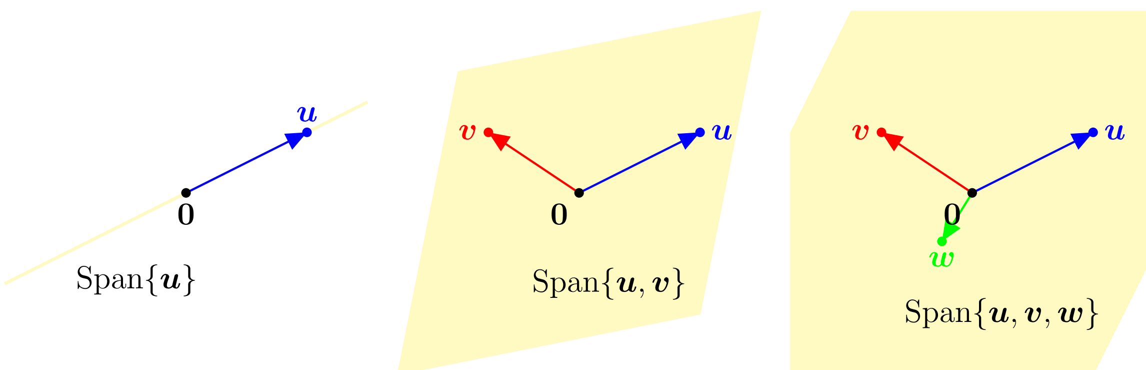

Figure 3.4: Geometric interpretation of span

Example 3.3.7. Let $\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\boldsymbol{v}_2 = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$. Then $\boldsymbol{u} = \begin{bmatrix} -1 \\ 4 \end{bmatrix}$ can be written as $2\boldsymbol{v}_1 + 3\boldsymbol{v}_2$, so $\boldsymbol{u}$ is a linear combination of $\boldsymbol{v}_1, \boldsymbol{v}_2$, or in other words $\boldsymbol{u} \in \mathrm{Span}\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$. $\mathbf{0} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ is in $\mathrm{Span}\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$ as well, because $\begin{bmatrix} 0 \\ 0 \end{bmatrix} = 0\boldsymbol{v}_1 + 0\boldsymbol{v}_2$.

Example 3.3.8. Let $A = \begin{bmatrix} | & | & & | \\ \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \\ | & | & & | \end{bmatrix}$ be an $m \times n$ matrix so that $\boldsymbol{v}_i \in \mathbb{R}^m$ for $1 \le i \le n$. Let $\boldsymbol{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$, then $A\boldsymbol{x} = x_1\boldsymbol{v}_1 + x_2\boldsymbol{v}_2 + \cdots x_n\boldsymbol{v}_n$. In other words, $A\boldsymbol{x}$ is a linear combination of columns of A.

Problem 3.3.9. Let $\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ -2 \\ -5 \end{bmatrix}$ and $\boldsymbol{v}_2 = \begin{bmatrix} 2 \\ 5 \\ 6 \end{bmatrix}$. Is $\boldsymbol{u} = \begin{bmatrix} 7 \\ 4 \\ -3 \end{bmatrix}$ in $\mathrm{Span}\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$?

Answer Note that

$$
\begin{aligned}
& \boldsymbol{u} \text{ is in } \operatorname{Span}\{\boldsymbol{v}_1, \boldsymbol{v}_2\} \\
\iff & \text{there are numbers } x_1, x_2 \text{ such that } \boldsymbol{u} = x_1\boldsymbol{v}_1 + x_2\boldsymbol{v}_2 \\
\iff & \text{the system } \left\{ \begin{array}{rrrr} x_1 & +2x_2 & = & 7 \\ -2x_2 & +5x_2 & = & 4 \\ -5x_1 & +6x_2 & = & -3 \end{array} \right. \text{ is consistent} \\
\iff & \text{the augmented matrix } \left[\begin{array}{rr|r} 1 & 2 & 7 \\ -2 & 5 & 4 \\ -5 & 6 & -3 \end{array}\right] \text{ is consistent}
\end{aligned}
$$

Since $\left[\begin{array}{rr|r} 1 & 2 & 7 \\ -2 & 5 & 4 \\ -5 & 6 & -3 \end{array}\right]$ reduces to $\left[\begin{array}{rr|r} 1 & 0 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{array}\right]$, we conclude that $\boldsymbol{u} \in \operatorname{Span}\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$. Moreover, we get $\boldsymbol{u} = 3\boldsymbol{v}_1 + 2\boldsymbol{v}_2$. ■

Theorem 3.3.10. *If $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_p$ are in $\mathbb{R}^n$, then Span$\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_p\}$ is a subspace of $\mathbb{R}^n$.*

Remark 3.3.11.

(a) The subspace in Theorem 3.3.10 is called the <u>subspace spanned by $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_p$</u>.

(b) In fact, if W is a subspace of $\mathbb{R}^n$, then there exists vectors $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k$ in $\mathbb{R}^n$ such that $W = \operatorname{Span}\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k\}$.

Definition 3.3.12. Let $S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p\}$ be a subset of $\mathbb{R}^n$. Let's consider a linear combination $c_1\boldsymbol{v}_1 + \cdots + c_p\boldsymbol{v}_p$ of $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p$. We can ask ourselves the following question: for which values of $c_1, \ldots, c_p$ does the linear combination $c_1\boldsymbol{v}_1 + \cdots + c_p\boldsymbol{v}_p$ become $\mathbf{0}$? Of course, if we take $c_1 = c_2 = \cdots = c_p = 0$, then clearly $c_1\boldsymbol{v}_1 + \cdots + c_p\boldsymbol{v}_p = \mathbf{0}$. So we have two possibilities:

Case 1: Taking $c_1 = c_2 = \cdots = c_p = 0$ is the *only* way to make $c_1\boldsymbol{v}_1 + \cdots + c_p\boldsymbol{v}_p = \mathbf{0}$.

Case 2: There are other choices of $c_1, c_2, \ldots, c_p$, *not all zero*, such that $c_1\boldsymbol{v}_1 + \cdots + c_p\boldsymbol{v}_p = \mathbf{0}$.

When Case 1 happens, the set $S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p\}$ is said to be <u>linearly independent</u>. When Case 2 happens, the set $S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p\}$ is said to be <u>linearly dependent</u>.

Problem 3.3.13. Show that $\left\{ \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \\ 2 \\ 0 \\ 1 \end{bmatrix} \right\}$ is linearly independent.

Answer Suppose $c_1 \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ 0 \\ -2 \\ 1 \\ 0 \end{bmatrix} + c_3 \begin{bmatrix} -3 \\ 0 \\ 2 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$, then

$$2c_1 + c_2 - 3c_3 = c_1 = -2c_2 + 2c_3 = c_2 = c_3 = 0,$$

from which it follows that $c_1 = c_2 = c_3 = 0$. We therefore conclude that the given set of vectors is linearly independent. ■

Problem 3.3.14. Show that $\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 4 \\ 6 \end{bmatrix} \right\}$ is linearly dependent.

Answer Since $4 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 3 \begin{bmatrix} 0 \\ 2 \end{bmatrix} - \begin{bmatrix} 4 \\ 6 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, the given set of vectors is linearly dependent. ■

Remark 3.3.15.

(a) If $S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p\}$ is linearly dependent, then we can find a vector $\boldsymbol{v}_i \in S$ such that $\boldsymbol{v}_i$ can be expressed as a linear combination of remaining $p - 1$ vectors.

(b) If $S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p\}$ is linearly independent, no vector in S can be written as a linear combination of remaining $p - 1$ vectors.

(c) If S is linearly independent and $T \subseteq S$, then T is linearly independent, too.

(d) If S is linearly dependent and $S \subseteq T$, then T is linearly dependent, too.

Definition 3.3.16. Let H be a subspace of $\mathbb{R}^n$ (H could be the whole $\mathbb{R}^n$). A set $\mathscr{B} = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p\}$ in $\mathbb{R}^n$ is said to be a <u>basis</u> for H if

(a) $\mathscr{B}$ is linearly independent, and

(b) $\operatorname{Span} \mathscr{B} = H$.

Remark 3.3.17. Condition (a) in Definition 3.3.16 states that $\mathscr{B}$ should not be too large, as large sets tend to be more likely to be linearly dependent. On the contrary, condition (b) in Definition 3.3.16 states that $\mathscr{B}$ should be large enough to cover all vectors in H. Thus a basis of H is an optimal set of vectors whose span covers all of H.

Example 3.3.18. Let $\mathscr{B} = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$. We will show that this is a basis for $\mathbb{R}^2$. First, we claim that the set $\mathscr{B}$ is linearly independent. Suppose that $c_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, then it follows that $c_1 = c_2 = 0$. Now we show that $\operatorname{Span} \mathscr{B} = \mathbb{R}^2$. Indeed, for any $\begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2$, we get $\begin{bmatrix} x \\ y \end{bmatrix} = x \begin{bmatrix} 1 \\ 0 \end{bmatrix} + y \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, a linear combination of $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$.

Problem 3.3.19. Show that $\mathscr{B}' = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$ is a basis for $\mathbb{R}^2$.

Answer Suppose that $c_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, then it follows that $c_1 + c_2 = c_2 = 0$, so consequently $c_1 = c_2 = 0$. This shows that $\mathscr{B}'$ is linearly independent. Now for any $\begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2$, we get $\begin{bmatrix} x \\ y \end{bmatrix} = (x - y) \begin{bmatrix} 1 \\ 0 \end{bmatrix} + y \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, a linear combination of $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$. This completes the proof that $\mathscr{B}'$ is a basis for $\mathbb{R}^2$. ■

Example 3.3.20. The set $S = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} \right\}$ is not a basis for $\mathbb{R}^3$, because, for example, $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ cannot be written as a linear combination of $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$.

Problem 3.3.21. Is the set $T = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 4 \\ 6 \end{bmatrix} \right\}$ a basis for $\mathbb{R}^2$?

Answer No. It is enough to show that T is linearly dependent, which was already shown in Problem 3.3.14. ■

Note that T in Problem 3.3.21 spans $\mathbb{R}^2$, so T fails to be a basis for $\mathbb{R}^2$ because it has too many vectors in it. One way to get a basis for $\mathbb{R}^2$ from T is to remove some unnecessary vectors, which is described in the following in more general setting.

Theorem 3.3.22 (The Spanning Set Theorem)**.** *Let $S = \{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_p\}$ be a subset of $\mathbb{R}^n$, and let $H = \mathrm{Span}\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_p\}$. If one of vectors in S, say $\boldsymbol{v}_k$, is a linear combination of the remaining vectors in S, then the set formed from S by removing $\boldsymbol{v}_k$ still spans H. That is,*

$$H = \mathrm{Span}\, S = \mathrm{Span}(S - \{\boldsymbol{v}_k\}) = \mathrm{Span}\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_{k-1}, \boldsymbol{v}_{k+1}, \ldots, \boldsymbol{v}_p\}.$$

Remark 3.3.23. Let S be as in the Spanning Set Theorem. If $\boldsymbol{v}_k$ in S is a linear combination of the remaining vectors in S, then S must be linearly dependent. So S cannot be a basis for $H = \mathrm{Span}\, S$. However, $S - \{\boldsymbol{v}_k\}$ still spans H. Keep removing vectors in S which are unnecessary to span H until no element can be written as a linear combination of others, and this will produce a basis for $H = \mathrm{Span}\, S$. This, combined with Remark 3.3.11, shows that any subspace W of $\mathbb{R}^n$ can be written as $W = \mathrm{Span}\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k\}$ for some linearly independent set $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k\}$ in $\mathbb{R}^n$ and in this case $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k\}$ is a basis for W.

Problem 3.3.24. Let's consider a set $S = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix} \right\}$ in $\mathbb{R}^3$. Find a basis for $\operatorname{Span} S$.

Answer Since $\begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix}$ can be written as a linear combination of $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, by the Spanning Set Theorem, $\operatorname{Span} S = \operatorname{Span} \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$. Now one can easily verify that $\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$ is linearly independent, so $\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$ is a basis for $\operatorname{Span} S$. ■

As seen in Example 3.3.18 and Problem 3.3.19, basis is not unique. However, both $\mathscr{B}$ and $\mathscr{B}'$ in Example 3.3.18 and Problem 3.3.19 contain the same number of vectors. This is not a coincidence. In fact, we have the following result.

Theorem 3.3.25. *Let H be a subspace of $\mathbb{R}^n$. If H has a basis of n vectors, then every basis of H must consist of exactly n vectors.*

Definition 3.3.26. The number of vectors in a basis for H is called the <u>dimension</u> of H and is denoted by $\dim H$.

Example 3.3.27. It is easy to check that $\left\{ \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \dots, \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \right\}$ is a basis for $\mathbb{R}^n$, so it follows that $\dim \mathbb{R}^n = n$.

Example 3.3.28. Let $H = \operatorname{Span} \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix} \right\}$. What is the dimension of H?

Answer Note that the answer is *not* 4: since $\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix} \right\}$ is linearly dependent, so it is not a basis for H. Note that in Problem 3.3.24 we show that $H = \operatorname{Span} \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$ and $\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$ is a basis for H. It follows that $\dim H = 2$.
■

Remark 3.3.29. In general, if H is a subspace of $\mathbb{R}^n$, then $\dim H \leq n$ and $\dim H = n$ if and only if $H = \mathbb{R}^n$.

Now we study vector spaces associated with a given matrix. We begin with the null space of a given matrix.

Definition 3.3.30. The null space of an $m \times n$ matrix A, written as $\operatorname{Nul} A$, is the set of all solutions to the linear system $A\boldsymbol{x} = \mathbf{0}$. In set notations,

$$\operatorname{Nul} A = \{\boldsymbol{x} \in \mathbb{R}^n : A\boldsymbol{x} = \mathbf{0}\}.$$

Example 3.3.31. Let $A = \begin{bmatrix} 1 & -3 & -2 \\ -5 & 9 & 1 \end{bmatrix}$, then $\boldsymbol{u} = \begin{bmatrix} 5 \\ 3 \\ -2 \end{bmatrix}$ is in $\operatorname{Nul} A$ since $A\boldsymbol{u} = \mathbf{0}$.

The following can be easily shown using Definition 3.3.1.

Theorem 3.3.32. *The null space of an $m \times n$ matrix is a subspace of $\mathbb{R}^n$.*

Problem 3.3.33. Let $A = \begin{bmatrix} -3 & 6 & -1 & 1 & -7 \\ 1 & -2 & 2 & 3 & -1 \\ 2 & -4 & 5 & 8 & -4 \end{bmatrix}$. Find $\operatorname{Nul} A$. Describe $\operatorname{Nul} A$ as the span of a set of vectors. What is the $\dim \operatorname{Nul} A$?

Answer We start with solving a linear system $A\boldsymbol{x} = \mathbf{0}$:

$$\left\{ \begin{array}{rrrrrl} -3x_1 & +6x_2 & -x_3 & +x_4 & -7x_5 = & 0 \\ x_1 & -2x_2 & +2x_3 & +3x_4 & -x_5 = & 0 \\ 2x_1 & -4x_2 & +5x_3 & +8x_4 & -4x_5 = & 0 \end{array} \right. .$$

Since the reduced row echelon form of A is $\begin{bmatrix} 1 & -2 & 0 & -1 & 3 \\ 0 & 0 & 1 & 2 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$, $\operatorname{Nul} A$ is given by

$$\left\{ \begin{array}{ll} x_1 & = 2x_2 + x_4 - 3x_5 \\ x_2 & \text{ is free} \\ x_3 & = -2x_4 + 2x_5 \\ x_4 & \text{ is free} \\ x_5 & \text{ is free} \end{array} \right. .$$

Note that this solution set can be more compactly described in the following way:

$$\begin{aligned}
\operatorname{Nul} A &= \left\{ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} : \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 2r+s-3t \\ r \\ -2s+2t \\ s \\ t \end{bmatrix}, r, s, t \in \mathbb{R} \right\} \\
&= \left\{ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} : \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = r \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + s \begin{bmatrix} 1 \\ 0 \\ -2 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} -3 \\ 0 \\ 2 \\ 0 \\ 1 \end{bmatrix}, r, s, t \in \mathbb{R} \right\} \\
&= \operatorname{Span} \left\{ \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \\ 2 \\ 0 \\ 1 \end{bmatrix} \right\}.
\end{aligned}$$

It is easy to check that $\left\{ \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \\ 2 \\ 0 \\ 1 \end{bmatrix} \right\}$ is linearly independent (see Problem 3.3.13), so we conclude that $\dim \operatorname{Nul} A = 3$. ∎

Remark 3.3.34. For a matrix A, $\dim \operatorname{Nul} A$ is called the nullity of A and denoted by $\operatorname{nullity}(A)$. In Problem 3.3.33, the nullity of A turns out to be the number of free variables, 3, of the system $A\boldsymbol{x} = \mathbf{0}$. In fact, this is true in general, that is,

$$\begin{aligned}
\operatorname{nullity}(A) &= \text{number of free variables of the system } A\boldsymbol{x} = \mathbf{0} \\
\text{(Problem 3.2.23, Remark 3.1.40)} &= \text{number of non-pivot columns of } A.
\end{aligned}$$

Next we consider another important vector space associated with a matrix.

Definition 3.3.35. For an $m \times n$ matrix $A = \begin{bmatrix} | & | & & | \\ \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \\ | & | & & | \end{bmatrix}$, the column space of A, written as $\operatorname{Col} A$, is $\operatorname{Span}\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$, where $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n$ are in $\mathbb{R}^m$.

Remark 3.3.36. By Theorem 3.3.10, the column space of an $m \times n$ matrix is a subspace of $\mathbb{R}^m$.

Example 3.3.37. Let $A = \begin{bmatrix} 5 & 4 \\ 0 & 1 \\ -1 & 0 \end{bmatrix}$, then $\operatorname{Col} A = \operatorname{Span}\left\{\begin{bmatrix} 5 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 4 \\ 1 \\ 0 \end{bmatrix}\right\}$, which is a subspace of $\mathbb{R}^3$. To see if $\boldsymbol{b} = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$ is in $\operatorname{Col} A$, we check if the system $A\boldsymbol{x} = \boldsymbol{b}$ is consistent. Note that the augmented matrix $\left[\begin{array}{cc|c} 5 & 4 & -1 \\ 0 & 1 & 1 \\ -1 & 0 & 1 \end{array}\right]$ reduces to $\left[\begin{array}{cc|c} 1 & 0 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{array}\right]$, so by Theorem 3.1.39, the system $A\boldsymbol{x} = \boldsymbol{b}$ is consistent and hence we conclude that $\boldsymbol{b}$ is sitting in $\operatorname{Col} A$. Indeed, $\begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} = -\begin{bmatrix} 5 \\ 0 \\ -1 \end{bmatrix} + \begin{bmatrix} 4 \\ 1 \\ 0 \end{bmatrix}$.

Remark 3.3.38. What we have learned so far can be summarized in the following. Let A be an $m \times n$ matrix. Then

(a) $\boldsymbol{u} \in \mathbb{R}^n$ is in $\operatorname{Nul} A$ if and only if $A\boldsymbol{u} = \boldsymbol{0}$.

(b) $\boldsymbol{u} \in \mathbb{R}^m$ is in $\operatorname{Col} A$ if and only if the system $A\boldsymbol{x} = \boldsymbol{u}$ is consistent.

If A is an $m \times n$ matrix, then $\operatorname{Col} A$ is a subspace of $\mathbb{R}^m$. How can we find a basis for $\operatorname{Col} A$? Roughly speaking, what are the *essential* columns of A when constructing $\operatorname{Col} A$?

Theorem 3.3.39. *The pivot columns of a matrix A form a basis for* $\operatorname{Col} A$. *In particular,* $\dim \operatorname{Col} A = \operatorname{rank}(A)$ *(see Definition 3.1.36).*

Example 3.3.40. Let $A = \begin{bmatrix} 1 & 4 & 0 & 2 & -1 \\ 3 & 12 & 1 & 5 & 5 \\ 2 & 8 & 1 & 3 & 2 \\ 5 & 20 & 2 & 8 & 8 \end{bmatrix}$, then it reduces to $\begin{bmatrix} 1 & 4 & 0 & 2 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$. Therefore, by Theorem 3.3.39, $S = \left\{\begin{bmatrix} 1 \\ 3 \\ 2 \\ 5 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ 5 \\ 2 \\ 8 \end{bmatrix}\right\}$ is a basis for $\operatorname{Col} A$. In other words, S is linearly independent and $\operatorname{Span} S = \operatorname{Col} A$.

We collect some properties of rank.

Theorem 3.3.41. *We have the following properties involving rank.*

(a) For an $m \times n$ matrix A, $\operatorname{rank}(A) \le \min\{m, n\}$.

(b) If $[A|B]$ denotes a matrix that is formed by attaching a matrix B to A side-by-side (so A and B have a same number of rows), then rank$(A) \le \operatorname{rank}([A|B])$.

(c) *If A is an $m \times k$ matrix and B is a $k \times n$ matrix, then*

$$\operatorname{rank}(A) + \operatorname{rank}(B) - k \le \operatorname{rank}(AB) \le \min\{\operatorname{rank}(A), \operatorname{rank}(B)\}.$$

(d) *If A, B are matrices of the same dimensions, then* $\operatorname{rank}(A+B) \le \operatorname{rank}(A) + \operatorname{rank}(B)$.

(e) $\operatorname{rank}(A) = \operatorname{rank}(A^t)$.

(f) $\operatorname{rank}(A^t A) = \operatorname{rank}(A)$.

Example 3.3.42. How to find a basis for $\operatorname{Nul} A$? Note that this was already covered in Problem 3.3.33. First of all, to find a basis for $\operatorname{Nul} A$, one needs to find $\operatorname{Nul} A$. For example, for the matrix in Example 3.3.40 above, we see that the general solution is given by

$$\begin{cases} x_1 &= -4x_2 - 2x_4 \\ x_2 & \text{is free} \\ x_3 &= x_4 \\ x_4 & \text{is free} \\ x_5 &= 0 \end{cases}.$$

That is, $\operatorname{Nul} A = \left\{ x_2 \begin{bmatrix} -4 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -2 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} : x_2, x_4 \text{ are free} \right\} = \operatorname{Span} \left\{ \begin{bmatrix} -4 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} \right\}$.

