Honors Linear Algebra and Multivariable Calculus

Math 340

R. Ebrahimian
University of Maryland, College Park

October 25, 2024

Notations

- $\in$, belongs to.
- $\forall$, for all.
- $\exists$, there exists or for some.
- D_f, the domain of function f.
- $\operatorname{Im} f$ or R_f, the image of function f.
- $\mathbb{N}$, the set of nonnegative integers.
- $\mathbb{Z}^+$, the set of positive integers.
- $\mathbb{Q}$, the set of rational numbers.
- $\mathbb{R}$, the set of real numbers.
- $A \subseteq B$, set A is a subset of set B.
- $A \subsetneqq B$, set A is a proper subset of set B.
- $A \cup B$, the union of sets A and B.
- $A \cap B$, the intersection of sets A and B.
- $\bigcup\limits_{i=1}^{n} A_i$, the union of sets $A_1, A_2, \ldots, A_n$.
- $\bigcap\limits_{i=1}^{n} A_i$, the intersection of sets $A_1, A_2, \ldots, A_n$.
- $A_1 \times A_2 \times \cdots \times A_n$, the Cartesian product of sets $A_1, A_2, \ldots, A_n$.
- $\emptyset$, the empty set.
- $f^{-1}(T)$, the inverse image (or pre-image) of set T under function f.
- $f(S)$, the image of set S under function f.
- $\operatorname{span} \mathcal{S}$, the subspace spanned by set $\mathcal{S}$.
- $\dim V$, the dimension of vector space V.
- $\langle \mathbf{v}, \mathbf{w} \rangle$, the inner product of vectors $\mathbf{v}$ and $\mathbf{w}$.
- $\mathbf{v} \cdot \mathbf{w}$, the standard inner product of vectors $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$.
- $||\mathbf{v}||$, the norm of vector $\mathbf{v}$.
- $\det A$, the determinant of a square matrix A.

- $D_{\mathbf{u}}f(\mathbf{x}_0)$, the directional derivative of f at $\mathbf{x}_0$ with respect to the nonzero vector $\mathbf{u}$.
- $f_x, D_1f, \dfrac{\partial f}{\partial x}$, the partial derivative of f with respect to x.
- $\mathbf{u} \times \mathbf{v}$, the cross product of $\mathbf{u}$ and $\mathbf{v}$.
- ∇f, the gradient of a scalar function f.
- curl $\mathbf{F}$, the curl of a vector field $\mathbf{F}$.
- div $\mathbf{F}$, the divergence of a vector field $\mathbf{F}$.

Contents

These note may contain occasional typos or errors. Feel free to email me at ebrahimi@umd.edu if you notice a typo or an error.

Week 1

1.1 Sets

YouTube Video: https://youtu.be/nU8U1G6BNqk

A **set** is a well-defined collection of unordered **elements**. Each set is usually defined either by listing all of its elements or by a property as below:

$$S = \{s_1, s_2, \dots, s_n\} \qquad \text{or} \qquad S = \{s \mid s \text{ satisfies property } \mathcal{P}\}$$

Note that the order of elements in a set and repetition do not matter. So, $\{1, 1, 2\}$, $\{1, 2\}$, $\{1, 2, 1, 2, 2\}$ and $\{2, 1\}$ are all the same set.

Notation: Instead of "x is an element of the set A" or "x belongs to the set A", we write "$x \in A$".

Definition 1.1. Let A and B be two sets for which the following statement is true:

"If $x \in A$, then $x \in B$."

Then, we say A is a **subset** of B, in which case we write $A \subseteq B$.
We say a subset A of a set B is **proper** if $A \neq B$, in which case we write $A \subsetneqq B$ or $A \subsetneq B$.
The **union** of A and B, denoted by $A \cup B$, is the set consisting of all elements that are in A or B (or both).
The **intersection** of A and B, denoted by $A \cap B$, is the set consisting of all elements that are in both A and B. In other words

$$A \cup B = \{x \mid x \in A, \text{ or } x \in B\}, \qquad \text{and} \qquad A \cap B = \{x \mid x \in A, \text{ and } x \in B\}.$$

The union and intersection of n sets is defined similarly:

$$\bigcup_{i=1}^{n} A_i = \{x \mid x \in A_i, \text{ for some } i\}, \qquad \text{and} \qquad \bigcap_{i=1}^{n} A_i = \{x \mid x \in A_i, \text{ for all } i\}.$$

The union and intersection of infinitely many sets $A_1, A_2, \dots$ are defined similarly and they are denoted by $\bigcup\limits_{n=1}^{\infty} A_n$ and $\bigcap\limits_{n=1}^{\infty} A_n$.

The **empty set** or the **null set** is the set with no elements. It is denoted by $\emptyset$ or $\{\}$.

Remark. The word "or" in mathematics is not exclusive. In other words, for two statements p and q, the statement "p or q" means "p or q or both". For example, "$x \in A$ or $x \in B$" means, "x is an element of A or x is an element of B or x is an element of both A and B."

Remark. Sometimes sets are labeled by elements of another set. For example, instead of $\bigcup_{n=0}^{\infty} A_n$ we may write $\bigcup_{n\in\mathbb{N}} A_n$ and instead of $\bigcup_{n=-\infty}^{\infty} A_n$ we may write $\bigcup_{n\in\mathbb{Z}} A_n$. This is especially useful when there are too many sets to label them using only integers. For example, in the union $\bigcup_{r\in\mathbb{R}} A_r$, there is a set A_r corresponding to every real number r.

Definition 1.2. We say two sets A and B are **equal** if and only if $A \subseteq B$ and $B \subseteq A$, in which case we write $A = B$.

Example 1.1. Prove that for every three sets A, B, and C we have $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$.

Definition 1.3. An **ordered pair** (a, b) is two objects a and b with a specified order. Two ordered pairs (a, b) and (c, d) are the same if and only if $a = c$ and $b = d$. An n**-tuple** $(a_1, a_2, \ldots, a_n)$ is n objects $a_1, a_2, \ldots, a_n$ with a specified order. Two n-tuples $(a_1, a_2, \ldots, a_n)$ and $(b_1, b_2, \ldots, b_n)$ are equal if and only if $a_i = b_i$ for $i = 1, \ldots, n$.

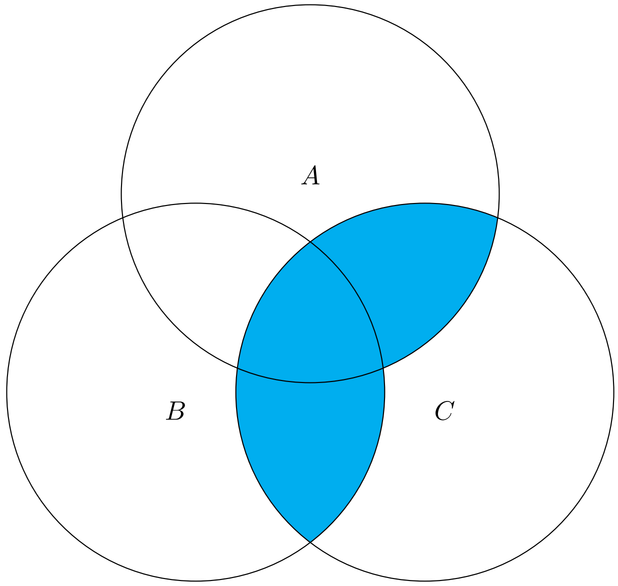
Definition 1.4. The **Cartesian product** of n sets $A_1, A_2, \ldots, A_n$, denoted by $A_1 \times A_2 \times \cdots \times A_n$, is the set of all n-tuples $(a_1, a_2, \ldots, a_n)$ for which $a_i \in A_i$ for all i. The Cartesian product of n copies of a set A is denoted by A^n.

Example 1.2. Every point on the plane can be represented by an element of the set $\mathbb{R}^2 = \{(x, y) \mid x, y \in \mathbb{R}\}$. Every point on the n-dimensional space can be represented by an element of the set

$$\mathbb{R}^n = \{(x_1, \ldots, x_n) \mid x_1, \ldots, x_n \in \mathbb{R}\}.$$

Definition 1.5. We say two sets A and B are **disjoint** if $A \cap B = \emptyset$. We say sets $A_1, A_2, \ldots, A_n$ are **disjoint** if $\bigcap_{i=1}^{n} A_i = \emptyset$. We say $A_1, A_2, \ldots, A_n$ are **pairwise disjoint** if for every $i \neq j$, A_i and A_j are disjoint. Similarly, these notions are defined for an infinite collection of sets A_i with $i \in I$.

To understand sets we often picture them as ovals or circles. The following shows the **Venn diagram** of $(A \cup B) \cap C$.



Definition 1.6. For two sets A, B, the **difference** $A - B$ consists of all elements of A that are not in B.

$$A - B = \{x \in A \mid x \notin B\}.$$

When dealing with sets, we often assume all of our sets are subsets of a given larger set U. This set is called a **universal set**. For example in number theory the universal set is usually $\mathbb{Z}$, the set of integers. In calculus we deal with real numbers and thus our universal set is typically $\mathbb{R}$.

Assume A is a subset of the universal set U. The **complement** of A in U is the set consisting of all elements of U that are not in A. The complement of A is denoted by A^c.

Theorem 1.1 (De Morgan's Laws)**.** *Given n subsets $A_1, \ldots, A_n$ of a universal set U we have:*

(a) $\left(\bigcap_{j=1}^{n} A_j\right)^c = \bigcup_{j=1}^{n} A_j^c.$

(b) $\left(\bigcup_{j=1}^{n} A_j\right)^c = \bigcap_{j=1}^{n} A_j^c.$

Remark. Similar to above, given a nonempty set of indices I and a collection of subsets A_i of a universal set U, for every $i \in I$, we have:

(a) $\left(\bigcap_{i\in I} A_i\right)^c = \bigcup_{i\in I} A_i^c.$

(b) $\left(\bigcup_{i\in I} A_i\right)^c = \bigcap_{i\in I} A_i^c.$

1.2 Functions

YouTube Video: https://youtu.be/lwQG_dMXQOA

Definition 1.7. Given two nonempty sets A and B, a **function** or a **mapping** $f : A \to B$ is a rule that assigns to every element $a \in A$ an element $f(a) \in B$. The set A is called the **domain** of f and is denoted by D_f. The set B is called the **co-domain** of f. The **range** or **image** of f, denoted by R_f or Im f, is the set Im $f = \{f(a) \mid a \in A\}$.

Two functions f and g are called **equal** if they have the same domain, the same co-domain, and $f(x) = g(x)$ for all x in their common domain.

f is called **surjective** or **onto** if for every $b \in B$ there is $a \in A$ for which $f(a) = b$.

f is called **injective** or **one-to-one** if whenever $f(a_1) = f(a_2)$ we also have $a_1 = a_2$.

f is called **bijective** if it is injective and surjective.

The **composition** $f \circ g$ of two functions f, g with $R_g \subseteq D_f$, is a function from D_g to the co-domain of f given by $f \circ g(x) = f(g(x))$, for all $x \in D_g$.

The function $\mathrm{id}_A : A \to A$ defined by $\mathrm{id}_A(a) = a$, for all $a \in A$ is called the **identity** function of A.

A function $f : A \to B$ is called **invertible** if and only if there is a function $g : B \to A$ for which $f \circ g = \mathrm{id}_B$ and $g \circ f = \mathrm{id}_A$. The function g is called the **inverse** of f and is denoted by f^{-1}.

Example 1.3 (Projection)**.** The function $\pi_1 : A \times B \to A$ defined by $\pi_1(a, b) = a$ is called the projection onto the first component. Similarly, the function $\pi_i : A_1 \times \cdots \times A_n \to A_i$ defined by $\pi_i(a_1, \ldots, a_n) = a_i$ is called the **projection** onto the i-th component.

Definition 1.8. Given a function $f : A \to B$, and a subset S of A, the **image** of S under f is the set $f(S) = \{f(s) \mid s \in S\}$. If T is a subset of B, then the **pre-image** or **inverse image** of T under f is the set $f^{-1}(T) = \{a \in A \mid f(a) \in T\}$.

Note that the pre-image and image are both sets.

Example 1.4. Let $f : A \times B \to B$ be the projection onto the second component. For every $b \in B$ find the pre-image of $\{b\}$ under f.

Example 1.5. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a function defined by $f(x, y) = 2x + 3y$. For every real number b, evaluate and describe $f^{-1}(\{b\})$. How do these pre-images change as b changes?

Theorem 1.2 (Properties of Pre-image)**.** *Suppose $f : A \to B$ is a function, $S \subseteq A$, $T \subseteq B$, and $T_i \subseteq B$ for every $i \in I$, where I is a nonempty set of indices. Then*

(a) $S \subseteq f^{-1}(f(S))$, *and* $f(f^{-1}(T)) \subseteq T$.

(b) $f^{-1}(\bigcup_{i \in I} T_i) = \bigcup_{i \in I} f^{-1}(T_i)$.

(c) $f^{-1}(\bigcap_{i \in I} T_i) = \bigcap_{i \in I} f^{-1}(T_i)$.

Proof. (a) Suppose $s \in S$. By definition of image, $f(s) \in f(S)$. Therefore, by definition of pre-image $s \in f^{-1}(f(S))$. This completes the proof of $S \subseteq f^{-1}(f(S))$.

Suppose $x \in f(f^{-1}(T_i))$. By definition $x = f(y)$, for some $y \in f^{-1}(T_i)$. Therefore, $f(y) \in T_i$, which means $x \in T_i$. This means $f(f^{-1}(T_i)) \subseteq T_i$.

Parts (b) and (c) are left as exercises. $\square$

1.3 Proofs

YouTube Video: https://youtu.be/sW823lmew64

In writing proofs you should note the following:

- You cannot prove a *universal statement* (statements involving *for every* or *for all*) by examples. For example if you are asked to prove "The sum of every two odd integers is even." your proof may not be "3 is odd, 5 is odd, 3+5=8 is even. Therefore, the sum of every two odd integers is even."

  On the other hand, for *existential statements* (when a statement is asking you to show something exists), giving an example and showing that the example satisfies all the required conditions is enough.

- Do not use the same variable for two different things.

- You may not assume anything but what is given in the assumptions.

- All steps must be justified and the justifications must all be clearly stated.

- You may only use known facts. These are typically things that have been previously proved as theorems or are facts stated in definitions.

- To prove a statement of the form "p if and only if q" we will need to prove both "If p, then q" and "If q, then p".

To prove a *statement* (usually of the form "If p then q"), there are three main methods of proof. We will look at each one via examples.

1.3.1 Direct Proof

In this method we start from the assumption and by taking logical steps we end up with the conclusion.

Example 1.6. Prove that the function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^3 + 2x$ is one-to-one.

Solution. By definition of one-to-one, we need to prove if $f(x) = f(y)$, then $x = y$.

Suppose $f(x) = f(y)$. Then $x^3 + 2x = y^3 + 2y$. Therefore, $x^3 - y^3 + 2(x - y) = 0$, which implies $(x - y)(x^2 + xy + y^2 + 2) = 0$. This means either $x = y$ or $x^2 + xy + y^2 + 2 = 0$. If the second equality holds, by the quadratic formula we obtain $x = \dfrac{-y \pm \sqrt{y^2 - 4(y^2 + 2)}}{2}$. The discriminant is $-3y^2 - 8$ which is negative. Therefore, this equality is impossible, and hence $x = y$. This means f is one-to-one. $\square$

1.3.2 Proof by Contradiction

In this method, we assume the conclusion is false while the assumption is true. After taking logical steps we obtain a contradiction. A contradiction is a statement that is false: Either it violates a fact in math such as a theorem or a definition or it violates the assumptions. When using proof by contradiction make sure to clearly specify you are using this method. You could say " On the contrary assume..." or "By way of contradiction assume..." or simply state "We will use proof by contradiction."

Example 1.7. Prove that there are infinitely many primes.

Solution. On the contrary assume there are only a finite number of primes, and let $p_1, p_2, \ldots, p_n$ be the list of all primes. Since the integer $d = p_1 \cdots p_n + 1$ is more than one, d has a prime factor. Since $p_1, p_2, \ldots, p_n$ is the list of all primes, one of the p_i's must divide d. On the other hand p_i divides $p_1 p_2 \cdots p_n$. Therefore, p_i must divide $d - p_1 p_2 \cdots p_n = 1$. This is a contradiction. Therefore, the initial assumption must be false, and thus there must exist infinitely many primes. $\square$

1.3.3 Proof by Induction

To prove a statement $P(n)$ (i.e. a statement that depends on a positive integer n) we will:

- Prove $P(1)$ (basis step); and
- Assume $P(n)$ holds for some $n \geq 1$, and then prove $P(n+1)$ (inductive step).

If you need to use $P(n-1)$ in your proof of $P(n+1)$, then the basis step must involve two consecutive integers, e.g. $P(1)$ and $P(2)$.

Often times we use what is called **strong induction** which involves assuming $P(1), \ldots, P(n)$ and then proving $P(n+1)$ in addition to proving the basis step.

When employing the method of mathematical induction keep in mind to always start your proof by "We will prove *the statement* by induction on *the variable*". Replace "the statement" and "the variable" accordingly. Also, clearly separate the basis step and the inductive step.

Example 1.8. Prove that the sum of the first n positive odd integers is n^2.

1.4 $\mathbb{R}^n$ as a Vector Space

YouTube Video: https://youtu.be/Bwpk4fPJmoU

As we saw earlier, elements of $\mathbb{R}^n$ are n-tuples of the form $(x_1, x_2, \ldots, x_n)$, where x_j's are real numbers. Each one of these elements is called a **vector** and these vectors can be added componentwise as follows:

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

Each vector can also be scaled by any real number c (also called a **scalar**) as follows:

$$c(x_1, x_2, \ldots, x_n) = (cx_1, cx_2, \ldots, cx_n)$$

This vector addition and scalar multiplication satisfy the following properties.

(I) (Closure) For every two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, and every scalar $c \in \mathbb{R}$, both $\mathbf{x} + \mathbf{y}$ and $c\mathbf{x}$ are in $\mathbb{R}^n$.

(II) (Associativity) For every $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n$, and every $a, b \in \mathbb{R}$, we have $\mathbf{x} + (\mathbf{y} + \mathbf{z}) = (\mathbf{x} + \mathbf{y}) + \mathbf{z}$, and $a(b\mathbf{x}) = (ab)\mathbf{x}$.

(III) (Commutativity) For every $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, we have $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$.

(IV) (Additive Identity) For every $\mathbf{x} \in \mathbb{R}^n$ we have $\mathbf{x} + (0, 0, \ldots, 0) = \mathbf{x}$. (The vector $(0, 0, \ldots, 0)$ is called the **zero vector** and is denoted by $\mathbf{0}$.)

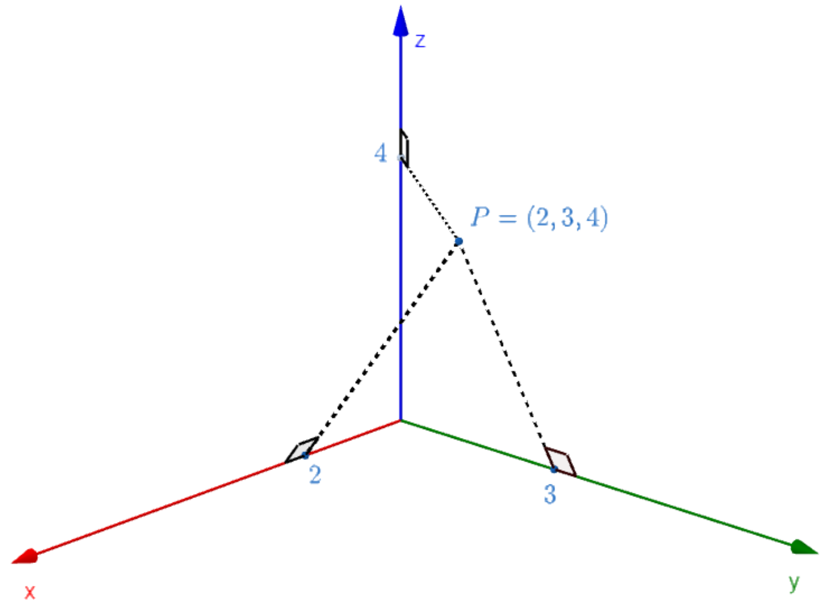
(V) (Additive Inverse) For every $\mathbf{x} \in \mathbb{R}^n$, there is a vector $\mathbf{y} \in \mathbb{R}^n$ for which $\mathbf{x} + \mathbf{y} = \mathbf{0}$. (This vector $\mathbf{y}$ is called the **additive inverse** of $\mathbf{x}$ and is denoted by $-\mathbf{x}$. It is given by $-(x_1, x_2, \ldots, x_n) = (-x_1, -x_2, \ldots, -x_n)$.)

(VI) (Distributivity) For every $a, b \in \mathbb{R}$ and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, we have $(a+b)\mathbf{x} = a\mathbf{x} + b\mathbf{x}$, and $a(\mathbf{x}+\mathbf{y}) = a\mathbf{x} + a\mathbf{y}$.

(VII) (Multiplicative Identity) For every $\mathbf{x} \in \mathbb{R}^n$ we have $1\mathbf{x} = \mathbf{x}$.

The seven properties I-VII listed above are called **vector space** properties of $\mathbb{R}^n$. This is often phrased as "$\mathbb{R}^n$ is a vector space." Note that sometimes we refer to elements of $\mathbb{R}^n$ as **points**. This is only for conceptualizing these objects. The math does not change. When elements of $\mathbb{R}^n$ are seen as points, the zero vector is referred to as the **origin**.

Geometrically, elements of $\mathbb{R}^2$ can be represented by points on a plane. Elements of $\mathbb{R}^3$ can be represented by points in a 3D space. To do that, we need three axes, x-, y-, and z-**axes**. These three axes must satisfy the right-hand rule. The coordinates of each point can be found by dropping perpendiculars to the axes.



The set of all points with positive coordinates, is called the **first octant**.

There are three planes each containing two of the x-, y-, and z- axes. Each of the three xy-, xz- and yz-planes is called a **coordinate plane**.

Theorem 1.3. *The distance between two points* (x_1, y_1, z_1) *and* (x_2, y_2, z_2) *in* $\mathbb{R}^3$ *is given by*

$$\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}.$$

1.5 Warm-ups

Example 1.9. How many elements does the set $\{2, 1, 3, 2\}$ have? How about the set $\{3, 2, 1\}$? How are these two sets related?

Solution. Since repetition and order does not matter in a set, these two sets are the same sets:

$$\{2, 1, 3, 2\} = \{3, 2, 1\}.$$

So, these sets both have three elements. □

Example 1.10. Let E be the set of all even integers and O be the set of all odd integers. Describe $E \cup O$ and $E \cap O$.

Solution. $E \cup O$ is the set of all integers that are odd or even. Since every integer is either odd or even, $E \cup O = \mathbb{Z}$.

By definition of intersection, $E \cap O$ is the set of all integers that are both even and odd. Since no integer is both even and odd, $E \cap O = \emptyset$. $\square$

Example 1.11. Consider the function $f : \{1, 2, 3\} \to \{1, 2, 3, 4\}$ defined by $f(1) = 2$, $f(2) = 2$, and $f(3) = 4$. Find the domain of f, the co-domain of f, the image of f, $f(\{1, 2\})$, and $f^{-1}(\{2, 3\})$.

Solution. The domain of f is $\{1, 2, 3\}$. The co-domain of f is $\{1, 2, 3, 4\}$. The image of f is $\{f(1), f(2), f(3)\} = \{2, 4\}$.

$$f(\{1, 2\}) = \{f(1), f(2)\} = \{2, 2\} = \{2\},$$

and

$$f^{-1}(\{2, 3\}) = \{x \in \{1, 2, 3\} \mid f(x) \in \{2, 3\}\}.$$

Thus, $f^{-1}(\{2, 3\}) = \{1, 2\}$. $\square$

1.6 More Examples

Example 1.12. Given sets $A = \{1, 2\}, B = \{0, 1, -1\}$, write each of the following sets by listing all of its elements:

(a) $A \cup B$

(b) $A \cap B$

(c) $A \times B$

Solution. (a) $A \cup B$ consists of all elements that are in A or B. Thus, $A \cup B = \{1, 2, 0, -1\}$.

(b) $A \cap B$ consists of all elements that are in both A and B. Thus, $A \cap B = \{1\}$.

(c) $A \times B$ consists of all elements of the form (a, b), where $a \in A$ and $b \in B$. Thus,

$$A \times B = \{(1, 0), (1, 1), (1, -1), (2, 0), (2, 1), (2, -1)\}$$

$\square$

Example 1.13. Prove that for all sets $A, B_1, B_2, \dots$, we have $A \bigcap (\bigcup\limits_{n=1}^{\infty} B_n) = \bigcup\limits_{n=1}^{\infty} (A \cap B_n)$.