We can show that the set $\left\{ \begin{bmatrix} -4 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} \right\}$ is linearly independent and hence we conclude that $\left\{ \begin{bmatrix} -4 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} \right\}$ is a basis for $\operatorname{Nul} A$.

We close this section with the relation between $\operatorname{nullity}(A)$ and $\operatorname{rank}(A)$, which is an immediate consequence of Remark 3.2.23 and Theorem 3.3.39, since any column of A is either pivotal or non-pivotal.

Theorem 3.3.43 (Rank-Nullity Theorem)**.** *If A is an $m \times n$ matrix, then*

$$\operatorname{nullity}(A) + \operatorname{rank}(A) = n.$$

Exercise 3.3.

1. Let $L = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} : x \geq 0 \text{ and } y \geq 0 \right\}$. Is this a subspace of $\mathbb{R}^2$?

2. Let S and T be finite subsets of $\mathbb{R}^n$ with $S \subseteq T$. Show that $\operatorname{Span} S \subseteq \operatorname{Span} T$.

3. Let $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p\}$ be a linearly independent set of vectors in $\mathbb{R}^n$. Let A be an $n \times n$ invertible matrix. Show that $\{A\boldsymbol{x}_1, \ldots, A\boldsymbol{x}_p\}$ is linearly independent.

4. Let $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k\}$ be a linearly independent subset of $\mathbb{R}^n$. Suppose $\boldsymbol{v} \in \mathbb{R}^n - \operatorname{Span}\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k\}$. Show that $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k, \boldsymbol{v}\}$ is also linearly independent.

5. Let $K = \operatorname{Span} \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} \right\}$. What is the dimension of K?

6. Find the null space of $\begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.

7. Let $B = \begin{bmatrix} 1 & 4 & 0 & 2 \\ 3 & 12 & 1 & 5 \\ 2 & 8 & 1 & 3 \\ 5 & 20 & 2 & 8 \end{bmatrix}$. Find a basis for $\operatorname{Col} B$.

8. Let $S = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p\}$ be a subset of $\mathbb{R}^n$ with $p > n$. Show that S is linearly dependent.

Hint: Suppose that S is linearly independent. Let $A = \begin{bmatrix} | & | & & | \\ \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_p \\ | & | & & | \end{bmatrix}$, then A is an $n \times p$ matrix and the system $A\boldsymbol{x} = \boldsymbol{0}$ has a unique solution, say $\boldsymbol{x} = \boldsymbol{0}$. This means that all columns of A are pivotal, implying that $\operatorname{rank}(A) = p$ (see Remark 3.1.40). You may want to apply Remark 3.3.29 to derive a contradiction.

9. (a) Show that $\operatorname{rank}(A^t A) = \operatorname{rank}(A A^t)$.

(b) Show that $\operatorname{rank}(AB) \neq \operatorname{rank}(BA)$ in general.

3.4 Determinants

Definition 3.4.1. Let $A = [a_{ij}]_{i,j=1}^n$ be an $n \times n$ matrix. The <u>determinant</u> of A, denoted by $\det A$ or $|A|$, is the number defined by

$$\det A = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^n a_{i\sigma(i)},$$

where S_n is the symmetric group on n letters and $\operatorname{sgn}(\sigma)$ denotes the <u>signature</u> of σ, that is, $\operatorname{sgn}(\sigma) = (-1)^{\text{the number of inversions of } \sigma}$ (see Remark 1.3.12).

Example 3.4.2. Let $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, then $\det A = a_{11}a_{22} - a_{12}a_{21}$. To see that this is consistent with Definition 3.4.1, we note that $S_2 = \{\iota, \tau\}$, where $\iota(1) = 1$, $\iota(2) = 2$, $\tau(1) = 2$, and $\tau(2) = 1$. Since ι has no inversions and τ has only one inversion, it follows that $\operatorname{sgn}(\iota) = (-1)^0 = 1$ and $\operatorname{sgn}(\tau) = (-1)^1 = -1$, and that

$$\det A = \operatorname{sgn}(\iota)a_{1\iota(1)}a_{2\iota(2)} + \operatorname{sgn}(\tau)a_{1\tau(1)}a_{2\tau(2)} = a_{11}a_{22} - a_{12}a_{21}.$$

Exercise 3.4.1 asks to compute the determinant of a 3×3 matrix using Definition 3.4.1.

For the determinant of a square matrix of size 4×4 or bigger, Definition 3.4.1 has little practical value, in which case we can make use of an inductive property of the determinant.

Definition 3.4.3. Let A be an $n \times n$ matrix. The $\underline{(i, j)\text{-minor}}$ of A is the determinant of the $(n-1) \times (n-1)$ submatrix formed by deleting the i^{th} row and j^{th} column of A. The $\underline{(i, j)\text{-cofactor}}$ is the number defined by $(-1)^{i+j}M_{ij}$, where M_{ij} denotes the (i, j)-minor of A.

Theorem 3.4.4 (Cofactor Expansion of the Determinant)**.** *Let A be an $n \times n$ matrix. Let C_{ij} denote the (i, j)-cofactor of A, then*

(a) For any i, $\det A = \sum_{j=1}^{n} a_{ij}C_{ij}$. That is, $\det A$ is obtained by the cofactor expansion along the i^{th} row, for any i.

(b) For any j, $\det A = \sum_{i=1}^{n} a_{ij}C_{ij}$. That is, $\det A$ is obtained by the cofactor expansion along the j^{th} column, for any j.

Problem 3.4.5. Let $A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$. Compute $\det A$ using the cofactor expansion along

(a) the first row.

(b) the second column.

Answer Using the formula for the determinant of a 2×2 matrix (Example 3.4.2) and the cofactor expansion along the first row, we have

$$\begin{aligned}
\det A &= a \cdot (-1)^{1+1} \det \begin{bmatrix} e & f \\ h & i \end{bmatrix} + b \cdot (-1)^{1+2} \det \begin{bmatrix} d & f \\ g & i \end{bmatrix} + c \cdot (-1)^{1+3} \det \begin{bmatrix} d & e \\ g & h \end{bmatrix} \\
&= a(ei - fh) - b(di - fg) + c(dh - eg) \\
&= aei + bfg + cdh - afh - bdi - ceg.
\end{aligned}$$

Similarly, the cofactor expansion along the second column gives

$$\begin{aligned}\det A &= b\cdot(-1)^{1+2}\det\begin{bmatrix} d & f \\ g & i\end{bmatrix} + e\cdot(-1)^{2+2}\det\begin{bmatrix} a & c \\ g & i\end{bmatrix} + h\cdot(-1)^{3+2}\det\begin{bmatrix} a & c \\ d & f\end{bmatrix}\\ &= -b(di-fg)+e(ai-cg)-h(af-cd)\\ &= aei+bfg+cdh-afh-bdi-ceg.\end{aligned}$$

Note that these two expansions yield the same result, as expected. ■

Problem 3.4.6. Compute $\det A$, where

$$A = \begin{bmatrix} 4 & 1 & 2 & 0 \\ 0 & 0 & -2 & 1 \\ 0 & 3 & 0 & 0 \\ 5 & -2 & 1 & 2 \end{bmatrix}.$$

Answer Note that the third row contains three zeros, so it will be most beneficial to use the cofactor expansion along with the third row. Indeed,

$$\det A = 3\cdot(-1)^{3+2}\det\begin{bmatrix} 4 & 2 & 0 \\ 0 & -2 & 1 \\ 5 & 1 & 2\end{bmatrix} = (-3)\cdot\det\begin{bmatrix} 4 & 2 & 0 \\ 0 & -2 & 1 \\ 5 & 1 & 2\end{bmatrix}$$

and the problem boils down to finding $\det\begin{bmatrix} 4 & 2 & 0 \\ 0 & -2 & 1 \\ 5 & 1 & 2\end{bmatrix}$. Using the cofactor expansion along with the first row, we obtain

$$\det\begin{bmatrix} 4 & 2 & 0 \\ 0 & -2 & 1 \\ 5 & 1 & 2\end{bmatrix} = 4\cdot((-2)\cdot 2 - 1\cdot 1) - 2\cdot(0\cdot 2 - 1\cdot 5) = -10$$

and finally we have $\det A = (-3)\cdot(-10) = 30$. ■

Example 3.4.7. Let $A = \begin{bmatrix} 4 & -1 & 9 \\ 0 & 1 & 2 \\ 0 & 0 & -2\end{bmatrix}$, an upper triangular matrix. The cofactor expansion along the first column gives $\det A = 4\cdot\det\begin{bmatrix} 1 & 2 \\ 0 & -2\end{bmatrix} = 4\cdot 1\cdot(-2)$, the product of all diagonal elements of A. More generally, an application of mathematical induction combined with the cofactor expansion along the first column shows that the determinant of an upper triangular matrix is the product of all diagonal elements. In particular, the determinant of a diagonal matrix is the product of all diagonal elements. Similarly, the cofactor expansion along the first row shows that the determinant of a lower triangular matrix is the product of all diagonal elements.

When a given matrix does not contain many zeros, even the cofactor expansion has limited use. This difficulty, however, can be overcome by transforming the matrix into one for which the determinant is easy to compute. To explain this idea, we start with the following result that investigates the affect of elementary row operations on the determinant.

Theorem 3.4.8. *Let A be a square matrix.*

(a) *If one row (or one column) of A is multiplied by c to produce B, then* $\det B = c\cdot\det A$.

(b) *If two rows (or two columns) of A are interchanged to produce B, then* $\det B = -\det A$.

(c) *If a multiple of one row (respectively, column) of A is added to another row (respectively, column) to produce a matrix B, then* $\det B = \det A$.

Problem 3.4.9. Compute $\det A$, where $A = \begin{bmatrix} 3 & -2 & 0 & 5 \\ -3 & 5 & 5 & 2 \\ 6 & -1 & -1 & 13 \\ 3 & -2 & 0 & 7 \end{bmatrix}$.

Answer Let B be the matrix obtained by applying replacement operations $R_2 \mapsto R_2 + R_1$, $R_3 \mapsto R_3 - 2R_1$, and $R_4 \mapsto R_4 - R_1$ to A, then $B = \begin{bmatrix} 3 & -2 & 0 & 5 \\ 0 & 3 & 5 & 7 \\ 0 & 3 & -1 & 3 \\ 0 & 0 & 0 & 2 \end{bmatrix}$ and $\det B = \det A$ by Theorem 3.4.8.(c). Let C be the matrix obtained from B by applying the replacement operation $R_3 \mapsto R_3 - R_2$, then $C = \begin{bmatrix} 3 & -2 & 0 & 5 \\ 0 & 3 & 5 & 7 \\ 0 & 0 & -6 & -4 \\ 0 & 0 & 0 & 2 \end{bmatrix}$ and $\det C = \det B$. Since C is an upper triangular matrix, by Example 3.4.7, we see that $\det A = \det B = \det C = 3 \cdot 3 \cdot (-6) \cdot 2 = -108$.

■

Theorem 3.4.10. *For any $n \times n$ square matrices A and B, we have the following properties:*

(a) $\det(AB) = \det(A)\det(B) = \det(BA)$.

(b) $\det A^t = \det A$.

(c) $\det(cA) = c^n \det A$ *for any scalar c.*

Theorem 3.4.11 (The Invertible Matrix Theorem). *Let A be an $n \times n$ matrix. The following statements are equivalent:*

(a) *A is invertible.*

(b) $\det A \neq 0$.

(c) The system $A\boldsymbol{x} = \boldsymbol{b}$ *has a unique solution for all* n *vector* $\boldsymbol{b}$.

(d) $\mathbf{0}$ *is the only solution to the homogeneous system* $A\boldsymbol{x} = \mathbf{0}$.

(e) A^t *is invertible.*

(f) Columns of A *are linearly independent.*

(g) $\operatorname{rank}(A) = n$.

Problem 3.4.12. Suppose that A is an invertible $n \times n$ matrix. Show that $\det A^{-1} = \dfrac{1}{\det A}$.

Answer First we note that $\det A \neq 0$. Since $A^{-1}A = I_n$, we have $\det(A^{-1}A) = \det I_n = 1$. By Theorem 3.4.10, $\det(A^{-1}) \cdot \det A = \det(A^{-1}A) = 1$ and the result follows. ∎

Problem 3.4.13. Suppose that A is an invertible $n \times n$ matrix. What is $\operatorname{Nul} A$? What is $\operatorname{Col} A$?

Answer If A is invertible, then $\mathbf{0}$ is the only solution to the system $A\boldsymbol{x} = \mathbf{0}$, so $\operatorname{Nul} A = \{\mathbf{0}\}$. On the other hand, if A is invertible, then for any $\boldsymbol{b} \in \mathbb{R}^n$, the system $A\boldsymbol{x} = \boldsymbol{b}$ is consistent, since $\boldsymbol{x}$ is then given by $\boldsymbol{x} = A^{-1}\boldsymbol{b}$. This shows that $\operatorname{Col} A = \mathbb{R}^n$. ∎

Exercise 3.4.

1. Let $A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$. Using Definition 3.4.1, show that

$$\det A = aei + bfg + cdh - afh - bdi - ceg.$$

2. Verify the following <u>Vandermonde determinant</u>:

$$\det \begin{bmatrix} 1 & a & a^2 \\ 1 & b & b^2 \\ 1 & c & c^2 \end{bmatrix} = (a-b)(b-c)(c-a).$$

Hint: Use Theorem 3.4.8 to show that

$$\det \begin{bmatrix} 1 & a & a^2 \\ 1 & b & b^2 \\ 1 & c & c^2 \end{bmatrix} = \det \begin{bmatrix} 1 & a & a^2 \\ 0 & b-a & b^2-a^2 \\ 0 & c-b & c^2-b^2 \end{bmatrix} = (b-a)(c-b) \det \begin{bmatrix} 1 & a & a^2 \\ 0 & 1 & b+a \\ 0 & 1 & c+b \end{bmatrix}.$$

3. Compute $\det\begin{bmatrix} 1 & -3 & 1 & -2 \\ 2 & -5 & -1 & -2 \\ 0 & -4 & 5 & 1 \\ -3 & 10 & -6 & 9 \end{bmatrix}$.

4. This exercise is related with the use of dummy variables in multiple linear regression with categorical predictors. Determine whether X^tX is invertible where

$$X = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix}.$$

What if $X = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$?

Hint: You may want to use the Invertible Matrix Theorem combined with Theorem 3.3.41.

3.5 Orthogonal Projection

In this section, we look further into the geometric interpretation of statistics. First we begin with a projection of a vector onto another. Let $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$. Consider a problem of finding the projection $\mathrm{proj}_{\boldsymbol{y}}$ of $\boldsymbol{x}$ onto $\boldsymbol{y}$ (see Figure 3.5).

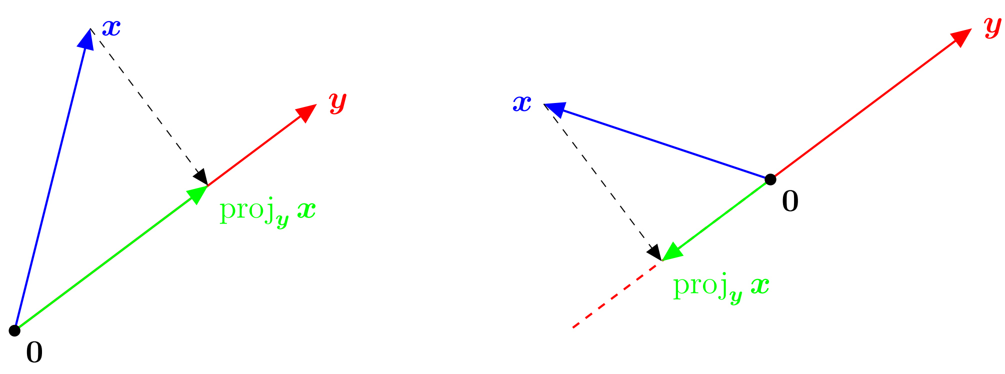

Figure 3.5: Projection onto a vector

Let $\boldsymbol{z}$ be the projection of $\boldsymbol{x}$ onto $\boldsymbol{y}$. Since $\boldsymbol{z}$ is parallel to $\boldsymbol{y}$, there is a constant c such that $\boldsymbol{z} = c\boldsymbol{y}$. Since $\boldsymbol{z} - \boldsymbol{x}$ (the broken arrow in Figure 3.5) should be orthogonal to $\boldsymbol{z}$, we get

$$\langle \boldsymbol{z}-\boldsymbol{x}, \boldsymbol{z}\rangle = 0 \quad \text{that is} \quad c^2\langle \boldsymbol{y}, \boldsymbol{y}\rangle = c\langle \boldsymbol{x}, \boldsymbol{y}\rangle$$

from which a nontrivial solution $c = \frac{\langle \boldsymbol{x},\boldsymbol{y}\rangle}{\langle \boldsymbol{y},\boldsymbol{y}\rangle}$ follows. In summary, $\operatorname{proj}_{\boldsymbol{y}} \boldsymbol{x} = \frac{\langle \boldsymbol{x},\boldsymbol{y}\rangle}{\langle \boldsymbol{y},\boldsymbol{y}\rangle}\boldsymbol{y}$.

The projection onto a subspace of $\mathbb{R}^n$ is now in order. Let W be a subspace of $\mathbb{R}^n$. The orthogonal complement of W, denoted $W^\perp$, is defined to be

$$W^\perp = \{\boldsymbol{x} \in \mathbb{R}^n : \langle \boldsymbol{x}, \boldsymbol{w}\rangle = 0 \text{ for all } \boldsymbol{w} \in W\}.$$

It is not difficult to show that $W^\perp$ is also a subspace of $\mathbb{R}^n$.

Example 3.5.1. Let $W = \operatorname{Span}\left\{\begin{bmatrix}1\\0\\0\end{bmatrix}\right\}$, then it is a subspace of $\mathbb{R}^3$ and

$$\begin{aligned} W^\perp &= \left\{\begin{bmatrix}x\\y\\z\end{bmatrix} : \begin{bmatrix}1\\0\\0\end{bmatrix}\cdot\begin{bmatrix}x\\y\\z\end{bmatrix} = 0, x, y, z \in \mathbb{R}\right\} \\ &= \left\{\begin{bmatrix}x\\y\\z\end{bmatrix} : x = 0, y, z \in \mathbb{R}\right\} \\ &= \left\{\begin{bmatrix}0\\y\\z\end{bmatrix} : y, z \in \mathbb{R}\right\} \\ &= \operatorname{Span}\left\{\begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}\right\}. \end{aligned}$$

Theorem 3.5.2. *Let W be a subspace of $\mathbb{R}^n$ and $W^\perp$ be its orthogonal complement. Let $\boldsymbol{x} \in \mathbb{R}^n$, then there is a unique pair $(\boldsymbol{w}, \boldsymbol{w}^\perp)$ such that $\boldsymbol{w} \in W$, $\boldsymbol{w}^\perp \in W^\perp$, and $\boldsymbol{x} = \boldsymbol{w}+\boldsymbol{w}^\perp$.*

Definition 3.5.3. Let $\boldsymbol{x}, \boldsymbol{w}$, and $\boldsymbol{w}^\perp$ be as in Theorem 3.5.2. $\boldsymbol{w}$ (respectively, $\boldsymbol{w}^\perp$) is called the orthogonal projection of $\boldsymbol{x}$ onto W (respectively, onto $W^\perp$) and is denoted by $\boldsymbol{w} = \operatorname{proj}_W \boldsymbol{x}$ (respectively, $\boldsymbol{w}^\perp = \operatorname{proj}_{W^\perp} \boldsymbol{x}$). The unique pair $(\boldsymbol{w}, \boldsymbol{w}^\perp)$ is called the orthogonal decomposition of $\boldsymbol{x}$ with respect to $(W, W^\perp)$. See Figure 3.6.

Example 3.5.4. Let W be as in Example 3.5.1, then the orthogonal decomposition of $\begin{bmatrix}2\\3\\1\end{bmatrix}$ with respect to $(W, W^\perp)$ is $\left(\begin{bmatrix}2\\0\\0\end{bmatrix}, \begin{bmatrix}0\\3\\1\end{bmatrix}\right)$.

We now consider a problem of finding the orthogonal decomposition in a more general setting.

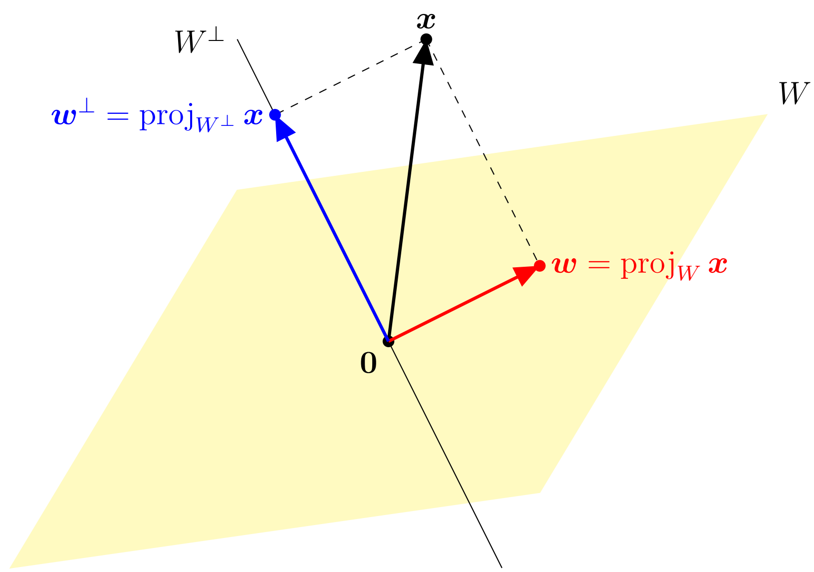

Figure 3.6: Orthogonal decomposition

Problem 3.5.5. Let W be a subspace of $\mathbb{R}^n$. Show that there is a unique $n \times n$ matrix P such that $P\boldsymbol{z} = \mathrm{proj}_W\,\boldsymbol{z}$ for all $\boldsymbol{z} \in \mathbb{R}^n$.

Answer By Remark 3.3.23, there is a linearly independent set $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k\}$ such that $W = \mathrm{Span}\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k\}$ (so it is necessary that $k \le n$ – see Exercise 3.3.8). Define $A = \begin{bmatrix} | & | & & | \\ \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_k \\ | & | & & | \end{bmatrix}$, then $W = \mathrm{Col}\,A$. Let $\boldsymbol{z} \in \mathbb{R}^n$. Since $\boldsymbol{w} = \mathrm{proj}_W\,\boldsymbol{z} \in W$, there is $\boldsymbol{x} \in \mathbb{R}^k$ such that $A\boldsymbol{x} = \boldsymbol{w}$. Since $\boldsymbol{u}_i \in W$ for all i, $1 \le i \le k$, it follows that $\langle \boldsymbol{u}_i, \boldsymbol{w}^\perp \rangle = \langle \boldsymbol{u}_i, \boldsymbol{z} - A\boldsymbol{x} \rangle = \langle \boldsymbol{u}_i, \boldsymbol{z} \rangle - \langle \boldsymbol{u}_i, A\boldsymbol{x} \rangle = 0$ for all i, or equivalently, $A^t\boldsymbol{z} = A^tA\boldsymbol{x}$. Since $\mathrm{rank}(A^tA) = \mathrm{rank}(A) = \dim W = k$, A^tA is an invertible $k \times k$ matrix and hence $\boldsymbol{w} = A\boldsymbol{x} = P\boldsymbol{z}$, where $P = A(A^tA)^{-1}A^t$. Uniqueness of P follows from the fact that the i^{th} column of P is given by $P\boldsymbol{e}_i$, where $\boldsymbol{e}_i$ is as in Example 3.2.24, and equals $\mathrm{proj}_W\,\boldsymbol{e}_i$. ■

Remark 3.5.6. Let P_W denote the matrix $A(A^tA)^{-1}A^t$ as in the answer of Problem 3.5.5. P_W is called the projection matrix corresponding to W. Problem 3.5.5 shows that the orthogonal projection of $\boldsymbol{z} \in \mathbb{R}^n$ onto W can be obtained by taking the product between P_W and $\boldsymbol{z}$.

Problem 3.5.7. Let W be a subspace of $\mathbb{R}^n$ and let P_W denote the corresponding projection matrix. Show that $P_W = P_W^t$ and $P_W^2 = P_W$.

Answer By Remark 3.3.23, there is a linearly independent set $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k\}$ such that

$W = \text{Span}\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k\}$. Let $A = \begin{bmatrix} | & | & & | \\ \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_k \\ | & | & & | \end{bmatrix}$, then $P_W = A(A^tA)^{-1}A^t$ (see Remark 3.5.6). Now it is easy to check that (see Exercise 3.2.5 and Problem 3.2.35)

$$P_W^t = (A(A^tA)^{-1}A^t)^t = (A^t)^t((A^tA)^{-1})^tA^t = A((A^tA)^t)^{-1}A^t = P_W$$

and

$$P_W^2 = (A(A^tA)^{-1}A^t)(A(A^tA)^{-1}A^t) = A(A^tA)^{-1}(A^tA)(A^tA)^{-1}A^t = A(A^tA)^{-1}A^t = P_W.$$

■

The orthogonal projection has a nice geometric characterization that makes it a useful tool in linear regression.

Theorem 3.5.8. *Let W be a subspace of $\mathbb{R}^n$ and let $\boldsymbol{x} \in \mathbb{R}^n$. Then*

$$\|\boldsymbol{x} - \text{proj}_W \boldsymbol{x}\| < \|\boldsymbol{x} - \boldsymbol{w}\|$$

for all $\boldsymbol{w} \in W$, $\boldsymbol{w} \neq \text{proj}_W \boldsymbol{x}$. In other words, $\text{proj}_W \boldsymbol{x}$ is the closest point in W to $\boldsymbol{x}$.

Theorem 3.5.8 is visualized in Figure 3.7.

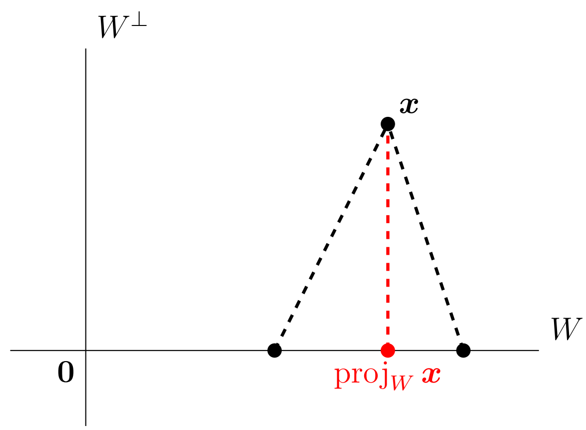

Figure 3.7: Geometric meaning of the orthogonal projection

Example 3.5.9 (Simple Linear Regression)**.** Given a set of points in $\mathbb{R}^2$, consider a problem of finding β_0 and β_1 so that the line $y = \beta_0 + \beta_1 x$ describes the trend in the set (see Figure 3.8). The criterion to determine β_0 and β_1 is to minimize the residual sum of squares

$$\sum_{i=1}^{n} |y_i - (\beta_0 + \beta_1 x_i)|^2. \tag{3.5.1}$$

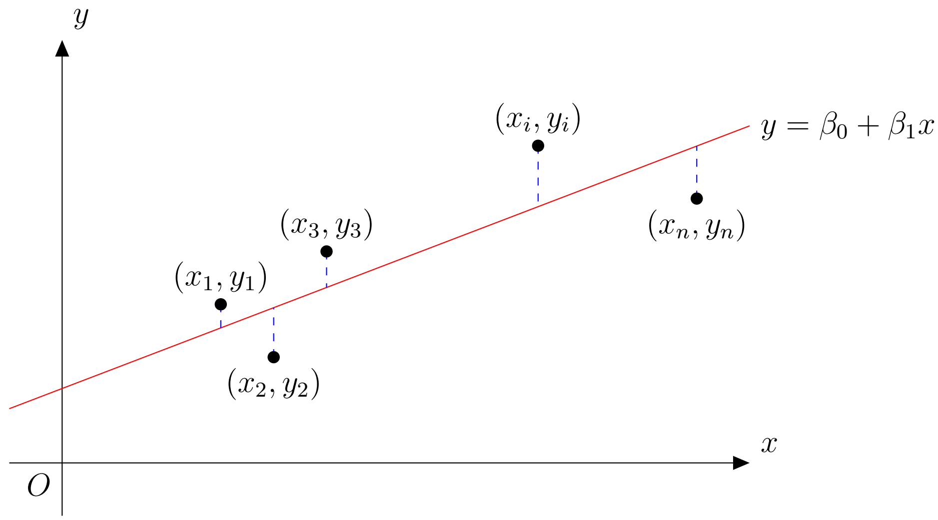

Figure 3.8: Least square fitting line

Define n-vectors $\mathbf{1} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$, $\boldsymbol{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$, and $\boldsymbol{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$, then (3.5.1) can be viewed as $\|\boldsymbol{y} - (\beta_0 \mathbf{1} + \beta_1 \boldsymbol{x})\|^2$ and the problem of finding β_0 and β_1 boils down to finding a vector $\boldsymbol{w}_0$ in W that is closest to $\boldsymbol{y}$, where $W = \operatorname{Span}\{\mathbf{1}, \boldsymbol{x}\}$. By Theorem 3.5.8 and Problem 3.5.5, we get that $\boldsymbol{w}_0 = \operatorname{proj}_W \boldsymbol{y} = A(A^t A)^{-1} A^t \boldsymbol{y}$, where $A = \begin{bmatrix} | & | \\ \mathbf{1} & \boldsymbol{x} \\ | & | \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$. Let $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$ and $\bar{y} = \frac{\sum_{i=1}^n y_i}{n}$, then a simple calculation shows that

$$A^t A = \begin{bmatrix} n & n\bar{x} \\ n\bar{x} & \sum_{i=1}^n x_i^2 \end{bmatrix}$$

and that

$$\begin{aligned}
\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} &= (A^t A)^{-1} A^t \boldsymbol{y} \\
&= \frac{1}{n \sum_{i=1}^n x_i^2 - n^2 \bar{x}^2} \begin{bmatrix} \sum_{i=1}^n x_i^2 & -n\bar{x} \\ -n\bar{x} & n \end{bmatrix} \begin{bmatrix} n\bar{y} \\ \sum_{i=1}^n x_i y_i \end{bmatrix} \\
&= \begin{bmatrix} \bar{y} - \frac{SXY}{SXX}\bar{x} \\ \frac{SXY}{SXX} \end{bmatrix},
\end{aligned}$$

where

$$SXX = \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - n\bar{x}^2$$

and

$$SXY = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}.$$

See Example 4.2.7 for a different approach to the problem.

We can extend the idea in Example 3.5.9.

Example 3.5.10 (Multiple Linear Regression)**.** Let $\boldsymbol{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$. Define an $n \times (p+1)$ matrix X by $X = \begin{bmatrix} 1 & x_{1,1} & \cdots & x_{1,p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & \cdots & x_{n,p} \end{bmatrix}$. We want to find $\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_p \end{bmatrix}$ that minimizes the residual sum of squares

$$\sum_{i=1}^{n} |y_i - (\beta_0 + \beta_1 x_{i,1} + \cdots + \beta_p x_{i,p})|^2 = \|\boldsymbol{y} - X\boldsymbol{\beta}\|^2.$$

Following the argument as in Example 3.5.9, one can see that the minimizing $\boldsymbol{\beta}$ is given by

$$\boldsymbol{\beta} = (X^t X)^{-1} X^t \boldsymbol{y}.$$

Exercise 3.5.

1. Let $\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ and $\mathbf{1} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$.

(a) Find $\operatorname{proj}_{\mathbf{1}} \boldsymbol{x}$.

(b) Compute $\boldsymbol{x} - \operatorname{proj}_{\mathbf{1}} \boldsymbol{x}$.

(c) Express the sample mean of $x_1, x_2, \ldots, x_n$ in terms of $\|\boldsymbol{x} - \operatorname{proj}_{\mathbf{1}} \boldsymbol{x}\|$.

2. Let W be a subspace of $\mathbb{R}^n$ and $W^\perp$ be its orthogonal complement. Show that $W \cap W^\perp = \{\mathbf{0}\}$.

3. Let W be a subspace of $\mathbb{R}^n$. Show that $W^{\perp\perp}$, the orthogonal complement of $W^\perp$, equals W.

4. Let P be an $n \times n$ matrix such that $P^2 = P = P^t$. Let $W = \{P\boldsymbol{x} : \boldsymbol{x} \in \mathbb{R}^n\}$. Show that $P_W = P$.
Hint: For any $\boldsymbol{x} \in \mathbb{R}^n$, show that $\boldsymbol{x} = P\boldsymbol{x} + (I_n - P)\boldsymbol{x}$ and that $\langle P\boldsymbol{x}, (I_n - P)\boldsymbol{x} \rangle = 0$.

5. Let W be a subspace of $\mathbb{R}^n$. Let P_W denote the projection matrix corresponding to W.

(a) Prove that $\boldsymbol{w} \in W$ if and only if $\boldsymbol{w} = P_W \boldsymbol{w}$.

(b) Prove that $W = \{P_W \boldsymbol{x} : \boldsymbol{x} \in \mathbb{R}^n\}$.

(c) Prove that $\dim W = \operatorname{rank}(P_W)$.
Hint: Use Theorem 3.3.41.

6. Let V and W be subspaces of $\mathbb{R}^n$. Show that $\langle \boldsymbol{v}, \boldsymbol{w} \rangle = 0$ for all $\boldsymbol{v} \in V$ and for all $\boldsymbol{w} \in W$ if and only if $P_V P_W = O$, the zero matrix. Here P_V (respectively, P_W) denotes the projection matrix corresponding to V (respectively, W).
Hint: Note that $\langle \boldsymbol{v}, \boldsymbol{w} \rangle = \langle P_V \boldsymbol{v}, P_W \boldsymbol{w} \rangle$ for $\boldsymbol{v} \in V$ and $\boldsymbol{w} \in W$ (see Exercise 3.5.5). You may want to use Example 3.2.24 to get the (i, j)-element of $P_V P_W$.

7. Let W be a subspace of $\mathbb{R}^n$ and $\boldsymbol{x} \in \mathbb{R}^n$. Prove that

$$\|\boldsymbol{x} - \operatorname{proj}_W \boldsymbol{x}\| = \max\{|\langle \boldsymbol{x}, \boldsymbol{y} \rangle| : \boldsymbol{y} \in W^{\perp}, \|\boldsymbol{y}\| \leq 1\}.$$

3.6 Eigenvectors and Eigenvalues

Definition An eigenvector of an $n \times n$ matrix A is a nonzero vector $\boldsymbol{\xi}$ such that $A\boldsymbol{\xi} = \lambda\boldsymbol{\xi}$ for some scalar λ. A scalar λ is called an eigenvalue of A if there is a nontrivial solution $\boldsymbol{\xi}$ of $A\boldsymbol{\xi} = \lambda\boldsymbol{\xi}$; such a $\boldsymbol{\xi}$ is called an eigenvector corresponding to λ.

Remark 3.6.1. By definition, an eigenvector must be a *nonzero* vector, but eigenvalue could be zero.

Example 3.6.2. Let $A = \begin{bmatrix} 3 & -2 \\ 1 & 0 \end{bmatrix}$ and $\boldsymbol{\xi} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$, then $A\boldsymbol{\xi} = \begin{bmatrix} 4 \\ 2 \end{bmatrix} = 2\boldsymbol{\xi}$. Therefore 2 is an eigenvalue of A and $\boldsymbol{\xi}$ is an eigenvector of A corresponding to the eigenvalue 2. Note that $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ is *not* the only eigenvector of A corresponding to the eigenvalue 2. For example, $\begin{bmatrix} 6 \\ 3 \end{bmatrix}$ is another eigenvector of A corresponding to 2. In fact, if $\boldsymbol{\xi}$ is an eigenvector of A corresponding to an eigenvalue λ, then so is any nonzero multiple $c\boldsymbol{\xi}$ of $\boldsymbol{\xi}$.

Problem 3.6.3. Let $B = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}$. Is $\boldsymbol{\xi} = \begin{bmatrix} 6 \\ -5 \end{bmatrix}$ an eigenvector of B? How about $\boldsymbol{\eta} = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$?

Answer Since $B\boldsymbol{\xi} = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix} \begin{bmatrix} 6 \\ -5 \end{bmatrix} = \begin{bmatrix} -24 \\ 20 \end{bmatrix} = -4 \begin{bmatrix} 6 \\ -5 \end{bmatrix} = -4\boldsymbol{\xi}$, we see that $\boldsymbol{\xi}$ is an eigenvector of B corresponding to the eigenvalue -4. However, $B\boldsymbol{\eta} = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ -2 \end{bmatrix} = \begin{bmatrix} -9 \\ 11 \end{bmatrix}$ is not a scalar multiple of $\boldsymbol{\eta}$, so $\boldsymbol{\eta}$ is *not* an eigenvector of B. ■

A matrix may have more than one eigenvalue.

Example 3.6.4. In Problem 3.6.3 above, we observed that $\lambda = -4$ is an eigenvalue of B. However, $\boldsymbol{\zeta} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is an eigenvector of B corresponding to the eigenvalue 7, since

$$B\boldsymbol{\zeta} = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 7 \\ 7 \end{bmatrix} = 7 \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 7\boldsymbol{\zeta}.$$

Problem 3.6.5. Let A and B be $n \times n$ matrices. Suppose that λ is an eigenvalue of AB. Show that λ is an eigenvalue of BA.

Answer First, suppose $\lambda = 0$, then there is $\boldsymbol{\xi} \neq \mathbf{0}$ such that $AB\boldsymbol{\xi} = \mathbf{0}$. By Invertible Matrix Theorem, $\det(AB) = 0$. Therefore, $\det(BA) = 0$ and again by Invertible Matrix Theorem, there is $\boldsymbol{\eta} \neq \mathbf{0}$ such that $BA\boldsymbol{\eta} = \mathbf{0}$, in other words, 0 is an eigenvalue of BA. Now suppose $\lambda \neq 0$, then there is $\boldsymbol{\xi} \neq \mathbf{0}$ such that $AB\boldsymbol{\xi} = \lambda\boldsymbol{\xi}$. Let $\boldsymbol{\eta} = B\boldsymbol{\xi}$, then $\boldsymbol{\eta} \neq \mathbf{0}$ and it follows that $BA\boldsymbol{\eta} = BAB\boldsymbol{\xi} = \lambda B\boldsymbol{\xi} = \lambda\boldsymbol{\eta}$, and this shows that λ is an eigenvalue of BA. ■

Suppose λ is an eigenvalue of A. How do we find all the eigenvectors corresponding to λ?

Definition 3.6.6. Let A be an $n \times n$ square matrix and λ be an eigenvalue of A. $\mathrm{Nul}(A - \lambda I_n)$, the null space of $A - \lambda I_n$, is called the <u>eigenspace</u> of A corresponding to λ. Therefore, the eigenspace of A corresponding to λ is a subspace of $\mathbb{R}^n$. The <u>geometric multiplicity</u> of λ is the dimension of the eigenspace of A corresponding to λ.

Remark 3.6.7.

(a) The geometric multiplicity of an eigenvalue is always greater than or equal to 1.

(b) Any nonzero vector in the eigenspace corresponding to λ is an eigenvector corresponding to λ.

Theorem 3.6.8. *If $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \cdots, \boldsymbol{\xi}_r$ are eigenvectors that correspond to distinct eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_r$ of an $n \times n$ matrix A, then the set $\{\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \cdots, \boldsymbol{\xi}_r\}$ is linearly independent.*

Example 3.6.9. From Problem 3.6.3 and Example 3.6.4 above, we see that $\begin{bmatrix} 6 \\ -5 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ are eigenvectors of $B = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}$ corresponding to -4 and 7, respectively. So $\left\{ \begin{bmatrix} 6 \\ -5 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$ is linearly independent.

Exercise 3.6.

1. Let $A = \begin{bmatrix} 4 & -1 & 6 \\ 2 & 1 & 6 \\ 2 & -1 & 8 \end{bmatrix}$. Show that 2 is an eigenvalue of A with an eigenvector $\begin{bmatrix} 3 \\ 0 \\ -1 \end{bmatrix}$.
Find a basis for the eigenspace of A corresponding to 2.

2. An $n \times n$ matrix $A = [a_{ij}]_{i,j=1}^{n}$ is called a Markov matrix if each row sum equals 1, that is, $\displaystyle\sum_{j=1}^{n} a_{ij} = 1$ for all i. Show that 1 is an eigenvalue for every Markov matrix.
Hint: Try to find an eigenvector.

3. Suppose that λ is an eigenvalue of A. For any polynomial p, show that $p(\lambda)$ is an eigenvalue of $p(A)$.
Hint: Let $\boldsymbol{\xi}$ be an eigenvector of A corresponding to λ. Show that $\boldsymbol{\xi}$ is an eigenvector of $p(A)$ corresponding to $p(\lambda)$.

4. Let S be a square matrix such that $S^2 = S$. Show that every eigenvalue of S must be either 1 or 0.

5. Suppose that λ is an eigenvalue of an invertible matrix A.

(a) Explain why λ cannot be equal to 0.

(b) Show that $\frac{1}{\lambda}$ is an eigenvalue of A^{-1}.
Hint: Let $\boldsymbol{\xi}$ be an eigenvector of A corresponding to λ. Show that $\boldsymbol{\xi}$ is an eigenvector of A^{-1} corresponding to $\frac{1}{\lambda}$.

6. Let A be an $m \times n$ matrix. Show that all eigenvalues of $B = A^t A$ are nonnegative real numbers.
Hint: You may want to use Exercise 3.2.7.

3.7 The Characteristic Equation

Let $A = \begin{bmatrix} 2 & 3 \\ 3 & -6 \end{bmatrix}$. We want to find all eigenvalues of A. By definition, λ is an eigenvalue of A if and only if $A\boldsymbol{\xi} = \lambda\boldsymbol{\xi}$ for some nonzero vector $\boldsymbol{\xi}$, which is the same as saying that the homogeneous system $(A - \lambda I_2)\boldsymbol{\xi} = \mathbf{0}$ has a nontrivial solution. Therefore by the Invertible Matrix Theorem, we see that λ is an eigenvalue of A if and only if the matrix $A - \lambda I_2$ is singular, that is, if and only if $\det(A - \lambda I_2) = 0$. Now $A - \lambda I_2 = \begin{bmatrix} 2-\lambda & 3 \\ 3 & -6-\lambda \end{bmatrix}$ and hence $\det(A - \lambda I_2) = (2-\lambda)(-6-\lambda) - 9 = \lambda^2 + 4\lambda - 21 = (\lambda + 7)(\lambda - 3)$. Therefore exactly -7 and 3 are eigenvalues of A. In general, to find all the eigenvalues of a given $n \times n$ matrix A,

(a) First, find $A - \lambda I_n$;

(b) Second, compute $\det(A - \lambda I_n)$, which is a polynomial in λ of degree n;

(c) Third, finally solve the equation $\det(A - \lambda I_n) = 0$. The solution set of this equation is exactly the set of eigenvalues of A.

Remark 3.7.1. Since eigenvalues are roots of a polynomial, they may not be real numbers. However, all eigenvalues of a *symmetric* matrix are real. See Theorem 3.9.2.

Definition 3.7.2. The polynomial $\det(A - \lambda I_n)$ is called the characteristic polynomial of A. The equation $\det(A - \lambda I_n) = 0$ is called the characteristic equation of A.

Problem 3.7.3. Find all the eigenvalues of $A = \begin{bmatrix} 5 & -2 & 6 & -1 \\ 0 & 3 & -8 & 0 \\ 0 & 0 & 5 & 4 \\ 0 & 0 & 0 & 1 \end{bmatrix}$.

Answer The characteristic polynomial of A is given by the determinant of

$$A - \lambda I_n = \begin{bmatrix} 5-\lambda & -2 & 6 & -1 \\ 0 & 3-\lambda & -8 & 0 \\ 0 & 0 & 5-\lambda & 4 \\ 0 & 0 & 0 & 1-\lambda \end{bmatrix}.$$

Since $A - \lambda I_n$ is an upper triangle matrix, by Example 3.4.7, we have $\det(A - \lambda I_n) = (5-\lambda)^2(3-\lambda)(1-\lambda)$. This gives three distinct eigenvalues 5, 3, and 1. ■

Remark 3.7.4. In general, the eigenvalues of a triangular matrix are precisely the diagonal entries of the matrix.