Solution. Suppose $x \in A \bigcap (\bigcup\limits_{n=1}^{\infty} B_n)$. By definition of intersection, $x \in A$ and $x \in \bigcup\limits_{n=1}^{\infty} B_n$. By definition of union, $x \in B_n$ for some n. This means $x \in A \cap B_n$ and thus $x \in \bigcup\limits_{n=1}^{\infty} (A \cap B_n)$. Therefore,

$$A \bigcap (\bigcup_{n=1}^{\infty} B_n) \subseteq \bigcup_{n=1}^{\infty} (A \cap B_n). \qquad (*)$$

Suppose $x \in \bigcup_{n=1}^{\infty}(A \cap B_n)$. By definition of union, $x \in A \cap B_n$ for some n. Thus, by definition of intersection, $x \in A$ and $x \in B_n$. Therefore, $x \in A$ and $x \in \bigcup_{n=1}^{\infty} B_n$. This implies $x \in A \bigcap (\bigcup_{n=1}^{\infty} B_n)$. Therefore,

$$\bigcup_{n=1}^{\infty}(A \cap B_n) \subseteq A \bigcap (\bigcup_{n=1}^{\infty} B_n). \qquad (**)$$

Combining $(*)$ and $(**)$ we obtain the result. $\square$

Example 1.14. Describe each set as a subset of $\mathbb{R}^2$.

(a) $[0,1] \times \{1\}$.

(b) $[1,2] \times [0,1]$.

Solution. (a) This is the set of all (x, y), where $x \in [0,1]$ and $y = 1$. This is a horizontal segment connecting $(0,1)$ and $(1,1)$.

(b) This set consists of all points (x, y) for which $x \in [1,2]$ and $y \in [0,1]$. This is a filled square with vertices $(1,0), (2,0), (1,1)$, and $(2,1)$. $\square$

Example 1.15. Let C be the unit circle $x^2 + y^2 = 1$ in the xy-plane. Geometrically describe the set $C \times \mathbb{R}$.

Solution. $C \times \mathbb{R}$ is the set of all (x, y, z) for which $x^2 + y^2 = 1$. This means $C \times \mathbb{R}$ is the union of the translation of the unit circle C in the direction of the z-axis. This is a right circular cylinder. $\square$

Example 1.16. Suppose X and Y are finite nonempty sets of sizes m and n respectively. Let Y^X be the set of all function $f : X \to Y$. What is the size of Y^X? (This should tell you why we use the notation "Y^X".)

Solution. Let $f : X \to Y$ be a function. For each $x \in X$, $f(x)$ could be any element of Y. Thus, there are n possible values for $f(x)$. Since this is true for each element of X, there are n^m functions $f : X \to Y$. $\square$

Example 1.17. Define the Fibonacci sequence F_n by $F_0 = 0, F_1 = 1$, and $F_{n+2} = F_{n+1} + F_n$ for all $n \geq 0$. Prove that $F_n < 2^n$ for all $n \geq 0$.

Sketch. The fact that each term of the sequence depends on the previous terms reminds us of the method of Mathematical Induction. So, we will employ this method. However since each term depends on the previous two terms, we will have to start with proving the given statement for two values of n.

Solution. We will prove $F_n < 2^n$ by induction on n.

Basis step: $F_0 = 0 < 2^0 = 1$, and $F_1 = 1 < 2^1$.

Inductive step: Suppose for some $n \geq 1$, $F_k < 2^k$ for $k = 0, \ldots, n$. By assumption $F_{n+1} = F_n + F_{n-1} < 2^n + 2^{n-1} = 2^{n-1}(2+1) < 2^{n+1}$, as desired. This completes the solution. $\square$

Example 1.18. Prove that if a real number x satisfies $|x| + x > 0$, then x is positive.

Solution. On the contrary assume x is not positive. Therefore, we have two cases:

Case I: $x = 0$. This means $|x| + x = 0$, which is a contradiction.

Case II: $x < 0$. This implies $|x| = -x$ and thus, $|x| + x = 0$, which is a contradiction.
Therefore x must be positive. $\square$

Definition 1.9. Suppose D is a subset of $\mathbb{R}$. A function $f : D \to \mathbb{R}$ is said to be **periodic** if there is a positive real number T for which:

- For every $x \in \mathbb{R}$, the number x is in D if and only if $x + T \in D$, and
- $f(x + T) = f(x)$ for every $x \in D$.

Example 1.19. Prove $\sin x$ and $\sin(\pi x)$ are both periodic, but their sum is not.

Solution. First note that the domain of both functions is $\mathbb{R}$. Thus, $x \in \mathbb{R}$ iff $x + T \in \mathbb{R}$ is valid for every $x, T \in \mathbb{R}$.

By properties of $\sin x$ we know $\sin(x + 2\pi) = \sin x$ and $\sin(\pi(x + 2)) = \sin(\pi x)$. Thus, both functions $\sin x$ and $\sin(\pi x)$ are periodic.

To prove $\sin x + \sin(\pi x)$ is not periodic we will use proof by contradiction. Assume on the contrary there is a positive real number T for which for every $x \in \mathbb{R}$ we have

$$\sin(x + T) + \sin(\pi(x + T)) = \sin x + \sin(\pi x) \qquad (*)$$

Differentiating $(*)$ with respect to x twice, we obtain

$$-\sin(x + T) - \pi^2 \sin(\pi(x + T)) = -\sin x - \pi^2 \sin(\pi x) \qquad (**)$$

Adding $(*)$ and $(**)$ and dividing both sides by $1 - \pi^2$ we conclude $\sin(\pi(x+T)) = \sin(\pi x)$ for every $x \in \mathbb{R}$. Combining this with $(*)$ we conclude that $\sin(x + T) = \sin x$ for all $x \in \mathbb{R}$. This implies $T = 2k\pi$ for some positive integer k. Substituting x by x/π in $\sin(\pi(x + T)) = \sin(\pi x)$ we conclude $\sin(x + \pi T) = \sin(x)$. Therefore, $\pi T = 2n\pi$ for some positive integer n. Therefore, $\pi(2k\pi) = 2n\pi$. Hence $\pi = n/k$ is rational, which is a contradiction. $\square$

Example 1.20. Prove that for every positive integer n, there is a polynomial $p_n(x)$ for which the n-th derivative of e^{x^2} at x is equal to $p_n(x)e^{x^2}$.

Solution. We will prove this by induction on n.
Basis step. By the Chain Rule, the first derivative of e^{x^2} is $2xe^{x^2}$. The fact that $p_1(x) = 2x$ is a polynomial proves the claim for $n = 1$.

Inductive step. Suppose the n-th derivative of e^{x^2} is $p_n(x)e^{x^2}$ for a polynomial p_n. Differentiating this using the Product Rule and the Chain Rule we conclude that the $(n+1)$-th derivative of e^{x^2} is equal to $p_n'(x)e^{x^2} + p_n(x)2xe^{x^2} = (p_n'(x) + 2xp_n(x))e^{x^2}$. Since the derivative of a polynomial is a polynomial and the product and sum of polynomials are polynomials, $p_n'(x) + 2xp_n(x)$ is a polynomial. So, setting $p_{n+1}(x) = p_n'(x) + 2xp_n(x)$ we conclude the $(n+1)$-th derivative of e^{x^2} is equal to $p_{n+1}(x)e^{x^2}$ for some polynomial p_{n+1}, as desired. $\square$

Example 1.21. Let $f : A \to B$ be a function, $S \subseteq A$, and $T \subseteq B$. Prove that:

(a) If f is one-to-one, then $S = f^{-1}(f(S))$.

(b) If f is onto, then $T = f(f^{-1}(T))$.

Solution. (a) By Theorem 1.2, $S \subseteq f^{-1}(f(S))$. It is enough to show $f^{-1}(f(S)) \subseteq S$. Suppose $x \in f^{-1}(f(S))$. By definition of pre-image, $f(x) \in f(S)$. By definition of $f(S)$ we conclude $f(x) = f(s)$ for some $s \in S$. Since f is one-to-one, $x = s$ and thus $x \in S$. This shows $f^{-1}(f(S)) \subseteq S$, as desired.

(b) By Theorem 1.2, $f(f^{-1}(T)) \subseteq T$. Thus, it is enough to prove $T \subseteq f(f^{-1}(T))$. Let $x \in T$. Since f is onto, there is $a \in A$ such that $f(a) = x$. Thus, by definition of pre-image $a \in f^{-1}(T)$. Therefore, by definition of image $f(a) \in f(f^{-1}(T))$. Since $f(a) = x$, we obtain $x \in f(f^{-1}(T))$. Therefore, $T \subseteq f(f^{-1}(T))$, as desired. $\square$

Example 1.22. Let $f : \mathbb{R} \to \mathbb{R}$ be defined as $f(x) = x^2$. Find each of the following:

(a) $f([0, 1))$.

(b) $f^{-1}([-1, 0))$.

(c) $f^{-1}((0, 2))$.

Solution. (a) Note that if $x \in [0, 1)$, then $0 \leq x^2 < 1$ and thus $f([0, 1)) \subseteq [0, 1)$. Furthermore, if $y \in [0, 1)$, then $\sqrt{y} \in [0, 1)$ and $f(\sqrt{y}) = y$. Therefore, $[0, 1) \subseteq f([0, 1))$. This shows $f([0, 1)) = [0, 1)$.

(b) By definition of pre-image, $x \in f^{-1}([-1, 0))$ if and only if $f(x) \in [-1, 0)$ if and only if $-1 \leq x^2 < 0$, which is impossible. Therefore, $f^{-1}([-1, 0)) = \emptyset$.

(c) By definition of pre-image, $x \in f^{-1}(0,2)$ if and only if $f(x) \in (0,2)$, i.e. $0 < x^2 < 2$. This holds if and only if $0 < x < \sqrt{2}$ or $-\sqrt{2} < x < 0$. Therefore, $f^{-1}((0,2)) = (0,\sqrt{2}) \cup (-\sqrt{2},0)$. □

Example 1.23. Let $f : A \to B$ be a function. Find each of the following:

(a) $f(\emptyset)$.

(b) $f^{-1}(\emptyset)$.

(c) $f^{-1}(B)$.

Solution. (a) $f(\emptyset)$ consists of all elements of the form $f(x)$, where $x \in \emptyset$, but since $\emptyset$ contains no elements, $f(\emptyset) = \emptyset$.

(b) $f^{-1}(\emptyset)$ consists of all elements $a \in A$ for which $f(a) \in \emptyset$. Since $\emptyset$ contains no elements $f^{-1}(\emptyset) = \emptyset$.

(c) $f^{-1}(B)$ consists of all elements $a \in A$ for which $f(a) \in B$, but since B is the co-domain, $f(a)$ is always in B, and thus $f^{-1}(B) = A$. □

Example 1.24. Let $f : A \to B$ be a function, and S_i with $i \in I$ be a collection of subsets of A. Prove that

(a) $f\left(\bigcup\limits_{i\in I} S_i\right) = \bigcup\limits_{i\in I} f(S_i)$.

(b) $f\left(\bigcap\limits_{i\in I} S_i\right) \subseteq \bigcap\limits_{i\in I} f(S_i)$. By an example show that the equality does not always hold.

Solution. (a) Let $x \in f\left(\bigcup\limits_{i\in I} S_i\right)$. By definition of image, $x = f(s)$ for some $s \in \bigcup\limits_{i\in I} S_i$. By definition of union, $s \in S_j$ for some $j \in I$. Therefore, $x = f(s) \in f(S_j)$, which implies $x \in \bigcup\limits_{i\in I} f(S_i)$, by definition of union. The other inclusion is similar and is left as an exercise.

(b) Let $x \in f\left(\bigcap\limits_{i\in I} S_i\right)$. By definition of image, $x = f(s)$ for some $s \in \bigcap\limits_{i=1}^{n} S_i$. By definition of intersection, $s \in S_i$ for all $i \in I$, and thus $x = f(s) \in f(S_i)$ for all $i \in I$, by definition of image. Therefore, $x \in \bigcap\limits_{i\in I} f(S_i)$, by definition of intersection. This completes the proof.

Consider $f : \{1,2\} \to \{1\}$ given by $f(1) = f(2) = 1$. Let $S_1 = \{1\}$ and $S_2 = \{2\}$. Then, $S_1 \cap S_2 = \emptyset$ and thus $f(S_1 \cap S_2) = \emptyset$. On the other hand $f(S_1) = f(S_2) = \{1\}$ and thus $f(S_1) \cap f(S_2) \neq \emptyset$. □

Example 1.25. Determine if each function below is one-to-one, onto, both or neither.

(a) $f : \mathbb{R}^2 \to \mathbb{R}^2$ given by $f(x,y) = (x+2y, x-y)$.

(b) $f : \mathbb{Z}^2 \to \mathbb{Q}^+$ given by $f(m,n) = 2^m \cdot 3^n$.

Solution. (a) Suppose $f(x, y) = f(a, b)$. This implies

$$\begin{cases} x + 2y = a + 2b \\ x - y = a - b \end{cases}$$

Subtracting the two equations above we obtain $3y = 3b$ and thus $y = b$. Substituting into the first equation we obtain $x = a$. Therefore, f is one-to-one.

Given $(a, b) \in \mathbb{R}^2$, we will need to determine if there is $(x, y) \in \mathbb{R}^2$ for which $f(x, y) = (a, b)$. Solving the system

$$\begin{cases} x + 2y = a \\ x - y = b \end{cases}$$

we obtain $x = (a + 2b)/3$ and $y = (a - b)/3$. Therefore, this function is also onto.

(b) Suppose $f(m, n) = f(r, s)$ for some integers m, n, r, s. Therefore, $2^m \cdot 3^n = 2^r \cdot 3^s$. Without loss of generality assume $m \geq r$. We see that $2^{m-r} = 3^{s-n}$. If the exponent $m - r$ is positive, then the left side is even, while the right side is not. This contradiction shows $m = r$ and thus $n = s$. Therefore, f is one-to-one.

This function is not onto. For example $f(m, n) = 5$ has no solutions, because $2^m \cdot 3^n = 5$ is impossible by the uniqueness of prime factorization. $\square$

Example 1.26. For a function $f : A \to B$ prove that the equality $f(S_1 \cap S_2) = f(S_1) \cap f(S_2)$ holds for all subsets S_1, S_2 of A if and only if f is one-to-one.

Solution. First, note that by Example 1.24, we know

$$f(S_1 \cap S_2) \subseteq f(S_1) \cap f(S_2).$$

($\Leftarrow$) Suppose f is one-to-one. Let $x \in f(S_1) \cap f(S_2)$. By definition, $x = f(s_1) = f(s_2)$ for some $s_1 \in S_1$, and some $s_2 \in S_2$. Since f is one-to-one, we have $s_1 = s_2$. Therefore, $s_1 \in S_1 \cap S_2$. This means $x \in f(S_1 \cap S_2)$. This shows $f(S_1) \cap f(S_2) = f(S_1 \cap S_2)$.

($\Rightarrow$) Assume $f(S_1 \cap S_2) = f(S_1) \cap f(S_2)$ for every two subsets S_1, S_2 of A. Suppose $f(a) = f(b)$, and let $S_1 = \{a\}, S_2 = \{b\}$. We know $f(S_1) = \{f(a)\}$ and $f(S_2) = \{f(b)\} = \{f(a)\}$. Therefore, $f(S_1) \cap f(S_2) = \{f(a)\}$. If $a \neq b$, then $S_1 \cap S_2 = \emptyset$, which means $f(S_1 \cap S_2) = \emptyset \neq \{f(a)\}$. Therefore, $a = b$. This shows f is one-to-one. $\square$

Example 1.27. Suppose c is a real number and $\mathbf{v}$ is a vector in $\mathbb{R}^n$. Prove that if $c\mathbf{v} = \mathbf{0}$, then $c = 0$ or $\mathbf{v} = \mathbf{0}$.

Solution. Let $\mathbf{v} = (x_1, \dots, x_n)$. On the contrary, assume neither c is zero, nor $\mathbf{v}$ is the zero vector. Therefore, x_i is not zero for some i. Since $c \neq 0$, we have $cx_i \neq 0$. Thus,

$$c\mathbf{v} = (cx_1, \dots, cx_i, \dots, cx_n) \neq \mathbf{0}.$$

This contradiction shows $c = 0$ or $\mathbf{v} = \mathbf{0}$. □

Further Reading: Click here for further reading on Sets, Maps, and Vector Spaces.

1.7 Exercises

Exercise 1.1. *Given the following sets A, B, C write down each of the sets $(A \times B) \cap C$, $A \cup (B \cap C)$ and $A \times B \times C$ by listing all of their elements in braces.*

$$A = \{1, -1\}, B = \{1, 0\}, C = \{(1, 1), (1, 0)\}.$$

Exercise 1.2. *For n sets $A_1, A_2, \dots, A_n$, prove that $A_1 \times A_2 \times \cdots \times A_n = \emptyset$ if and only if $A_i = \emptyset$ for some i.*

Hint: Proof by contradiction might be useful.

Exercise 1.3. *Suppose for two nonempty sets A, B we know $A \times B = B \times A$. Prove that $A = B$.*

Exercise 1.4. *Prove or disprove:*

(a) For every three sets A, B, C we have $A - (B \cup C) = (A - B) \cap (A - C)$.

(b) For every three sets A, B, C we have $A - (B \cap C) = (A - B) \cup (A - C)$.

Exercise 1.5. *Determine (with full justification) which of the following statements are true.*

(a) $(\mathbb{Z} \times \mathbb{R}) \cup (\mathbb{R} \times \mathbb{Z}) = \mathbb{R} \times \mathbb{R}$.

(b) $(\mathbb{Z} \times \mathbb{R}) \cap (\mathbb{R} \times \mathbb{Z}) = \mathbb{Z} \times \mathbb{Z}$.

(c) $(\mathbb{R} - \mathbb{Z}) \times \mathbb{Z} = (\mathbb{R} \times \mathbb{Z}) - (\mathbb{Z} \times \mathbb{Z})$.

(d) $\{x \in \mathbb{R} \mid \sin x \in \mathbb{Z}\} = \{x \in \mathbb{R} \mid \cos x \in \mathbb{Z}\}$.

(e) $\{n \in \mathbb{Z} \mid n = 6m + 1 \text{ for some } m \in \mathbb{Z}\} = \{n \in \mathbb{Z} \mid n = 6m - 5 \text{ for some } m \in \mathbb{Z}\}$.

Definition 1.10. The **power set** of a set A, denoted by $\mathcal{P}(A)$ is the set consisting of all subsets of A.

Exercise 1.6. *Prove or disprove each of the following:*

(a) For every two sets A, B, we have $\mathcal{P}(A \cap B) = \mathcal{P}(A) \cap \mathcal{P}(B)$.

(b) For every two sets A, B, we have $\mathcal{P}(A \cup B) = \mathcal{P}(A) \cup \mathcal{P}(B)$.

(c) For every two sets A, B, we have $\mathcal{P}(A \times B) = \mathcal{P}(A) \times \mathcal{P}(B)$.

Exercise 1.7. *Suppose $J \subseteq I$ are two nonempty sets of indices, and each A_i with $i \in I$ is a set. Prove each of the following:*

(a) $\bigcup_{j \in J} A_i \subseteq \bigcup_{i \in I} A_i$.

(b) $\bigcap_{i \in I} A_i \subseteq \bigcap_{j \in J} A_i$.

Exercise 1.8. *Prove that*

$$\bigcup_{x \in [0,1]} ([x, 1] \times [0, x^2]) = \{(x, y) \mid 0 \leq x \leq 1 \text{ and } 0 \leq y \leq x^2\}$$

Exercise 1.9. *Prove that*

$$\bigcap_{x \in [0,1]} ([x, 1] \times [0, x^2]) = \{(1, 0)\}$$

Exercise 1.10. *Given a nonempty set A, what is the set $\bigcup_{a \in A} \{a\}$? How about $\bigcup_{a \in A} \{(a, 1)\}$? Prove your claims.*

Exercise 1.11. *Let X be a nonempty set with n elements. How many one-to-one functions $f : X \to X$ are there? How many onto functions $f : X \to X$ are there?*

Exercise 1.12. *The graph of a function $f : X \to Y$ is defined by $\Gamma_f = \{(x, f(x)) \mid x \in X\}$. Prove that two functions $f, g : X \to Y$ are equal if and only if $\Gamma_f = \Gamma_g$.*

Exercise 1.13. *Suppose f, g are two functions for which $R_g \subseteq D_f$. Prove or disprove each statement.*

(a) If both f and g are injective, then so is $f \circ g$.

(b) If both f and g are surjective, then so is $f \circ g$.

Exercise 1.14. *Determine if each function is injective, surjective or neither.*

(a) $f : \mathbb{R}^2 \to \mathbb{R}^2$ given by $f(x, y) = (x + y, xy)$.

(b) $f : \mathbb{R}^n \to \mathbb{R}$ given by $f(x_1, \ldots, x_n) = x_1 + \cdots + x_n$.

(c) $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^3 - 5x$.

Exercise 1.15. *Prove parts (b) and (c) of Theorem 1.2.*

Exercise 1.16. *Prove each of the following:*

(a) $\sum_{i=1}^{n} \frac{1}{i^2 + i} = \frac{n}{n+1}$ *for every $n \in \mathbb{Z}^+$.*

(b) $2^n > n$ for every $n \in \mathbb{Z}^+$.

(c) $2^n \geq n^2$ for every $n \in \mathbb{N}$ with $n \geq 4$.

Exercise 1.17. *Prove that if for a real number x, the number x^2 is irrational, then so is x.*

Exercise 1.18. *Let E, D be the set of all even and odd integers, respectively. Find a bijection $f : \mathbb{N} \to E$ and another bijection $g : D \to \mathbb{N}$.*

Exercise 1.19. *Suppose $p(x)$ is a polynomial. Prove that for every positive integer n, there is a polynomial $q(x)$ for which the n-th derivative of $e^{p(x)}$ is equal to $q(x)e^{p(x)}$ for every $x \in \mathbb{R}$.*

Exercise 1.20. *Prove that for every positive integer n, there is a polynomial p_n for which the n-th derivative of e^{-1/x^2} at $x \neq 0$ is equal to $p_n(1/x)e^{-1/x^2}$.*

Exercise 1.21. *Carefully prove all vector space properties I-VII of $\mathbb{R}^n$.*

Exercise 1.22. *Let $f : A \to B$ be a function, T be a subset of B. Prove that $f^{-1}(T^c) = (f^{-1}(T))^c$. (Note: For a subset S of A and a subset T of B we have $T^c = B - T$ and $S^c = A - S$.)*

Definition 1.11. A function $f : D \to \mathbb{R}$ is said to be **even** (resp. **odd**) if:

- D is a subset of $\mathbb{R}$ that satisfies $x \in D$ if and only if $-x \in D$, and
- $f(-x) = f(x)$ (resp. $f(-x) = -f(x)$) for every $x \in D$.

Exercise 1.23. *Prove that every function $f : \mathbb{R} \to \mathbb{R}$ can be written as sum of two functions $g, h : \mathbb{R} \to \mathbb{R}$, where g is even and h is odd. Prove the representation $f = g + h$ into sum of an even and an odd function is unique.*

Exercise 1.24. *Suppose functions $f, g : \mathbb{R} \to \mathbb{R}$ are n-times differentiable. Prove*

$$(fg)^{(n)}(x) = \sum_{k=0}^{n} \binom{n}{k} f^{(k)}(x) g^{(n-k)}(x), \text{ for all } x \in \mathbb{R}.$$

Note: $f^{(0)}(x) = f(x)$ and $g^{(0)}(x) = g(x)$.

1.8 Challenge Problems

Exercise 1.25. *Let $r \geq 2$ be a fixed positive integer, and let $\mathcal{F}$ be an infinite family of distinct sets, each of size r, no two of which are disjoint. Prove that there exists a set of size $r - 1$ that intersects each set in $\mathcal{F}$.*

Exercise 1.26. *Let A be a nonempty set. Suppose $f : \mathcal{P}(A) \to \mathcal{P}(A)$ is a bijection for which for every subsets X and Y of A:*

$$\text{If } X \subseteq Y, \text{ then } f(X) \subseteq f(Y).$$

(a) If A is finite, show that if $f(X) \subseteq f(Y)$, then $X \subseteq Y$.

(b) Show part (a) does not necessarily hold when A is infinite.