Definition 3.7.5. Let A and B be $n \times n$ matrices. A is said to be similar to B if there is an invertible matrix P such that $P^{-1}AP = B$. When A is similar to B, then we write $A \sim B$.

Example 3.7.6. $A = \begin{bmatrix} 1 & 2 \\ 1 & 4 \end{bmatrix}$ is similar to $B = \begin{bmatrix} 2 & 4 \\ 1 & 3 \end{bmatrix}$ because with $P = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$ (so $P^{-1} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$)

$$P^{-1}AP = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 1 & 4 \end{bmatrix}\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 6 \\ 1 & 4 \end{bmatrix}\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 4 \\ 1 & 3 \end{bmatrix} = B.$$

Remark 3.7.7.

(a) If A is similar to B, then B is similar to A (i.e., $A \sim B \Longrightarrow B \sim A$).

(b) A is similar to A itself ($A \sim A$).

(c) If A is similar to B and B is similar to C, then A is similar to C ($A \sim B$ and $B \sim C \Longrightarrow A \sim C$).

Theorem 3.7.8. *If A and B are similar $n \times n$ matrices, then they have the same characteristic polynomials and hence the same eigenvalues.*

Exercise 3.7.

1. Find all eigenvalues of each of the following matrices:

(a) $A = \begin{bmatrix} 5 & 1 \\ 3 & 3 \end{bmatrix}$

(b) $B = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 8 \\ 0 & 2 & 1 \end{bmatrix}$

2. Let $\lambda_1, \dots, \lambda_n$ (not necessarily distinct) be eigenvalues of A. Prove that $\det A = \prod_{i=1}^n \lambda_i$.

3. Let A be an $n \times n$ matrix. Let M be an invertible $n \times n$ matrix.

(a) Show that λ is an eigenvalue of A if and only if λ is an eigenvalue of $M^{-1}AM$ (see Theorem 3.7.8).

(b) Show that $tr(A) = tr(M^{-1}AM)$.
Hint: See Exercise 3.2.6.

3.8 Diagonalization

There are many practical applications in which we need to compute A^k for large k. This requires a lot of computations for general $n \times n$ matrix A. When A is diagonal, however, the computation is quite simple.

Example 3.8.1. If $D = \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix}$, then $D^2 = \begin{bmatrix} 25 & 0 \\ 0 & 9 \end{bmatrix}$. In general, $D^n = \begin{bmatrix} 5^n & 0 \\ 0 & 3^n \end{bmatrix}$.

Example 3.8.2. We will compute A^{10}, where $A = \begin{bmatrix} 7 & 2 \\ -4 & 1 \end{bmatrix}$. Note that computing A^{10} with bare hands requires considerable amount of time, so we need a trick to get this done. To this end, define a matrix $P = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix}$, then P is invertible with $P^{-1} = \begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix}$ and

$$P^{-1}AP = \begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix} \begin{bmatrix} 7 & 2 \\ -4 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix} = \begin{bmatrix} 10 & 5 \\ -3 & -3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix} = \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix}.$$

In other words, A is similar to a diagonal matrix $D = \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix}$. Taking the 10^{th} power of the both sides gives

RHS: $D^{10} = \begin{bmatrix} 5^{10} & 0 \\ 0 & 3^{10} \end{bmatrix}$.

LHS: $(P^{-1}AP)^{10} = \underbrace{(P^{-1}AP)(P^{-1}AP)\cdots(P^{-1}AP)}_{10 \text{ times}} = P^{-1}A(PP^{-1})A(PP^{-1})A\cdots(PP^{-1})AP$, which reduces to $P^{-1}A^{10}P$.

Therefore we conclude that $P^{-1}A^{10}P = \begin{bmatrix} 5^{10} & 0 \\ 0 & 3^{10} \end{bmatrix}$, or

$$\begin{aligned} A^{10} &= P \begin{bmatrix} 5^{10} & 0 \\ 0 & 3^{10} \end{bmatrix} P^{-1} \\ &= \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix} \begin{bmatrix} 5^{10} & 0 \\ 0 & 3^{10} \end{bmatrix} \begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix} \\ &= \begin{bmatrix} 2 \cdot 5^{10} - 3^{10} & 5^{10} - 3^{10} \\ -2 \cdot 5^{10} + 2 \cdot 3^{10} & -5^{10} + 2 \cdot 3^{10} \end{bmatrix}. \end{aligned}$$

In general,

$$A^n = \begin{bmatrix} 2 \cdot 5^n - 3^n & 5^n - 3^n \\ -2 \cdot 5^n + 2 \cdot 3^n & -5^n + 2 \cdot 3^n \end{bmatrix}$$

for any $n \in \mathbb{N}$.

Given A, is it always possible to find an invertible matrix P such that $P^{-1}AP$ is diagonal? The answer is *no*.

Definition 3.8.3. A square matrix A is said to be diagonalizable if A is similar to a diagonal matrix, that is, if $P^{-1}AP = D$ for some invertible matrix P and some diagonal matrix D.

Remark 3.8.4. Suppose that A is diagonalizable with $P^{-1}AP = D$. By Theorem 3.7.8, A and D have the same eigenvalues and it follows that the diagonal entries of D are eigenvalues of A.

Example 3.8.5. The matrix $A = \begin{bmatrix} 7 & 2 \\ -4 & 1 \end{bmatrix}$ in Example 3.8.2 is diagonalizable.

If A is diagonalizable, how can we find a matrix P such that $P^{-1}AP$ is diagonal? The next theorems characterize diagonalizable matrices and states how to find P such that $P^{-1}AP$ is diagonal for a given diagonalizable matrix A.

Theorem 3.8.6 (The Diagonalization Theorem, Part 1)**.** *An $n \times n$ matrix A is diagonalizable if and only if A has n linearly independent eigenvectors. In fact, $P^{-1}AP = D$, with D diagonal, if and only if the columns of P are n linearly independent eigenvectors of A. In this case, the diagonal entries of D are eigenvalues of A that correspond, respectively, to the eigenvectors (i.e. columns) in P.*

Theorem 3.8.7 (The Diagonalization Theorem, Part 2)**.** *Let A be an $n \times n$ matrix with distinct eigenvalues $\lambda_1, \cdots, \lambda_p$.*

(a) For $1 \le k \le p$, the dimension of the eigenspace corresponding to λ_k (that is, the geometric multiplicity of λ_k) is always greater than equal to 1 and less than or equal to the algebraic multiplicity of the eigenvalue λ_k as a zero of the characteristic polynomial of A.

(b) The matrix A is diagonalizable if and only if the geometric multiplicity of λ_k equals the algebraic multiplicity of λ_k for each k. In particular, if A has n distinct eigenvalues, then A is diagonalizable.

(c) If $\mathcal{B}_j$ (resp. $\mathcal{B}_k$) is a basis for the eigenspace corresponding to λ_j (resp. λ_k), then the set $\mathcal{B}_j \cup \mathcal{B}_k$ is linearly independent.

Example 3.8.8. We go back to Example 3.8.2 and see how we can obtain the matrix P. $A = \begin{bmatrix} 7 & 2 \\ -4 & 1 \end{bmatrix}$, the eigenvalues of A are the zeros of the characteristic polynomial of A: $\det(A - \lambda I_2) = (7-\lambda)(1-\lambda) + 8 = \lambda^2 - 8\lambda + 15 = (\lambda - 3)(\lambda - 5)$. To Find an eigenvector corresponding to $\lambda = 5$, we solve $(A - 5I_2)\boldsymbol{\xi} = \mathbf{0}$ and this gives $\boldsymbol{\xi} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$. Similarly, an eigenvector $\begin{bmatrix} 1 \\ -2 \end{bmatrix}$ corresponding to $\lambda = 3$ is obtained by solving $(A - 3I_2)\boldsymbol{\xi} = \mathbf{0}$. Therefore, we can take $P = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix}$.

Problem 3.8.9. Diagonalize A (that is, find an invertible matrix P and a diagonal matrix D such that $P^{-1}AP = D$), if possible, where $A = \begin{bmatrix} 1 & 3 & 3 \\ -3 & -5 & -3 \\ 3 & 3 & 1 \end{bmatrix}$.

Answer Note that $A - \lambda I_3 = \begin{bmatrix} 1-\lambda & 3 & 3 \\ -3 & -5-\lambda & -3 \\ 3 & 3 & 1-\lambda \end{bmatrix}$ does not contain zeros, so computing its determinant is not handy. To compute the determinant easily, we produce zeros by applying row operations $R_3 \mapsto R_3 + R_2$ and $R_2 \mapsto R_2 + R_1$. These row operations and the cofactor expansion along the first column gives

$$\begin{aligned}
\det(A - \lambda I_3) &= \det\left(\begin{bmatrix} 1-\lambda & 3 & 3 \\ -2-\lambda & -2-\lambda & 0 \\ 0 & -2-\lambda & -2-\lambda \end{bmatrix}\right) \\
&= (1-\lambda)\det\begin{bmatrix} -2-\lambda & 0 \\ -2-\lambda & -2-\lambda \end{bmatrix} + (2+\lambda)\underbrace{\det\begin{bmatrix} 3 & 3 \\ -2-\lambda & -2-\lambda \end{bmatrix}}_{0} \\
&= (1-\lambda)(2+\lambda)^2,
\end{aligned}$$

so we have two distinct eigenvalues 1 and -2. Since $\lambda = 1$ has algebraic multiplicity 1, its geometric multiplicity is also equal to 1. Note, however, that $\lambda = -2$ has algebraic multiplicity 2, so we need to check the dimension of $\operatorname{Nul}(A+2I_3)$. Indeed, $A+2I_3 = \begin{bmatrix} 3 & 3 & 3 \\ -3 & -3 & -3 \\ 3 & 3 & 3 \end{bmatrix}$ reduces to $\begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ and it follows that $\operatorname{Nul}(A+2I_3) = \operatorname{Span}\left\{ \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \right\}$ and that $\dim \operatorname{Nul}(A+2I_3) = 2$. This shows that A is diagonalizable. An eigenvector $\boldsymbol{\xi}_1$ corresponding to $\lambda = 1$ is obtained by finding a nonzero vector in the null space of $A - I_3 = \begin{bmatrix} 0 & 3 & 3 \\ -3 & -6 & -3 \\ 3 & 3 & 0 \end{bmatrix}$. Since $\operatorname{Nul}(A - I_3) = \operatorname{Span}\left\{ \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \right\}$, one can simply take $\boldsymbol{\xi}_1 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$. For eigenvectors corresponding to $\lambda = -2$, we can choose $\boldsymbol{\xi}_{2,1} = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$ and $\boldsymbol{\xi}_{2,2} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$. Finally, with $P = \begin{bmatrix} 1 & -1 & -1 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$ and $D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{bmatrix}$, we get $P^{-1}AP = D$. ■

Example 3.8.10. Not all matrices are diagonalizable. We show that $B = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ is *not* diagonalizable. Suppose for a contradiction that B is diagonalizable. Since the characteristic equation of B is $(1-\lambda)^2 = 0$, 1 is the only eigenvalue of B, so it follows by the Diagonalization Theorem, Part 1, that B is similar to the identity matrix I_2, which would imply that $B = I_2$.

Example 3.8.11 (Application to the Fibonacci sequence)**.** Consider the sequence defined by

$$a_1 = 1, \quad a_2 = 1, \quad a_n = a_{n-1} + a_{n-2} \text{ for } n \geq 3.$$

We are interested in the closed form of a_n. Note that for $n \geq 3$

$$\begin{bmatrix} a_n \\ a_{n-1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} a_{n-1} \\ a_{n-2} \end{bmatrix}.$$

Similarly, if n is large enough, then

$$\begin{bmatrix} a_{n-1} \\ a_{n-2} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} a_{n-2} \\ a_{n-3} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} a_{n-2} \\ a_{n-3} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} a_{n-3} \\ a_{n-4} \end{bmatrix}$$

so

$$\begin{bmatrix} a_n \\ a_{n-1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^2 \begin{bmatrix} a_{n-2} \\ a_{n-3} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^3 \begin{bmatrix} a_{n-3} \\ a_{n-4} \end{bmatrix} = \cdots = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^{n-2} \begin{bmatrix} a_2 \\ a_1 \end{bmatrix}.$$

Consider $\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^{n-2}$. The characteristic polynomial of $\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$ is $\lambda^2 - \lambda - 1$ and eigenvalues are $\frac{1\pm\sqrt{5}}{2}$. Thus the matrix $\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$ is diagonalizable. One eigenvector corresponding to $\lambda_1 = \frac{1+\sqrt{5}}{2}$ is $\begin{bmatrix} 1 \\ -\frac{1-\sqrt{5}}{2} \end{bmatrix}$. Similarly, $\begin{bmatrix} 1 \\ -\frac{1+\sqrt{5}}{2} \end{bmatrix}$ is an eigenvector corresponding to $\lambda_2 = \frac{1-\sqrt{5}}{2}$. Let $P = \begin{bmatrix} 1 & 1 \\ -\frac{1-\sqrt{5}}{2} & -\frac{1+\sqrt{5}}{2} \end{bmatrix}$ then $P^{-1}\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}P = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$. Finally, $\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^{n-2} = P\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}^{n-2} P^{-1} = \frac{1}{\sqrt{5}}\begin{bmatrix} \lambda_1^{n-1} - \lambda_2^{n-1} & \lambda_1^{n-2} - \lambda_2^{n-2} \\ \lambda_1^{n-2} - \lambda_2^{n-2} & \lambda_1^{n-3} - \lambda_2^{n-3} \end{bmatrix}$ and hence

$$a_n = \frac{1}{\sqrt{5}}(\lambda_1^{n-1} - \lambda_2^{n-1} + \lambda_1^{n-2} - \lambda_2^{n-2}) = \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right]$$

after calculation.

Exercise 3.8.

1. Let $A = \begin{bmatrix} -2 & 12 \\ -1 & 5 \end{bmatrix}$.

(a) Find eigenvalues and corresponding eigenvectors of A.

(b) Find a diagonal matrix D and an invertible matrix P such that $A = PDP^{-1}$.

(c) Compute A^{100}.

2. Let A be an $n \times n$ square matrix with n distinct positive eigenvalues. Show that there is an $n \times n$ matrix B such that $B^2 = A$.
Hint: If A is a diagonal matrix, then B can be easily obtained. Use diagonalization for general case.

3. Let A be an $n \times n$ diagonalizable matrix. Show that there is an $n \times n$ matrix B such that $A = B^3$.

3.9 Symmetric Matrices and Quadratic Forms

For a random vector $\boldsymbol{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}$, its variance-covariance matrix Σ is defined to be an $n\times n$ matrix whose (i,j)-entry Σ_{ij} is given by the covariance between X_i and X_j. It is known that the variance-covariance matrix of a random vector has many desirable properties, including positive definiteness. In this section, we study properties of symmetric and positive definite matrices and their roles in geometric interpretation of multidimensional data.

Definition 3.9.1. Let A be a square matrix.

(a) A is said to be symmetric if $A = A^t$.

(b) A is said to be orthogonal if $A^t = A^{-1}$.

Theorem 3.9.2. *If A is a symmetric matrix, then all eigenvalues of A are real.*

Theorem 3.9.3. *If A is a symmetric matrix, then there exists an orthogonal matrix U such that $U^tAU = D$, where D is a diagonal matrix. Moreover, $A = \sum_{i=1}^n \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i^t$, where $\boldsymbol{u}_i$ is the i^{th} column of U and λ_i is the i^{th} diagonal element of D.*

Remark 3.9.4. Let A be an $n \times n$ symmetric matrix. By Theorem 3.9.2, we can list all eigenvalues of A in a decreasing order, say, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. By Theorem 3.9.3, one can write $A = UDU^t$, and this is called a spectral decomposition of A.

Definition 3.9.5. A quadratic form Q is a map from $\mathbb{R}^n$ to $\mathbb{R}$ of the form $Q(\boldsymbol{x}) = \boldsymbol{x}^tA\boldsymbol{x}$, where A is a $n \times n$ matrix and $\boldsymbol{x} \in \mathbb{R}^n$. A quadratic form $Q(\boldsymbol{x}) = \boldsymbol{x}^tA\boldsymbol{x}$, or the matrix A, is said to be positive definite (respectively, positive semidefinite) if $\boldsymbol{x}^tA\boldsymbol{x} > 0$ (respectively, $\boldsymbol{x}^tA\boldsymbol{x} \geq 0$) for all $\boldsymbol{x} \neq \boldsymbol{0}$. Similarly, a quadratic form $Q(\boldsymbol{x}) = \boldsymbol{x}^tA\boldsymbol{x}$, or the matrix A, is said to be negative definite (respectively, negative semidefinite) if $\boldsymbol{x}^tA\boldsymbol{x} < 0$ (respectively, $\boldsymbol{x}^tA\boldsymbol{x} \leq 0$) for all $\boldsymbol{x} \neq \boldsymbol{0}$.

Remark 3.9.6. For a quadratic form $Q(\boldsymbol{x}) = \boldsymbol{x}^tA\boldsymbol{x}$, we may assume that A is symmetric. See Exercise 3.9.4. So from now on, when it comes to a quadratic form Q associated with a matrix A, we always assume that A is symmetric.

Theorem 3.9.7.

(a) A is positive definite (respectively, positive semidefinite) if and only if $-A$ is negative definite (respectively, negative semidefinite).

(b) A is positive semidefinite if and only if there exists an M such that $A = M^tM$.

(c) A is positive definite if and only if A is positive semidefinite and $\det A > 0$.

M in Theorem 3.9.7.(b) can be assumed to be a lower triangular matrix.

Theorem 3.9.8 (Cholesky Decomposition)**.** *Let A be a positive semidefinite matrix. Then there is a lower triangular matrix L such that $A = LL^t$. In particular, if A is positive definite, then all diagonal entries in L are positive.*

Problem 3.9.9. Let A be a symmetric matrix. Show that A is positive definite (respectively, positive semidefinite) if and only if all eigenvalues of A are positive (respectively, nonnegative).

Answer Let $\lambda_1, \ldots, \lambda_n$ be eigenvalues of A. Suppose that $\lambda_i > 0$ for all i, $1 \le i \le n$. By Theorem 3.9.3, A has a spectral decomposition, that is, there exists an orthogonal matrix U such that $A = UDU^t$, where D is a diagonal matrix whose diagonal entries are $\lambda_1, \ldots, \lambda_n$. Let R be a diagonal matrix with diagonal entries $\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n}$, then $D = R^2 = RR^t$ and it follows that $A = (UR)(UR)^t$. Since $\det A > 0$ (see Exercise 3.7.2), by Theorem 3.9.7, A is positive definite. See Exercise 3.9.2 for the other implication. ■

Example 3.9.10 (Classical Multidimensional Scaling)**.** Let $\boldsymbol{X}_i$ $(1 \le i \le N)$, $\bar{\boldsymbol{X}}$, X, and B be as in Remark 3.2.32. Let d_{ij} denote the distance between $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$, then it is obvious that d_{ij} is obtained from X. For example, if one defines an $n \times n$ matrix $S = [s_{ij}]_{i,j=1}^n$ by $S = B^tB$, then

$$\begin{aligned} d_{ij}^2 &= \|\boldsymbol{X}_i - \boldsymbol{X}_j\|^2 \\ &= \|\boldsymbol{X}_i - \bar{\boldsymbol{X}} - (\boldsymbol{X}_j - \bar{\boldsymbol{X}})\|^2 \\ &= (\boldsymbol{X}_i - \bar{\boldsymbol{X}} - (\boldsymbol{X}_j - \bar{\boldsymbol{X}})) \cdot (\boldsymbol{X}_i - \bar{\boldsymbol{X}} - (\boldsymbol{X}_j - \bar{\boldsymbol{X}})) \\ &= \|\boldsymbol{X}_i - \bar{\boldsymbol{X}}\|^2 + \|\boldsymbol{X}_j - \bar{\boldsymbol{X}}\|^2 - 2(\boldsymbol{X}_i - \bar{\boldsymbol{X}}) \cdot (\boldsymbol{X}_j - \bar{\boldsymbol{X}}) \\ &= s_{ii} + s_{jj} - 2s_{ij}. \end{aligned} \tag{3.9.1}$$

Conversely, consider a problem of obtaining a $p \times N$ matrix X from the distance information $\{d_{ij} : 1 \le i, j \le N\}$. First we note that X is not uniquely determined, since distance remains invariant under, for example, identical translations of $\boldsymbol{X}_i$'s, so we assume that $\bar{\boldsymbol{X}} = \boldsymbol{0}$. Letting $S = [s_{ij}]_{i,j=1}^n = X^tX$, we get

$$\sum_{i=1}^N s_{ij} = \sum_{i=1}^n \langle \boldsymbol{X}_i, \boldsymbol{X}_j \rangle = \langle N\bar{\boldsymbol{X}}, \boldsymbol{X}_j \rangle = 0$$

and similarly

$$\sum_{j=1}^N s_{ij} = \sum_{j=1}^n \langle \boldsymbol{X}_i, \boldsymbol{X}_j \rangle = \langle \boldsymbol{X}_i, N\bar{\boldsymbol{X}} \rangle = 0.$$

It follows from (3.9.1) that

$$\sum_{j=1}^N d_{ij}^2 = Ns_{ii} + tr(S), \quad \sum_{i=1}^N d_{ij}^2 = tr(S) + Ns_{jj}, \quad \sum_{i,j=1}^N d_{ij}^2 = 2Ntr(S)$$

from which we obtain

$$\begin{aligned} tr(S) &= \frac{1}{2N} \sum_{i,j=1}^N d_{ij}^2, \\ s_{ii} &= \frac{1}{N}\left(\sum_{j=1}^N d_{ij}^2 - tr(S)\right) = \frac{1}{N}\left(\sum_{j=1}^N d_{ij}^2 - \frac{1}{2N}\sum_{i,j=1}^N d_{ij}^2\right), \text{and} \\ s_{jj} &= \frac{1}{N}\left(\sum_{i=1}^N d_{ij}^2 - tr(S)\right) = \frac{1}{N}\left(\sum_{i=1}^N d_{ij}^2 - \frac{1}{2N}\sum_{i,j=1}^N d_{ij}^2\right). \end{aligned} \tag{3.9.2}$$

From (3.9.1) and (3.9.2), it follows that

$$\begin{aligned} s_{ij} &= -\frac{1}{2}\left(d_{ij}^2 - s_{ii} - s_{jj}\right) \\ &= -\frac{1}{2}\left(d_{ij}^2 - \frac{1}{N}\sum_{j=1}^{N} d_{ij}^2 - \frac{1}{N}\sum_{i=1}^{N} d_{ij}^2 + \frac{1}{N^2}\sum_{i,j=1}^{N} d_{ij}^2\right), \end{aligned}$$

which means that $S = X^tX$ is obtained using d_{ij}'s. To recover a $p \times N$ matrix X from S, we write $S = UDU^t$, where $U = \begin{bmatrix} | & | & & | \\ \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_N \\ | & | & & | \end{bmatrix}$ is an $N \times N$ orthogonal matrix and $D = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_N \end{bmatrix}$. Note that S is positive semidefinite and $\lambda_i \geq 0$ for $1 \leq i \leq N$. Since $\operatorname{rank}(D) = \operatorname{rank}(S) = \operatorname{rank}(X)$ is at most p, it follows that $\lambda_{p+1} = \lambda_{p+2} = \cdots = \lambda_N = 0$ and that $S = U_0D_0U_0^t$, where $U_0 = \begin{bmatrix} | & | & & | \\ \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_p \\ | & | & & | \end{bmatrix}$ and $D_0 = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_p \end{bmatrix}$.