1.9 Summary

- To prove $A \subseteq B$, start with $x \in A$ and prove $x \in B$.
- To prove two sets A and B are equal we need to show if $x \in A$, then $x \in B$ and vice-versa.
- For a function $f : A \to B$, a subset S of A, and a subset T of B, we have the following:

$$x \in f(S) \text{ iff } x = f(s) \text{ for some } s \in S, \text{ and } y \in f^{-1}(T) \text{ iff } f(y) \in T.$$

- $f^{-1}(\bigcup_{i \in I} T_i) = \bigcup_{i \in I} f^{-1}(T_i)$ and $f^{-1}(\bigcap_{i \in I} T_i) = \bigcap_{i \in I} f^{-1}(T_i)$.
- $f(f^{-1}(T)) \subseteq T$ and $S \subseteq f^{-1}(f(S))$.
- To prove a statement by contradiction, assume the conclusion is false and after taking logical steps obtain a contradiction.
- To prove a statement depending on a positive integer n, first prove the statement for $n = 1$ (basis step), then prove that if the statement is true for n it must be true for $n + 1$ (inductive step).

Week 2

2.1 Subspaces

YouTube Video: https://youtu.be/3iWdoRqTpE0

Definition 2.1. A subset W of $\mathbb{R}^n$ is called a **subspace** of $\mathbb{R}^n$ if W along with the same operations of $\mathbb{R}^n$ satisfies all properties of a vector space, i.e. properties I-VII listed in the previous section.

Theorem 2.1 (Subspace Criterion). *A subset W of $\mathbb{R}^n$ is a subspace if and only if it satisfies all of the following:*

- *W contains the zero vector, and*
- *for all $\mathbf{x}, \mathbf{y} \in W$ and $c \in \mathbb{R}$, we have $\mathbf{x} + \mathbf{y} \in W$ and $c\mathbf{x} \in W$. [We say W is closed under vector addition and scalar multiplication.]*

Example 2.1. Here are some examples of subspaces:

(a) The set of all points (x, y) on a given line $y = mx$ is a subspace of $\mathbb{R}^2$.

(b) The sets $\{\mathbf{0}\}$ and $\mathbb{R}^n$ are subspaces of $\mathbb{R}^n$.

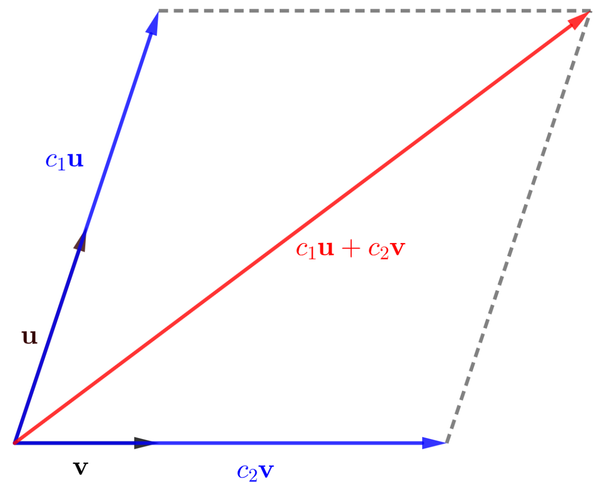
Example 2.2. If W and U are subspaces of $\mathbb{R}^n$, then so is $W \cap U$.

Solution. We will use the subspace criterion Theorem (i.e. Theorem 2.1). First note that $\mathbf{0}$ belongs to both U and W and thus it is in $U \cap W$.

Next, suppose $\mathbf{x}, \mathbf{y} \in U \cap W$ and $c \in \mathbb{R}$. By definition of intersection, $\mathbf{x}$, and $\mathbf{y}$ are in both U and W. Since U and W are both subspaces, by Theorem 2.1, we have $\mathbf{x} + \mathbf{y} \in U$, $\mathbf{x} + \mathbf{y} \in W$, $c\mathbf{x} \in U$ and $c\mathbf{x} \in W$. Therefore, by definition of intersection, $\mathbf{x} + \mathbf{y} \in U \cap W$, and $c\mathbf{x} \in U \cap W$, as desired. $\square$

2.2 Linear Dependence, Spanning, and Basis

Definition 2.2. Let $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ be a subset of the vector space $\mathbb{R}^n$, and $\mathbf{w}$ be a vector in $\mathbb{R}^n$. We say $\mathbf{w}$ is a **linear combination** of elements of S if $\mathbf{w} = c_1\mathbf{v}_1 + \cdots + c_m\mathbf{v}_m$ for some $c_1, \ldots, c_m \in \mathbb{R}$. By definition, if S is the empty set, then the only linear combination of elements of S is $\mathbf{0}$, the zero vector.



We note that every vector $\mathbf{v} = (x, y, z)$ in $\mathbb{R}^3$ can be written as:

$$(x, y, z) = x(1, 0, 0) + y(0, 1, 0) + z(0, 0, 1).$$

The vectors $(1, 0, 0), (0, 1, 0)$, and $(0, 0, 1)$ are in some way "independent" of one another. The next definition allows us to formalize this idea of "independence".

Definition 2.3. We say vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m \in \mathbb{R}^n$ are **linearly dependent** if one of these vectors can be written as a linear combination of the others. Otherwise, we say $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ are **linearly independent**.

Example 2.3. Check if each of the following vectors are linearly dependent or linearly independent.

(a) $(1, 0, 0), (0, 1, 0)$, and $(0, 0, 1)$.

(b) $(1, 2, 4), (3, 1, 2)$, and $(4, 3, 6)$.

Theorem 2.2. *The vectors* $\mathbf{v}_1, \ldots, \mathbf{v}_m \in \mathbb{R}^n$ *are linearly dependent if and only if there are real numbers* $c_1, \ldots, c_m$*, not all zero, such that* $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_m\mathbf{v}_m = \mathbf{0}$.
In other words, vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ *are linearly independent if and only if the following statement is true*

$$\text{If } c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_m\mathbf{v}_m = \mathbf{0} \text{ for some scalars } c_1, c_2, \ldots, c_m, \text{ then } c_1 = c_2 = \cdots = c_m = 0.$$

Definition 2.4. Given a subspace V of $\mathbb{R}^n$, we say a subset $\mathcal{S}$ of V is **spanning** (or **generating**) if every $\mathbf{v} \in V$ is a linear combination of some vectors in $\mathcal{S}$.

Definition 2.5. We say a subset $\mathcal{B}$ of a subspace V of $\mathbb{R}^n$ is a **basis** if $\mathcal{B}$ is both linearly independent and spanning.

Example 2.4. Prove that $(1,0,0),(0,1,0),(0,0,1)$ form a basis for $\mathbb{R}^3$.

Theorem 2.3. *Let V be a subspace of $\mathbb{R}^n$. Vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m \in V$ form a basis for V if and only if every vector $\mathbf{w} \in V$ can be uniquely written as $\mathbf{w} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_m\mathbf{v}_m$.*

2.3 Some Examples of Subspaces

YouTube Video: https://youtu.be/Qw4qSq0lgJ4

Example 2.5 (Span of vectors)**.** Let $\mathcal{A}$ be a set of vectors in $\mathbb{R}^n$, and let "span $\mathcal{A}$" be the set consisting of all vectors that are linear combinations of some vectors of $\mathcal{A}$. Then span $\mathcal{A}$ is a subspace of $\mathbb{R}^n$.

Definition 2.6. Let A be an $m \times n$ matrix. The **row space of** A denoted by $\mathrm{Row}(A)$ is the subspace of $\mathbb{R}^n$ spanned by the rows of A, and the **column space of** A denoted by $\mathrm{Col}(A)$ is the subspace of $\mathbb{R}^m$ spanned by the columns of A.

Example 2.6. Consider the matrix

$$\begin{pmatrix} 1 & 2 & -1 \\ 0 & 4 & 2 \end{pmatrix}$$

Describe the row and column space of the matrix above.

Example 2.7 (Row space and column space)**.** Prove that row space and column space of every matrix are vector spaces.

2.4 Systems of Linear Equations

YouTube Video: https://youtu.be/a6rYNze3Zfw

Suppose we would like to solve the following system of equations:

$$\begin{cases} 3x + 2y - z = 4 \\ x + 3y - 2z = 1 \\ 5x + y - z = 4 \end{cases}$$

In high school algebra, we learn two methods for solving systems of linear equations: substitution and elimination. Substitution could typically get too computational, especially when the number of variables is too large. Elimination often works better, but we still need to keep track of too many things. Our objective is to keep track of all of the work in a more organized fashion. We will keep all coefficients and constants in a single matrix, separating the coefficients from constants by a vertical bar. This matrix is called the **augmented matrix** of the given system. For the example, the augmented matrix of the above system is as follows:

$$\left(\begin{array}{ccc|c} 3 & 2 & -1 & 4 \\ 1 & 3 & -2 & 1 \\ 5 & 1 & -1 & 4 \end{array}\right)$$

In the elimination method, we will add an appropriate multiple of one of the equations to another equation. This means we are doing the same thing to the rows of the augmented matrix. We note that each step is reversible and thus we are not inserting or eliminating any solutions. In this process, three operations are used. The operations (listed below) are called **elementary row operations**.

- **Row Addition:** Adding a scalar multiple of a row to another row.
- **Row Interchange:** Interchanging two rows.
- **Row Scale:** Multiplying a row by a nonzero number.

The objective is to obtain a matrix that satisfies all of the following.

- All zero rows are at the bottom.
- The entries below the first nonzero entry of each row are all zero.
- The leading nonzero entry of each row is to the left of the leading nonzero entry of all rows below it.

Such a matrix is called a matrix in **(row) echelon form**.

If in addition to the above, we also have the following two conditions:

- the first nonzero entry of each row is 1, and
- these 1's are the only nonzero entry of their column.

Then, we say the matrix is in **reduced (row) echelon form**.

To apply this method:

- Interchange rows so that the first entry of the first row is nonzero. (If the first column is all zero, apply this to the first nonzero column.)
- Using the first row and the row addition operation, make all entries below the first nonzero entry zero.
- If possible, by interchanging rows, make the second entry of the second row nonzero. If not, move on to the next entry.
- Repeat this process so that you obtain a matrix in echelon form.
- Scale all rows to obtain 1's as the leading nonzero entries.
- Turn the rest of the entries in columns of each leading 1 into zero to obtain a matrix in reduced echelon form.

Theorem 2.4. *Every matrix can be turned into a matrix in reduced echelon form by applying the three elementary row operations. Furthermore, the reduced echelon form for any matrix is unique.*

Definition 2.7. The leading nonzero entries in a matrix in echelon form are called **pivot** entries. Each column that contains a pivot entry is called a **pivot column**.

Definition 2.8. A system of linear equations is called **homogeneous** if the right hand side of the system is all zeros. In other words, any homogeneous system is of the following form:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1k}x_k = 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2k}x_k = 0 \\ \qquad\qquad\vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nk}x_k = 0 \end{cases}$$

Here, all a_{ij}'s are constants. Note that every homogeneous system has a **trivial** solution

$$x_1 = x_2 = \cdots = x_k = 0.$$

Intuitively, in a homogeneous system if the number of equations is less than the number of variables, we must have infinitely many solutions. Let's test this hypothesis with an example.

Example 2.8. Find all solutions of the system:

$$\begin{cases} 2x_1 - x_2 + 3x_3 + x_4 = 0 \\ x_1 - 3x_2 + x_4 = 0 \\ x_2 - x_3 + 4x_4 = 0 \end{cases}$$

With the method used in the solution of the above example we can prove the following theorem:

Theorem 2.5. *Any homogeneous system that has less equations than variables has a nontrivial solution.*

Corollary 2.1. Every $n+1$ vectors in $\mathbb{R}^n$ are linearly dependent.

2.5 More Examples

Example 2.9. Determine if each of the following is a subspace of $\mathbb{R}^2$.

(a) The set of points on the line $3x + 2y = 1$.

(b) The set of points on the line $4x - 3y = 0$.

(c) The set of points on the unit circle $x^2 + y^2 = 1$.

Solution. (a) This is not a subspace of $\mathbb{R}^2$ since $(0,0)$ does not lie on this line, but the origin lies on every subspace.

(b) This is a subspace. To prove that we will use the Subspace Criterion. First, note that $(0,0)$ is on this line. Suppose (a,b) and (c,d) lie on this line and $r \in \mathbb{R}$. By assumption,

$$4a - 3b = 0, \text{ and } 4c - 3d = 0.$$

We have

$$4(a+c)-3(b+d)=(4a-3b)+(4c-3d)=0+0=0, \text{ and } 4(ra)-3(rb)=r(4a-3b)=r0=0.$$

Therefore, $(a+c,b+d)$ and (ra,rb) both belong to the same line. Thus, this line is a subspace of $\mathbb{R}^2$.

(c) This is not a subspace since it does not contain $(0,0)$. $\square$

Example 2.10. Prove that every set of vectors that contains the zero vector is linearly dependent.

Solution. Let $\mathcal{S}$ be a set of vectors containing $\mathbf{0}$. We see that $1\mathbf{0}=\mathbf{0}$ and the coefficient 1 is nonzero. Therefore, by Theorem 2.2, the set $\mathcal{S}$ is linearly dependent. $\square$

Example 2.11. Prove the vectors $\mathbf{x}=(1,2)$, and $\mathbf{y}=(-1,2)$ form a basis for $\mathbb{R}^2$.

Solution. We need to show $\mathbf{x}$ and $\mathbf{y}$ are linearly independent and spanning.
For linear independence, suppose $c_1\mathbf{x}+c_2\mathbf{y}=\mathbf{0}$, for some real numbers c_1,c_2. Thus, $(c_1-c_2,2c_1+2c_2)=(0,0)$, which implies $c_1=c_2$ and $c_1=-c_2$. This yields $c_1=c_2=0$. Therefore, $\mathbf{x}$ and $\mathbf{y}$ are linearly independent.
For spanning, suppose $(a,b)\in\mathbb{R}^2$. We will have to find $c_1,c_2\in\mathbb{R}$ for which $c_1\mathbf{x}+c_2\mathbf{y}=(a,b)$. This means we need to solve the system:

$$\begin{aligned} c_1-c_2&=a\\ 2c_1+2c_2&=b \end{aligned}$$

Now solve this and find c_1 and c_2 in terms of a and b, and your solution would be complete. $\square$

Example 2.12. Let S and T be two subsets of $\mathbb{R}^n$. Then span $S=$ span T if and only if $S\subseteq$ span T and $T\subseteq$ span S.

Solution. $\Rightarrow$: Suppose span $S=$ span T. By definition of span, $S\subseteq$ span $S=$ span T. Similarly $T\subseteq$ span $T=$ span S, as desired.

$\Leftarrow$: Now, suppose $S\subseteq$ span T, and $T\subseteq$ span S. By definition of span, every element $\mathbf{v}\in$ span T is of the form $\mathbf{v}=c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n$ for some $\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_m\in T$. Since $T\subseteq$ span S, we have $\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_m\in$ span S. Since span S is a subspace, $\mathbf{v}=c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n\in$ span S. Therefore, span $T\subseteq$ span S. Similarly span $S\subseteq$ span T. This implies span $S=$ span T, as desired. $\square$

Example 2.13. Let $\mathbf{v}_1,\ldots,\mathbf{v}_m\in\mathbb{R}^n$ be linearly independent. Consider arbitrary vectors $\mathbf{w}_1,\ldots,\mathbf{w}_m\in\mathbb{R}^k$, and let $\mathbf{x}_1=(\mathbf{v}_1,\mathbf{w}_1),\ldots,\mathbf{x}_m=(\mathbf{v}_m,\mathbf{w}_m)\in\mathbb{R}^{n+k}$ be vectors created by placing components of $\mathbf{v}_j$ followed by components of $\mathbf{w}_j$. Prove that $\mathbf{x}_1,\ldots,\mathbf{x}_m$ are linearly independent.

Solution. Let $c_1,\ldots,c_m\in\mathbb{R}$ be scalars for which

$$c_1\mathbf{x}_1+\cdots+c_m\mathbf{x}_m=\mathbf{0}.$$

Using the way $\mathbf{x}_j$'s are created we have

$$c_1(\mathbf{v}_1,\mathbf{w}_1)+\cdots+c_m(\mathbf{v}_m,\mathbf{w}_m)=\mathbf{0}\Rightarrow(c_1\mathbf{v}_1+\cdots+c_m\mathbf{v}_m,c_1\mathbf{w}_1+\cdots+c_m\mathbf{w}_m)=\mathbf{0}\Rightarrow c_1\mathbf{v}_1+\cdots+c_m\mathbf{v}_m=\mathbf{0}.$$

Since $\mathbf{v}_1,\ldots,\mathbf{v}_m$ are linearly independent we obtain $c_1=\cdots=c_m=0$, and hence $\mathbf{x}_1,\ldots,\mathbf{x}_m$ are linearly independent. $\square$

Example 2.14. Suppose $\mathbf{v}_1,\ldots,\mathbf{v}_n$ are linearly independent vectors of $\mathbb{R}^k$ and $\mathbf{w}_1,\mathbf{w}_2,\ldots,\mathbf{w}_m$ are also linearly independent vectors of $\mathbb{R}^k$. Prove that $\mathbf{v}_1,\ldots,\mathbf{v}_n,\mathbf{w}_1,\ldots,\mathbf{w}_m$ are linearly independent if and only if

$$\text{span}\,\{\mathbf{v}_1,\ldots,\mathbf{v}_n\}\cap\text{span}\,\{\mathbf{w}_1,\ldots,\mathbf{w}_m\}=\{\mathbf{0}\}.$$

Solution. For simplicity, let $V=\text{span}\,\{\mathbf{v}_1,\ldots,\mathbf{v}_n\}$, and $W=\text{span}\,\{\mathbf{w}_1,\ldots,\mathbf{w}_m\}$.

$\Rightarrow$: Suppose $\mathbf{v}_1,\ldots,\mathbf{v}_n,\mathbf{w}_1,\ldots,\mathbf{w}_m$ are linearly independent and $\mathbf{x}\in V\cap W$. Thus $\mathbf{x}=\sum_{i=1}^{n}a_i\mathbf{v}_i=\sum_{j=1}^{m}b_j\mathbf{w}_j$, for some $a_i,b_j\in\mathbb{R}$. Therefore, $\sum_{i=1}^{n}a_i\mathbf{v}_i-\sum_{j=1}^{m}b_j\mathbf{w}_j=\mathbf{0}$. Since $\mathbf{v}_1,\ldots,\mathbf{v}_n,\mathbf{w}_1,\ldots,\mathbf{w}_m$ are linearly independent, we must have $a_i=b_j=0$ and thus $\mathbf{x}=\mathbf{0}$. So, we proved every element of $V\cap W$ is the zero vector. On the other hand $\mathbf{0}\in V\cap W$. Therefore, $V\cap W=\{\mathbf{0}\}$.

$\Leftarrow$: Now assume $V\cap W=\{\mathbf{0}\}$. Suppose $\sum_{i=1}^{n}a_i\mathbf{v}_i+\sum_{j=1}^{m}b_j\mathbf{w}_j=\mathbf{0}$. This implies $\sum_{i=1}^{n}a_i\mathbf{v}_i=-\sum_{j=1}^{m}b_j\mathbf{w}_j\in V\cap W$, which implies $\sum_{i=1}^{n}a_i\mathbf{v}_i=-\sum_{j=1}^{m}b_j\mathbf{w}_j=\mathbf{0}$. Since $\mathbf{v}_1,\ldots,\mathbf{v}_n$ and $\mathbf{w}_1,\ldots,\mathbf{w}_m$ are linear independent we must have $a_i=b_j=0$ for all i,j. Therefore, $\mathbf{v}_1,\ldots,\mathbf{v}_n,\mathbf{w}_1,\ldots,\mathbf{w}_m$ are linearly independent. $\square$

Example 2.15. Determine if each of the following matrices is in echelon form, reduced echelon form or neither. If the matrix is not in reduced echelon form, turn it into reduced echelon form by appropriate elementary row operations. In each step make sure you specify which row operation is used.

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

(b) $\begin{pmatrix} -1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 5 \end{pmatrix}$

Solution. (a) This is not in echelon form. Applying R_3+2R_2, then, $-R_2$ we obtain the following:

$$\begin{pmatrix} 1 & 0 & -1 \\ 0 & -1 & 0 \\ 0 & 2 & 3 \end{pmatrix}\xrightarrow{R_3+2R_2}\begin{pmatrix} 1 & 0 & -1 \\ 0 & -1 & 0 \\ 0 & 0 & 3 \end{pmatrix}\xrightarrow{-R_2}\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

Now, if we apply $R_3/3$ followed by R_1+R_3 we obtain a matrix in reduced echelon form as shown below:

$$\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{pmatrix} \xrightarrow{R_3/3} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \xrightarrow{R_1+R_3} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

(b) This is in echelon form but is not in reduced echelon form. Applying $-R_1$ and $R_3/5$ yields a matrix in reduced echelon form.

$$\begin{pmatrix} -1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 5 \end{pmatrix} \xrightarrow{-R_1} \begin{pmatrix} 1 & 0 & -2 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 5 \end{pmatrix} \xrightarrow{R_3/5} \begin{pmatrix} 1 & 0 & -2 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

□

Example 2.16. Find all values of h for which the following system has a solution.

$$\begin{cases} x_1 + 2x_2 - x_3 = 7 + h \\ x_2 - 2x_3 = 3 \\ 2x_1 + 5x_2 - 4x_3 = h \end{cases}$$

Solution. We will row reduce the augmented matrix associated with the above system:

$$\left(\begin{array}{ccc|c} 1 & 2 & -1 & 7+h \\ 0 & 1 & -2 & 3 \\ 2 & 5 & -4 & h \end{array}\right) \xrightarrow{R_3-2R_1} \left(\begin{array}{ccc|c} 1 & 2 & -1 & 7+h \\ 0 & 1 & -2 & 3 \\ 0 & 1 & -2 & -14-h \end{array}\right) \xrightarrow{R_3-R_2} \left(\begin{array}{ccc|c} 1 & 2 & -1 & 7+h \\ 0 & 1 & -2 & 3 \\ 0 & 0 & 0 & -17-h \end{array}\right)$$

Note that the $(1,2)$ entry can be easily made zero by applying $R_1 - 2R_2$. This means, from the first two equations, we can find x_1, x_2 in terms of x_3. For this system to have a solution we need $0 = -17 - h$, which is obtained from the last equation. Therefore, $\boxed{h = -17}$. □

Further Reading: Click here and here for further reading on systems of linear equations and echelon forms.

2.6 Exercises

Exercise 2.1. *Determine if each of the following is a subspace of $\mathbb{R}^n$ once by checking if they satisfy all vector spaces properties I-VII, and once by using the subspace criterion.*

(a) The set of all vectors $(x_1, \ldots, x_n) \in \mathbb{R}^n$ satisfying $x_1 + 2x_2 + \cdots + nx_n = 0$.

(b) The empty set.

(c) The set of all vectors $(x_1, \ldots, x_n) \in \mathbb{R}^n$ satisfying $x_1^2 + x_2^2 + \cdots + x_n^2 = 0$.

(d) The set of all vectors $(x_1, \ldots, x_n) \in \mathbb{R}^n$ satisfying $x_1^2 + x_2^2 + \cdots + x_n^2 = 1$.

Exercise 2.2. *Determine if the following set is a subspace of* $\mathbb{R}^3$*:*

$$\{(x, y, z) \in \mathbb{R}^3 \mid (x^2 + 2y^2 + 3z^2)(z - 3y) = 0\}.$$

Exercise 2.3. *Suppose* U *and* W *are subspaces of* $\mathbb{R}^n$ *for which* $U \cup W$ *is also a subspace. Prove that* $U \subseteq W$ *or* $W \subseteq U$*.*

Hint: Use proof by contradiction.

Exercise 2.4. *Consider the homogeneous system*

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1k}x_k = 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2k}x_k = 0 \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nk}x_k = 0 \end{cases}$$

Prove that the set of vectors $(x_1, x_2, \ldots, x_k) \in \mathbb{R}^k$ *satisfying the system above is a subspace of* $\mathbb{R}^k$*.*

Exercise 2.5. *Suppose* V *and* W *are subspaces of* $\mathbb{R}^n$*. Define*

$$V + W = \{\mathbf{v} + \mathbf{w} \mid \mathbf{v} \in V, \text{ and } \mathbf{w} \in W\}.$$

Prove that $V + W$ *is a subspace of* $\mathbb{R}^n$*.*

Exercise 2.6. *Suppose* V *is the subset of* $\mathbb{R}^3$ *consisting of all points* (x, y, z) *for which*

$$x + 2y - z = 0, \text{ and } 2x - 4y + 7z = 0.$$

Prove that V *is a subspace of* $\mathbb{R}^3$*.*

Exercise 2.7. *Suppose* $A = (x_1, y_1)$*, and* $B = (x_2, y_2)$ *are two distinct points on the plane. Let* S *be the set of all points that are equidistant from* A *and* B*. Find the necessary and sufficient condition on points* A, B *for which* S *is a subspace of* $\mathbb{R}^2$*.*

Exercise 2.8. *Prove that the only finite subspace of* $\mathbb{R}^n$ *is the trivial subspace* $\{\mathbf{0}\}$ *containing only the zero vector.*

Exercise 2.9. *Suppose* V, W *are two subspaces of* $\mathbb{R}^n$ *for which* $V \cap W$ *contains at least one nonzero vector. Prove that* $V \cap W$ *is an infinite set.*

Exercise 2.10. *Show the only proper subspace of* $\mathbb{R}$ *is* $\{0\}$*.*

Exercise 2.11. *Prove that if* $n > 1$*, then* $\mathbb{R}^n$ *can be written as the union of all of its proper subspaces.*

Exercise 2.12. *Prove the following set is a subspace of* $\mathbb{R}^3$*, once by showing it satisfies all vector space properties I-VII, and once by applying the Subspace Criterion.*

$$\{(x, y, z) \in \mathbb{R}^3 \mid x + y + 2z = 0, \text{ and } z - 2y + 3x = 0\}$$

Exercise 2.13. *Determine if each of the following matrices is in echelon form, reduced echelon form or neither. If a matrix is not in reduced echelon form, turn it into reduced echelon form by appropriate elementary row operations. In each step make sure you specify which row operation is used.*

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

(b) $\begin{pmatrix} 1 & 2 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 5 \end{pmatrix}$

Exercise 2.14. *Using elementary row operations, find all solutions of each system or show the system has no solutions.*

(a) $\begin{cases} x_1 + 3x_2 + x_4 = 5 \\ x_2 - x_3 + 5x_4 = 1 \\ 2x_1 - x_3 + x_4 = 0 \end{cases}$

(b) $\begin{cases} x_1 + x_2 + 3x_3 - x_4 = 5 \\ x_2 - x_3 + 5x_4 = -2 \\ 2x_1 + 3x_2 + 5x_3 + 3x_4 = 0 \end{cases}$

Exercise 2.15. *Show that if a matrix B is obtained by applying an elementary row operation to a matrix A, then $\mathrm{Row}(A) = \mathrm{Row}(B)$. (Hint: Check each of the three row operations separately. You could use Example 2.12.) By an example show that $\mathrm{Col}(A) = \mathrm{Col}(B)$ does not always hold.*

Exercise 2.16. *Describe all 2×2 matrices that are in reduced echelon form.*

Exercise 2.17. *Let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be vectors in a subspace V of $\mathbb{R}^n$ for which some of them are linearly dependent. Prove that all of them are linearly dependent.*

Exercise 2.18. *Prove that if two vectors in $\mathbb{R}^n$ are linearly dependent, then one of them is a scalar multiple of the other. By an example show that it is not necessarily true that both must be scalar multiples of the other.*

Exercise 2.19. *Find three vectors in $\mathbb{R}^3$ that are linearly dependent but each pair of them are linearly independent.*

Exercise 2.20. *Find all values of real number h for which each equation has a solution or show no such h exists.*

(a) $\begin{cases} x_1 + 3x_2 - x_3 = h + 2 \\ 2x_1 + x_2 - x_3 = h \\ -3x_1 + x_2 + x_3 = h + 1 \end{cases}$

(b) $\begin{cases} x_1 + x_2 - 2x_3 = h + 2 \\ x_1 + x_3 = 5 \\ -3x_1 + x_2 = 3h \end{cases}$

(c) $\begin{cases} x_1 + x_2 - 2x_3 + x_4 = h \\ x_1 + x_3 - 2x_4 = 5 \\ 3x_1 + 2x_2 - 3x_3 = 2h + 9 \end{cases}$

Exercise 2.21. *For a subspace V of $\mathbb{R}^n$ and a vector $\mathbf{x} \in \mathbb{R}^n$, define the set $\mathbf{x} + V$ by*

$$\mathbf{x} + V = \{\mathbf{x} + \mathbf{v} \mid \mathbf{v} \in V\}.$$

Prove that $\mathbf{x} + V$ is a subspace of $\mathbb{R}^n$ if and only if $\mathbf{x} \in V$.

2.7 Challenge Problems

Exercise 2.22. *Let $0 < k < m \leq n$ be integers. Prove that there are m linearly dependent vectors in $\mathbb{R}^n$, every k of which are linearly independent.*

2.8 Summary

- To prove $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly independent vectors, start with $c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n = \mathbf{0}$ and prove $c_1 = \cdots = c_n = 0$.
- To prove $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are generating, start with an arbitrary vector in the vector space and show it can be written as a linear combination of $\mathbf{v}_j$'s.
- A basis is a set of vectors that are linearly independent and generating.
- Every matrix can be turned into a matrix in echelon form by using three row operations: row addition, row interchange, and row scale.
- W is a subspace of $\mathbb{R}^n$ if W along with the operations of $\mathbb{R}^n$ itself satisfies all properties I-VII of a vector space.
- To prove W is a subspace of $\mathbb{R}^n$ we use the Subspace Criterion: W contains the zero vector, and W is closed under addition and scalar multiplication.

Week 3

3.1 Dimension of a Vector Space

YouTube Video: https://youtu.be/r_HB3Mop058

Theorem 3.1. *Assume V is a subspace of $\mathbb{R}^n$. Then, there is an integer $m \leq n$ for which V has a basis consisting of m distinct vectors. Furthermore, every basis of V contains precisely m vectors.*

Definition 3.1. A subspace V of $\mathbb{R}^n$ is said to have **dimension** m, written as $\dim V = m$, if it has a basis of size m.

Example 3.1. Find the dimension of each of the following vector spaces.

(a) $\mathbb{R}^n$

(b) $\{\mathbf{0}\}$ as a subspace of $\mathbb{R}^n$.

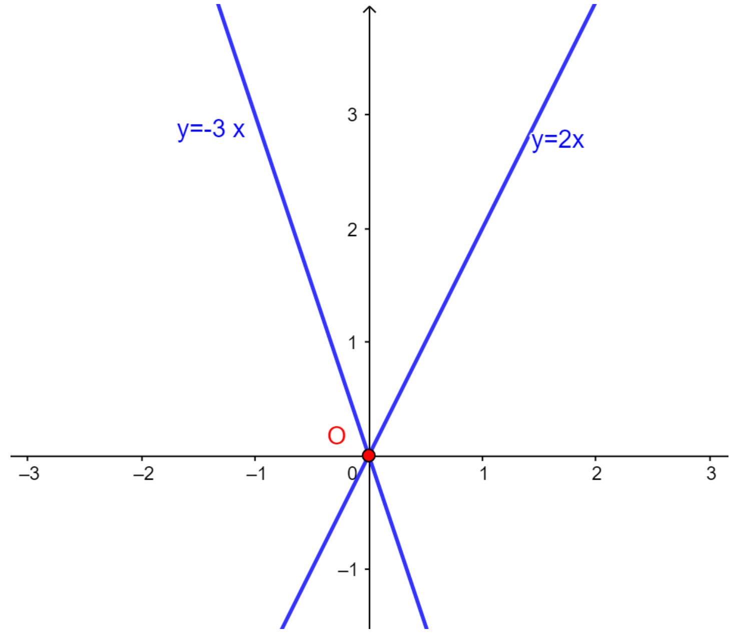
(c) The set of points on the line $y = 3x$ in the xy-plane.

Theorem 3.2. *Let V be a subspace of $\mathbb{R}^n$ of dimension m. Then,*

(a) Every m linearly independent vectors in V form a basis for V.

(b) Every m spanning vectors in V form a basis for V.

Example 3.2. All subspaces of $\mathbb{R}^2$ are either $\{\mathbf{0}\}$, lines through the origin or $\mathbb{R}^2$ itself.



Subspaces of $\mathbb{R}^2$: The origin; All lines through the origin; and $\mathbb{R}^2$ itself.

3.2 Analyzing Pivots

YouTube Video: https://youtu.be/3HKq1MtAUVk

Theorem 3.3. *Let A be a matrix.*

- *The dimension of* $\mathrm{Row}(A)$ *is equal to the number of pivot entries of the echelon form of A. Furthermore, the nonzero rows of the echelon form of A form a basis for* $\mathrm{Row}(A)$.
- *The dimension of* $\mathrm{Col}(A)$ *is equal to the number of pivot entries of the echelon form of A. Furthermore, the pivot columns of A form a basis for* $\mathrm{Col}(A)$.

Example 3.3. Find a basis for $\mathrm{Row}(A)$ and $\mathrm{Col}(A)$, where

$$A = \begin{pmatrix} 0 & 1 & 3 & 0 \\ -1 & -1 & 3 & -1 \\ 1 & 2 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$

Remark. Note that to find a basis for $\mathrm{Col}(A)$ we **must** look at pivot columns of A, and not those of its echelon form, but to find a basis for $\mathrm{Row}(A)$ we **must** look at rows of its echelon form with pivot entries, and not those of A.

Definition 3.2. The **rank** of a matrix A, denoted by rank A, is the dimension of $\mathrm{Row}(A)$ (which is the same as the dimension of $\mathrm{Col}(A)$).

Definition 3.3. The **transpose** of an $m \times n$ matrix A is an $n \times m$ matrix denoted by A^T whose every (i, j) entry is the (j, i) entry of A.

Theorem 3.4. *For every matrix A, we have* $\operatorname{rank} A = \operatorname{rank} A^T$.

Example 3.4. Find a basis for the subspace of $\mathbb{R}^4$ generated by $(1,2,0,1), (-1,1,2,1), (1,5,2,3), (1,1,-2,0)$.

Example 3.5 (Null space)**.** Given an $m \times n$ matrix A whose columns are $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \in \mathbb{R}^m$, the set of all vectors $\mathbf{v} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ for which

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_n\mathbf{v}_n = \mathbf{0}$$

is a subspace of $\mathbb{R}^n$.

Definition 3.4. The subspace defined in the previous example is called the **null space** or the **kernel** of A.

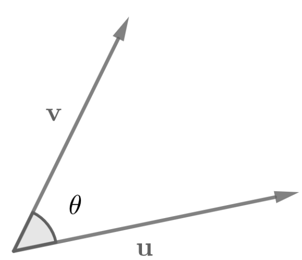
3.3 Inner Products and Angles

YouTube Video: https://youtu.be/ysLQ7RmvSsU

To better understand the geometry of $\mathbb{R}^n$, we need to define the notion of "angles" between vectors.

Example 3.6. Consider the vectors $\mathbf{u} = (x_1, y_1)$ and $\mathbf{v} = (x_2, y_2)$ in $\mathbb{R}^2$. Let θ be the angle between $\mathbf{u}$ and $\mathbf{v}$. Using the law of cosines, prove that

$$x_1x_2 + y_1y_2 = \sqrt{x_1^2 + y_1^2}\,\sqrt{x_2^2 + y_2^2}\cos\theta.$$



Definition 3.5. An **inner product** (or **scalar product**) on $\mathbb{R}^n$ is a function that assigns a real number $\langle \mathbf{x}, \mathbf{y} \rangle$ to every pair of vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ that satisfies the following for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n$ and all $a, b \in \mathbb{R}$:

(a) $\langle \mathbf{x}, \mathbf{x} \rangle > 0$ if $\mathbf{x} \neq \mathbf{0}$ (Positivity),

(b) $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$ (Symmetry),

(c) $\langle a\mathbf{x} + b\mathbf{y}, \mathbf{z} \rangle = a\langle \mathbf{x}, \mathbf{z} \rangle + b\langle \mathbf{y}, \mathbf{z} \rangle$ (Linearity).

Note that by symmetry and linearity with respect to the first vector we can obtain linearity with respect to the second vector:

$$\langle \mathbf{z}, a\mathbf{x} + b\mathbf{y} \rangle = a\langle \mathbf{z}, \mathbf{x} \rangle + b\langle \mathbf{z}, \mathbf{y} \rangle$$

Example 3.7. The following are two examples of inner products in $\mathbb{R}^n$.

(a) $\langle (x_1, x_2, \ldots, x_n), (y_1, y_2, \ldots, y_n) \rangle = x_1y_1 + x_2y_2 + \cdots + x_ny_n$.

(b) $\langle (x_1, x_2, \ldots, x_n), (y_1, y_2, \ldots, y_n) \rangle = x_1y_1 + 2x_2y_2 + \cdots + nx_ny_n$.

Remark: The first inner product of $\mathbb{R}^n$ in the example above is called the **standard inner product of** $\mathbb{R}^n$. It is also sometimes called the **dot product of** $\mathbb{R}^n$, and is denoted by "$\cdot$". In other words:

$$(x_1, x_2, \ldots, x_n) \cdot (y_1, y_2, \ldots, y_n) = \sum_{j=1}^{n} x_j y_j.$$

When a particular inner product for $\mathbb{R}^n$ is not specified we will use the dot product above.

The **length** of a vector $\mathbf{v} \in \mathbb{R}^n$ relative to an arbitrary inner product is given by $||\mathbf{v}|| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$. Therefore, the length of a vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ relative to the standard inner product is given by

$$||\mathbf{x}|| = \sqrt{\mathbf{x} \cdot \mathbf{x}} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2},$$

which matches the familiar Euclidean distance formulas in $\mathbb{R}^2$ and $\mathbb{R}^3$ (See Theorem 1.3).

By Example 3.6 we notice that, in $\mathbb{R}^2$, when $\theta = \dfrac{\pi}{2}$ we have $\mathbf{v} \cdot \mathbf{w} = 0$. This suggests the following definition:

Definition 3.6. Given an inner product on $\mathbb{R}^n$, we say two vectors $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ are **orthogonal** (or **perpendicular**) iff $\langle \mathbf{v}, \mathbf{w} \rangle = 0$. We say nonzero vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are **orthogonal** iff $\mathbf{v}_i$, and $\mathbf{v}_j$ are orthogonal for every $i \neq j$.

Example 3.8. Show that $(1, 3, -1)$ and $(-2, 1, 1)$ are orthogonal vectors of $\mathbb{R}^3$ using the standard inner product of $\mathbb{R}^3$, but not relative to the second inner product stated in Example 3.7.

Example 3.9. Let $\mathbf{e}_i \in \mathbb{R}^n$ be the vector whose i-th component is 1 and whose all other components are zero. Then, $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ is an orthogonal basis for $\mathbb{R}^n$ relative to the dot product of $\mathbb{R}^n$.

YouTube Video: https://youtu.be/8zbPkCAyDNY

Theorem 3.5 (Pythagorean Theorem)**.** *If vectors* $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ *are orthogonal relative to an inner product of* $\mathbb{R}^n$*, then*

$$||\mathbf{v}||^2 + ||\mathbf{w}||^2 = ||\mathbf{v} + \mathbf{w}||^2.$$

Example 3.6 suggests we should define the angle θ between two vectors $\mathbf{v}, \mathbf{w}$ by $\cos\theta = \dfrac{\langle \mathbf{v}, \mathbf{w} \rangle}{||\mathbf{v}|| \, ||\mathbf{w}||}$. In order for us to be able to define the angle between two vectors by $\cos\theta = \dfrac{\langle \mathbf{v}, \mathbf{w} \rangle}{||\mathbf{v}|| \, ||\mathbf{w}||}$ we need the following:

Theorem 3.6 (Cauchy-Schwarz Inequality)**.** *Given an inner product* $\langle , \rangle$ *of* $\mathbb{R}^n$*, we have*

$$|\langle \mathbf{v}, \mathbf{w} \rangle| \leq ||\mathbf{v}|| \, ||\mathbf{w}||.$$

Definition 3.7. The **angle** between two nonzero vectors $\mathbf{v}$ and $\mathbf{w}$ in $\mathbb{R}^n$ relative to a given inner product $\langle , \rangle$ is defined by

$$\theta = \cos^{-1}\left(\frac{\langle \mathbf{v}, \mathbf{w} \rangle}{||\mathbf{v}|| \, ||\mathbf{w}||}\right).$$

Example 3.10. Find the angle between $(1,2,-1)$ and $(1,1,3)$, once relative to the standard inner product and once relative to the inner product given by

$$\langle (x_1,y_1,z_1),(x_2,y_2,z_2)\rangle = x_1x_2 + 2y_1y_2 + 3z_1z_2.$$

Theorem 3.7. *Suppose $\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_k \in \mathbb{R}^n$ are orthogonal (nonzero) vectors with respect to some inner product of $\mathbb{R}^n$. Then, they are linearly independent.*

YouTube Video: https://youtu.be/Rtkvgp3EdyU

The following theorem allows us to find an orthogonal basis for any subspace of $\mathbb{R}^n$.

Theorem 3.8 (Gram-Schmidt Orthogonalization Process)**.** *Let $\langle,\rangle$ be an inner product on $\mathbb{R}^n$, and let $\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_m$ be linearly independent vectors in $\mathbb{R}^n$. Define vectors $\mathbf{w}_1,\mathbf{w}_2,\ldots,\mathbf{w}_m$ recursively as follows:*

$$\begin{aligned}
\mathbf{w}_1 &= \mathbf{v}_1\\
\mathbf{w}_2 &= \mathbf{v}_2 - \frac{\langle \mathbf{v}_2,\mathbf{w}_1\rangle}{\langle \mathbf{w}_1,\mathbf{w}_1\rangle}\mathbf{w}_1\\
\mathbf{w}_3 &= \mathbf{v}_3 - \frac{\langle \mathbf{v}_3,\mathbf{w}_1\rangle}{\langle \mathbf{w}_1,\mathbf{w}_1\rangle}\mathbf{w}_1 - \frac{\langle \mathbf{v}_3,\mathbf{w}_2\rangle}{\langle \mathbf{w}_2,\mathbf{w}_2\rangle}\mathbf{w}_2\\
&\vdots\\
\mathbf{w}_m &= \mathbf{v}_m - \frac{\langle \mathbf{v}_m,\mathbf{w}_1\rangle}{\langle \mathbf{w}_1,\mathbf{w}_1\rangle}\mathbf{w}_1 - \frac{\langle \mathbf{v}_m,\mathbf{w}_2\rangle}{\langle \mathbf{w}_2,\mathbf{w}_2\rangle}\mathbf{w}_2 - \cdots - \frac{\langle \mathbf{v}_m,\mathbf{w}_{m-1}\rangle}{\langle \mathbf{w}_{m-1},\mathbf{w}_{m-1}\rangle}\mathbf{w}_{m-1}
\end{aligned}$$

Then $\mathbf{w}_1,\mathbf{w}_2,\ldots,\mathbf{w}_m$ form an orthogonal basis for the subspace of $\mathbb{R}^n$ spanned by $\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_m$.

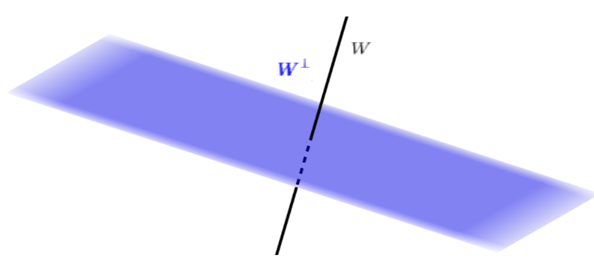
Corollary 3.1. Every subspace of $\mathbb{R}^n$ has an orthogonal basis.

Definition 3.8. We say vectors $\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_n$ are **orthonormal** relative to an inner product $\langle,\rangle$ if they are orthogonal and $\langle \mathbf{v}_i,\mathbf{v}_i\rangle = 1$ for every i. (i.e. all of them have length 1.)

Example 3.11. Find an orthogonal basis for the subspace of $\mathbb{R}^4$ generated by $(1,2,0,-1),(0,1,0,2),(0,0,2,1)$.

Definition 3.9. Let W be a subspace of $\mathbb{R}^n$. The **orthogonal complement of** W relative to an inner product $\langle,\rangle$, denoted by $W^\perp$, is defined as

$$W^\perp = \{\mathbf{v} \in \mathbb{R}^n \mid \langle \mathbf{v},\mathbf{w}\rangle = 0 \text{ for all } \mathbf{w} \in W\}.$$



Theorem 3.9. *Let W be a subspace of $\mathbb{R}^n$. Then $W^\perp$ is a subspace of $\mathbb{R}^n$ and*

$$\dim W + \dim W^\perp = n.$$

Proof. The fact that $W^\perp$ is a subspace is left as an exercise.

Let $\{\mathbf{w}_1, \dots, \mathbf{w}_k\}$ be an orthogonal basis for W, and $\{\mathbf{w}_{k+1}, \dots, \mathbf{w}_m\}$ be an orthogonal basis for $W^\perp$. Since each element of W is orthogonal to each element of $W^\perp$, $\mathcal{B} = \{\mathbf{w}_1, \dots, \mathbf{w}_m\}$ is an orthogonal set (of nonzero vectors) and thus it is linearly independent. It is left to prove $\mathcal{B}$ is generating. Let $\mathbf{v} \in \mathbb{R}^n$. Using a method similar to Gram-Schmidt process, we see that $\mathbf{x} = \mathbf{v} - \frac{\langle \mathbf{v}, \mathbf{w}_1 \rangle}{\langle \mathbf{w}_1, \mathbf{w}_1 \rangle}\mathbf{w}_1 - \cdots - \frac{\langle \mathbf{v}, \mathbf{w}_k \rangle}{\langle \mathbf{w}_k, \mathbf{w}_k \rangle}\mathbf{w}_k$ is orthogonal to $\mathbf{w}_1, \dots, \mathbf{w}_k$ and thus all elements of W, by linearity of inner product. Therefore, $\mathbf{x} \in W^\perp$. This implies there are scalars $c_{k+1}, \dots, c_m$ for which

$$\mathbf{x} = c_{k+1}\mathbf{w}_{k+1} + \cdots + c_m\mathbf{w}_m,$$

This means

$$\mathbf{v} = \frac{\langle \mathbf{v}, \mathbf{w}_1 \rangle}{\langle \mathbf{w}_1, \mathbf{w}_1 \rangle}\mathbf{w}_1 + \cdots + \frac{\langle \mathbf{v}, \mathbf{w}_k \rangle}{\langle \mathbf{w}_k, \mathbf{w}_k \rangle}\mathbf{w}_k + c_{k+1}\mathbf{w}_{k+1} + \cdots + c_m\mathbf{w}_m \in \operatorname{span} \mathcal{B}.$$

We showed $\mathcal{B}$ is a basis for $\mathbb{R}^n$. Since $\dim \mathbb{R}^n = n$ we conclude that $m = n$, as desired. □

3.4 Warm-ups

Example 3.12. Find the angle between vectors $(1, 0, -1)$ and $(2, 1, 2)$ in $\mathbb{R}^3$ using the standard inner product.

Solution. If the angle between these two vectors is θ, then we have

$$\cos\theta = \frac{(1, 0, -1) \cdot (2, 1, 2)}{||(1, 0, -1)||\ ||(2, 1, 2)||} = \frac{2 + 0 - 2}{\sqrt{1 + 0 + 1}\sqrt{4 + 1 + 4}} = 0 \Rightarrow \theta = \frac{\pi}{2}.$$

Therefore, the angle is $\pi/2$. □

3.5 More Examples

Example 3.13. Find a vector in $\mathbb{R}^3$ in the direction of $(1, -2, 2)$ that has length 4 with the Euclidean length.

Solution. The vector must be of the form $c(1, -2, 2)$ where c is a positive constant. The length must be four and thus $c^2 + 4c^2 + 4c^2 = 4^2$, which means $c = \frac{4}{3}$. The answer is $\left(\frac{4}{3}, \frac{-8}{3}, \frac{8}{3}\right)$.

Another method would be to notice that $||(1, -2, 2)|| = \sqrt{1 + 4 + 4} = 3$. Thus, by properties of length we have

$$\left|\left|\frac{4}{3}(1, -2, 2)\right|\right| = \frac{4}{3} \cdot 3 = 4.$$

This, of course, yields the same answer. □

Example 3.14. Suppose $\{\mathbf{v}, \mathbf{w}\}$ is a basis for a 2-dimensional subspace V of $\mathbb{R}^n$. Let a, b be two real numbers. Prove that $\{\mathbf{v} + a\mathbf{w}, \mathbf{v} + b\mathbf{w}\}$ is a basis for V if and only if $a \neq b$.

Solution. $\Rightarrow$: Suppose $\{\mathbf{v}+a\mathbf{w}, \mathbf{v}+b\mathbf{w}\}$ is a basis for V. Thus, $\mathbf{v}+a\mathbf{w}$ and $\mathbf{v}+b\mathbf{w}$ cannot be scalar multiples and thus $\mathbf{v}+a\mathbf{w} \neq \mathbf{v}+b\mathbf{w}$, which means $a \neq b$, as desired. (Since $a = b$ implies $\mathbf{v}+a\mathbf{w} = \mathbf{v}+b\mathbf{w}$.)