Finally, one can take $X = D_0^{1/2}U_0^t$, where $D_0^{1/2} = \begin{bmatrix} \sqrt{\lambda_1} & & & \\ & \sqrt{\lambda_2} & & \\ & & \ddots & \\ & & & \sqrt{\lambda_p} \end{bmatrix}$.

The following is another criterion for positive definiteness. See Exercise 3.9.6 for a proof of a special case.

Theorem 3.9.11 (Sylvester's Criterion)**.** *Let A be an $n \times n$ symmetric matrix. For $1 \leq k \leq n$, let Δ_k denote the determinant of the upper left $k \times k$ submatrix of A. Then*

(a) A is positive definite if and only if $\Delta_k > 0$ for all k, $1 \leq k \leq n$.

(b) A is negative definite if and only if $\Delta_k < 0$ for all odd k and $\Delta_k > 0$ for all even k.

In many applications, we need to maximize a quadratic form under certain conditions. We start with the next theorem.

Theorem 3.9.12. *Let A be a symmetric matrix with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Let $M = \max\{\boldsymbol{x}^tA\boldsymbol{x} : \|\boldsymbol{x}\| = 1\}$ and $m = \min\{\boldsymbol{x}^tA\boldsymbol{x} : \|\boldsymbol{x}\| = 1\}$, then $M = \boldsymbol{\xi}_1^tA\boldsymbol{\xi}_1$ and $m = \boldsymbol{\xi}_n^tA\boldsymbol{\xi}_n$, where $\boldsymbol{\xi}_i$ is a unit eigenvector of A corresponding to λ_i.*

Theorem 3.9.13. *Let A, λ_i, and $\boldsymbol{\xi}_i$ be as in Theorem 3.9.12. Then the maximum value of*

$$\{\boldsymbol{x}^t A\boldsymbol{x} : \|\boldsymbol{x}\| = 1, \langle \boldsymbol{x}, \boldsymbol{\xi}_1\rangle = 0\}$$

is the second largest eigenvalue, λ_2, and this is obtained when $\boldsymbol{x} = \boldsymbol{\xi}_2$. In general, for $2 \le i \le n$, the maximum value of

$$\{\boldsymbol{x}^t A\boldsymbol{x} : \|\boldsymbol{x}\| = 1, \langle \boldsymbol{x}, \boldsymbol{\xi}_1\rangle = \langle \boldsymbol{x}, \boldsymbol{\xi}_2\rangle = \cdots = \langle \boldsymbol{x}, \boldsymbol{\xi}_{i-1}\rangle = 0\}$$

is λ_i, and this is obtained when $\boldsymbol{x} = \boldsymbol{\xi}_i$.

Example 3.9.14. Let A be a 2×2 positive definite matrix, $\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \in \mathbb{R}^2$, and $c > 0$. Consider the problem of determining the trace of points $\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^2$ such that $(\boldsymbol{x}-\boldsymbol{\mu})^t A(\boldsymbol{x}-\boldsymbol{\mu}) = c$. First, we assume that $\boldsymbol{\mu} = \mathbf{0}$. If $A = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$ is a diagonal matrix with $\lambda_1 \ge \lambda_2 > 0$, then the equation $\boldsymbol{x}^t A\boldsymbol{x} = c$ becomes $\lambda_1 x_1^2 + \lambda_2 x_2^2 = c$ and the trace turns out to be an ellipse centered at the origin with half length of minor axis equal to $\frac{\sqrt{c}}{\sqrt{\lambda_1}}$ and half length of major axis equal to $\frac{\sqrt{c}}{\sqrt{\lambda_2}}$ (see Figure 3.9 (a)).

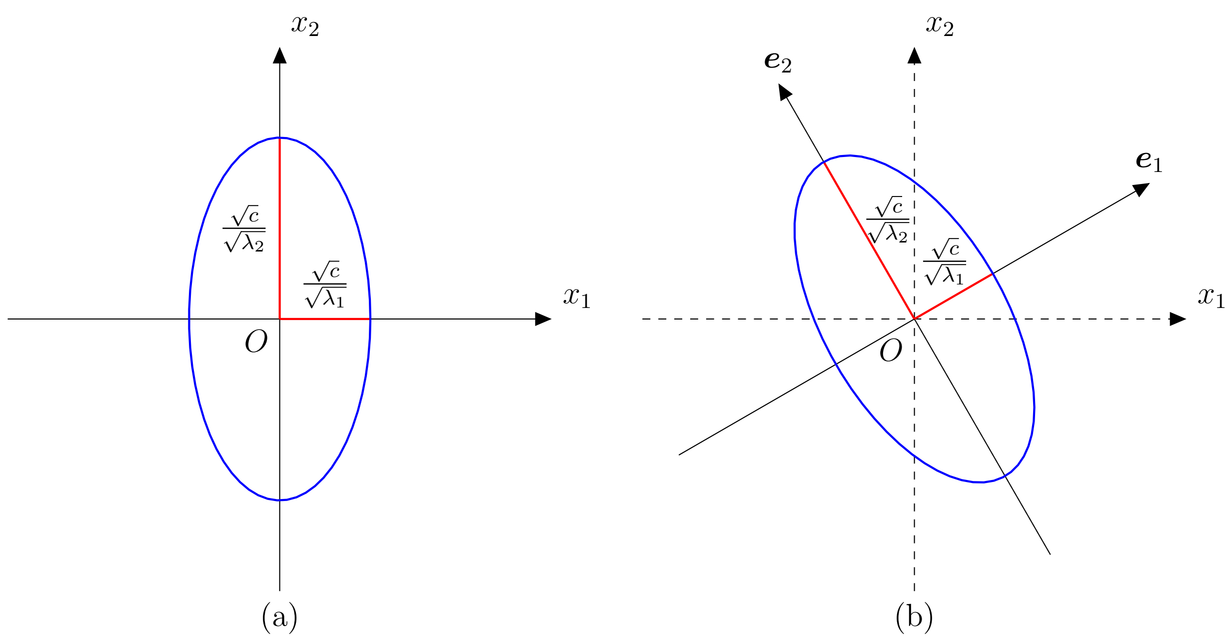

Figure 3.9: Positive definite matrix and ellipse

In general, let $A = UDU^t$ be a spectral decomposition of A with $D = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$,

$\lambda_1 \geq \lambda_2 > 0$, and $U = \begin{bmatrix} | & | \\ \boldsymbol{e}_1 & \boldsymbol{e}_2 \\ | & | \end{bmatrix}$, where $\boldsymbol{e}_i$ is a unit eigenvector corresponding to λ_i for $i = 1, 2$. Let $\boldsymbol{y} = U^t\boldsymbol{x}$. Since $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ are orthogonal to each other, $U^t\boldsymbol{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $U^t\boldsymbol{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, so $\boldsymbol{y}$ can be viewed as a new coordinate system that transforms $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ to $(1, 0)$ and $(0, 1)$, respectively. Since the equation $\boldsymbol{x}^tA\boldsymbol{x} = c$ becomes $\boldsymbol{y}^tD\boldsymbol{y} = c$, one obtains a standard ellipse in this new coordinate system (see Figure 3.9 (b)). Finally, for general $\boldsymbol{\mu}$, the trace is given by simple translation (see Figure 3.10).

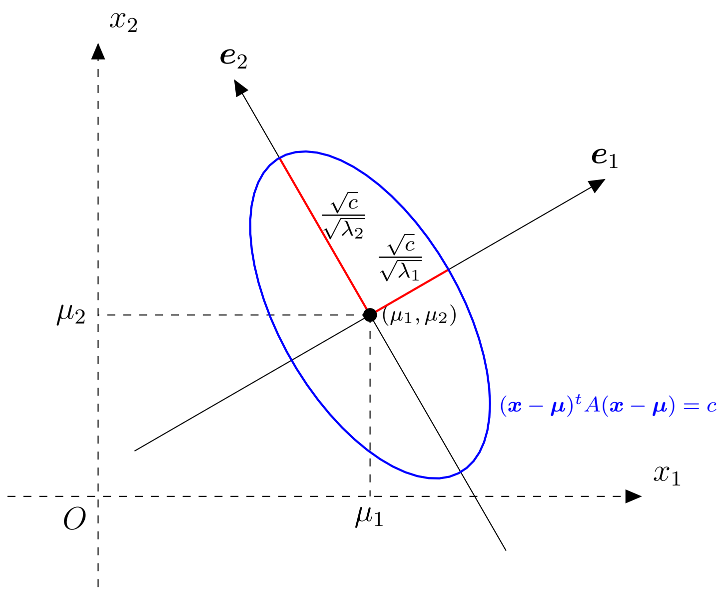

Figure 3.10: Ellipse described by the equation $(\boldsymbol{x} - \boldsymbol{\mu})^tA(\boldsymbol{x} - \boldsymbol{\mu}) = c$. $\lambda_1 > \lambda_2$ are eigenvalues of a positive definite matrix A and $\boldsymbol{e}_1$, $\boldsymbol{e}_2$ are eigenvectors corresponding to λ_1, λ_2, respectively.

Remark 3.9.15. Suppose that $X_1, X_2, \ldots, X_n$ denote a random sample from a multivariate normal population. When making an inference on the population mean, $\boldsymbol{\mu}_0$, the statistic

$$T^2 = (\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0)^t \left(\frac{\boldsymbol{S}}{n}\right)^{-1} (\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0) = n(\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0)^t\boldsymbol{S}^{-1}(\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0),$$

called Hotelling's T^2, plays a crucial role. The arguments used in Example 3.9.14, for example, can be used to determine the confidence region of $\boldsymbol{\mu}_0$ in bivariate normal case.

Theorem 3.9.16 (Singular Value Decomposition). *Let A be an $m \times n$ matrix with $m \leq n$. Then there are an $m \times m$ orthogonal matrix U, an $n \times n$ orthogonal matrix V, and an $m \times n$ rectangular diagonal matrix Σ such that $A = U\Sigma V^T$ and Σ is of the form*

$$\Sigma = \begin{bmatrix} \sigma_1 & & & & \\ & \sigma_2 & & & \\ & & \ddots & & \\ & & & \sigma_m & \end{bmatrix}$$

with

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_m \geq 0.$$

Remark 3.9.17.

(a) The factorization in Theorem 3.9.16 is called a singular value decomposition of A.

(b) The numbers $\sigma_1, \ldots, \sigma_m$, called the singular values of A, are square roots of eigenvalues of AA^t (see Exercise 3.6.6) and hence uniquely determined.

(c) The columns of U are eigenvectors of AA^t and the columns of V are eigenvectors of A^tA.

(d) The rank of A is r if and only if $\sigma_1 \geq \cdots \geq \sigma_r > \sigma_{r+1} = 0$.

The next theorem states how to approximate a given matrix with a matrix with a smaller rank. Recall the definitions of the Hilbert-Schmidt norm and the operator norm of a matrix (see Remark 3.2.29).

Theorem 3.9.18 (Eckart-Young Theorem). *Let $m \leq n$. Let A be an $m \times n$ matrix with a singular value decomposition $A = U\Sigma V^t$ and $\operatorname{rank}(A) = r$. Let $k < r$ and $A_k = U\Sigma_k V^t$, where Σ_k is a truncation of Σ defined by*

$$\Sigma_k = \begin{bmatrix} \sigma_1 & & & & & & & \\ & \sigma_2 & & & & & & \\ & & \ddots & & & & & \\ & & & \sigma_k & & & & \\ & & & & 0 & & & \\ & & & & & \ddots & & \\ & & & & & & 0 & \end{bmatrix}.$$

Then

$$\min\{\|A - B\|_{HS} : \operatorname{rank}(B) \leq k\} = \|A - A_k\|_{HS} = \sqrt{\sigma_{k+1}^2 + \sigma_{k+2}^2 + \cdots + \sigma_r^2}$$

and

$$\min\{\|A - B\|_{op} : \operatorname{rank}(B) \leq k\} = \|A - A_k\|_{op} = \sigma_{k+1}.$$

Example 3.9.19 (Principal Component Analysis)**.** For $1 \le i \le N$, let $\boldsymbol{X}_i$ be a multivariate data point in $\mathbb{R}^p$ and let $X = \begin{bmatrix} | & | & & | \\ \boldsymbol{X}_1 & \boldsymbol{X}_2 & \cdots & \boldsymbol{X}_N \\ | & | & & | \end{bmatrix}$, as in Remark 3.2.32. For simplicity, we assume that the mean $\bar{\boldsymbol{X}} = \dfrac{1}{N}\displaystyle\sum_{i=1}^{N} \boldsymbol{X}_i = \mathbf{0}$ – otherwise, one can simply replace $\boldsymbol{X}_i$ by $\boldsymbol{X}_i - \bar{\boldsymbol{X}}$. Let $k < p$. The goal of Principal Component Analysis is to find a k-dimensional subspace W_0 of $\mathbb{R}^p$ and the corresponding projection matrix P_{W_0} so that $\{P_{W_0}\boldsymbol{X}_1, \dots, P_{W_0}\boldsymbol{X}_N\}$ preserves as much information of the distribution of $\{\boldsymbol{X}_1, \dots, \boldsymbol{X}_N\}$ as possible, by which we mean that the total variance (see Remark 3.2.32) of $\{P_{W_0}\boldsymbol{X}_1, \dots, P_{W_0}\boldsymbol{X}_N\}$ is kept as large as possible (see Figure 3.11). Let W be a k-dimensional subspace of $\mathbb{R}^p$. Define $\boldsymbol{Y}_i = P_W\boldsymbol{X}_i$ and $Y = \begin{bmatrix} | & | & & | \\ \boldsymbol{Y}_1 & \boldsymbol{Y}_2 & \cdots & \boldsymbol{Y}_N \\ | & | & & | \end{bmatrix}$, then it follows that $Y = P_W X$. By Problem 3.5.7,

$$\|Y\|_{HS}^2 = tr(Y^tY) = tr(X^tP_W^tP_WX) = tr(X^tP_WX) = tr(X^tY)$$

and

$$\begin{aligned} \|X-Y\|_{HS}^2 &= tr((X-Y)^t(X-Y)) \\ &= tr(X^tX) - tr(X^tY) - tr(Y^tX) + tr(Y^tY) \\ \text{(Exercise 3.2.6)} &= tr(X^tX) - tr(X^tY) - tr((Y^tX)^t) + tr(Y^tY) \\ &= tr(X^tX) - 2tr(X^tY) + tr(Y^tY) \\ &= \|X\|_{HS}^2 - \|Y\|_{HS}^2. \end{aligned} \tag{3.9.3}$$

Since

$$\bar{\boldsymbol{Y}} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{Y}_i = \frac{1}{N}\sum_{i=1}^{N}P_W\boldsymbol{X}_i = P_W\bar{\boldsymbol{X}} = \mathbf{0},$$

by (3.9.3), the total variance of $\{\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_N\}$ is given by

$$\frac{1}{N-1}\|Y\|_{HS}^2 = \frac{1}{N-1}\left(\|X\|_{HS}^2 - \|X-Y\|_{HS}^2\right) \tag{3.9.4}$$

and is maximized when Y is chosen so that $\|X-Y\|_{HS}$ is minimized. To this end, let $X = U\Sigma V^t$ be a singular value decomposition of X as in Theorem 3.9.18, and let $P = UQU^t$, where

$$Q = \begin{bmatrix} 1 & & & & & & \\ & 1 & & & & & \\ & & \ddots & & & & \\ & & & 1 & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{bmatrix}$$

is a $p \times p$ diagonal matrix whose only nonzero elements are 1's placed at the first k diagonal entries. Note that $P^2 = P = P^t$. By Exercise 3.5.4, if we define $W_0 = \{P\boldsymbol{x} : \boldsymbol{x} \in \mathbb{R}^p\}$, then it follows that $P_{W_0} = P$, $\dim W_0 = \operatorname{rank}(P_{W_0}) = k$, and $P_{W_0}X = U\Sigma_k V^t$, where Σ_k is as in Theorem 3.9.18. By Eckart-Young Theorem, the largest possible total variance is obtained by $\{P_{W_0}\boldsymbol{X}_1, \ldots, P_{W_0}\boldsymbol{X}_N\}$.

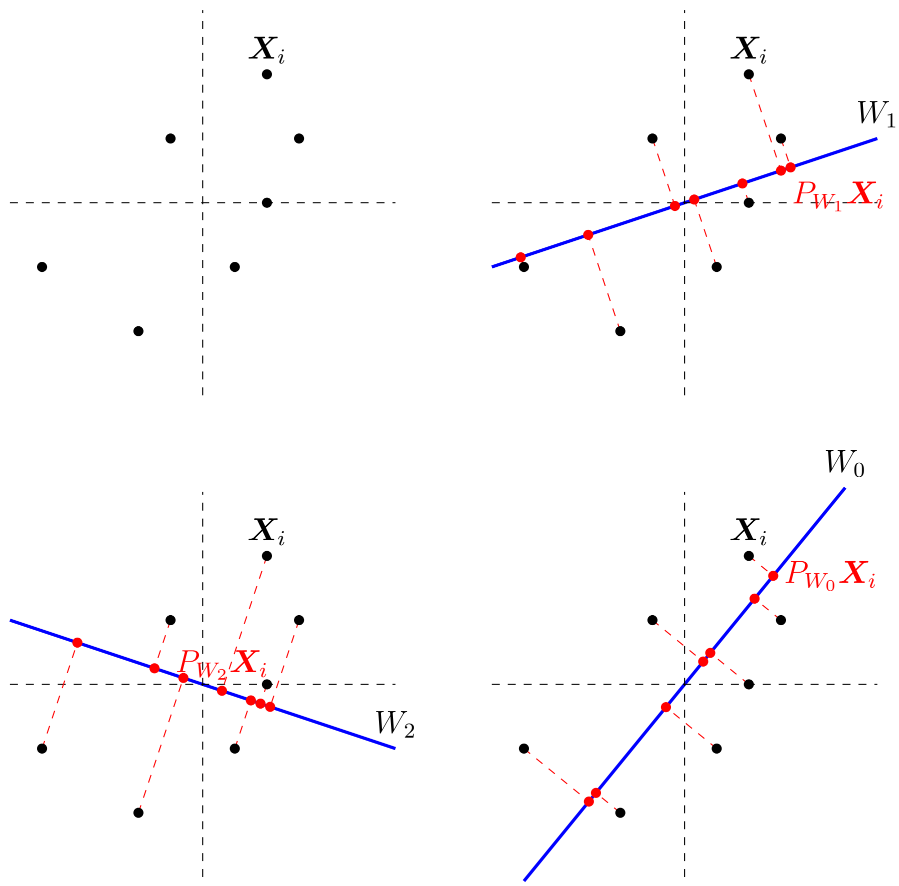

Figure 3.11: Principal component analysis when $p = 2$ and $k = 1$. $W = W_0$ yields the largest total variance of $\{P_W\boldsymbol{X}_1, \ldots, P_W\boldsymbol{X}_N\}$ (red dots).

Remark 3.9.20. Let X and Y be as in Example 3.9.19. Since

$$\|X - Y\|_{HS}^2 = \sum_{i=1}^{N} \|\boldsymbol{X}_i - \boldsymbol{Y}_i\|^2 = \sum_{i=1}^{N} \|\boldsymbol{X}_i - P_W\boldsymbol{X}_i\|^2,$$

by (3.9.4), the total variance of $\{\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_N\}$ is maximized when the sum of the squares of the distances between $\boldsymbol{X}_i$ and $P_W\boldsymbol{X}_i$ is minimized. See Figure 3.12.

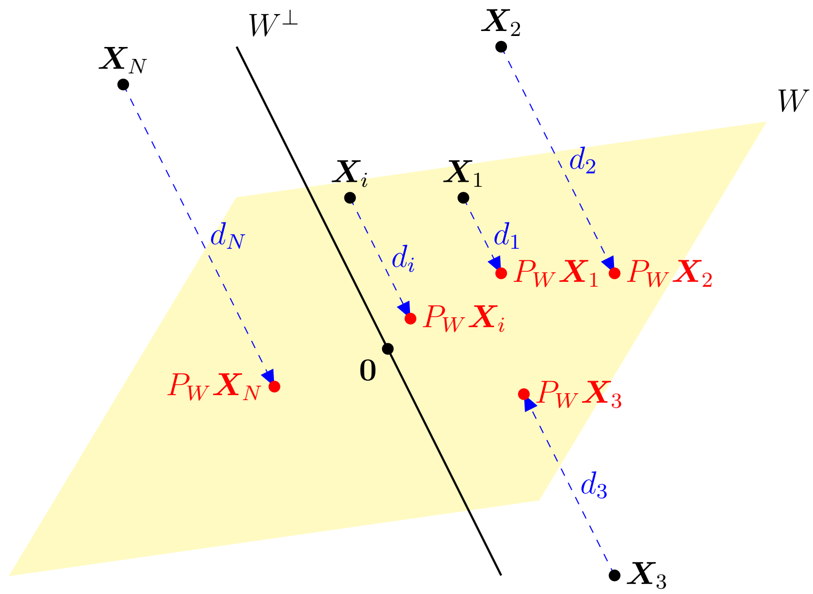

Figure 3.12: Principal component analysis when $p = 3$ and $k = 2$. The largest total variance of $\{P_W\boldsymbol{X}_1, \ldots, P_W\boldsymbol{X}_N\}$ is obtained when $\|X - P_WX\|_{HS} = \sqrt{\sum_{i=1}^N {d_i}^2}$ is minimized.

Exercise 3.9.

1. Let λ and μ be distinct eigenvalues of a symmetric matrix A. Let $\boldsymbol{\xi}$ (respectively, $\boldsymbol{\eta}$) be an eigenvector of A corresponding to λ (respectively, μ). Show that $\langle\boldsymbol{\xi}, \boldsymbol{\eta}\rangle = 0$.
Hint: Compare $\lambda\langle\boldsymbol{\xi}, \boldsymbol{\eta}\rangle$ and $\mu\langle\boldsymbol{\xi}, \boldsymbol{\eta}\rangle$. You might want to use Problem 3.2.25.

2. Let λ be an eigenvalue of a positive definite (respectively, positive semidefinite) matrix A. Show that $\lambda > 0$ (respectively, $\lambda \geq 0$).

3. Let A be an $n \times n$ positive semidefinite matrix and $c > 0$ be a constant. Show that $A + cI_n$ is invertible.
Hint: Show that 0 cannot be an eigenvalue of $A + cI_n$. You may want to use Exercise 3.6.3.

4. Let $Q(\boldsymbol{x}) = \boldsymbol{x}^t A\boldsymbol{x}$ be a quadratic form. Show that there is a *symmetric* matrix B such that $Q(\boldsymbol{x}) = \boldsymbol{x}^t B\boldsymbol{x}$.
Hint: Since $Q(\boldsymbol{x}) = \boldsymbol{x}^t A\boldsymbol{x}$ is a scalar, $Q(\boldsymbol{x}) = (\boldsymbol{x}^t A\boldsymbol{x})^t = \boldsymbol{x}^t A^t\boldsymbol{x}$.

5. Let $M = \begin{bmatrix} 4-b & -2 \\ a & b \end{bmatrix}$. Determine a and b so that M is positive semidefinite.

6. Let $A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$. If $a > 0$ and $ac - b^2 > 0$, show that A is positive definite.

7. Let $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ be linearly independent vectors in $\mathbb{R}^m$ (so $m \geq n$ by Exercise 3.3.8). Let $\langle \cdot, \cdot \rangle$ denote the inner product in $\mathbb{R}^n$ and define $v_{ij} = \langle \boldsymbol{y}_i, \boldsymbol{y}_j \rangle$. Show that the matrix $V = [v_{ij}]$ is positive definite.

8. The goal of this exercise is to show that the compound symmetry structure of the variance-covariance matrix for longitudinal studies is positive definite. Let $0 \leq \rho < 1$.

Show that $V = \begin{bmatrix} 1 & \rho & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & \rho & \cdots & \rho \\ \rho & \rho & 1 & \rho & \cdots & \rho \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \rho & \rho & \cdots & 1 \end{bmatrix}$ is positive definite.

Hint: For $1 \leq i \leq n$, define $\boldsymbol{y}_i \in \mathbb{R}^{n+1}$ by

$$\boldsymbol{y}_1 = \frac{1}{\sqrt{1+\alpha^2}} \begin{bmatrix} \alpha \\ 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \boldsymbol{y}_2 = \frac{1}{\sqrt{1+\alpha^2}} \begin{bmatrix} \alpha \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \cdots, \boldsymbol{y}_n = \frac{1}{\sqrt{1+\alpha^2}} \begin{bmatrix} \alpha \\ 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix},$$

where $\alpha = \sqrt{\frac{\rho}{1-\rho}}$ so that $\frac{\alpha^2}{1+\alpha^2} = \rho$. Use Exercise 3.9.7.

9. Show that $\begin{bmatrix} 1 & -\frac{2}{3} & -\frac{2}{3} \\ -\frac{2}{3} & 1 & -\frac{2}{3} \\ -\frac{2}{3} & -\frac{2}{3} & 1 \end{bmatrix}$ is not positive definite. This shows that the condition $\rho \geq 0$ is indispensable in Exercise 3.9.8.
Hint: You may want to use Sylvester's Criterion.

10. Let S be an $n \times n$ positive definite matrix. For $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, define $[\boldsymbol{x}, \boldsymbol{y}]_S$ by $\langle \boldsymbol{x}, S\boldsymbol{y} \rangle$. Show that $[\cdot, \cdot]$ satisfies all properties in Theorem 3.2.12.
Hint: You may want to use Theorem 3.9.7 and Problem 3.2.25.

11. Let S be a square matrix such that $S^2 = S = S^t$. Show that $\text{rank}(S) = tr(S)$.
Hint: Consider a diagonalization of S. You may find Exercise 3.6.4 useful.

12. This problem is about the optimization of the Rayleigh quotient. Let A be an $n \times n$ positive definite matrix with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$.

(a) Show that $\max\{\frac{\boldsymbol{x}^t A \boldsymbol{x}}{\boldsymbol{x}^t \boldsymbol{x}} : \boldsymbol{x} \neq \boldsymbol{0}\} = \lambda_1$ and $\min\{\frac{\boldsymbol{x}^t A \boldsymbol{x}}{\boldsymbol{x}^t \boldsymbol{x}} : \boldsymbol{x} \neq \boldsymbol{0}\} = \lambda_n$.

(b) Let B be an $n \times n$ positive definite matrix. Show that

$$\max\left\{\frac{\boldsymbol{x}^t A \boldsymbol{x}}{\boldsymbol{x}^t B \boldsymbol{x}} : \boldsymbol{x} \neq \boldsymbol{0}\right\}$$

is the largest eigenvalue of AB^{-1}.
Hint: By Theorem 3.9.7.(b), there is a matrix C such that $B = C^t C$ and it follows that

$$\max\left\{\frac{\boldsymbol{x}^t A\boldsymbol{x}}{\boldsymbol{x}^t B\boldsymbol{x}} : \boldsymbol{x} \neq \mathbf{0}\right\} = \max\left\{\frac{\boldsymbol{y}^t (C^{-1})^t A C^{-1}\boldsymbol{y}}{\boldsymbol{y}^t \boldsymbol{y}} : \boldsymbol{y} \neq \mathbf{0}\right\}.$$

Use part (a) above. You may want to use Problem 3.6.5 and Problem 3.2.35.

13. Let A be an $m \times n$ matrix with $m \leq n$. Show that there are an $m \times n$ matrix U, an $n \times n$ orthogonal matrix V, and an $n \times n$ diagonal matrix Σ such that $A = U\Sigma V^T$ and columns of U are orthonormal (this factorization is called the reduced singular value decomposition of A).

Chapter 4

Multivariable Calculus

4.1 Partial Derivatives and Interchange of Operations

Let $f : \mathbb{R}^n \to \mathbb{R}$. f is said to be continuous at $\boldsymbol{x}_0$ if for $\epsilon > 0$, there exists $\delta > 0$ such that $\|\boldsymbol{x} - \boldsymbol{x}_0\| < \delta$ implies that $|f(\boldsymbol{x}) - f(\boldsymbol{x}_0)| < \epsilon$. Here $\|\boldsymbol{x} - \boldsymbol{x}_0\|$ denotes the distance between $\boldsymbol{x}$ and $\boldsymbol{x}_0$ in $\mathbb{R}^n$: see Definition 3.2.13 for more detail regarding the geometry of $\mathbb{R}^n$.

Let $\boldsymbol{x} = (x_1, \ldots, x_n)$. The partial derivative $\frac{\partial f}{\partial x_i}$ ($D_i f$ and f_{x_i} are also used) of f with respect to x_i is a map from $\mathbb{R}^n$ to $\mathbb{R}$ defined by

$$\frac{\partial f}{\partial x_i}(\boldsymbol{x}) = \lim_{h \to 0} \frac{f(x_1, \ldots, x_{i-1}, x_i + h, x_{i+1}, \ldots, x_n) - f(x_1, \ldots, x_i, \ldots, x_n)}{h}.$$

Roughly speaking, $\frac{\partial f}{\partial x_i}$ is obtained by differentiating f with respect to the i^{th} variable x_i, treating other variables as constants.
One can also consider higher order partial derivatives. For example,

$$\begin{aligned}\frac{\partial^2 f}{\partial x_i \partial x_j}(\boldsymbol{x}) &= \frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}\right)(\boldsymbol{x}) \\ &= \lim_{h \to 0} \frac{\frac{\partial f}{\partial x_j}(x_1, \ldots, x_{i-1}, x_i + h, x_{i+1}, \ldots, x_n) - \frac{\partial f}{\partial x_j}(x_1, \ldots, x_i, \ldots, x_n)}{h}.\end{aligned}$$

In general, the order of differentiation matters and it can happen that $\frac{\partial^2 f}{\partial x_i \partial x_j} \neq \frac{\partial^2 f}{\partial x_j \partial x_i}$. Under some mild conditions, however, we have that $\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$, as the next result shows.

Theorem 4.1.1. *Let $f : \mathbb{R}^n \to \mathbb{R}$. If the second partial derivatives $\frac{\partial^2 f}{\partial x_j \partial x_i}$ and $\frac{\partial^2 f}{\partial x_i \partial x_j}$ are continuous, then $\frac{\partial^2 f}{\partial x_j \partial x_i} = \frac{\partial^2 f}{\partial x_i \partial x_j}$.*

In many applications (e.g. moment generating function), we need to differentiate a function defined by integration and a natural question arises: is the derivative of an integral the same as the integral of a derivative? The following result deals with differentiation under integral sign.

Theorem 4.1.2 (Leibniz Integral Rule). *Let $f(t,x)$ be a function such that $f_t(x) = f(t,x)$ is integrable for all t and $\frac{\partial f}{\partial t}$ is continuous in t for each x. Then*

$$\frac{d}{dt}\int_{a(t)}^{b(t)} f(t,x)\,dx = f(t,b(t))b'(t) - f(t,a(t))a'(t) + \int_{a(t)}^{b(t)} \frac{\partial f}{\partial t}(t,x)\,dx.$$

In particular, if $a(t) = a$ and $b(t) = b$ are constant functions, then

$$\frac{d}{dt}\int_a^b f(t,x)\,dx = \int_a^b \frac{\partial f}{\partial t}(t,x)\,dx. \tag{4.1.1}$$

Remark 4.1.3. (4.1.1) is valid even when a and/or b is/are $\pm\infty$.

Example 4.1.4. Let X be a continuous random variable with the pdf f. For a real number t, define the moment generating function $\varphi_X(t)$ of X by

$$\varphi_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f(x)\,dx. \tag{4.1.2}$$

Differentiating (4.1.2) with respect to t and applying Leibniz Integral Rule,

$$\frac{d}{dt}\varphi_X(t)\Big|_{t=0} = \int_{-\infty}^{\infty} x e^{tx} f(x)\,dx\Big|_{t=0} = \int_{-\infty}^{\infty} x f(x)\,dx = E(X).$$

Differentiating (4.1.2) twice with respect to t, we get

$$\frac{d^2}{dt^2}\varphi_X(t)\Big|_{t=0} = \int_{-\infty}^{\infty} x^2 e^{tx} f(x)\,dx\Big|_{t=0} = \int_{-\infty}^{\infty} x^2 f(x)\,dx = E(X^2).$$

In general, it follows that

$$\frac{d^n}{dt^n}\varphi_X(t)\Big|_{t=0} = E(X^n). \tag{4.1.3}$$

Remark 4.1.5. The concept of the moment generating function extends to discrete random variables and (4.1.3) is still valid provided that the term-by-term differentiation

$$\frac{d^n}{dt^n}\left(\sum_k e^{tk} P(X=k)\right) = \sum_k \frac{d^n}{dt^n}\left(e^{tk} P(X=k)\right) \tag{4.1.4}$$

is valid. In fact, under some conditions (which include uniform convergence of the right hand side of (4.1.4)), one can show that (4.1.4) is true.

Example 4.1.6 (See Example 2.7.11 and Exercise 2.7.5)**.** Let $X \sim Exp(\lambda)$, then

$$\varphi_X(t) = \int_0^\infty e^{tx}\frac{1}{\lambda}e^{-\frac{x}{\lambda}}\,dx = \frac{1}{\lambda}\int_0^\infty e^{(t-\frac{1}{\lambda})x}\,dx = \frac{1}{1-\lambda t}$$

for $t < \frac{1}{\lambda}$, so

$$\varphi_X'(0) = \left.\frac{\lambda}{(1-\lambda t)^2}\right|_{t=0} = \lambda \quad \text{and} \quad \varphi_X''(0) = \left.\frac{2\lambda^2}{(1-\lambda t)^3}\right|_{t=0} = 2\lambda^2.$$

It follows that $E(X) = \lambda$ and $Var(X) = E(X^2) - E(X)^2 = 2\lambda^2 - \lambda^2 = \lambda^2$.

We close this section with a couple of results that justifies the interchange of the order of limit and integration.

Theorem 4.1.7 (Lebesgue's Dominated Convergence Theorem)**.** *Let (f_n) be a sequence of functions on $[a, b]$. Suppose that there exists a function g such that*

(a) $|f_n(x)| \le g(x)$ *for all* $x \in [a, b]$ *and* n *and*

(b) $\int_a^b g(x)\,dx < \infty$.

Then

$$\lim_{n\to\infty}\int_a^b f_n(x)\,dx = \int_a^b \left(\lim_{n\to\infty} f_n(x)\right)\,dx.$$

Theorem 4.1.8 (Lebesgue's Monotone Convergence Theorem)**.** *Let (f_n) be a sequence of functions on $[a, b]$ such that $0 \le f_k(x) \le f_{k+1}(x)$ for all k and for all $x \in [a, b]$. Then*

$$\lim_{n\to\infty}\int_a^b f_n(x)\,dx = \int_a^b \left(\lim_{n\to\infty} f_n(x)\right)\,dx.$$

Exercise 4.1.

1. Let $f(x, y) = \ln(x^2y + y^3)$. Compute $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$. Show that $\frac{\partial^2 f}{\partial x\partial y} = \frac{\partial^2 f}{\partial y\partial x}$.

4.2 Finding Extreme Values

As in the case of one variable function, there is a multivariate Taylor series expansion. Before we proceed, we need a multivariate version of the first and second derivatives.

Definition 4.2.1. Let $f:\mathbb{R}^n \to \mathbb{R}$. The gradient ∇f of f is a (row) n-vector defined by $(\frac{\partial f}{\partial x_1}(\boldsymbol{x}),\ldots,\frac{\partial f}{\partial x_n}(\boldsymbol{x}))$. The Hessian $H(f)(\boldsymbol{x})$ of f is an $n\times n$ matrix given by the Jacobian of ∇f, that is,

$$H(f)(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(\boldsymbol{x}) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(\boldsymbol{x}) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(\boldsymbol{x}) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(\boldsymbol{x}) & \frac{\partial^2 f}{\partial x_2^2}(\boldsymbol{x}) & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(\boldsymbol{x}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(\boldsymbol{x}) & \frac{\partial^2 f}{\partial x_n \partial x_2}(\boldsymbol{x}) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(\boldsymbol{x}) \end{bmatrix}.$$

Remark 4.2.2.

(a) A point $\boldsymbol{x}_0$ is called a critical point of f if $\nabla f(\boldsymbol{x}_0) = (0,0,\ldots,0)$.

(b) By Theorem 4.1.1, if the second partial derivatives of f are continuous, then $H(f)$ is symmetric (see Definition 3.9.1).

Theorem 4.2.3. *Let $f:\mathbb{R}^n \to \mathbb{R}$. If all second order partial derivatives of f exist, then*

$$f(\boldsymbol{x}) = f(\boldsymbol{x}_0) + \nabla f(\boldsymbol{x}_0)(\boldsymbol{x}-\boldsymbol{x}_0) + \frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}_0)^t Hf(\boldsymbol{x}_1)(\boldsymbol{x}-\boldsymbol{x}_0)$$

for some point $\boldsymbol{x}_1$ lying on the line segment that connects $\boldsymbol{x}_0$ and $\boldsymbol{x}$.

Remark 4.2.4. Theorem 4.2.3 can be viewed as a multivariate version of Taylor expansion:

$$f(x) = f(x_0) + f'(x_0)(x-x_0) + \frac{1}{2}f''(x_1)(x-x_0)^2$$

for some x_1 between x_0 and x.

Now we are ready to introduce a multivariate version of the second derivative test.

Theorem 4.2.5 (Multivariate Second Derivative Test)**.** *Let $f:\mathbb{R}^n \to \mathbb{R}$ and suppose that $\nabla f(\boldsymbol{x}_0) = (0,0,\ldots,0)$.*

(a) If $Hf(\boldsymbol{x}_0)$ is positive definite, then f achieves a local minimum at $\boldsymbol{x}_0$.

(b) If $Hf(\boldsymbol{x}_0)$ is negative definite, then f achieves a local maximum at $\boldsymbol{x}_0$.

Example 4.2.6. Let $f(x,y) = x^3 + y^3 - 3xy$. Since $\nabla f = (3x^2 - 3y, 3y^2 - 3x)$, we have two critical points $\boldsymbol{x}_0 = (0,0)$ and $\boldsymbol{x}_1 = (1,1)$. Since $Hf(\boldsymbol{x}) = \begin{bmatrix} 6x & -3 \\ -3 & 6y \end{bmatrix}$, we have $Hf(\boldsymbol{x}_0) = \begin{bmatrix} 0 & -3 \\ -3 & 0 \end{bmatrix}$ and $Hf(\boldsymbol{x}_1) = \begin{bmatrix} 6 & -3 \\ -3 & 6 \end{bmatrix}$. Using Problem 3.9.9, one can easily check that $Hf(\boldsymbol{x}_0)$ is neither positive definite nor negative definite, while $Hf(\boldsymbol{x}_1)$ is positive definite. Therefore, we conclude that f has a local minimum at $(1,1)$.

Example 4.2.7. Recall Example 3.5.9. We will find β_0 and β_1 that minimize (3.5.1) using the Multivariate Second Derivative Test. Let $f(\beta_0, \beta_1) = \sum_{i=1}^n |y_i - (\beta_0 + \beta_1 x_i)|^2 = \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2$, then

$$\frac{\partial f}{\partial \beta_0} = \sum_{i=1}^n 2(y_i - (\beta_0 + \beta_1 x_i))(-1)$$

and

$$\frac{\partial f}{\partial \beta_1} = \sum_{i=1}^n 2(y_i - (\beta_0 + \beta_1 x_i))(-x_i),$$

so the equation$\nabla f = (0,0)$ gives (using the notations as in Example 3.5.9)

$$\begin{cases} n\bar{y} - n\beta_0 - n\beta_1 \bar{x} = 0 \\ \sum_{i=1}^n x_i y_i - n\beta_0 \bar{x} - \beta_1 \sum_{i=1}^n x_i^2 = 0 \end{cases},$$

which in turn yields $\beta_1 = \frac{SXY}{SXX}$ and $\beta_0 = \bar{y} - \frac{SXY}{SXX}\bar{x}$. To show that this is a local minimum, we compute $H(f) = \begin{bmatrix} 2n & 2n\bar{x} \\ 2n\bar{x} & 2\sum_{i=1}^n x_i^2 \end{bmatrix}$. Since $\sum_{i=1}^n x_i^2 - n\bar{x}^2 = \sum_{i=1}^n (x_i - \bar{x})^2$, by Exercise 3.9.6, $H(f)$ is easily shown to be positive definite.

Exercise 4.2.

1. For a real number x and a positive real number y, define

$$f(x,y) = \frac{1}{y} e^{-\frac{(x-1)^2}{2y^2}}.$$

(a) Compute the gradient vector ∇f and the Hessian matrix $H(f)$.

(b) Find a local maximum of f.

4.3 Matrix Differentiation

In this section, we deal with differentiation that involves matrices. Just as the differentiation of a single variable function plays an important role in finding extreme values of the function, matrix differentiation gives a convenient way to compute the extreme values of a multivariable function, which is essential in multiple linear regressions. We start with a convention that will be used throughout the section.

Definition 4.3.1. Let $\boldsymbol{y} = \boldsymbol{y}(\boldsymbol{x}) = \begin{bmatrix} y_1(\boldsymbol{x}) \\ \vdots \\ y_m(\boldsymbol{x}) \end{bmatrix}$ such that each y_i is a differentiable function of $\boldsymbol{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$. We define $\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}}$ to be an $m \times n$ matrix given by

$$\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \frac{\partial y_m}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix}.$$

In particular, if $m = 1$, then $\frac{\partial y}{\partial \boldsymbol{x}}$ is nothing but the gradient ∇y of y.

It is easy to verify that the operator $\frac{\partial}{\partial \boldsymbol{x}}$ is linear. To be more precise, if $\boldsymbol{y}, \boldsymbol{z}$ are $n \times 1$ vectors whose components are functions of $\boldsymbol{x}$, and if c is a constant, then

$$\frac{\partial(\boldsymbol{y}+\boldsymbol{z})}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}} + \frac{\partial \boldsymbol{z}}{\partial \boldsymbol{x}} \quad \text{and} \quad \frac{\partial(c\boldsymbol{y})}{\partial \boldsymbol{x}} = c\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}}. \tag{4.3.1}$$

Example 4.3.2. Let $A = [a_{ij}]$ be an $m \times n$ constant matrix and $\boldsymbol{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$. Let $\boldsymbol{y} = A\boldsymbol{x}$ so that $\boldsymbol{y}$ is an $m \times 1$ vector with component $y_i(\boldsymbol{x}) = \sum_{j=1}^n a_{ij}x_j$, $1 \le j \le m$. Since $\frac{\partial y_i}{\partial x_j} = a_{ij}$, we see that $\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}} = A$.

Problem 4.3.3. Let $\boldsymbol{c} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$ be an $n \times 1$ constant vector and $\boldsymbol{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$. Let $y = \boldsymbol{x}^t\boldsymbol{c}$ (note that y is a scalar). Compute $\frac{\partial y}{\partial \boldsymbol{x}}$.

Answer First we note that the result must be a $1 \times n$ vector. Since $y = \sum_{j=1}^n c_jx_j$, we get $\frac{\partial y}{\partial x_j} = c_j$, it follows that $\frac{\partial y}{\partial \boldsymbol{x}} = \boldsymbol{c}^t$. ■

The next result can be regarded as a matrix version of the Chain Rule.