$\Leftarrow$: Now, assume $a \neq b$. We will show $\mathbf{v}+a\mathbf{w}, \mathbf{v}+b\mathbf{w}$ are linearly independent. Suppose $c_1(\mathbf{v}+a\mathbf{w}) + c_2(\mathbf{v}+b\mathbf{w}) = \mathbf{0}$, for some scalars c_1, c_2. This means $(c_1+c_2)\mathbf{v}_1 + (ac_1+bc_2)\mathbf{w} = \mathbf{0}$. Since $\mathbf{v}, \mathbf{w}$ are linearly independent we must have $c_1 + c_2 = ac_1 + bc_2 = 0$. Eliminating c_1 from the two equations we obtain $(b-a)c_2 = 0$, which implies $c_2 = 0$ since $a \neq b$. This implies $c_1 = 0$ from the equation $c_1 + c_2 = 0$. This means $\mathbf{v}+a\mathbf{w}, \mathbf{v}+b\mathbf{w}$ are linearly independent. Since the dimension of V is 2, $\{\mathbf{v}+a\mathbf{w}, \mathbf{v}+b\mathbf{w}\}$ is a basis for V. $\square$

Example 3.15. Prove that if $||\cdot||$ is a length relative to an inner product of $\mathbb{R}^n$ and $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$, then

$$||\mathbf{v}+\mathbf{w}||^2 + ||\mathbf{v}-\mathbf{w}||^2 = 2(||\mathbf{v}||^2 + ||\mathbf{w}||^2).$$

Solution. By definition we have $||\mathbf{v} \pm \mathbf{w}||^2 = \langle \mathbf{v} \pm \mathbf{w}, \mathbf{v} \pm \mathbf{w} \rangle$. By linearity and symmetry this simplifies to

$$\langle \mathbf{v} \pm \mathbf{w}, \mathbf{v} \pm \mathbf{w} \rangle = \langle \mathbf{v}, \mathbf{v} \rangle \pm \langle \mathbf{v}, \mathbf{w} \rangle \pm \langle \mathbf{w}, \mathbf{v} \rangle + \langle \mathbf{w}, \mathbf{w} \rangle = \langle \mathbf{v}, \mathbf{v} \rangle \pm 2\langle \mathbf{v}, \mathbf{w} \rangle + \langle \mathbf{w}, \mathbf{w} \rangle.$$

Summing the two together and using the fact that $\langle \mathbf{v}, \mathbf{v} \rangle = ||\mathbf{v}||^2$ and $\langle \mathbf{w}, \mathbf{w} \rangle = ||\mathbf{w}||^2$ we obtain the result. $\square$

Example 3.16. Consider the subspace V of $\mathbb{R}^4$ spanned by $\mathbf{v} = (1,1,1,0)$ and $\mathbf{w} = (0,1,2,0)$. Find a basis for the orthogonal complement of V relative to the dot product.

Solution. Note that since $\mathbf{v}$ and $\mathbf{w}$ are not multiples of each other, $\dim V = 2$. By Theorem 3.9, we have $\dim V^{\perp} = 4 - 2 = 2$.

We will find a basis for $\mathbb{R}^4$ containing $\mathbf{v}$ and $\mathbf{w}$. To do that, we will place these vectors in rows of a matrix, and row reduce the matrix as below:

$$\begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 2 & 0 \end{pmatrix}$$

This matrix is already in echelon form. By adding $\mathbf{e}_3$ and $\mathbf{e}_4$ to the rows of this matrix, we obtain a matrix in echelon form. Thus, $\mathbf{v}, \mathbf{w}, \mathbf{e}_3, \mathbf{e}_4$ form a basis for $\mathbb{R}^4$. Now, can apply the Gram-Schmidt process to $\mathbf{v}, \mathbf{w}, \mathbf{e}_3, \mathbf{e}_4$.

$$\begin{aligned}
\mathbf{w}_1 &= \mathbf{v} = (1,1,1,0) \\
\mathbf{w}_2 &= \mathbf{w} - \frac{\mathbf{w} \cdot \mathbf{w}_1}{\mathbf{w}_1 \cdot \mathbf{w}_1}\mathbf{w}_1 = (-1,0,1,0) \\
\mathbf{w}_3 &= \mathbf{e}_3 - \frac{\langle \mathbf{e}_3, \mathbf{w}_1 \rangle}{\langle \mathbf{w}_1, \mathbf{w}_1 \rangle}\mathbf{w}_1 - \frac{\langle \mathbf{e}_3, \mathbf{w}_2 \rangle}{\langle \mathbf{w}_2, \mathbf{w}_2 \rangle}\mathbf{w}_2 = (\frac{1}{6}, -\frac{1}{3}, \frac{1}{6}, 0) \\
\mathbf{w}_4 &= \mathbf{e}_4 - \frac{\langle \mathbf{e}_4, \mathbf{w}_1 \rangle}{\langle \mathbf{w}_1, \mathbf{w}_1 \rangle}\mathbf{w}_1 - \frac{\langle \mathbf{e}_4, \mathbf{w}_2 \rangle}{\langle \mathbf{w}_2, \mathbf{w}_2 \rangle}\mathbf{w}_2 - \frac{\langle \mathbf{e}_4, \mathbf{w}_3 \rangle}{\langle \mathbf{w}_3, \mathbf{w}_3 \rangle}\mathbf{w}_3 = \mathbf{e}_4
\end{aligned}$$

The vectors $\mathbf{w}_3, \mathbf{w}_4$ are linearly independent and are in $V^\perp$. Since $\dim V^\perp = 2$, the two vectors $\mathbf{w}_3$ and $\mathbf{w}_4$ form a basis for $V^\perp$. $\square$

Example 3.17. Let $c \in \mathbb{R}$ be a constant. For which constants c does the function below,

$$\langle (x_1, y_1), (x_2, y_2) \rangle = x_1 x_2 + c y_1 y_2$$

defines an inner product on $\mathbb{R}^2$?

Scratch: Positivity means, from $x^2 + cy^2 = 0$ we need to be able to imply $x = y = 0$. This means c cannot be nonpositive.

Solution. We claim that the given expression is an inner product if and only if c is positive.
If $c \leq 0$, then $\langle (0,1), (0,1) \rangle = c \leq 0$, violating the positivity property. Thus, it is not an inner product.

Now assume $c > 0$ and let $\mathbf{x} = (x_1, y_1), \mathbf{y} = (x_2, y_2), \mathbf{z} = (x_3, y_3) \in \mathbb{R}^2$, $a, b \in \mathbb{R}$. If $(x_1, y_1) \neq \mathbf{0}$, then $x_1^2 + cy_1^2 > 0$, since x_1^2 and y_1^2 are both nonnegative, c is positive and not both x_1 and y_1 are zero. This means we obtain the positivity.
$\langle \mathbf{x}, \mathbf{y} \rangle = x_1 x_2 + c y_1 y_2 = x_2 x_1 + c y_2 y_1 = \langle \mathbf{y}, \mathbf{x} \rangle$. This proves the symmetry.
$\langle a\mathbf{x} + b\mathbf{y}, \mathbf{z} \rangle = (ax_1 + bx_2)x_3 + c(ay_1 + by_2)y_3 = a(x_1 x_3 + c y_1 y_3) + b(x_2 x_3 + c y_2 y_3) = a\langle \mathbf{x}, \mathbf{z} \rangle + b\langle \mathbf{y}, \mathbf{z} \rangle$. This proves the linearity. Therefore, when c is positive, the given function defines an inner product. $\square$

Example 3.18. Suppose $\mathbf{v}_1, \mathbf{v}_2$ form an orthogonal basis for $\mathbb{R}^2$ with respect to some inner product. Prove that if $\mathbf{w}$ is orthogonal to $\mathbf{v}_1$, then $\mathbf{w} = c\mathbf{v}_2$ for some scalar c.

Solution. Since $\mathbf{v}_1, \mathbf{v}_2$ is a basis for $\mathbb{R}^2$, there are scalars c_1, c_2 for which $\mathbf{w} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2$. By assumption and linearity of inner product we obtain the following:

$$\langle \mathbf{w}, \mathbf{v}_1 \rangle = 0 \Rightarrow c_1 \langle \mathbf{v}_1, \mathbf{v}_1 \rangle + c_2 \langle \mathbf{v}_2, \mathbf{v}_1 \rangle = 0.$$

Since $\mathbf{v}_1$ and $\mathbf{v}_2$ are orthogonal we obtain $c_1 \langle \mathbf{v}_1, \mathbf{v}_1 \rangle = 0$. Since $\mathbf{v}_1$ is an element of a basis, we know $\mathbf{v}_1 \neq \mathbf{0}$, and by positivity of inner products we conclude that $c_1 = 0$, which means $\mathbf{w} = c_2\mathbf{v}_2$, as desired. $\square$

Example 3.19. Prove that for all real numbers $x_1, \ldots, x_n, y_1, \ldots, y_n$ we have

$$(x_1 y_1 + \cdots + x_n y_n)^2 \leq (x_1^2 + \cdots + x_n^2)(y_1^2 + \cdots + y_n^2).$$

Solution. We will use the Cauchy-Schwarz Inequality for the standard inner product of $\mathbb{R}^n$. Consider the two vectors

$$\mathbf{v} = (x_1, \ldots, x_n), \text{ and } \mathbf{w} = (y_1, \ldots, y_n) \text{ in } \mathbb{R}^n.$$

We have

$$\mathbf{v} \cdot \mathbf{w} = x_1 y_1 + \cdots + x_n y_n, ||\mathbf{v}|| = \sqrt{x_1^2 + \cdots + x_n^2}, \text{ and } ||\mathbf{w}|| = \sqrt{y_1^2 + \cdots + y_n^2}.$$

Applying the Cauchy-Schwarz Inequality we obtain:

$$|x_1y_1 + \cdots + x_ny_n| \leq \sqrt{x_1^2 + \cdots + x_n^2} \cdot \sqrt{y_1^2 + \cdots + y_n^2}.$$

Squaring both sides we obtain the result. □

Example 3.20. Find a basis for the orthogonal complement of $V = \operatorname{span}\{(1,2,-1),(0,1,1)\}$ under the standard inner product.

Solution. Placing these vectors into rows of a matrix we obtain the following matrix:

$$\begin{pmatrix} 1 & 2 & -1 \\ 0 & 1 & 1 \end{pmatrix}$$

This matrix is in echelon form and adding $(0,0,1)$ gives us another matrix in echelon form:

$$\begin{pmatrix} 1 & 2 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

This means the vectors $\mathbf{v}_1 = (1,2,-1), \mathbf{v}_2 = (0,1,1), \mathbf{v}_3 = (0,0,1)$ are linearly independent and thus they form a basis for $\mathbb{R}^3$. Since the dimension of V is 2, by Theorem 3.9 the dimension of its orthogonal complement is 1. Using the Gram-Schmidt process we will find the following vectors:

$$\mathbf{w}_1 = \mathbf{v}_1 = (1,2,-1),$$

$$\mathbf{w}_2 = \mathbf{v}_2 - \frac{\mathbf{v}_2 \cdot \mathbf{w}_1}{\mathbf{w}_1 \cdot \mathbf{w}_1}\mathbf{w}_1 = \left(-\frac{1}{6}, \frac{2}{3}, \frac{7}{6}\right) = \frac{1}{6}(-1,4,7),$$

$$\mathbf{w}_3 = \mathbf{v}_3 - \frac{\mathbf{v}_3 \cdot \mathbf{w}_1}{\mathbf{w}_1 \cdot \mathbf{w}_1}\mathbf{w}_1 - \frac{\mathbf{v}_3 \cdot \mathbf{w}_2}{\mathbf{w}_2 \cdot \mathbf{w}_2}\mathbf{w}_2 = \left(\frac{3}{11}, -\frac{1}{11}, \frac{1}{11}\right) = \frac{1}{11}(3,-1,1)$$

Thus, $\mathbf{w}_3$ is orthogonal to every element of V. Therefore, $\mathbf{w}_3$ forms a basis for the orthoghonal complement of V. □

3.6 Exercises

Exercise 3.1. *Determine which of the following vectors form a basis for the appropriate* $\mathbb{R}^n$.

(a) $(1,2),(-2,1)$.

(b) $(1,0,1),(1,1,2),(-1,-2,-3)$.

(c) $(1,0),(2,3),(1,1)$.

(d) $(1,0,0),(0,1,1),(0,1,2)$.

Exercise 3.2. *Suppose the homogeneous system*

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = 0 \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n = 0 \end{cases}$$

has only the trivial solution. Prove that for every $b_1, b_2, \ldots, b_n \in \mathbb{R}$*, the system*

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n = b_n \end{cases}$$

has a unique solution.

Hint: Use $\dim \mathbb{R}^n = n$, and consider the vectors $(a_{11}, a_{21}, \ldots, a_{n1}), \ldots, (a_{1n}, a_{2n}, \ldots, a_{nn})$.

Exercise 3.3. *Determine the dimension of each vector space.*

(a) The subspace of $\mathbb{R}^3$ *generated by vectors* $(1, 2, -1), (2, 3, 4)$*, and* $(4, 10, 2)$*.*

(b) The subspace of $\mathbb{R}^3$ *generated by* $(1, 2, 0), (-1, 1, 1)$*, and* $(1, 5, 1)$*.*

Exercise 3.4. *Let* V *be a subspace of* $\mathbb{R}^n$*. Prove that if* $\mathcal{A} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ *is a linearly independent set of vectors in* V*, then there is a basis for* V *that contains* $\mathcal{A}$*.*

Hint: Consider the subspace generated by $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$. If this subspace is not V, and then add an element $\mathbf{v}_{k+1}$ from V, but outside of span $\mathcal{A}$ to the set $\mathcal{A}$. Show this new larger set is linearly independent. Repeat this until you get a basis. You must show this process ends. This is where you should use the fact that every $n + 1$ vectors in $\mathbb{R}^n$ are linearly dependent.

Exercise 3.5. *Suppose* W *and* V *are subspaces of* $\mathbb{R}^n$ *for which* $W \subseteq V$*. Prove that if* $\dim W = \dim V$*, then* $W = V$*.*

Exercise 3.6. *Let* V *be a subspace of* $\mathbb{R}^n$*. Prove that if* $\mathcal{A}$ *is a spanning subset of* V*, then there is a basis for* V *that is a subset of* $\mathcal{A}$*.*

Exercise 3.7. *Find the dimension of the vector space spanned by* $(0, 0, 1, 1), (1, 1, 0, 0)$*, and* $(1, 1, 0, 1)$*.*

Exercise 3.8. *Find the angle between:*

(a) $(1, 2, -1)$ *and* $(0, 2, -1)$ *in* $\mathbb{R}^3$ *with the standard inner product.*

(b) $(1, 1, 5)$ *and* $(1, -1, 0)$ *with the inner product given by* $\langle (x_1, y_1, z_1), (x_2, y_2, z_2) \rangle = x_1x_2 + 2y_1y_2 + 3z_1z_2$*.*

Exercise 3.9. *Determine if the triangle whose vertices are* $A = (1, 2, 2), B = (-1, 1, 0), C = (2, -2, 1)$ *is a right triangle.*

Exercise 3.10. *Consider $\mathbb{R}^3$ with the standard inner product. Find an orthogonal basis for $\mathbb{R}^3$ for which one of the elements of this basis is $(1, 2, -1)$.*

Hint: Use the idea of echelon form to extend this vector to a basis. Then apply Gram-Schmidt. See Example 3.20.

Exercise 3.11. *Find all real numbers c for which the vectors $(1, 2, c)$ and $(-1, -c, c+1)$ are orthogonal with respect to the standard inner product. For this value of c, give an example of an inner product where these two vectors are not orthogonal.*

Exercise 3.12. *Find all inner products of $\mathbb{R}^2$, if any, for which $||\mathbf{e}_1|| = 4$, $||\mathbf{e}_2|| = 3$ and the angle between $\mathbf{e}_1$ and $\mathbf{e}_2$ is $\frac{\pi}{3}$.*

Hint: First, use the given assumptions to find $\langle \mathbf{e}_1, \mathbf{e}_2 \rangle$. Next, write (x_1, x_2) and (y_1, y_2) as linear combinations of $\mathbf{e}_1, \mathbf{e}_2$. Then use linearity, symmetry and the given assumptions to evaluate $\langle (x_1, x_2), (y_1, y_2) \rangle$. Finally, show the result is in fact an inner product.

Exercise 3.13. *Suppose $c_1, \ldots, c_n$ are real constants. Define a function $\langle , \rangle$ by*

$$\langle (x_1, \ldots, x_n), (y_1, \ldots, y_n) \rangle = \sum_{j=1}^{n} c_j x_j y_j.$$

(a) Show $\langle , \rangle$ is linear and symmertic.

(b) Show $\langle , \rangle$ is an inner product iff $c_1, \ldots, c_n$ are all positive.

Exercise 3.14. *Suppose $\langle , \rangle$ is an inner product of $\mathbb{R}^n$. Let $\mathbf{v} = (x_1, \ldots, x_n)$ and $\mathbf{w} = (y_1, \ldots, y_n)$ be two vectors in $\mathbb{R}^n$.*

(a) By writing $\mathbf{v}$ and $\mathbf{w}$ as linear combinations of $\mathbf{e}_1, \ldots, \mathbf{e}_n$, and applying linearity prove that $\langle \mathbf{v}, \mathbf{w} \rangle = \sum_{k=1}^{n} \sum_{j=1}^{n} \langle \mathbf{e}_k, \mathbf{e}_j \rangle x_k y_j$.

(b) Using the previous part and Exercise 3.13, deduce that every inner product of $\mathbb{R}^n$ for which $\mathbf{e}_1, \ldots, \mathbf{e}_n$ are orthogonal is of the form $\langle \mathbf{v}, \mathbf{w} \rangle = \sum_{j=1}^{n} c_j x_j y_j$ for some positive real numbers $c_1, \ldots, c_n$, and every such function is an inner product.

Exercise 3.15. *Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \in \mathbb{R}^n$ be orthogonal vectors with respect to some inner product of $\mathbb{R}^n$. Prove that*

$$||\mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_n||^2 = ||\mathbf{v}_1||^2 + ||\mathbf{v}_2||^2 + \cdots + ||\mathbf{v}_n||^2.$$

Hint: Use the Pythagorean Theorem and proof by induction.

Exercise 3.16. *Suppose $\langle , \rangle$ is an inner product of $\mathbb{R}^n$. Using linearity prove that for every $\mathbf{w} \in \mathbb{R}^n$ we have $\langle \mathbf{0}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{0} \rangle = 0$. Deduce the Cauchy-Schwarz inequality in the case when $\mathbf{v} = \mathbf{0}$. (In class we assumed $\mathbf{v} \neq \mathbf{0}$.)*

Exercise 3.17. *Let $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ be a basis for $\mathbb{R}^n$. For every two vectors*

$$\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n, \text{ and } \mathbf{w} = b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + \cdots + b_n\mathbf{v}_n \text{ in } \mathbb{R}^n,$$

define $\langle \mathbf{v}, \mathbf{w} \rangle = a_1b_1 + a_2b_2 + \cdots + a_nb_n$. Prove that this defines an inner product on $\mathbb{R}^n$.

Exercise 3.18. *Prove the converse of the Pythagorean Theorem stated below:*

Given an inner product $\langle , \rangle$ of $\mathbb{R}^n$ and $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$, if $||\mathbf{v} + \mathbf{w}||^2 = ||\mathbf{v}||^2 + ||\mathbf{w}||^2$, then $\langle \mathbf{v}, \mathbf{w} \rangle = 0$.

Exercise 3.19. *Let $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$, and $\langle , \rangle$ be an inner product of $\mathbb{R}^n$. Prove that $|\langle \mathbf{v}, \mathbf{w} \rangle| = ||\mathbf{v}||\ ||\mathbf{w}||$ if and only if $\mathbf{w}$ is a scalar multiple of $\mathbf{v}$ or $\mathbf{v} = \mathbf{0}$.*

Hint: Follow the proof of Cauchy-Schwarz inequality and see when the equality holds.

Exercise 3.20. *Let θ be the angle between two nonzero vectors $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$. Prove that*

(a) if $\theta = 0$, then $\mathbf{v} = c\mathbf{w}$ for some positive real number c.

(b) if $\theta = \pi$, then $\mathbf{v} = c\mathbf{w}$ for some negative real number c.

Exercise 3.21. *Let A be an $m \times n$ matrix with real entries. We have shown that $\text{Row}(A)$ and $\text{Ker } A$ are both subspaces of $\mathbb{R}^n$. What is the relationship between $\text{Ker } A$ and $(\text{Row}(A))^\perp$? Justify your answer.*

Hint: Show that a vector is in $(\text{Row}(A))^\perp$ if and only if it is orthogonal to all rows of A.

Exercise 3.22. *Let S be a nonempty subset of $\mathbb{R}^n$, and $\langle , \rangle$ be an inner product of $\mathbb{R}^n$. Prove that $S^\perp$ defined by*

$$S^\perp = \{\mathbf{v} \in V \mid \langle \mathbf{v}, \mathbf{s} \rangle = 0 \text{ for all } \mathbf{s} \in S\}$$

is a subspace of $\mathbb{R}^n$.

Exercise 3.23. *Suppose W is a subspace of $\mathbb{R}^n$. Prove that $(W^\perp)^\perp = W$.*

Hint: Show the dimension of both sides are the same, and the right hand side is a subset of the left hand side.

Exercise 3.24. *Suppose $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ satisfy $\mathbf{v} \cdot \mathbf{w} = 0$. Prove that $||\mathbf{v} + \mathbf{w}|| = ||\mathbf{v} - \mathbf{w}||$ once using properties of inner product, once using the definition of dot product, and once using geometry. (For the geometric proof, you may use known facts from Euclidean geometry.)*

Exercise 3.25. *For every inner product $\langle , \rangle$ on $\mathbb{R}^n$, its corresponding length, and every two vectors $\mathbf{u}, \mathbf{v}$ prove the polarization identity stated below:*

$$\langle \mathbf{u}, \mathbf{v} \rangle = \frac{1}{2}\left(||\mathbf{u} + \mathbf{v}||^2 - ||\mathbf{u} - \mathbf{v}||^2\right).$$

Exercise 3.26. *Suppose $\mathbf{u}, \mathbf{v}$ are nonzero vectors in $\mathbb{R}^n$ and c is a nonzero scalar. Suppose the angle between $\mathbf{u}$ and $\mathbf{v}$ is α and the angle between $c\mathbf{u}$ and $\mathbf{v}$ is β. Prove that:*

(a) if $c > 0$, then $\alpha = \beta$.

(b) if $c < 0$, then $\alpha + \beta = \pi$.

Exercise 3.27. *Suppose V, W are subspaces of $\mathbb{R}^m$ and $\mathbb{R}^n$, respectively. Prove that $V \times W$ is a subspace of $\mathbb{R}^{m+n}$ and that*

$$\dim(V \times W) = \dim V + \dim W.$$

Note that vectors in $V \times W$ are obtained by placing coordinates of each vector of V followed by coordinates of each vector of W. In other words, in order to obtain a vector in R^{m+n} we drop the brackets for vectors in V and W.

Definition 3.10. Given two points $A, B \in \mathbb{R}^n$, the **vector** $\overrightarrow{AB}$ is defined as $B - A$. The **line segment** AB is the set of all points $X \in \mathbb{R}^n$ for which $X = tA + (1-t)B$ for some $t \in [0, 1]$. The set $AB \cup BC \cup CA$, denoted by ABC, is called a **triangle** if $\overrightarrow{AB}$ and $\overrightarrow{AC}$ are linearly independent vectors. The three segments AB, BC and CA are called sides of triangle ABC. A **side length** in triangle ABC is the length of each of the three vectors $\overrightarrow{AB}$, $\overrightarrow{AC}$ and $\overrightarrow{BC}$. The **angle** between vectors $\overrightarrow{AB}$ and $\overrightarrow{AC}$ is denoted by $\angle BAC$ or $\angle CAB$. Similarly $\angle ABC, \angle CBA, \angle ACB, \angle BCA$ are all defined. Note that to define this angle $\mathbb{R}^n$ must be equipped with an inner product. Each of the three points A, B, C is called a **vertex** of triangle ABC. A triangle ABC is called a **right triangle** if one is its angles is $\dfrac{\pi}{2}$. A triangle is called **isosceles** if two of its side lengths are equal. A triangle is called **equilateral** if all three side lengths are equal.

Exercise 3.28. *Prove that for every three points $A, B, C \in \mathbb{R}^n$ we have $\overrightarrow{AB} + \overrightarrow{BC} = \overrightarrow{AC}$.*

Exercise 3.29. *Prove that if ABC is a triangle, then vectors $\overrightarrow{AB}, \overrightarrow{BC}, \overrightarrow{AC}$ are pairwise linearly independent, but all three of them are linearly dependent.*

Exercise 3.30. *Prove the Law of Cosines:*

> *Consider $\mathbb{R}^n$ equipped with an inner product. Suppose ABC is a triangle in $\mathbb{R}^n$. Then,*
> $$||\overrightarrow{BC}||^2 = ||\overrightarrow{AB}||^2 + ||\overrightarrow{AC}||^2 - 2||\overrightarrow{AB}||\ ||\overrightarrow{AC}|| \cos(\angle BAC).$$

Exercise 3.31. *Prove the Law of Sines:*

> *Consider $\mathbb{R}^n$ equipped with an inner product. Suppose ABC is a triangle in $\mathbb{R}^n$. Then,*
> $$\frac{||\overrightarrow{AB}||}{\sin(\angle ACB)} = \frac{||\overrightarrow{AC}||}{\sin(\angle ABC)}.$$

Exercise 3.32. *Assume $\mathbb{R}^n$ is equipped with an inner product. Show that a triangle ABC in $\mathbb{R}^n$ is isosceles if and only if two of its angles are congruent.*

Exercise 3.33. *Assume $\mathbb{R}^n$ is equipped with an inner product. Show that a triangle ABC in $\mathbb{R}^n$ is equilateral if and only if all of its angles are $\dfrac{\pi}{3}$.*

3.7 Challenge Problems

Exercise 3.34. *Let k be a positive integer. Find the smallest positive integer n for which there are k nonzero vectors $v_1, v_2, \ldots, v_k \in \mathbb{R}^n$ for which v_i and v_j are orthogonal for every i and j for which $i > j+1$.*

Exercise 3.35. *Prove the Angle Bisector Theorem:*

Consider $\mathbb{R}^n$ equipped with an inner product. Suppose ABC is a triangle in $\mathbb{R}^n$, and D is a point on side BC for which $\angle BAD = \angle DAC$. Then, $\dfrac{||\overrightarrow{AB}||}{||\overrightarrow{AC}||} = \dfrac{||\overrightarrow{BD}||}{||\overrightarrow{DC}||}$.

Exercise 3.36. *Consider $\mathbb{R}^n$ equipped with an inner product. Prove the altitudes of ABC are concurrent. In other words, prove there is a unique point $H \in \mathbb{R}^n$ for which:*

$$\langle \overrightarrow{AH}, \overrightarrow{BC} \rangle = \langle \overrightarrow{BH}, \overrightarrow{AC} \rangle = \langle \overrightarrow{CH}, \overrightarrow{AB} \rangle = 0.$$

Exercise 3.37. *Consider $\mathbb{R}^n$ equipped with an inner product. Assume D is a point on the segment BC of triangle ABC. Prove $\angle BAC = \angle BAD + \angle DAC$.*

Exercise 3.38. *Consider $\mathbb{R}^n$ equipped with an inner product. Prove that the sum of the three angles of every triangle is π.*

3.8 Summary

- The number of pivot entries is the same as both the dimension of row space and the dimension of column space.
- To find a basis for a space spanned by a set of vectors in $\mathbb{R}^n$:
  - Place these vectors in rows of a matrix.
  - Row reduce this matrix.
  - The nonzero rows of the echelon form, create a basis for the desired space.
- The dimension of a vector space is the number of vectors in a basis of that vector space.
- In a vector space of dimension n every $n+1$ (or more) vectors are linearly dependent.
- Rank of a matrix is the dimension of its column space.
- To show $\mathbf{v}_1, \ldots, \mathbf{v}_n$ form a basis for a vector space V we can do one of the following:
  - $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly independent and spanning.
  - $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly independent and $\dim V = n$.
  - $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are spanning and $\dim V = n$.

- To check if $\langle \mathbf{v}, \mathbf{w} \rangle$ is an inner product we need to check if it satisfies three properties: Positivity, Symmetry, and Linearity.
- The angle θ between two vectors $\mathbf{v}, \mathbf{w}$ is given by $\cos\theta = \dfrac{\langle \mathbf{v}, \mathbf{w} \rangle}{||\mathbf{v}||\ ||\mathbf{w}||}$.
- If the angle between two vectors is $\pi/2$ we say the two vectors are orthogonal.
- Pythagorean Theorem: If $\mathbf{v}$ and $\mathbf{w}$ are orthogonal, then $||\mathbf{v}+\mathbf{w}||^2 = ||\mathbf{v}||^2 + ||\mathbf{w}||^2$.
- Cauchy-Schwarz Inequality: In any inner product space $|\langle \mathbf{v}, \mathbf{w} \rangle| \leq ||\mathbf{v}||\ ||\mathbf{w}||$.
- Given linearly independent vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ in an inner product space, we can use the Gram-Schmidt process to find an orthogonal basis for the subspace spanned by $\mathbf{v}_i$'s.
- The orthogonal complement of a subspace W is the subspace consisting of all vectors that are orthogonal to all vectors in W.
- To find a basis for $W^\perp$:
  - First find a basis for W.
  - Extend that basis to a basis of $\mathbb{R}^n$ using echelon form.
  - Start from vectors in W and apply the Gram-Schmidt process. This produces an orthogonal basis for W followed by an orthogonal basis for $W^\perp$.

Week 4

4.1 Norms

YouTube Video: https://youtu.be/UgwMfs9w91s

The definition of length, $||\mathbf{v}|| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$, in the previous section relied on an inner product of $\mathbb{R}^n$, however the concept of "length" can be defined independently. We will define a "norm" (or length) to be an assignment of nonnegative real numbers to vectors that satisfies certain properties that we expect from a geometric length.

Definition 4.1. A **norm** on $\mathbb{R}^n$ is a function that assigns to any vector $\mathbf{v} \in \mathbb{R}^n$ a nonnegative real number $||\mathbf{v}||$ that satisfies all of the following:

(a) $||\mathbf{v}|| > 0$ for every nonzero $\mathbf{v} \in \mathbb{R}^n$ (Positivity),

(b) $||\mathbf{v} + \mathbf{w}|| \leq ||\mathbf{v}|| + ||\mathbf{w}||$ for every $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ (Triangle Inequality), and

(c) $||c\mathbf{v}|| = |c| \; ||\mathbf{v}||$ for every $\mathbf{v} \in \mathbb{R}^n$ and $c \in \mathbb{R}$ (Homogeneity).

The following theorem connects the two notions of inner product and norm.

Theorem 4.1. *If $\langle , \rangle$ is an inner product on $\mathbb{R}^n$, then the function defined by $||\mathbf{v}|| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$ is a norm.*

Example 4.1. The following are examples of norms on $\mathbb{R}^n$.

(a) $||(x_1, x_2, \ldots, x_n)|| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$.

(b) $||(x_1, x_2, \ldots, x_n)|| = \max\{|x_1|, |x_2|, \ldots, |x_n|\}$.

Remark. When a particular norm of $\mathbb{R}^n$ is not specified we use the Euclidean norm, stated in the example above.

4.2 Linear Transformations and Matrices

YouTube Video: https://youtu.be/WA12KFfJ-Uk

Remark. All vector spaces are subspaces of $\mathbb{R}^n$ for some positive integer n.

Definition 4.2. Let V, W be two vector spaces. (i.e. V is a subspace of $\mathbb{R}^m$ and W is a subspace of $\mathbb{R}^n$ for some positive integers m, n.) A function $L : V \to W$ is said to be **linear** (or a **linear transformation**) if for all $\mathbf{v}, \mathbf{w} \in V$ and $c \in \mathbb{R}$,

- $L(\mathbf{v} + \mathbf{w}) = L(\mathbf{v}) + L(\mathbf{w})$ (Additivity), and
- $L(c\mathbf{v}) = cL(\mathbf{v})$ (Homogeneity)

Example 4.2. Determine which of the following functions are linear:

(a) $f : \mathbb{R} \to \mathbb{R}, f(x) = cx$, where c is a constant.

(b) $g : \mathbb{R}^2 \to \mathbb{R}, g(x, y) = 2x + 3y$.

(c) $h : \mathbb{R} \to \mathbb{R}, h(x) = 2x + 3$.

(d) $k : \mathbb{R}^n \to \mathbb{R}$, $k(\mathbf{v}) = \langle \mathbf{w}, \mathbf{v} \rangle$, where $\mathbf{w}$ is a fixed vector and $\langle , \rangle$ in an inner product of $\mathbb{R}^n$.

Theorem 4.2. *Let $L : V \to W$ be a function between vector spaces. Then, the following are equivalent.*

(a) L is linear.

(b) $L(\mathbf{u} + c\mathbf{v}) = L(\mathbf{u}) + cL(\mathbf{v})$ for all $\mathbf{u}, \mathbf{v} \in V$ and all $c \in \mathbb{R}$.

(c) $L(a\mathbf{u} + b\mathbf{v}) = aL(\mathbf{u}) + bL(\mathbf{v})$ for all $\mathbf{u}, \mathbf{v} \in V$ and all $a, b \in \mathbb{R}$.

Example 4.3. Prove that all linear transformations $f : \mathbb{R}^n \to \mathbb{R}$ are given by $f(\mathbf{v}) = \mathbf{w} \cdot \mathbf{v}$, where $\mathbf{w}$ is a fixed vector in $\mathbb{R}^n$.

Example 4.4. Identify all linear transformations $f : \mathbb{R}^3 \to \mathbb{R}^2$.

Example 4.5. Prove that all linear transformations $f : \mathbb{R}^n \to \mathbb{R}^m$ are given by

$$f(\mathbf{v}) = \begin{pmatrix} \mathbf{w}_1 \cdot \mathbf{v} \\ \vdots \\ \mathbf{w}_m \cdot \mathbf{v} \end{pmatrix},$$

where $\mathbf{w}_1, \ldots, \mathbf{w}_m \in \mathbb{R}^n$ are fixed vectors.

Definition 4.3. Given an $m \times n$ matrix

$$A = \begin{pmatrix} \mathbf{w}_1 \\ \vdots \\ \mathbf{w}_m \end{pmatrix},$$

where $\mathbf{w}_j$'s are rows of A, and given a column vector $\mathbf{v} \in \mathbb{R}^n$. The product of A and $\mathbf{v}$, denoted by $A\mathbf{v}$, is defined by:

$$A\mathbf{v} = \begin{pmatrix} \mathbf{w}_1 \cdot \mathbf{v} \\ \vdots \\ \mathbf{w}_m \cdot \mathbf{v} \end{pmatrix}.$$

YouTube Video: https://youtu.be/SIClX9f9Bp0

Theorem 4.3. *A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is linear if and only if there is an $m \times n$ matrix A for which $f(\mathbf{v}) = A\mathbf{v}$. Furthermore, for every given linear transformation f the matrix A is unique, and the columns of A are given by $f(\mathbf{e}_1), \ldots, f(\mathbf{e}_n)$. In other words,*

$$A = \left(f(\mathbf{e}_1) \cdots f(\mathbf{e}_n)\right).$$

Definition 4.4. The matrix A of the linear transformation f in theorem above is called the matrix of f with respect to the standard basis and is denoted by M_f.

Example 4.6. Let α be an angle. Suppose $R_\alpha : \mathbb{R}^2 \to \mathbb{R}^2$ is the rotation with angle α about the origin. From geometry we know R_α is linear. Find the matrix of R_α with respect to the standard basis.

Remark. The set of $m \times n$ matrices with real entries is denoted by $M_{m\times n}(\mathbb{R})$. The set of (square) $n \times n$ matrices with real entries is denoted by $M_n(\mathbb{R})$.

Definition 4.5. Let $A \in M_{m\times n}(\mathbb{R})$ and $B \in M_{n\times k}(\mathbb{R})$ matrix. The matrix AB is an $m \times k$ matrix whose j-th column is obtained from multiplying A by the j-th column of B. In other words, the (i, j) entry of AB is obtained by finding the dot product of the i-th row of A with the j-th column of B.

Remark. Note that to be able to evaluate the multiplication AB of two matrices A and B, we need the number of columns of A to be the same as the number of rows of B.

Example 4.7. Evaluate the matrices AB and BA, where

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 1 \end{pmatrix}, \text{ and } B = \begin{pmatrix} 0 & -1 \\ 2 & 3 \\ 5 & -1 \end{pmatrix}.$$

Example 4.8. Consider a 2×3 matrix A and a vector $\mathbf{v}$ as follows:

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix}, \text{ and } \mathbf{v} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

Show that $A\mathbf{v}$ is the following linear combination of columns of A:

$$A\mathbf{v} = x_1 \begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ a_{22} \end{pmatrix} + x_3 \begin{pmatrix} a_{13} \\ a_{23} \end{pmatrix}.$$

Remark. For every $m \times n$ matrix A and every column vector $\mathbf{v} \in \mathbb{R}^n$ the vector $A\mathbf{v}$ is a linear combination of columns of A with coefficients from entries of $\mathbf{v}$.

Theorem 4.4. *If the functions $f : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^m \to \mathbb{R}^k$ are linear, then $g \circ f : \mathbb{R}^n \to \mathbb{R}^k$ is also linear and $M_{g\circ f} = M_g M_f$.*

Proof. The part that $g \circ f$ is linear is left as an exercise. We know the j-th column of the matrix of $g \circ f$ is $g \circ f(\mathbf{e}_j)$. This equals $g(f(\mathbf{e}_j)) = M_g f(\mathbf{e}_j)$. Since the j-th column of M_f is $f(\mathbf{e}_j)$, the j-th column of $M_g M_f$ is precisely $M_g f(\mathbf{e}_j)$. Therefore, the j-th column of $M_g M_f$ is precisely $g \circ f(\mathbf{e}_j)$. Therefore, the matrix of $M_{g\circ f}$ in standard basis is $M_g M_f$, as desired. $\square$

Example 4.9. The matrix of the identity function $I : \mathbb{R}^n \to \mathbb{R}^n$ defined by $I(\mathbf{x}) = \mathbf{x}$ is given by

$$\begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}.$$

The matrix above is called the **identity** matrix of size n and is denoted by I_n.

Definition 4.6. The **sum** of two matrices A, B is defined precisely when they have the same size. In which case, the addition is entry-wise. In other words, the (i, j) entry of $A + B$ is the sum of (i, j) entries of A and B. Any $m \times n$ matrix A can be multiplied by any scalar c. The result is also an $m \times n$ matrix, whose (i, j) entry is c times the (i, j) entry of A for every i, j.

Theorem 4.5. *For matrices A, B, C and a real number r we have the following:*

(a) $A(BC) = (AB)C$. (Associativity).

(b) $A(B + C) = AB + AC$, and $(B + C)A = BA + CA$. (Distributivity).

(c) $AI_n = A$, and $I_m A = A$. (Multiplicative Identity)

(d) $r(AB) = (rA)B = A(rB)$.

Provided that in each case the appropriate multiplication or addition is defined.

Note that in general for two matrices A, B the products AB and BA are not equal, even if both of them are defined.

Definition 4.7. We say two matrices A, B **commute** if $AB = BA$.

4.3 Kernel and Image

YouTube Video: https://youtu.be/pmU5t_1G1g8

Definition 4.8. Given a linear transformation $L : V \to W$, the **kernel** of L is defined as Ker $L = L^{-1}(\{\mathbf{0}\})$. In other words,

$$\text{Ker } L = \{\mathbf{v} \in V \mid L(\mathbf{v}) = \mathbf{0}\}.$$

The **image** (or range) of L is defined as

$$\text{Im } L = \{\mathbf{w} \in W \mid \mathbf{w} = L(\mathbf{v}) \text{ for some } \mathbf{v} \in V\}.$$

Theorem 4.6. *Let $L : V \to W$ be a linear transformation of vector spaces. Then* Ker L *is a subspace of* V *and* Im L *is a subspace of* W*.*

Proof. We will use the subspace criterion for both.

First, note that $L(\mathbf{0}) = L(0\mathbf{0}) = 0L(\mathbf{0}) = \mathbf{0}$, by homogeneity. Thus, $L(\mathbf{0}) = \mathbf{0}$. Therefore, $\mathbf{0} \in \text{Ker } L$. Now, assume $\mathbf{x}, \mathbf{y} \in \text{Ker } L$, and $c \in \mathbb{R}$. By definition, $L(\mathbf{x}) = L(\mathbf{y}) = \mathbf{0}$. By linearity we have

$$L(\mathbf{x}+\mathbf{y}) = L(\mathbf{x}) + L(\mathbf{y}) = \mathbf{0}+\mathbf{0} = \mathbf{0}, \text{ and } L(c\mathbf{x}) = cL(\mathbf{x}) = c\mathbf{0} = \mathbf{0} \Rightarrow \mathbf{x}+\mathbf{y}, c\mathbf{x} \in \text{Ker } L.$$

Therefore, Ker L is a subspace of V.

Since $L(\mathbf{0}) = \mathbf{0}$, the zero vector of W is in Im L. Assume $\mathbf{v}, \mathbf{w} \in \text{Im } L$. Thus, there are vectors $\mathbf{x}, \mathbf{y} \in V$ for which $\mathbf{v} = L(\mathbf{x})$ and $\mathbf{w} = L(\mathbf{y})$. Given a scalar $c \in \mathbb{R}$ we have

$$\mathbf{v}+\mathbf{w} = L(\mathbf{x}) + L(\mathbf{y}) = L(\mathbf{x}+\mathbf{y}) \in \text{Im } L, \text{ and } c\mathbf{v} = cL(\mathbf{x}) = L(c\mathbf{x}) \in \text{Im } L.$$

Therefore, Im L is a subspace of W. □

The last part of the theorem above can be generalized as follows:

Example 4.10. Find the kernel and image of the linear transformation $L : \mathbb{R}^3 \to \mathbb{R}^2$ given by

$$L(x, y, z) = (x + 2y + z, 2x - y - z).$$

Theorem 4.7. *Suppose $L : V \to W$ is a linear transformation of vector spaces. Then,* Ker $L = \{\mathbf{0}\}$ *if and only if L is one-to-one. If L is one-to-one, then* $\dim \text{Im } L = \dim V$*.*

Theorem 4.8. *Suppose $L : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation whose matrix in the standard basis is A. Then,*

(a) Im $L = \text{Col}(A)$.

(b) Ker $L = (\text{Row}(A))^{\perp}$.

(c) $\dim \text{Ker } L + \dim \text{Im } L = n$.

Theorem 4.9 (Rank-Nullity Theorem)**.** *Let V and W be vector spaces, and $L : V \to W$ be a linear transformation. Then,*

$$\dim \text{Ker } L + \dim \text{Im } L = \dim V.$$

4.4 More Examples

Example 4.11. Prove that if $||\cdot||$ is a norm on $\mathbb{R}^n$, then $||\mathbf{0}|| = 0$.

Solution. By homogeneity $||0\mathbf{0}|| = |0|\ ||\mathbf{0}|| = 0||\mathbf{0}|| = 0$. Since $0\mathbf{0} = \mathbf{0}$, we obtain $||\mathbf{0}|| = 0$, as desired. □

Example 4.12. Prove that if $\mathbf{v}_1, \ldots, \mathbf{v}_m$ are vectors in $\mathbb{R}^n$ with a norm $||\cdot||$, then

$$||\mathbf{v}_1 + \cdots + \mathbf{v}_m|| \leq ||\mathbf{v}_1|| + \cdots + ||\mathbf{v}_m||.$$

Solution. We will prove this by induction on m.

Basis step: For $m = 1$ both sides of the inequality are $||\mathbf{v}_1||$. This proves the basis step.

Inductive Step: Let $\mathbf{v}_1, \ldots, \mathbf{v}_{m+1}$ be vectors in $\mathbb{R}^n$. Suppose

$$||\mathbf{v}_1 + \cdots + \mathbf{v}_m|| \leq ||\mathbf{v}_1|| + \cdots + ||\mathbf{v}_m|| \qquad (*)$$

By the Triangle Inequality we obtain:

$$||\mathbf{v}_1 + \cdots + \mathbf{v}_{m+1}|| \leq ||\mathbf{v}_1 + \cdots + \mathbf{v}_m|| + ||\mathbf{v}_{m+1}||.$$

Combining this with $(*)$ completes the inductive step. $\square$

Example 4.13. Find Ker L, and Im L, where $L : \mathbb{R}^3 \to \mathbb{R}^2$ is defined by

$$L(\mathbf{v}) = \begin{pmatrix} 1 & 2 & 3 \\ -1 & -1 & 0 \end{pmatrix} \mathbf{v}.$$

Solution. We row reduce the given matrix to obtain.

$$\begin{pmatrix} 1 & 0 & -3 \\ 0 & 1 & 3 \end{pmatrix}$$

This means the first two columns of the original matrix are linearly independent. Therefore, the image is 2-dimensional. Since the image is a subspace of $\mathbb{R}^2$, by Exercise 3.5 the image must be equal to $\mathbb{R}^2$.

For the kernel, we must solve the system

$$\begin{cases} x - 3z &= 0 \\ y + 3z &= 0 \end{cases}$$

This gives us $x = 3z$, and $y = -3z$. Therefore,

$$\text{Ker } L = \{(3z, -3z, z) \mid z \in \mathbb{R}\} = \text{span } \{(3, -3, 1)\}.$$

$\square$

Example 4.14. Find the kernel and image of the rotation function $R_\alpha : \mathbb{R}^2 \to \mathbb{R}^2$.

Solution. For kernel, suppose $R_\alpha(\mathbf{v}) = \mathbf{0}$ for some $\mathbf{v} \in \mathbb{R}^2$. But that means if we rotate $\mathbf{0}$ with angle $-\alpha$ we should obtain the vector $\mathbf{v}$, and thus $\mathbf{v} = \mathbf{0}$. The zero vector is in Ker R_α. Therefore, Ker $R_\alpha = \{\mathbf{0}\}$.

By the Rank-Nullity Theorem $\dim \text{Ker } R_\alpha + \dim \text{Im } R_\alpha = 2$. Thus $\dim \text{Im } R_\alpha = 2$, and since Im R_α is a subspace of $\mathbb{R}^2$ and $\dim \mathbb{R}^2 = 2$, we conclude that Im $R_\alpha = \mathbb{R}^2$, as desired. $\square$

Example 4.15. Let $L : V \to W$ be a linear transformation. Prove that $L(\mathbf{0}) = \mathbf{0}$.

Solution. By linearity we have $L(\mathbf{0}) = L(0\mathbf{0}) = 0L(\mathbf{0}) = \mathbf{0}$, since the product of 0 and any vector is $\mathbf{0}$. □

Example 4.16. Prove Theorem 4.2.

Solution. $(a) \Rightarrow (b)$: Assume L is linear. By additivity and homogeneity we have

$$\begin{aligned} L(\mathbf{u} + c\mathbf{v}) &= L(\mathbf{u}) + L(c\mathbf{v}) \quad \text{Additivity} \\ &= L(\mathbf{u}) + cL(\mathbf{v}) \quad \text{Homogeneity} \end{aligned}$$

$(b) \Rightarrow (c)$: Assume L satisfies (b), and let $\mathbf{u}, \mathbf{v} \in V, a, b \in \mathbb{R}$. Applying (b) to vectors $a\mathbf{u}, \mathbf{v}$ and the scalar b we obtain the following:

$$\begin{aligned} L(a\mathbf{u} + b\mathbf{v}) &= L(a\mathbf{u}) + bL(\mathbf{v}) \\ &= L(\mathbf{0} + a\mathbf{u}) + bL(\mathbf{v}) \\ &= L(\mathbf{0}) + aL(\mathbf{u}) + bL(\mathbf{v}) \quad (*) \end{aligned}$$

On the other hand if we set $\mathbf{u} = \mathbf{v} = \mathbf{0}$ and $c = 1$ in (b), we obtain $L(\mathbf{0}+\mathbf{0}) = L(\mathbf{0}) + L(\mathbf{0})$. Thus, $L(\mathbf{0}) = \mathbf{0}$. Substituting this into $(*)$ we conclude $L(a\mathbf{u} + b\mathbf{v}) = aL(\mathbf{u}) + bL(\mathbf{v})$, which proves (c).

$(c) \Rightarrow (a)$: Letting $a = b = 1$ in (c) we obtain $L(\mathbf{u} + \mathbf{v}) = L(\mathbf{u}) + L(\mathbf{v})$, which is precisely the additivity. Letting $a = 0$ in (c), we obtain $L(0\mathbf{u} + b\mathbf{v}) = 0L(\mathbf{u}) + bL(\mathbf{v})$, which implies $L(b\mathbf{v}) = bL(\mathbf{v})$, which is the homogeneity. Therefore, L is linear. □

Example 4.17. Let c be a scalar, A, B be two matrices with real entries and $\mathbf{v}$ be a column vector. Prove or disprove each of the following:

(a) If $c\mathbf{v} = \mathbf{0}$, then $c = 0$ or $\mathbf{v} = \mathbf{0}$.

(b) If $AB = 0$, then $A = 0$ or $B = 0$.

(c) If $A\mathbf{v} = \mathbf{0}$, then $A = 0$ or $\mathbf{v} = \mathbf{0}$.

Solution. (a) This is true. Suppose $c\mathbf{v} = 0$. If $c \neq 0$, then multiplying both sides by $1/c$ we obtain $1\mathbf{v} = \mathbf{0}$ and thus $\mathbf{v} = \mathbf{0}$.