Theorem 4.3.4. *Let $\boldsymbol{y} = A\boldsymbol{x}$, where A is an $m \times n$ constant matrix and $\boldsymbol{x}$ is an $n \times 1$ vector. Let $\boldsymbol{z}$ be an $\ell \times 1$ vector, then*

$$\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{z}} = A\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{z}}.$$

Next, we consider differentiation of a quadratic form.

Theorem 4.3.5. *Let $y = \boldsymbol{x}^tA\boldsymbol{x}$, where A is an $n \times n$ constant matrix and $\boldsymbol{x}$ is an $n \times 1$ vector. Then $\frac{\partial y}{\partial \boldsymbol{x}} = \boldsymbol{x}^t(A + A^t)$. In particular, if A is symmetric, then $\frac{\partial y}{\partial \boldsymbol{x}} = 2\boldsymbol{x}^tA$.*

Problem 4.3.6. Let $y = \|A\boldsymbol{x} - \boldsymbol{b}\|^2$, where A is an $m \times n$ constant matrix and $\boldsymbol{x}, \boldsymbol{b}$ are $m \times 1$ vectors. Compute $\frac{\partial y}{\partial \boldsymbol{x}}$.

Answer Note that $y = (A\boldsymbol{x} - \boldsymbol{b})^t(A\boldsymbol{x} - \boldsymbol{b}) = \boldsymbol{x}^t A^t A\boldsymbol{x} - \boldsymbol{x}^t A^t \boldsymbol{b} - \boldsymbol{b}^t A\boldsymbol{x} + \boldsymbol{b}^t\boldsymbol{b}$. By (4.3.1), Example 4.3.2, Problem 4.3.3, and Theorem 4.3.5, we obtain

$$\begin{aligned}\frac{\partial y}{\partial \boldsymbol{x}} &= \frac{\partial(\boldsymbol{x}^t A^t A\boldsymbol{x})}{\partial \boldsymbol{x}} - \frac{\partial(\boldsymbol{x}^t A^t \boldsymbol{b})}{\partial \boldsymbol{x}} - \frac{\partial(\boldsymbol{b}^t A\boldsymbol{x})}{\partial \boldsymbol{x}} + \frac{\partial(\boldsymbol{b}^t\boldsymbol{b})}{\partial \boldsymbol{x}} \\ &= 2\boldsymbol{x}^t A^t A - \boldsymbol{b}^t A - \boldsymbol{b}^t A + \boldsymbol{0} \\ &= 2\boldsymbol{x}^t A^t A - 2\boldsymbol{b}^t A.\end{aligned}$$

■

Remark 4.3.7. Least squares estimators for multiple linear regressions can be found using Problem 4.3.6. To be more precise, it is known that $y = \|A\boldsymbol{x}-\boldsymbol{b}\|^2$ is maximized when $\frac{\partial y}{\partial \boldsymbol{x}} = \boldsymbol{0}^t$. In addition, if A^tA is invertible, then $\|A\boldsymbol{x} - \boldsymbol{b}\|^2$ is maximized when $\boldsymbol{x} = (A^tA)^{-1}A^tb$. See Exercise 4.3.3

There is also a matrix version of the Product Rule.

Theorem 4.3.8. *Let $\boldsymbol{x}, \boldsymbol{y}$ be $n \times 1$ vectors and both $\boldsymbol{x}$ and $\boldsymbol{y}$ are functions of $\boldsymbol{z}$. Let $w = \boldsymbol{y}^t\boldsymbol{x}$, then $\frac{\partial w}{\partial \boldsymbol{z}} = \boldsymbol{x}^t\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{z}} + \boldsymbol{y}^t\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{z}}$.*

Problem 4.3.9. Let $w = \boldsymbol{y}^t A\boldsymbol{x}$, where $\boldsymbol{y}$ is an $m \times 1$ vector, A is an $m \times n$ constant matrix, and $\boldsymbol{x}$ is an $n \times 1$ vector. If both $\boldsymbol{y}$ and $\boldsymbol{x}$ are functions of $\boldsymbol{z}$, show that

$$\frac{\partial w}{\partial \boldsymbol{z}} = \boldsymbol{x}^t A^t \frac{\partial \boldsymbol{y}}{\partial \boldsymbol{z}} + \boldsymbol{y}^t A\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{z}}.$$

Answer Note that $w = (A^t\boldsymbol{y})^t\boldsymbol{x}$. By Theorem 4.3.8,

$$\frac{\partial w}{\partial \boldsymbol{z}} = \boldsymbol{x}^t\frac{\partial(A^t\boldsymbol{y})}{\partial \boldsymbol{z}} + (A^t\boldsymbol{y})^t\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{z}}.$$

The result follows by Example 4.3.2. ■

Exercise 4.3.

1. Let $y = \boldsymbol{x}^t\boldsymbol{x}$, where $\boldsymbol{x}$ is an $n \times 1$ vector. Show that $\frac{\partial y}{\partial \boldsymbol{x}} = 2\boldsymbol{x}^t$.

2. Let $y = \boldsymbol{x}^t\boldsymbol{x}$, where $\boldsymbol{x}$ is an $n \times 1$ vector. If $\boldsymbol{x}$ is a function of $\boldsymbol{z}$, show that $\frac{\partial y}{\partial \boldsymbol{z}} = 2\boldsymbol{x}^t\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{z}}$.

3. Let A be an $m \times n$ constant matrix and $\boldsymbol{b}$ be an $m \times 1$ constant vector. Let $\boldsymbol{x}$ be an $n \times 1$ vector and $y = \|A\boldsymbol{x} - \boldsymbol{b}\|^2$. Show that $\frac{\partial y}{\partial \boldsymbol{x}} = \boldsymbol{0}^t$ if and only if $A^tA\boldsymbol{x} = A^tb$.
Hint: See Problem 4.3.6.

4. Let A be an $m \times n$ constant matrix and $\boldsymbol{b}$ be an $m \times 1$ constant vector. Let W be a constant $m \times m$ positive semidefinite matrix. Compute $\frac{\partial y}{\partial \boldsymbol{x}}$, where

$$y = (A\boldsymbol{x} - \boldsymbol{b})^tW(A\boldsymbol{x} - \boldsymbol{b}).$$

Hint: You may want to use Problem 4.3.6 together with Theorem 3.9.7.

4.4 Double Integrals

Suppose that a function $f : \mathbb{R}^2 \to \mathbb{R}$ is nonnegative over a region D in $\mathbb{R}^2$ (see Figure 4.1). The volume under the surface of $z = f(x, y)$ and above the region D is denoted by

$$\iint_D f(x,y)\, dxdy$$

and called the double integral of f over D.

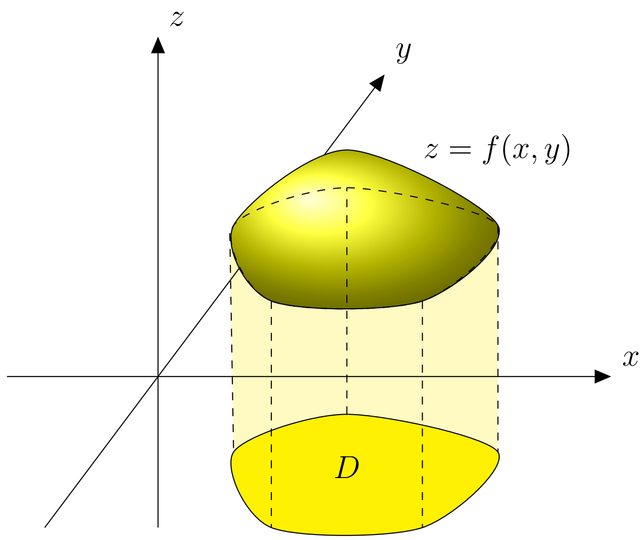

Figure 4.1: Double integral as volume under surface $z = f(x, y)$

Example 4.4.1. For $R > 0$, let $D = \{(x, y) : x^2 + y^2 \le R^2\}$. The function $f(x, y) = \sqrt{R^2 - (x^2 + y^2)}$ represents the upper hemisphere of radius R centered at $(0, 0, 0)$ in $\mathbb{R}^3$ and $\iint_D \sqrt{R^2 - (x^2 + y^2)}\, dxdy$ denotes its volume over D (see Figure 4.2). Therefore,

$$\iint_D \sqrt{R^2 - (x^2 + y^2)}\, dxdy = \frac{2}{3}\pi R^3.$$

As in the case of definite integral of a single variable function, the double integral of function that has both positive and negative parts should be interpreted as *signed* volume.

Example 4.4.2. When $f(x, y) = c$ is a constant function, $\iint_D f(x, y)\, dxdy = c \cdot \text{area}(D)$, where $\text{area}(D)$ denotes the area of the region D.

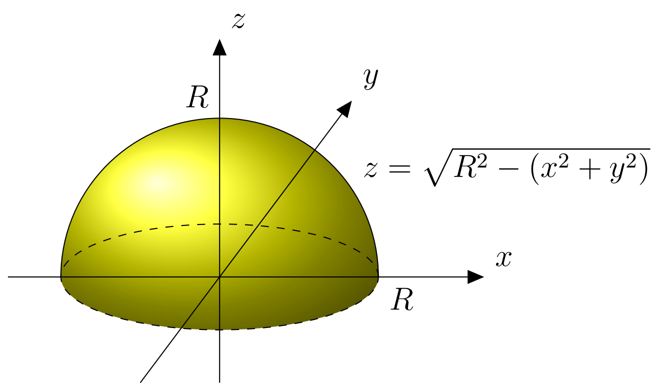

Figure 4.2: Graph of $z = \sqrt{R^2 - (x^2 + y^2)}$

As stated in the next two theorems, under certain conditions, double integrals can be calculated using iterated integrals.

Theorem 4.4.3 (Tonelli's Theorem)**.** *If f is nonnegative, then*

$$\iint_{\mathbb{R}^2} f(x,y)\,dxdy = \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} f(x,y)\,dy\right)\,dx = \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} f(x,y)\,dx\right)\,dy.$$

Theorem 4.4.4 (Fubini's Theorem)**.** *If* $\iint_{\mathbb{R}^2} |f(x,y)|\,dxdy < \infty$*, then*

$$\iint_{\mathbb{R}^2} f(x,y)\,dxdy = \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} f(x,y)\,dy\right)\,dx = \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} f(x,y)\,dx\right)\,dy.$$

Remark 4.4.5. Since Tonelli's Theorem applies only to nonnegative functions, when dealing with the double integral of a function that has both positive and negative parts, one usually relies on Fubini's Theorem. When applying Fubini's Theorem, to show that $\iint_{\mathbb{R}^2} |f(x,y)|\,dxdy < \infty$, one can use Tonelli's Theorem. To be more precise, since $|f(x,y)|$ is nonnegative, by Tonelli's Theorem it follows that

$$\iint_{\mathbb{R}^2} |f(x,y)|\,dxdy = \int_{\mathbb{R}}\left(\int_{\mathbb{R}} |f(x,y)|\,dy\right)\,dx = \int_{\mathbb{R}}\left(\int_{\mathbb{R}} |f(x,y)|\,dx\right)\,dy.$$

So if one can show that one of $\int_{\mathbb{R}}\left(\int_{\mathbb{R}} |f(x,y)|\,dy\right)\,dx$ or $\int_{\mathbb{R}}\left(\int_{\mathbb{R}} |f(x,y)|\,dx\right)\,dy$ is finite, then the double integral $\iint_{\mathbb{R}^2} f(x,y)\,dxdy$ can be evaluated via iterated integrals.

Example 4.4.6. Consider $\displaystyle\iint_D (3y+x)\,dxdy$, where $D = \{(x,y) : -3 \le x \le 1, 1 \le y \le 3\}$. For $(x,y) \in D$, we have $|3y+x| \le 3|y| + |x| \le 12$, so by Exercise 4.4.1, we see that $\displaystyle\iint_D |3y+x|\,dxdy \le 12 \cdot \text{area}(D)$. By Fubini's Theorem, it follows that

$$\iint_D (3y+x)\,dxdy = \int_1^3 \left(\int_{-3}^1 (3y+x)\ dx \right) dy = \int_1^3 (12y-4)\,dy = 40.$$

Problem 4.4.7. Let λ be a constant. Define $f : \mathbb{R}^2 \to \mathbb{R}$ by

$$f(x,y) = \begin{cases} \lambda x, & 0 < x < y < 1, \\ 0, & \text{otherwise.} \end{cases}$$

Determine λ so that $\displaystyle\iint_{\mathbb{R}^2} f(x,y)\,dxdy = 1$.

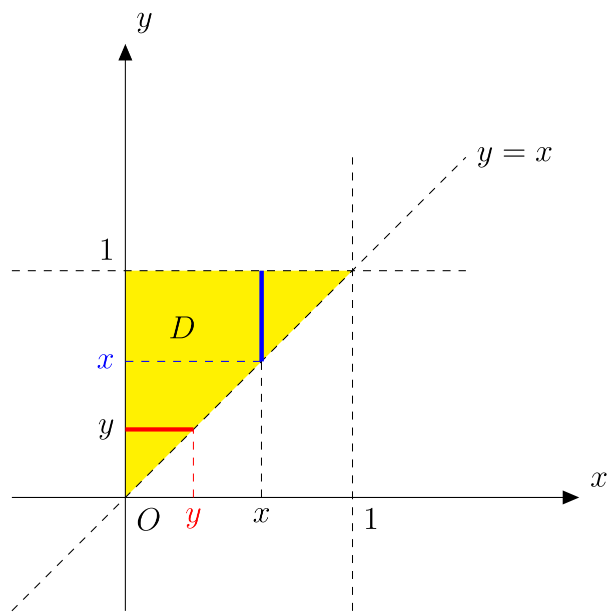

Figure 4.3: $D = \{(x,y) : 0 < x < y < 1\}$

Answer Let $D = \{(x,y) : 0 < x < y < 1\}$ (see Figure 4.3). Since f is zero outside D, $\displaystyle\iint_{\mathbb{R}^2} f(x,y)\,dxdy = \iint_D f(x,y)\,dxdy = \lambda \iint_D x\,dxdy$. Note that, in D, y ranges from 0 to 1 and corresponding x ranges from 0 to y, so

$$\iint_D x\,dxdy = \int_0^1 \left(\int_0^y x\,dx \right)\ dy = \int_0^1 \frac{y^2}{2}\,dy = \left[\frac{y^3}{6} \right]_0^1 = \frac{1}{6}$$

and it follows that $\lambda = 1$. Note that we could compute $\iint_D x\,dxdy$ using the other iterated integrals. Since x in D ranges from 0 to 1 and corresponding y ranges from x to 1, we get

$$\iint_D x\,dxdy = \int_0^1 \left(\int_x^1 x\,dy \right) dx = \int_0^1 x(1-x)\,dx = \left[\frac{x^2}{2} - \frac{x^3}{3} \right]_0^1 = \frac{1}{6}.$$

■

Problem 4.4.8. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by

$$f(x,y) = \begin{cases} 2e^{-x-y}, & 0 < x < y, \\ 0, & \text{otherwise.} \end{cases}$$

Compute $\iint_D f(x,y)\,dxdy$ and $\iint_E f(x,y)\,dxdy$, where

$$D = \{(x,y) : y < 1\} \quad \text{and} \quad E = \{(x,y) : x < 1\}.$$

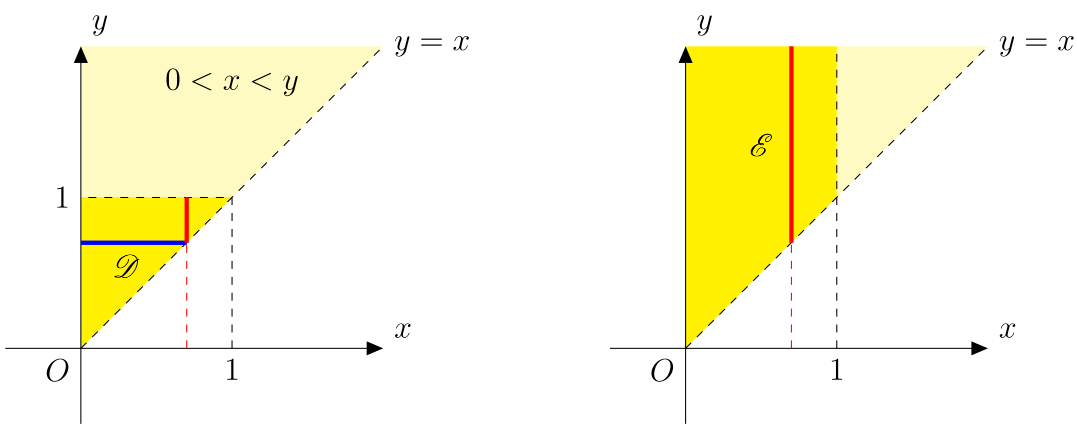

Figure 4.4: Iterated integrals

Answer Let $R = \{(x,y) : 0 < x < y\}$, $\mathscr{D} = D \cap R$, and $\mathscr{E} = E \cap R$ (see Figure 4.4). Then

$$\iint_D f(x,y)\,dxdy = \iint_{\mathscr{D}} f(x,y)\,dxdy = \iint_{\mathscr{D}} 2e^{-x-y}\,dxdy$$

and

$$\iint_E f(x,y)\,dxdy = \iint_{\mathscr{E}} f(x,y)\,dxdy = \iint_{\mathscr{E}} 2e^{-x-y}\,dxdy.$$

Here

$$
\begin{aligned}
\iint_{\mathscr{D}} 2e^{-x-y}\, dxdy &= \int_0^1 \int_0^y 2e^{-x-y}\, dxdy \\
&= \int_0^1 2e^{-y} \left[\int_0^y e^{-x}\, dx\right] dy \\
&= \int_0^1 2e^{-y}(1-e^{-y}) dy \\
&= \left[-2e^{-y} + e^{-2y}\right]_0^1 \\
&= 1 - 2e^{-1} + e^{-2}.
\end{aligned}
$$

Note that one could have used the other iterated integrals to compute $\iint_{\mathscr{D}} 2e^{-x-y}\, dxdy$, that is,

$$
\begin{aligned}
\iint_{\mathscr{D}} 2e^{-x-y}\, dxdy &= \int_0^1 \int_x^1 2e^{-x-y}\, dydx \\
&= \int_0^1 2e^{-x} \left[\int_x^1 e^{-y}\, dy\right] dx \\
&= \int_0^1 2e^{-x}(e^{-x} - e^{-1}) dx \\
&= \left[-e^{-2x} + 2e^{-1}e^{-x}\right]_0^1 \\
&= e^{-2} + 1 - 2e^{-1}.
\end{aligned}
$$

For the other integral, we have

$$
\begin{aligned}
\iint_{\mathscr{E}} 2e^{-x-y}\, dxdy &= \int_0^1 \int_x^\infty 2e^{-x-y}\, dydx \\
&= \int_0^1 2e^{-x} \left[\int_x^\infty e^{-y}\, dy\right] dx \\
&= \int_0^1 2e^{-2x} dx \\
&= 1 - e^{-2}.
\end{aligned}
$$

■

Example 4.4.9. Consider the area under the ROC curve (see Example 2.7.3) of a binary classification: Figure 4.5.

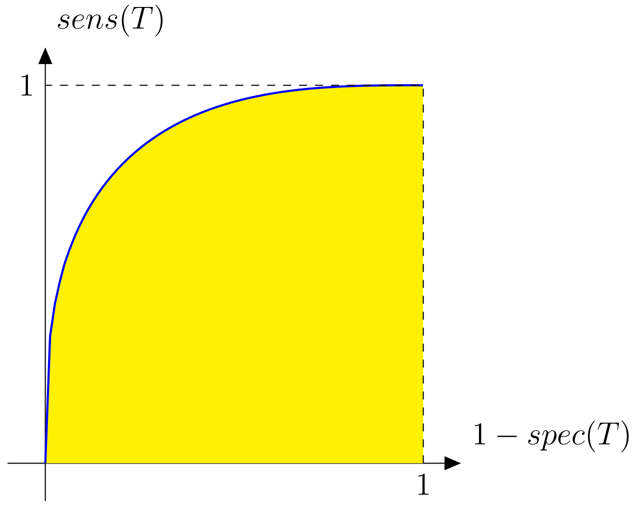

Figure 4.5: The area under the ROC curve. It can be interpreted as the probability that the underlying binary classifier will give a randomly chosen positive instance a higher score than an independently chosen negative instance.

With the same notation as in Example 2.7.3, it can be shown (area under a parametric curve) that the area under the ROC curve is given by

$$\begin{aligned}
&\int_{\infty}^{-\infty} sens(y) \frac{d}{dy} \left(1 - spec(y)\right)\, dy \\
&= \int_{-\infty}^{\infty} \left(\int_{y}^{\infty} f_+(x)\, dx \right) f_-(y)\, dy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I_{x>y}(x, y) f_+(x) f_-(y)\, dx\, dy.
\end{aligned}$$

Let X be the score of an instance that belongs to *positive* and Y be the score of an instance that is independent of the former and belongs to *negative*. Since the joint density of X and Y is $f_+(x)f_-(y)$, the probability that the score X is greater than Y is given by

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I_{x>y}(x, y) f_+(x) f_-(y)\, dx\, dy.$$

That is, the area under the ROC curve can be interpreted as the probability that the underlying classifier will give a randomly chosen positive instance a higher score than an independently chosen negative instance.

Our next result can be viewed as a higher-dimensional analogue of Theorem 2.6.5.

Theorem 4.4.10 (Change of Variables)**.** *Let $\varphi : \mathscr{D} \subseteq \mathbb{R}^2 \to D \subseteq \mathbb{R}^2$ be a differentiable one-to-one map of $\mathscr{D}$ onto D (see Figure 4.6). Then*

$$\iint_D f(x,y)\,dxdy = \iint_{\mathscr{D}} f\left(\varphi_1(u,v), \varphi_2(u,v)\right) \left|\det J_\varphi\right|\,dudv,$$

where $\det J_\varphi$ *denotes the determinant (see Definition 3.4.1) of the Jacobian matrix* $\begin{bmatrix} \frac{\partial\varphi_1}{\partial u} & \frac{\partial\varphi_1}{\partial v} \\ \frac{\partial\varphi_2}{\partial u} & \frac{\partial\varphi_2}{\partial v} \end{bmatrix}$.