(b) This is false. Consider the two matrices $A = (1\ 0)$ and $B = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. Neither A nor B is zero, but $AB = 0$.

(c) This is false. The same example as the one in part (b) works. □

Example 4.18. Prove that if a, b are real numbers with $b \neq 0$, then the function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = ax + b$ is not linear.

Solution. There are multiple ways of doing this. One way would be to note $f(0) = b \neq 0$ and thus f cannot be linear by Example 4.15.

Another way would be to use Theorem 4.3: If f were linear, then there would be a 1×1 matrix A for which $f(x) = Ax$. Note that 1×1 matrices are just real numbers. Thus, we must have $ax + b = Ax$ for all $x \in R$ and thus $b = 0$, which is a contradiction.

We could also check that such f does not satisfy the homogeneity (or the additivity) condition. For example $f(2) = 2a + b \neq 2f(1) = 2a + 2b$, since $b \neq 0$. $\square$

Example 4.19. Determine if each of the following functions is linear:

(a) $L : \mathbb{R}^2 \to \mathbb{R}$, given by $L(x, y) = xy$.

(b) $L : \mathbb{R}^2 \to \mathbb{R}$, given by $L(x, y) = x + 3y$.

(c) $L : \mathbb{R}^n \to \mathbb{R}$, given by $L(x_1, \dots, x_n) = x_1$.

Solution. (a) This is not linear. Note that $f(1, 0) = 0$, $f(0, 1) = 0$, and $f(1, 1) = 1$. This means $f(1, 1) \neq f(1, 0) + f(0, 1)$, which implies f is not additive and thus it is not linear.

(b) This is linear. We have

$$L(x, y) = (1\ 3)\begin{pmatrix} x \\ y \end{pmatrix}.$$

By Theorem 4.3 this function is linear.

(c) This is linear using Theorem 4.3 and the following:

$$L(x_1, \dots, x_n) = (1\ 0\ \cdots\ 0)\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

$\square$

Example 4.20. Let V, W be two vector spaces, and let $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ be a basis for V. Assume $S, T : V \to W$ are linear transformations. Prove that $S = T$ if and only if $S(\mathbf{v}_j) = T(\mathbf{v}_j)$ for $j = 1, \dots, n$.

Solution. $\Rightarrow$: If $S = T$, then $S(\mathbf{v}_j) = T(\mathbf{v}_j)$, by definition of equality of two functions.

$\Leftarrow$: Suppose $S(\mathbf{v}_j) = T(\mathbf{v}_j)$ for $j = 1, \dots, n$. Let $\mathbf{v} \in V$. Since $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ is a basis for V, there are scalars $c_1, c_2, \dots, c_n$ for which $\mathbf{v} = \sum_{j=1}^{n} c_j\mathbf{v}_j$. By linearity of S and T, and the fact that $S(\mathbf{v}_j) = T(\mathbf{v}_j)$ we have

$$S(\mathbf{v}) = S(\sum_{j=1}^{n} c_j\mathbf{v}_j) = \sum_{j=1}^{n} c_j S(\mathbf{v}_j) = \sum_{j=1}^{n} c_j T(\mathbf{v}_j) = T(\sum_{j=1}^{n} c_j\mathbf{v}_j) = T(\mathbf{v}).$$

Note also that S and T have the same domain and co-domain. Therefore $S = T$, as desired. $\square$

Example 4.21. Suppose $T : \mathbb{R}^2 \to \mathbb{R}^3$ is a linear transformation for which $T(1,2) = (1,0,1)$ and $T(2,1) = (1,1,0)$. Find the matrix M_T.

Solution. We need to find $T(\mathbf{e}_1)$ and $T(\mathbf{e}_2)$. We see

$$(1,0) = \frac{2}{3}(2,1) - \frac{1}{3}(1,2), \text{ and } (0,1) = \frac{2}{3}(1,2) - \frac{1}{3}(2,1).$$

By linearity of T we have

$$T(\mathbf{e}_1) = \frac{2}{3}T(2,1) - \frac{1}{3}T(1,2) = \frac{2}{3}(1,1,0) - \frac{1}{3}(1,0,1) = (1/3, 2/3, -1/3),$$

and

$$T(\mathbf{e}_2) = \frac{2}{3}T(1,2) - \frac{1}{3}T(2,1) = \frac{2}{3}(1,0,1) - \frac{1}{3}(1,1,0) = (1/3, -1/3, 2/3).$$

Therefore, by a theorem the matrix M_T is given by

$$M_T = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} \\ \frac{2}{3} & \frac{-1}{3} \\ \frac{-1}{3} & \frac{2}{3} \end{pmatrix}$$

$\square$

4.5 Exercises

Exercise 4.1. *Determine if each of the following is a linear transformation. If it is linear, provide a proof. If it is not, by an example prove that it fails to satisfy one of the conditions of linear transformation.*

(a) $L : \mathbb{R}^3 \to \mathbb{R}^3$, $L(x,y,z) = (x+y, z, x^2)$.

(b) $L : \mathbb{R}^2 \to \mathbb{R}^3$, $L(x,y) = (x+2y, y, -x)$.

(c) $L : \mathbb{R}^3 \to \mathbb{R}^2$, $L(x,y,z) = (xy, xz)$.

(d) $L : \mathbb{R}^3 \to \mathbb{R}^2$, $L(x,y,z) = (x+y, z-1)$.

Exercise 4.2. *Find all linear transformations* $T : \mathbb{R}^3 \to \mathbb{R}^2$ *satisfying all of the following:*

$$T(1,2,0) = (0,2), T(-1,1,1) = (-2,3), \text{ and } T(1,-2,-1) = (1,-3).$$

Exercise 4.3. *Let* $\alpha \in [0, 2\pi)$ *be an angle. Consider the transformation* $T_\alpha : \mathbb{R}^3 \to \mathbb{R}^3$ *which rotates every point around the* z*-axis with angle* α*. Assume we know* T_α *is linear. Find* M_{T_α}*.*

Exercise 4.4. *Find all* 2×2 *matrices* A *that commute with every other matrix. In other words, find all matrices* $A \in M_2(\mathbb{R})$*, for which* $AB = BA$*, for every* $B \in M_2(\mathbb{R})$*.*

Exercise 4.5. *True or false? If true provide a proof, and if false provide a counter-example.*

(a) If for a square matrix A *we have* $A^2 = 0$*, then* $A = 0$*.*

(b) If the two products AB *and* BA *are defined, then* A *and* B *must be square matrices.*

(c) $AB = BA$ *for every two* 2×2 *matrices* A *and* B

Exercise 4.6. *Find an example of three matrices* A, B, C *for which* $AB = BA, AC = CA$*, but* $BC \neq CB$*.*

Exercise 4.7. *Find an example of two matrices* A, B *for which* A^2 *and* B *commute but* A *and* B *do not commute.*

Exercise 4.8. *Prove that if two matrices* A *and* B *commute, then for every two positive integers* m, n *the two matrices* A^n *and* B^m *also commute.*

Exercise 4.9. *Using the definition of linearity, prove that if* $S : V \to W$ *and* $T : W \to U$ *are linear transformations of vector spaces, then* $T \circ S : V \to U$ *is also linear.*

Exercise 4.10. *Suppose* $T : V \to W$ *is a linear transformation between vector spaces. Using induction, prove that for every* $c_1, \ldots, c_n \in \mathbb{R}$ *and every* $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$*, we have*

$$T(c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n) = c_1T(\mathbf{v}_1) + \cdots + c_nT(\mathbf{v}_n).$$

Exercise 4.11. *Provide an example of a function* $f : \mathbb{R}^n \to \mathbb{R}^m$ *for some* n, m *that is homogeneous, but it is not linear.*

Exercise 4.12. *Let* V, W *be two vector spaces. Suppose* $L : V \to W$ *is a function that is additive and satisfies* $L(c\mathbf{v}) = cL(\mathbf{v})$*, for all* $\mathbf{v} \in V$ *and all positive* $c \in \mathbb{R}$*. Does* L *have to be linear?*

Exercise 4.13. *Suppose* $T : V \to W$ *is a function between vector spaces that is additive. Prove that:*

(a) $T(-\mathbf{v}) = -T(\mathbf{v})$*, for all* $\mathbf{v} \in V$*.*

(b) For every positive integer n *and every* $\mathbf{v} \in V$*,* $T(n\mathbf{v}) = nT(\mathbf{v})$*. (Hint: Use induction on* n*.)*

(c) Combining parts (a) and (b), prove $T(n\mathbf{v}) = nT(\mathbf{v})$ *for every* $\mathbf{v} \in V$ *and every* $n \in \mathbb{Z}$*.*

(d) Prove that for every $r \in \mathbb{Q}$ *and every* $\mathbf{v} \in V$*, we have* $T(r\mathbf{v}) = rT(\mathbf{v})$*.*

Exercise 4.14. *Let* V, W *be vector spaces. Assume* $\mathbf{v}_1, \ldots, \mathbf{v}_n$ *form a basis for* V*, and let* $\mathbf{w}_1, \ldots, \mathbf{w}_n \in W$*. Prove that* $T : V \to W$ *defined by*

$$T(c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n) = c_1\mathbf{w}_1 + c_2\mathbf{w}_2 + \cdots + c_n\mathbf{w}_n, \text{ for all } c_1, c_2, \ldots, c_n \in \mathbb{R}$$

is a linear transformation.

Exercise 4.15. *Suppose* $f, g : \mathbb{R}^n \to \mathbb{R}^m$ *are linear transformations. Prove that* $f + g : \mathbb{R}^n \to \mathbb{R}^m$ *is linear, and that* $M_{f+g} = M_f + M_g$*.*

Exercise 4.16. *Suppose $L : V \to W$ is a bijective linear transformation. Prove that $L^{-1} : W \to V$ is linear.*

Exercise 4.17. *Suppose A, B are matrices of size $m \times n$ and $n \times k$, respectively. Prove that $(AB)^T = B^T A^T$.*

Exercise 4.18. *Suppose $\mathbb{R}^n$ is equipped with a norm. Given every two distinct points $A, B \in \mathbb{R}^n$, prove there is a unique point M on segment AB that is equidistant from A and B. In other words, $||\overrightarrow{AM}|| = ||\overrightarrow{BM}||$.*

Definition 4.9. The point M in the above exercise is called the **midpoint** of segment AB.

Definition 4.10. Consider $\mathbb{R}^n$ equipped with a norm. Given a triangle ABC in $\mathbb{R}^n$, the segment connecting A and the midpoint of BC is called the **median** of ABC correspoding to vertex A.

Exercise 4.19. *Consider $\mathbb{R}^n$ equipped with a norm. Prove the medians of ABC are concurrent. In other words, prove there is a unique point $G \in \mathbb{R}^n$ for which G lies on all the three medians corresponding to vertices of triangle ABC.*

Exercise 4.20. *Suppose $\langle , \rangle$ is an inner product of $\mathbb{R}^n$. Prove that for every $\mathbf{w} \in \mathbb{R}^n$, there is a vector $\mathbf{x} \in \mathbb{R}^n$ for which $\langle \mathbf{w}, \mathbf{v} \rangle = \mathbf{x} \cdot \mathbf{v}$ for all $\mathbf{v} \in \mathbb{R}^n$.*

Exercise 4.21. *We know composition of every two linear transformations is linear. Is the converse true? In other words, suppose $f : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^m \to \mathbb{R}^k$ are functions for which $g \circ f$ is linear. Can we conclude f and g are linear?*

Definition 4.11. A **distance** on $\mathbb{R}^n$ is a function d that assigns a nonnegative real number $d(\mathbf{v}, \mathbf{w})$ to every pair of points $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ that satisfies the following for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$:

(a) $d(\mathbf{v}, \mathbf{w}) > 0$ if $\mathbf{v} \neq \mathbf{w}$ (Positivity),

(b) $d(\mathbf{v}, \mathbf{w}) = d(\mathbf{w}, \mathbf{v})$ (Symmetry),

(c) $d(\mathbf{v}, \mathbf{w}) + d(\mathbf{w}, \mathbf{u}) \geq d(\mathbf{v}, \mathbf{u})$ (Triangle ineuqality).

Exercise 4.22. *Prove that any norm on $\mathbb{R}^n$ provides a distance d between every two $\mathbf{v}, \mathbf{w} \in V$ defined by:*

$$d(\mathbf{v}, \mathbf{w}) = |\mathbf{v} - \mathbf{w}|.$$

4.6 Challenge Problems

Exercise 4.23. *Prove that for every $n \geq 2$ there is no inner product on $\mathbb{R}^n$ that gives us the norm*

$$||(x_1, \ldots, x_n)|| = \max\{|x_1|, \ldots, |x_n|\}.$$

Exercise 4.24. *Prove that for every $n \geq 2$ there is no inner product on $\mathbb{R}^n$ that gives us the norm*

$$||(x_1, \ldots, x_n)|| = |x_1| + \cdots + |x_n|.$$

Definition 4.12. For two subspaces U and W of $\mathbb{R}^n$, define

$$U + W = \{\mathbf{x} \in V \mid \mathbf{x} = \mathbf{u} + \mathbf{w} \text{ for some } \mathbf{u} \in U, \text{ and } \mathbf{w} \in W\}.$$

Exercise 4.25. *Suppose U and W are subspaces of a vector space V. Prove that*

$$\dim(U + W) + \dim(U \cap W) = \dim U + \dim W.$$

Exercise 4.26. *Suppose $A \in M_n(\mathbb{R})$ is not invertible. Prove that there is a nonzero matrix $B \in M_n(\mathbb{R})$ for which $AB = BA = 0$.*

Exercise 4.27. *Suppose $L : \mathbb{R}^2 \to \mathbb{R}^2$ is a bijective linear transformation. Prove that the image of every ellipse is also an ellipse. What can we say about images of parabolas and hyperbolas?*

4.7 Summary

- To check if $||\mathbf{v}||$ is a norm, we need to check if it satisfies three properties: Positivity, Triangle Inequality, and Homogeneity.
- To prove $L : V \to W$ is linear we need to prove two properties for all $\mathbf{v}, \mathbf{w} \in V$ and $c \in \mathbb{R}$:
  - Additivity: $L(\mathbf{v} + \mathbf{w}) = L(\mathbf{v}) + L(\mathbf{w})$, and
  - Homogeneity: $L(c\mathbf{v}) = cL(\mathbf{v})$
- To prove $L : V \to W$ is not linear, we need to show either the additivity or the homogeneity fails for at least *some* vectors and constants. We do not need to prove that both additivity and homogeneity fail.
- To find the product $A\mathbf{v}$, where A is an $m \times n$ matrix and $\mathbf{v}$ is an $n \times 1$ column we can use one of the following:
  - Using rows of A, write: $A = \begin{pmatrix} \mathbf{w}_1 \\ \vdots \\ \mathbf{w}_m \end{pmatrix}$. Then we have $A\mathbf{v} = \begin{pmatrix} \mathbf{w}_1 \cdot \mathbf{v} \\ \vdots \\ \mathbf{w}_m \cdot \mathbf{v} \end{pmatrix}$.
  - Using columns of A, write: $A = \begin{pmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{pmatrix}$ and $\mathbf{v} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$. Then,

$$A\mathbf{v} = x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n.$$

- Every linear transformation $L : \mathbb{R}^n \to \mathbb{R}^m$ is given by $L(\mathbf{v}) = A\mathbf{v}$, where A is an $m \times n$ matrix. The columns of the matrix A are $L(\mathbf{e}_1), \ldots, L(\mathbf{e}_n)$. Every such function is linear.
- The (i, j) entry of AB is obtained by finding the dot product of the i-th row of A and the j-th column of B.
- For the matrix AB to be defined we need the number of columns of A and the number of rows of B to be the same.
- If A is a matrix of size $m \times n$ and B is a matrix of size $n \times k$, then the matrix AB is of size $m \times k$.

- Note that in general AB and BA are not the same matrices.
- Image and kernel of every linear transformation are subspaces.
- The Rank-Nullity Theorem states that for every linear transformation $L : V \to W$ we have

$$\dim \operatorname{Ker} L + \dim \operatorname{Im} L = \dim V.$$

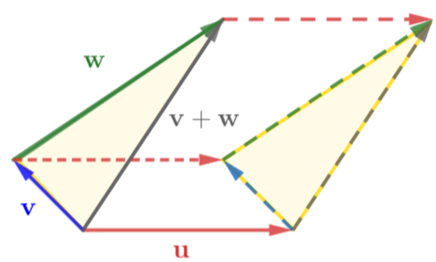
Week 5

5.1 Determinants

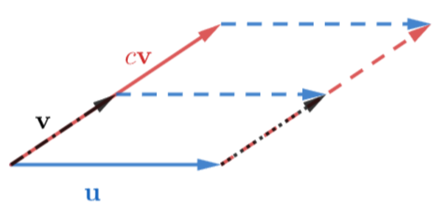
YouTube Video: https://youtu.be/_1PtPKcVMs4

In this section we would like to define the determinant of a square matrix. One interpretation of determinant is "volume". Given n vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{R}^n$, we want the $n \times n$ determinant corresponding to $\mathbf{v}_1, \ldots, \mathbf{v}_n$ to determine the volume of the parallelepiped determined by these n vectors. We expect any reasonable volume to satisfy the following properties:

$\text{Area}(\mathbf{u}, \mathbf{u}) = 0$



$\text{Area}(\mathbf{u}, \mathbf{v} + \mathbf{w}) = \text{Area}(\mathbf{u}, \mathbf{v}) + \text{Area}(\mathbf{u}, \mathbf{w})$



$\text{Area}(\mathbf{u}, c\mathbf{v}) = c\text{Area}(\mathbf{u}, \mathbf{v})$

Definition 5.1. Let $D : M_n(\mathbb{R}) \to \mathbb{R}$ be a function.

(a) We say D is **multi-linear** if D is linear with respect to each row. In other words, for every i we have

$$D\begin{pmatrix} \mathbf{v}_1 \\ \vdots \\ a\mathbf{v}_i + b\mathbf{w} \\ \vdots \\ \mathbf{v}_n \end{pmatrix} = aD\begin{pmatrix} \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_i \\ \vdots \\ \mathbf{v}_n \end{pmatrix} + bD\begin{pmatrix} \mathbf{v}_1 \\ \vdots \\ \mathbf{w} \\ \vdots \\ \mathbf{v}_n \end{pmatrix} \leftarrow i-\text{th row.}$$

(b) We say D is **alternating** if $D\begin{pmatrix} \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_n \end{pmatrix} = 0$ when $\mathbf{v}_i = \mathbf{v}_j$ for some $i \neq j$.

To keep the notations more compact, instead of writing $D\begin{pmatrix} \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_n \end{pmatrix}$ we write $D(\mathbf{v}_1, \ldots, \mathbf{v}_n)$; inserting commas to indicate $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are rows and not columns.

Example 5.1. Let $D : M_2(\mathbb{R}) \to \mathbb{R}$ be an alternating, multi-linear function. Prove that

$$D(\mathbf{u}, \mathbf{v}) = -D(\mathbf{v}, \mathbf{u}).$$

Example 5.2. Find all alternating, multi-linear functions $D : M_2(\mathbb{R}) \to \mathbb{R}$ satisfying $D(I) = 1$.

Theorem 5.1. *Let $D : M_n(\mathbb{R}) \to \mathbb{R}$ be alternating and multi-linear, then it satisfies the following properties.*

(a) Swapping two rows, negates D. In other words,

$$D(\mathbf{v}_1, \ldots, \mathbf{v}_i, \ldots, \mathbf{v}_j, \ldots, \mathbf{v}_n) = -D(\mathbf{v}_1, \ldots, \mathbf{v}_j, \ldots, \mathbf{v}_i, \ldots, \mathbf{v}_n).$$

(b) Scaling a row by c scales D by c. In other words,

$$D(\mathbf{v}_1, \ldots, c\mathbf{v}_i, \ldots, \mathbf{v}_n) = cD(\mathbf{v}_1, \ldots, \mathbf{v}_i, \ldots, \mathbf{v}_n).$$

(c) Adding a multiple of one row to a different row does not change D. In other words,

$$D(\mathbf{v}_1, \ldots, \mathbf{v}_i + c\mathbf{v}_j, \ldots, \mathbf{v}_n) = D(\mathbf{v}_1, \ldots, \mathbf{v}_i, \ldots, \mathbf{v}_n), \text{ for } i \neq j.$$

(d) $D(\mathbf{v}_1, \ldots, \mathbf{v}_n) = 0$ if $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly dependent.

Clearly the first three operations are very familiar. These are precisely the row operations that we explored when solving systems of linear equations.

Theorem 5.2. *For every positive integer n, there is a unique multi-linear, alternating function $D : M_n(\mathbb{R}) \to \mathbb{R}$ satisfying $D(I) = 1$.*

Definition 5.2. Let D be the function in the above theorem. Then the **determinant** of an $n \times n$ matrix A whose rows are $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is defined as $D(\mathbf{v}_1, \ldots, \mathbf{v}_n)$ and is denoted by $\det A$ or $\det(A)$.

Example 5.3. Evaluate

$$\det \begin{pmatrix} 1 & 2 & -1 \\ 2 & 0 & 1 \\ 3 & 2 & 1 \end{pmatrix}.$$

5.2 Row Operations, Matrix Multiplications, and Determinants

YouTube Video: https://youtu.be/aP17CwRMVac

The outcome of each row operation to matrix A is a matrix EA as follows:

- If the operation is interchanging rows i and j with $i < j$, then

$$E = \begin{array}{cl} \begin{pmatrix} \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_j \\ \vdots \\ \mathbf{e}_i \\ \vdots \\ \mathbf{e}_n \end{pmatrix} & \begin{array}{l} \\ \\ \leftarrow i\text{-th row} \\ \\ \leftarrow j\text{-th row} \\ \\ \\ \end{array} \end{array}.$$

- If the operation is scaling of the i-th row by a factor of c, then $E = \begin{pmatrix} \mathbf{e}_1 \\ \vdots \\ c\mathbf{e}_i \\ \vdots \\ \mathbf{e}_n \end{pmatrix} \begin{array}{l} \\ \\ \leftarrow i\text{-th row.} \\ \\ \\ \end{array}$

- If the operation is adding a multiple of the j-th row to the i-th row then $E = \begin{pmatrix} \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_i + c\mathbf{e}_j \\ \vdots \\ \mathbf{e}_n \end{pmatrix} \begin{array}{l} \\ \\ \leftarrow i\text{-th row.} \\ \\ \\ \end{array}$

Definition 5.3. Any matrix E of one the forms above is called an **elementary matrix**.

Combining the above and Theorem 5.1 we conclude that $\det(EA) = (\det E)(\det A)$, for every $n \times n$ matrix A and $n \times n$ elementary matrix E as above.

Theorem 5.3. *Let A and B be two $n \times n$ matrices, then* $\det(AB) = (\det A)(\det B)$

Corollary 5.1. For every square matrix A we have $\det(A^T) = \det A$.

Determinants can be evaluated using co-factor expansions. Here is an example.

$$\det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = a_{11} \det \begin{pmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{pmatrix} - a_{12} \det \begin{pmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{pmatrix} + a_{13} \det \begin{pmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix}.$$

In other words, we can write the determinant of a 3×3 matrix A as follows:

$$\det A = a_{11} \det A_{11} - a_{12} \det A_{12} + a_{13} \det A_{13},$$

where A_{ij} is obtained by removing the i-th row and the j-th row of A.

YouTube Video: https://youtu.be/OuQxN8uzrLU

Theorem 5.4. *(Cofactor Expansion Along a Row or a Column) Let $A = (a_{ij})_{n\times n}$ be an $n \times n$ matrix with a_{ij} as its (i, j) entry. Then, for every i with $1 \le i \le n$, we have*

$$\det A = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det A_{ij},$$

where A_{ij} is obtained by removing the i-th row and the j-th column of A. Similarly, for every j with $1 \le j \le n$, we have

$$\det A = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det A_{ij},$$

Definition 5.4. A square matrix A is called **invertible** or **nonsingular** if there is a square matrix B for which $AB = BA = I$. A matrix that is not invertible is said to be **singular**.

The inverse of an invertible matrix is unique and is denoted by A^{-1}. The proof is left as an exercise.

To find an inverse of an invertible matrix A we row reduce the augmented matrix $(A|I)$ to obtain $(I|B)$. This matrix B is the inverse of A.

Example 5.4. Find the inverse of $\begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 3 \\ 2 & 1 & 0 \end{pmatrix}$.

Theorem 5.5. *For a square matrix A the following are equivalent:*

(a) A is invertible.

(b) $\det A \neq 0$.

(c) Rows of A are linearly independent.

(d) Columns of A are linearly independent.