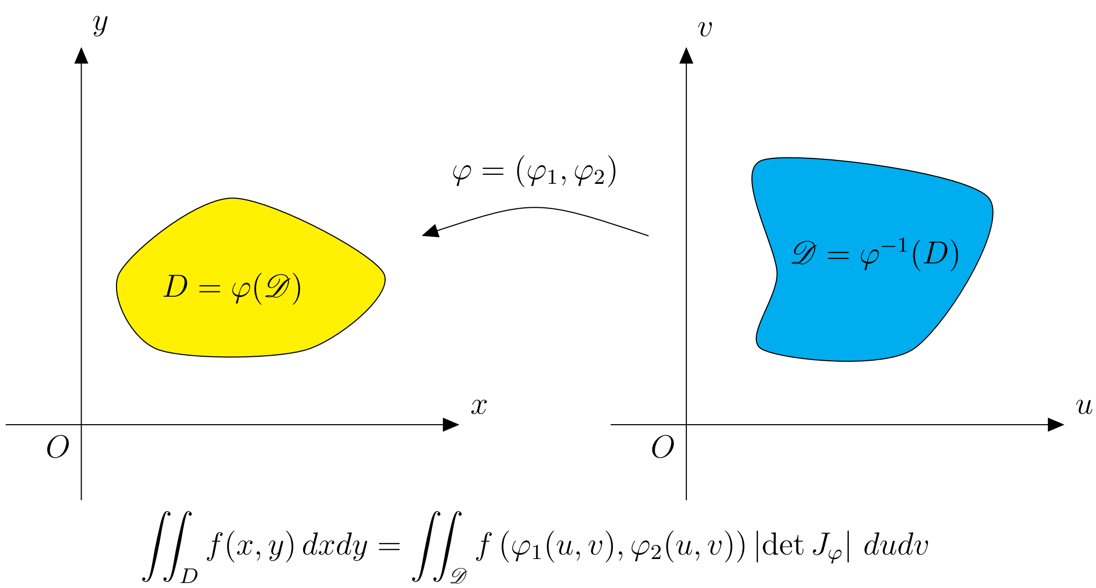

Figure 4.6: Change of variables formula in $\mathbb{R}^2$

Example 4.4.11. We compute $\iint_D (2x+4y)\,dxdy$, where D is a rectangle with vertices at $(0,1)$, $(1,2)$, $(3,0)$, and $(2,-1)$ (see Figure 4.7). It is natural to consider change of variables $u = y - x$ and $v = y + x$ so that $x = \varphi_1(u,v) = \frac{v-u}{2}$, $y = \varphi_2(u,v) = \frac{v+u}{2}$, and D is mapped onto $\mathscr{D}$ under φ^{-1}, where $\mathscr{D} = \{(u,v) : -3 \le u \le 1, 1 \le v \le 3\}$. Since $J_\varphi = \begin{bmatrix} -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$, it follows that $|\det J_\varphi| = |-\frac{1}{2}| = \frac{1}{2}$ and that

$$\begin{aligned} \iint_D (2x+4y)\,dxdy &= \iint_{\mathscr{D}} \left(2\cdot\frac{v-u}{2} + 4\cdot\frac{v+u}{2}\right)\,dudv \\ &= \iint_{\mathscr{D}} (3v+u)\,dudv \\ \text{(Example 4.4.6)} &= 40. \end{aligned}$$

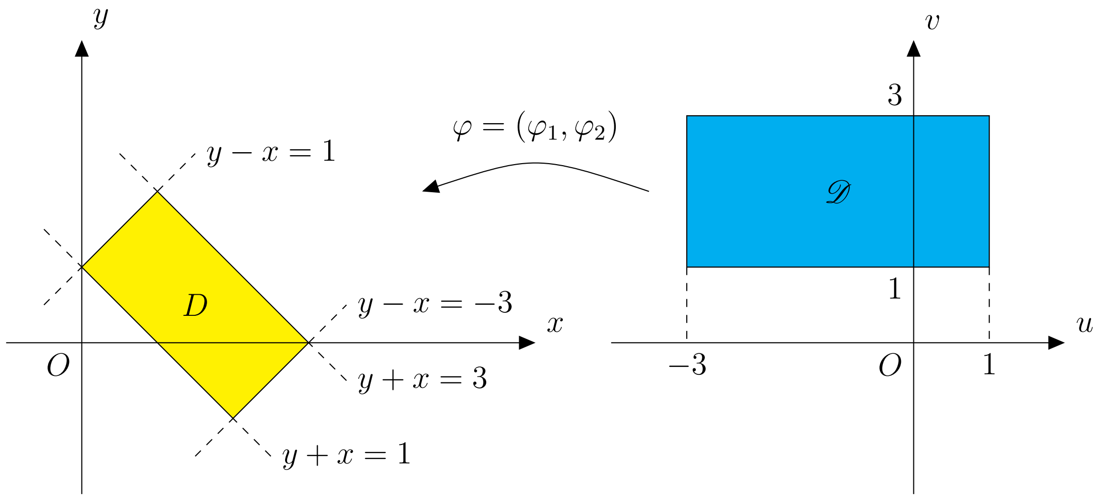

Figure 4.7: Change of variables in $\mathbb{R}^2$

Example 4.4.12. We compute $\iint_D e^{x^2+y^2}\,dxdy$, where D is the region between two arcs as described in Figure 4.8. With change of variables $x=\varphi_1(r,\theta)=r\cos\theta$ and $y=\varphi_2(r,\theta)=r\sin\theta$, D is mapped onto $\mathscr{D}$ under φ^{-1}, where $\mathscr{D}=\{(r,\theta):1\le r\le 4, \frac{\pi}{6}\le v\le\frac{\pi}{3}\}$. Since $J_\varphi = \begin{bmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{bmatrix}$, it follows that $|\det J_\varphi| = |r(\cos^2\theta+\sin^2\theta)| = r$ and that

$$\begin{aligned}\iint_D e^{x^2+y^2}\,dxdy &= \iint_{\mathscr{D}} e^{r^2} r\,drd\theta \\ &= \int_{\frac{\pi}{6}}^{\frac{\pi}{3}}\left(\int_1^4 re^{r^2}\,dr\right)d\theta \\ \text{(Integration by Substitution with } u=r^2) &= \int_{\frac{\pi}{6}}^{\frac{\pi}{3}} \frac{1}{2}\left(e^{16}-e\right)d\theta \\ &= \frac{\pi}{12}\left(e^{16}-e\right).\end{aligned}$$

Problem 4.4.13. Compute $\iint_D e^{-(x^2+y^2)}\,dxdy$, where D is the part of the unit disc in the first quadrant (see Figure 4.9).

Answer First we note that the change of variables $x=\varphi_1(r,\theta)=r\cos\theta$ and $y=\varphi_2(r,\theta)=r\sin\theta$ is *not* one-to-one from $\mathscr{D}=\{(r,\theta):0\le r\le 1, 0\le\theta\le\frac{\pi}{2}\}$ onto D, so we cannot apply the Change of Variables formula directly. However, we see that

$$\iint_D e^{-(x^2+y^2)}\,dxdy = \lim_{\epsilon\to 0}\iint_{D_\epsilon} e^{-(x^2+y^2)}\,dxdy,$$

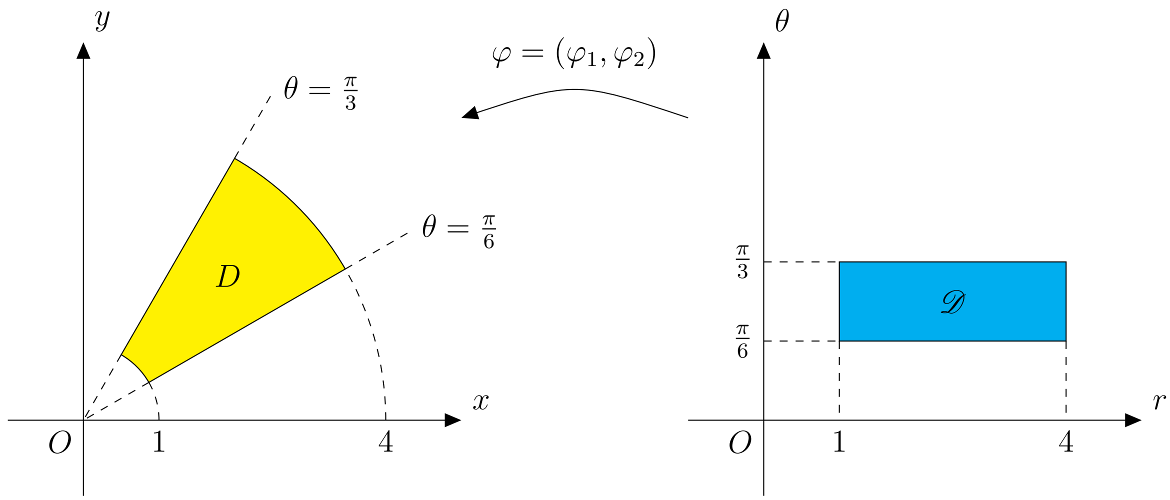

Figure 4.8: Area between two arcs

where D_ϵ is a subset of D constructed by taking a small sector around the origin from D (see Figure 4.9). Using the change of variables $x = \varphi_1(r,\theta) = r\cos\theta$ and $y = \varphi_2(r,\theta) = r\sin\theta$, we see that $\{(r,\theta) : \epsilon \le r \le 1, 0 \le \theta \le \frac{\pi}{2}\}$ is mapped onto D_ϵ, so it follows that

$$\iint_{D_\epsilon} e^{-(x^2+y^2)}\,dxdy = \int_0^{\frac{\pi}{2}} \left(\int_\epsilon^1 e^{-r^2} r dr\right) d\theta = \frac{\pi}{4}\left(e^{-\epsilon^2} - e^{-1}\right)$$

and that

$$\iint_D e^{-(x^2+y^2)}\,dxdy = \lim_{\epsilon\to 0} \frac{\pi}{4}\left(e^{-\epsilon^2} - e^{-1}\right) = \frac{\pi}{4}\left(1 - e^{-1}\right).$$

■

Example 4.4.14. For $b > 0$, let $I(b) = \displaystyle\int_0^b e^{-x^2}\,dx$. It is well known (Liouville's Theorem on Differential Algebra) that the antiderivatives of e^{-x^2} cannot be expressed in terms of elementary functions. By Tonelli's Theorem,

$$I(b)^2 = \int_0^b e^{-x^2}\,dx \int_0^b e^{-y^2}\,dy = \iint_{[0,b]\times[0,b]} e^{-(x^2+y^2)}\,dxdy.$$

Note that $R_1 \subseteq [0,b]\times[0,b] \subseteq R_2$, where R_1 and R_2 are sectors described in Figure 4.10. By Exercise 4.4.1, it follows that

$$\iint_{R_1} e^{-(x^2+y^2)}\,dxdy \le I(b)^2 \le \iint_{R_2} e^{-(x^2+y^2)}\,dxdy.$$

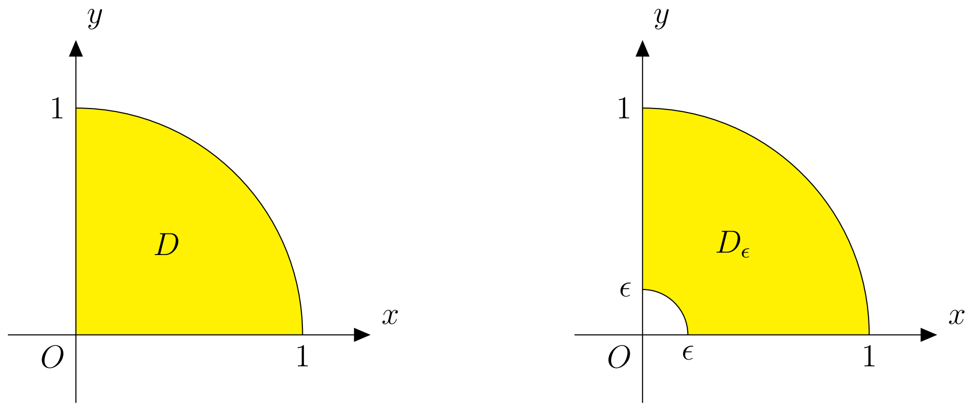

Figure 4.9: A sector in the first quadrant

By Problem 4.4.13, we get

$$\iint_{R_1} e^{-(x^2+y^2)}\,dxdy = \int_0^{\frac{\pi}{2}} \int_0^b e^{-r^2} r dr d\theta = \frac{\pi}{4}\left(1 - e^{-b^2}\right).$$

Similarly,

$$\iint_{R_2} e^{-(x^2+y^2)}\,dxdy = \int_0^{\frac{\pi}{2}} \int_0^{\sqrt{2}b} e^{-r^2} r dr d\theta = \frac{\pi}{4}\left(1 - e^{-2b^2}\right)$$

and it follows that $\lim\limits_{b\to\infty} \iint_{R_1} e^{-(x^2+y^2)}\,dxdy = \lim\limits_{b\to\infty} \iint_{R_2} e^{-(x^2+y^2)}\,dxdy = \frac{\pi}{4}$. By Squeeze Theorem, we finally get $\int_0^\infty e^{-x^2}\,dx = \lim\limits_{b\to\infty} I(b) = \frac{\sqrt{\pi}}{2}$.

Exercise 4.4.

1. (see Exercise 2.5.5) Suppose that $f(x,y) \geq g(x,y)$ for all $(x,y) \in D$. Show that $\iint_D f(x,y)\,dxdy \geq \iint_D g(x,y)\,dxdy$.

2. Let D be the region in $\mathbb{R}^2$ enclosed by two curves $y = x^2$ and $y = \sqrt{x}$. Compute $\iint_D \sqrt{y}\,dxdy$.

3. Prove that $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\,dx = 1$. In general, prove that $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx = 1$ for all μ and σ, $\sigma > 0$. This shows that the normal density is indeed a pdf.

4. Compute the double integral $\int_0^\infty \int_{-\infty}^\infty (x^2+y^2)e^{-(x^2+y^2)}\,dxdy$.

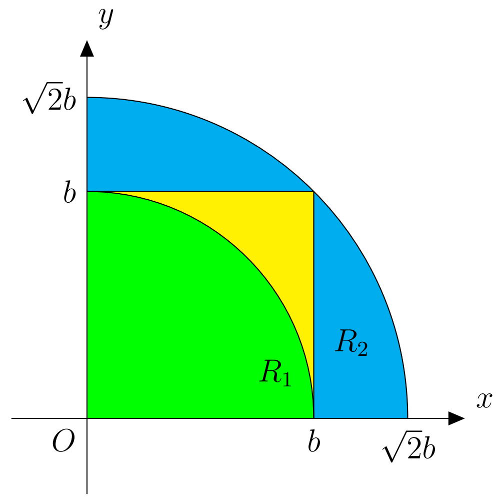

Figure 4.10: Calculation of $\displaystyle\int_0^\infty e^{-x^2}\,dx$

5. Show that $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ (see Example 2.7.8 for the definition of the gamma function). *Hint*: Recall that $\Gamma(\frac{1}{2}) = \displaystyle\int_0^\infty x^{-\frac{1}{2}} e^{-x}\,dx$. Use Integration by Substitution with $z = \sqrt{x}$, then the result follows from Example 4.4.14.

6. Compute $\displaystyle\iint_D e^{3(x+y)}\,dxdy$, where D is the parallelogram with vertices at $(1,0)$, $(0,2)$, $(1,3)$, and $(2,1)$.

Bibliography

[1] David C. Lay, *Linear Algebra and Its Applications*. Addison Wesley, 3rd Edition, 2006.

[2] Roger A. Horn and Charles R. Johnson, *Matrix Analysis*, Cambridge University Press, 1990.

Index

S_n, 17
$\mathbb{N}$, 8
$\mathbb{Q}$, 8
$\mathbb{R}$, 8
$\mathbb{Z}$, 8
n-vector, 82

angle between vectors, 93
antiderivative, 53
Associative laws, 13
augmented matrix, 83

backward substitution, 83
base of natural logarithm, 32
basic variable, 87
basis, 103
Beta Distribution, 67
beta function, 66
Binomial Distribution, 21
Binomial Theorem, 20
Bisection Method, 33
bounded from above, 34
bounded from below, 34

cardinality, 9
Cartesian product, 10
Cauchy Distribution, 68
Cauchy Mean Value Theorem, 45
Cauchy-Schwarz Inequality, 93
cdf, 29
center of a power series, 74
Chain Rule, 41
Change of Variables Formula, 153
    one-dimensional, 58
characteristic equation, 124
characteristic polynomial, 124
Chi-squared Distribution, 67
Classical Multidimensional Scaling, 131
coefficient matrix, 83
coefficients of linear combination, 100
coefficients of linear equation, 81
column space, 107
column vector, 82
combination, 17
common ratio, 72
Commutative laws, 13
complement, 12
component, 82
compound symmetry, 139
consistent, 82
Constant Multiple Rule, 40
continuous, 29, 141
convergence in distribution, 68
convergence in law, 68
convergence in probability, 68
critical point, 46, 144
Critical Point Theorem, 47
cumulative distribution function, 29
curvature, 41

De Morgan's laws, 13
definite integral, 50
derivative, 39
derivative function, 39
derivative of f at a, 37
determinant, 110
diagonalizable, 126
Diagonalization Theorem
    part 1, 126
    part 2, 126

difference of sets, 12
differentiable, 37
  twice, 41
dimension, 105
disjoint, 11
distance, 93
Distributive laws, 13
dot product, 92
double integral, 148

eigenspace, 122
eigenvalue, 121
eigenvector, 121
element, 7, 82
elementary row operations, 84
empty set, 8
entry, 82
equivalent linear systems, 81
Euclidean n-space, 10
Euler's Number, 73
Euler's number, 32
even function, 54
Exponential Distribution, 67
Extreme Value Theorem, 35

factorial, 17
familywise error rate, 48
free variable, 86
Frobenius norm, 97
Fubini's Theorem, 149
Fundamental Theorem of Calculus, 53
FWER, 48

Gamma Distribution, 66
gamma function, 65
Gauss-Jordan elimination, 85
Geometric Distribution, 72
geometric multiplicity, 122
geometric series, 72
global maximum, 47
global minimum, 47
gradient, 144

harmonic series, 72
Hessian, 144
Hilbert-Schmidt norm, 97
homogeneous, 95
Hotelling's T^2, 134
Hypergeometric Distribution, 22

improper integral, 62
  convergent, 62
  divergent, 62
inconsistent, 82
indefinite integral, 53
indicator function, 14
infimum, 34
infinite series, 71
inner product, 92
integer, 8
integrable function, 50
Integration by Parts, 59
Integration by Substitution, 58
interchange operation, 84
Intermediate Value Theorem, 33
intersection, 11
inversion, 17
Invertible Matrix Theorem, 113

L'Hopital's Theorem, 41
Lebesgue's Dominated Convergence Theorem, 143
Lebesgue's Monotone Convergence Theorem, 143
left-continuous, 30
Leibniz Integral Rule, 142
length of a vector, 93
limit at infinity, 31
limit from the left, 30
limit from the right, 30
limit of a function, 29
limit of a sequence, 69
linear combination, 100
linear equation, 81
linear system, 81

linearly dependent, 102
linearly independent, 102
Liouville's Theorem, 156
local extremum, 46
local maximum, 46
local minimum, 46
lower bound, 34
lower sum, 49

Markov matrix, 123
Mathematical Induction, 26
matrix, 82
  diagonal matrix, 91
  difference, 90
  equality, 90
  identity matrix, 91
  inverse, 98
  invertible, 98
  lower triangular matrix, 91
  orthogonal, 130
  product, 94
  square matrix, 91
  sum, 90
  symmetric, 130
  upper triangular matrix, 91
  zero matrix, 91
Mean Value Theorem, 43
mean-deviation form, 97
member, 7
memoryless property, 71
minor, 110
moment generating function, 142
multiple linear regression, 120
multivariate second derivative test, 144

natural number, 8
negative definite, 130
negative semidefinite, 130
nondecreasing, 70
nonincreasing, 70
norm, 96
Normal Distribution, 67

null space, 105
nullity, 107

odd function, 54
operator norm, 97
ordered n-tuple, 9
ordered pair, 9
ordered triple, 9
orthogonal complement, 115
orthogonal decomposition, 116
orthogonal projection, 116
orthogonal vectors, 93

partial derivative, 141
partial sum, 71
Pascal's Identity, 18
pdf, 37
permutation, 16
perpendicular vectors, 93
pivot column, 89
pivot position, 89
Poisson Distribution, 79
positive definite, 130
positive semidefinite, 130
power series, 74
Principal Component Analysis, 136
probability density function, 37
Product Rule, 16, 41
projection matrix, 117

quadratic form, 130
  differentiation, 146
Quotient Rule, 41

radius of convergence, 75
rank, 89
rational number, 8
Rayleigh quotient, 139
real number, 8
receiver operating characteristic curve, 63
reduced row echelon form, 87
reduced row echelon matrix, 87
reduced singular value decomposition, 140

replacement operation, 84
residual sum of squares, 118
right-continuous, 30
ROC, 63
Rolle's Theorem, 43
row equivalent, 87
row vector, 82

sample correlation coefficient, 94
scalar multiple of a matrix, 90
scalar product, 92
scaling operation, 84
Second Derivative Test, 47
sensitivity, 63
set, 7
signature, 110
similar, 124
simple linear regression, 118
singular value decomposition, 135
singular values, 135
solution of linear system, 81
solution set of linear system, 81
span, 100
Spanning Set Theorem, 104
Spearman correlation coefficient, 24
specificity, 63
spectral decomposition, 130
Squeeze Theorem, 32
Standard Normal Distribution, 67
subset, 8
subspace, 100
  spanned by a set of vectors, 102
  trivial subspace, 100
Sum Rule, 40
supremum, 34
symmetric group on n letters, 17

Taylor expansion with remainder, 78
Taylor polynomial, 77
Taylor Series Expansion, 78
Taylor series expansion, 78
Tonelli's Theorem, 149
total variance, 97
trace, 97
transpose, 90

union, 11
unit vector, 93
universal set, 12
upper bound, 34
upper sum, 49

Vandermonde determinant, 114
Vandermonde's Identity, 22

weights of linear combination, 100

Young's Inequality, 56

zero vector, 91