YouTube Video: https://youtu.be/NZqdBHdrBk8

Theorem 5.6. *Let A be an invertible matrix. Then the (i, j) entry of A^{-1} equals $\dfrac{(-1)^{i+j}\det(A_{ji})}{\det A}$, where A_{ji} is the matrix obtained from A by removing the j-th row and i-th column of A.*

Example 5.5. Find the inverse of the 2×2 matrix: $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$.

Theorem 5.7 (Cramer's Rule)**.** *Let $A = (\mathbf{a}_1 \cdots \mathbf{a}_n)$ be an invertible matrix. Then for every column vector $\mathbf{b}$, the only solution to $A\mathbf{x} = \mathbf{b}$ is*

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix},$$

where $x_j = \dfrac{\det(\mathbf{a}_1 \cdots \mathbf{a}_{j-1}\ \mathbf{b}\ \mathbf{a}_{j+1} \cdots \mathbf{a}_n)}{\det(A)}$.

Example 5.6. Solve the system of equations using Cramer's Rule:

$$\begin{cases} x + y - 2z = 1 \\ y + 2z = 1 \\ x - z = 3 \end{cases}$$

5.3 More Examples

Example 5.7. For real numbers $a_1, \ldots, a_n$ let $A = \operatorname{diag}(a_1, \ldots, a_n)$ be the $n \times n$ matrix whose diagonal entries are $a_1, \ldots, a_n$ in that order. Prove that $\det A = a_1 \cdots a_n$ in two ways:

(a) Using induction along with co-factor expansion.

(b) Using row operations

Solution. (a) We will prove this by induction on n.

Basis step. For $n = 1$, $A = (a_1)$, and we have $\det(a_1) = a_1$.

Inductive step. Expanding $\det A$ along the last row we obtain $\det A = (-1)^{n+n} a_n \det(\operatorname{diag}(a_1, \ldots, a_{n-1}))$ $(*)$, since the rest of the terms in the expansion are zero. By inductive hypothesis $\det(\operatorname{diag}(a_1, \ldots, a_{n-1})) = a_1 \cdots a_{n-1}$. Combining this with $(*)$ we obtain the result.

(b) Note that rows of the given matrix are $a_1\mathbf{e}_1, \ldots, a_n\mathbf{e}_n$. By the scaling row operation with a factor of a_1

and with respect to the first row we obtain the following:

$$\det\begin{pmatrix} a_1\mathbf{e}_1 \\ a_2\mathbf{e}_2 \\ \vdots \\ a_n\mathbf{e}_n \end{pmatrix} = a_1 \det\begin{pmatrix} \mathbf{e}_1 \\ a_2\mathbf{e}_2 \\ \vdots \\ a_n\mathbf{e}_n \end{pmatrix}$$

Repeating this we conclude that

$$\det(\operatorname{diag}(a_1, \dots, a_n)) = a_1 \cdots a_n \det I = a_1 \cdots a_n,$$

as desired. $\square$

Example 5.8. Suppose A is a square matrix such that A and A^{-1} both only have integer entries. Prove that $\det A = \pm 1$.

Solution. By co-factor expansion we know that $\det A$ is an integer. (This can be done by induction on the size of A.) Similarly $\det A^{-1}$ is also an integer. Since $\det(AA^{-1}) = \det I = 1$, we must have $(\det A)(\det A^{-1}) = 1$. Since both $\det A$ and $\det A^{-1}$ are integers, we must have $\det A = \pm 1$. $\square$

Example 5.9. Let a, b, c be three real numbers. Evaluate the following determinant:

$$\det\begin{pmatrix} a & a^2 & a^3 \\ b & b^2 & b^3 \\ c & c^2 & c^3 \end{pmatrix}$$

Solution. We will use Theorem 5.1.

$$\det\begin{pmatrix} a & a^2 & a^3 \\ b & b^2 & b^3 \\ c & c^2 & c^3 \end{pmatrix} = abc \det\begin{pmatrix} 1 & a & a^2 \\ 1 & b & b^2 \\ 1 & c & c^2 \end{pmatrix}$$

Use row operations $R_2 - R_1$ and $R_3 - R_1$ we obtain the following:

$$abc \det\begin{pmatrix} 1 & a & a^2 \\ 0 & b-a & b^2-a^2 \\ 0 & c-a & c^2-a^2 \end{pmatrix} = abc(b-a)(c-a) \det\begin{pmatrix} 1 & a & a^2 \\ 0 & 1 & b+a \\ 0 & 1 & c+a \end{pmatrix},$$

which is obtained by taking out scalars $b - a$ and $c - a$ from the second and third rows of the matrix. Using the row operation $R_3 - R_2$ we obtain the following:

$$abc(b-a)(c-a) \det\begin{pmatrix} 1 & a & a^2 \\ 0 & 1 & b+a \\ 0 & 0 & c-b \end{pmatrix} = abc(b-a)(c-a)(c-b) \det\begin{pmatrix} 1 & a & a^2 \\ 0 & 1 & b+a \\ 0 & 0 & 1 \end{pmatrix}$$

Expanding this along the first column and the fist column again we obtain $\boxed{abc(b-a)(c-a)(c-b)}$. $\square$

Example 5.10. Let A, B, C be three matrices of sizes $n \times m, m \times k$, and $k \times n$, respectively. Prove that:

(a) $\text{Row}(CA) \subseteq \text{Row}(A)$.

(b) $\text{Col}(AB) \subseteq \text{Col}(A)$.

Using the above, conclude that if P, Q are $n \times n$ and $m \times m$ invertible matrices, then

$$\text{rank }(PA) = \text{rank }(AQ) = \text{rank } A.$$

Solution. (a) Suppose rows of A are $\mathbf{a}_1, \dots, \mathbf{a}_n$. Assume the i-th row of C is $(c_{i1} \cdots c_{in})$. Then the i-th row of CA would be

$$c_{i1}\mathbf{a}_1 + \cdots + c_{in}\mathbf{a}_n \in \text{span }\{\mathbf{a}_1, \dots, \mathbf{a}_n\} = \text{Row}(A).$$

This means every row of CA is in $\text{Row}(A)$. Since $\text{Row}(A)$ is closed under linear combination, all linear combinations of rows of CA are in $\text{Row}(A)$. Therefore, $\text{Row}(CA) \subseteq \text{Row}(A)$.

(b) Let $\mathbf{v}_1, \dots, \mathbf{v}_m$ be all columns of A and $(b_{1j} \cdots b_{mj})^T$ be the j-th column of B. The j-th column of AB is then given by

$$b_{1j}\mathbf{v}_1 + \cdots + b_{mj}\mathbf{v}_m \in \text{span }\{\mathbf{v}_1, \dots, \mathbf{v}_m\} = \text{Col}(A).$$

Therefore, every column of AB is in $\text{Col}(A)$. Similar to above $\text{Col}(AB) \subseteq \text{Col}(A)$, as desired.

Now, assume P is an invertible $n \times n$ matrix. By (a) above we have:

$$\text{Row}(PA) \subseteq \text{Row}(A).$$

On the other hand if we write $A = P^{-1}PA$ and apply part (a) again we obtain the following:

$$\text{Row}(A) = \text{Row}(P^{-1}(PA)) \subseteq \text{Row}(PA).$$

Therefore, $\text{Row}(PA) = \text{Row}(A)$. Similarly we can show if Q is an invertible $m \times m$ matrix then $\text{Col}(AQ) = \text{Col}(A)$. $\square$

Example 5.11. Show that if the entries of an invertible matrix are all rational, then all entries of its inverse are also rational.

Solution. First, we prove by induction on n that the determinant of every $n \times n$ matrix with rational entries is rational.

Basis. If $A = (a)$ is a 1×1 matrix with $a \in \mathbb{Q}$, then $\det A = a$ is rational.

Inductive Step. Suppose $A = (a_{ij})$ is an $(n+1) \times (n+1)$ matrix with rational entries. By co-factor expansion along the first row, $\det A = \sum_{j=1}^{n+1} (-1)^{1+j} a_{1j} \det A_{1j}$, where A_{1j} is the matrix obtained by removing the first row and the j-th column of A. By inductive hypothesis, $\det A_{1j}$ is rational. Since a_{1j} is also rational,

$\det A$ is rational. This completes the proof.

Now, assume A is an invertible matrix with rational entries. The (i,j) entry of A^{-1} is $\dfrac{(-1)^{i+j}\det A_{ji}}{\det A}$. By what we showed above $\det A_{ji}$ and $\det A$ are rational. Since A is invertible $\det A$ is nonzero. Therefore, $(-1)^{i+j}\det A_{ji}$ is rational. This implies, all entries of A^{-1} are rational. ☐

5.4 Exercises

Exercise 5.1. *Evaluate the following determinant by each of the following methods.*

(a) Using row operations, i.e. Theorem 5.1.

(b) Using co-factor expansion.

$$\det\begin{pmatrix} 1 & -1 & 2 \\ 0 & 2 & -1 \\ 3 & 4 & 1 \end{pmatrix}$$

Exercise 5.2. *Prove that if A is an $n\times n$ matrix and E is an $n\times n$ elementary matrix corresponding to a row operation, then* $\det(EA) = (\det E)(\det A)$. *Use that to prove Theorem 5.3.*

Exercise 5.3. *Suppose a, b, c, d are real numbers for which $ad \neq bc$. Using the method of row reduction find the inverse of the 2×2 matrix*

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Hint: You may have to take cases.

Exercise 5.4. *A square matrix is called upper triangular iff all enteries below its main diagonal are zero. Prove that the determinant of an upper triangular matrix is the product of its diagonal entries.*

$$\det\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ & a_{22} & \cdots & a_{2n} \\ & & \ddots & \vdots \\ \mathit{0} & & & a_{nn} \end{pmatrix} = a_{11}a_{22}\cdots a_{nn}$$

Hint: Use induction and cofactor expansion.

Exercise 5.5. *A binary matrix is one whose entries are all 0 or 1. What is the largest number of zeros that an $n\times n$ invertible binary matrix can have? How about the smallest number of zeros? Your answer may be in terms of n.*

Exercise 5.6. *Prove that if A is an $n\times m$ matrix and B is an $m\times n$ matrix, where $m < n$, then AB is not invertible.*

Hint: Use Theorem 2.5.

Exercise 5.7. *Show that if the entries of an invertible matrix are all rational, then all entries of its inverse are also rational.*

Exercise 5.8. *Prove that the inverse of an $n \times n$ matrix is unique.*

Exercise 5.9. *A square matrix A is said to be orthogonal if $AA^T = I$. Prove that an $n \times n$ matrix A is orthogonal if and only if rows of A form an orthonormal basis for $\mathbb{R}^n$.*

Exercise 5.10. *Prove that the transpose of each elementary matrix is also an elementary matrix of the same type.*

Exercise 5.11. *Prove that the inverse of each elementary matrix is also an elementary matrix of the same type.*

Exercise 5.12. *Suppose A, B are square matrices for which $AB = I$. Prove A is invertible and $B = A^{-1}$.*

Exercise 5.13. *Let $A \in M_n(\mathbb{R})$ and let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be a basis for $\mathbb{R}^n$ for which $A\mathbf{v}_j = \mathbf{0}$ for $j = 1, \ldots, n$. Prove $A = 0$.*

Exercise 5.14. *Suppose $A(x)$ is an $n \times n$ matrix all of whose entries are continuous functions over an open interval I. Suppose $\det(A(x)) \neq 0$ for every $x \in I$. Prove that all entries of the inverse of $A(x)$ are continuous functions over I.*

5.5 Challenge Problems

Exercise 5.15. *Prove that for every positive integer n the $n \times n$ matrix whose (i, j) entry is $\dfrac{1}{i+j-1}$ is invertible.*

Exercise 5.16. *Is there a subspace $M_2(\mathbb{R})$ of dimension larger than 1 whose only noninvertible matrix is the zero matrix? How about $M_3(\mathbb{R})$? How about $M_n(\mathbb{R})$?*

Exercise 5.17. *For real numbers $a_0, a_1, \ldots, a_n$ let $S_k = \sum_{j=0}^{k} a_j$ for $k = 0, 1, \ldots, n$. Evaluate determinant of the following matrix:*

$$\begin{pmatrix} S_0 & S_0 & S_0 & S_0 & \cdots & S_0 \\ S_0 & S_1 & S_1 & S_1 & \cdots & S_1 \\ S_0 & S_1 & S_2 & S_2 & \cdots & S_2 \\ S_0 & S_1 & S_2 & S_3 & \cdots & S_3 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ S_0 & S_1 & S_2 & S_3 & \cdots & S_n \end{pmatrix}$$

5.6 Summary

- To evaluate determinants use row operations along with co-factor expansion.
- Swapping two rows (or columns) negates the determinant.
- Row (or column) additions do not change the determinant.
- Scaling a row (or a column) by a factor c multiplies the determinant by c.
- If the rows (or columns) of a matrix are linearly dependent then the matrix has zero determinant.
- $\det A = \det A^T$
- $\det(AB) = \det(A)\det(B)$
- A matrix A is invertible if and only if $\det A \neq 0$ if and only if rows (or columns) or A are linearly independent.
- To find the inverse of a square matrix A:
  - Create a matrix $(A|I)$ by placing the identity matrix next to the matrix A.
  - Row reduce this matrix to obtain a matrix of this form $(I|B)$.
  - B would be the inverse of A.
  - If row reducing A does not end up with the identity matrix and we end up with a zero row, then A would not be invertible.

Week 6

6.1 Limits

YouTube Video: https://youtu.be/eLNY4uPoWLI

In order to be able to define limit of a function at a given point, we need to be able to approach that point. For example consider the following function:

$$f(x) = \begin{cases} x+1 & \text{if } x < 0 \\ 2 & \text{if } x = 1 \\ 2x-1 & \text{if } 2 \leq x \end{cases}$$

The domain of this function is $(-\infty, 0) \cup \{1\} \cup [2, \infty)$. Since the only points close to zero inside the domain are less than 0, we can only talk about $\lim_{x \to 0^-} f(x)$, and not $\lim_{x \to 0^+} f(x)$. The point $x = 1$ is an isolated point, so we cannot talk about any meaningful limit at $x = 1$, and for $x = 2$ we can only talk about the limit from the right. This motivates the following definition:

Definition 6.1. Let $\mathbf{a}$ be a point in $\mathbb{R}^n$ and r be a positive real number. The **open ball of radius r centered at $\mathbf{a}$** is defined by

$$B_r(\mathbf{a}) = \{\mathbf{x} \in \mathbb{R}^n \mid ||\mathbf{x} - \mathbf{a}|| < r\},$$

and the **closed ball of radius r centered at $\mathbf{a}$** is defined by

$$\overline{B}_r(\mathbf{a}) = \{\mathbf{x} \in \mathbb{R}^n \mid ||\mathbf{x} - \mathbf{a}|| \leq r\}.$$

Here, we are using the standard Euclidean norm.

Note that since $||\mathbf{x} - \mathbf{a}|| < 0$ is impossible, the open ball $B_0(\mathbf{a})$ is empty. We never consider open balls where the radius is zero. Also note that $\overline{B}_0(\mathbf{a}) = \{\mathbf{a}\}$.

Definition 6.2. Let D be a subset of $\mathbb{R}^n$. A point $\mathbf{a}$ in $\mathbb{R}^n$ is called a **limit point** of D iff every open ball centered at $\mathbf{a}$ contains at least one point of D other than $\mathbf{a}$.

Example 6.1. Find all limit points of $(0, 1)$ in $\mathbb{R}$.

Definition 6.3. Let D be a subset of $\mathbb{R}^n$, $f : D \to \mathbb{R}^m$ be a function, $\mathbf{a} \in \mathbb{R}^n$ be a limit point of D, and $\mathbf{b} \in \mathbb{R}^m$. We say **$\mathbf{b}$ is the limit of f at $\mathbf{a}$**, written $\lim\limits_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) = \mathbf{b}$, iff for every $\epsilon > 0$, there is $\delta > 0$ such that if for some $\mathbf{x} \in D$, we have $0 < ||\mathbf{x} - \mathbf{a}|| < \delta$, then $||f(\mathbf{x}) - \mathbf{b}|| < \epsilon$. If no such $\mathbf{b}$ exists we say f does not have a limit at $\mathbf{a}$ or the limit does not exist.

Remark. Note that since the limit of a function at $\mathbf{a}$ only depends on the functional values near $\mathbf{a}$, i.e. on an open ball centered at $\mathbf{a}$, if two functions f and g are the same near $\mathbf{a}$, except possibly at $\mathbf{a}$, then their limits at $\mathbf{a}$ are the same.

Similar to above we may also define limits of sequences.

Definition 6.4. Let $\mathbf{x}_k \in \mathbb{R}^n$, for $k = 1, 2, \ldots$ be a sequence and $\mathbf{a} \in \mathbb{R}^n$. We say $\mathbf{x}_k$ converges to $\mathbf{a}$, written as $\mathbf{x}_k \to \mathbf{a}$ or $\lim\limits_{k\to\infty} \mathbf{x}_k = \mathbf{a}$, iff the following holds:

$$\forall\ \epsilon > 0\ \exists\ N \in \mathbb{N} \text{ such that } \forall\ k \in \mathbb{N},\ \text{ if } k \geq N,\ \text{ then } ||\mathbf{x}_k - \mathbf{a}|| < \epsilon.$$

Example 6.2. Prove that $\lim\limits_{x\to 1} 3x + 2 = 5$.

Example 6.3. Prove that $\lim\limits_{x\to 1} \dfrac{1+x}{1+2x} = \dfrac{2}{3}$.

Example 6.4. Prove that $\lim\limits_{(x,y)\to(1,-1)} x^2 + y^2 = 2$.

Example 6.5. Show that $\lim\limits_{(x,y)\to(0,0)} \dfrac{xy}{x^2+y^2}$ does not exist.

Theorem 6.1. *Suppose D is a subset of $\mathbb{R}^n$, $\mathbf{a}$ is a limit point of D, and $f : D \to \mathbb{R}^m$ is a function. If there are two sequences $\mathbf{x}_k, \mathbf{y}_k \in D - \{\mathbf{a}\}$ for which $\mathbf{x}_k \to \mathbf{a}$ and $\mathbf{y}_k \to \mathbf{a}$, but the limits $\lim\limits_{k\to\infty} f(\mathbf{x}_k)$ and $\lim\limits_{k\to\infty} f(\mathbf{y}_k)$ are not the same. Then $\lim\limits_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x})$ does not exist.*

YouTube Video: https://youtu.be/DFUTxABLeK0

Definition 6.5. Given a function $f : D \to \mathbb{R}^m$, where D is a subset of $\mathbb{R}^n$, we write $f = (f_1, \ldots, f_m)$ if $f(\mathbf{x}) = (f_1(\mathbf{x}), \ldots, f_m(\mathbf{x}))$ for all $\mathbf{x} \in D$, and we say functions $f_1, \ldots, f_m$ from D to $\mathbb{R}$ are **coordinate functions of f**.

Theorem 6.2. *Let D be a subset of $\mathbb{R}^n$, and $\mathbf{a}$ be a limit point of D, and $\mathbf{b} = (b_1, \ldots, b_m) \in \mathbb{R}^m$. Assume $f = (f_1, \ldots, f_m) : D \to \mathbb{R}^m$ is a function. Then $\lim\limits_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) = \mathbf{b}$ if and only if $\lim\limits_{\mathbf{x}\to\mathbf{a}} f_i(\mathbf{x}) = b_i$, for all $i = 1, \ldots, m$.*

Theorem 6.3 (Squeeze Theorem)**.** *Suppose D is a subset of $\mathbb{R}^n$, and $\mathbf{a}$ is a limit point of D. Let $f, g, h : D \to \mathbb{R}$ be functions for which*

$$f(\mathbf{x}) \leq g(\mathbf{x}) \leq h(\mathbf{x}) \quad \textit{for all } \mathbf{x} \in D - \{\mathbf{a}\}.$$

If $\lim\limits_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) = \lim\limits_{\mathbf{x}\to\mathbf{a}} h(\mathbf{x}) = L$ for some real number L, then $\lim\limits_{\mathbf{x}\to\mathbf{a}} g(\mathbf{x}) = L$.

6.2 Continuity

Definition 6.6. Let $\mathbf{a}$ be a limit point of a subset D of $\mathbb{R}^n$. We say a function $f : D \to \mathbb{R}^m$ is **continuous** at $\mathbf{a}$ if $\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) = f(\mathbf{a})$. If f is continuous at every point inside its domain we say f is **continuous** on its domain.

Example 6.6. Prove that the following functions are continuous on their domain.

(a) $\pi_i : \mathbb{R}^n \to \mathbb{R}$ defined by $\pi_i(x_1, \ldots, x_n) = x_i$, where $1 \le i \le n$ is fixed.

(b) $p : \mathbb{R}^2 \to \mathbb{R}$ defined by $p(x, y) = xy$.

(c) $s : \mathbb{R}^2 \to \mathbb{R}$ defined by $s(x, y) = x + y$.

Theorem 6.4. *Let D be a subset of $\mathbb{R}^n$, and $\mathbf{a}$ be a limit point of D. The function $f : D \to \mathbb{R}^m$ is continuous at $\mathbf{a}$ if and only if each coordinate function of f is continuous at $\mathbf{a}$.*

Theorem 6.5. *Suppose D_1 and D_2 are subsets of $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. Let $f : D_1 \to \mathbb{R}^m$, and $g : D_2 \to \mathbb{R}^k$ be two functions. Let $\mathbf{a}$ be a limit point of D_1, and a limit point of the domain of $g \circ f$. Suppose $\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) = \mathbf{b}$, and $\mathbf{b} \in D_2$, and that g is continuous at $\mathbf{b}$. Then $\lim_{\mathbf{x}\to\mathbf{a}} g \circ f(\mathbf{x}) = g(\mathbf{b})$.*

Theorem 6.6. *Let D be a subset of $\mathbb{R}^n$, and $\mathbf{a}$ be a limit point of D. Suppose $f, g : D \to \mathbb{R}$ are functions. Then*

$$\lim_{\mathbf{x}\to\mathbf{a}} (f(\mathbf{x}) + g(\mathbf{x})) = \lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) + \lim_{\mathbf{x}\to\mathbf{a}} g(\mathbf{x}),$$

and

$$\lim_{\mathbf{x}\to\mathbf{a}} (f(\mathbf{x})g(\mathbf{x})) = (\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}))(\lim_{\mathbf{x}\to\mathbf{a}} g(\mathbf{x})),$$

assuming both $\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x})$ and $\lim_{\mathbf{x}\to\mathbf{a}} g(\mathbf{x})$ exist.

Theorem 6.7. *All of the following single variable real-valued functions are continuous over their domains:*

Polynomials and Root functions, Rational functions, Trigonometric functions and their inverses, Exponential functions and their inverses.

Example 6.7. Prove that the function $f : \mathbb{R}^3 \to \mathbb{R}^2$ defined by $f(x, y, z) = (x + y, \sin(xy) + \cos(z))$ is continuous.

Example 6.8. Every polynomial $f(x_1, \ldots, x_n)$ is a continuous function from $\mathbb{R}^n$ to $\mathbb{R}$.

6.3 More Examples

Example 6.9. Suppose $\mathbf{a}$ is a limit point of D, where D is a subset of $\mathbb{R}^n$. Prove that every open ball centered at $\mathbf{a}$ contains infinitely many points of D.

Solution. Suppose on the contrary an open ball $B_r(\mathbf{a})$ contains only finitely many points of D. Let $\mathbf{b} \neq \mathbf{a}$ be the point in $B_r(\mathbf{a}) \cap D$ with the minimum distance to $\mathbf{a}$, and set $s = ||\mathbf{b} - \mathbf{a}||$. Note that this minimum distance exists and is positive, since there are finitely many $\mathbf{b}$'s and $\mathbf{b} \neq \mathbf{a}$.

We claim $B_s(\mathbf{a})$ contains no point of D that is different from $\mathbf{a}$.

Let $\mathbf{x} \in D \cap B_s(\mathbf{a})$ and $\mathbf{x} \neq \mathbf{a}$. Since $||\mathbf{x} - \mathbf{a}|| < s$, by the definition of s, $\mathbf{x} \notin D$. This contradicts the fact that $\mathbf{a}$ is a limit point of D. $\square$

Example 6.10. Find all limit points of the set $A = \{\frac{1}{n} \mid n = 1, 2, 3, \ldots\}$.

Solution. We will show that 0 is the only limit point of this set. First note that for every ball $(-r, r)$ (with $r > 0$) around 0, there is a positive integer n for which $r > 1/n$ and thus $(-r, r)$ has a point in A other than 0. Therefore, by definition, 0 is a limit point of A.

Now, assume $0 \neq x \in \mathbb{R}$. If x is negative then the ball $B_{-x}(x)$ contains no element of A, since if $y \in B_{-x}(x)$, then $|y - x| < -x$ or $y - x < -x$, and thus $y < 0$, which means $y \notin A$.

If $x = \frac{1}{n}$ for some positive integer n, then we will show that $B_{\frac{1}{n} - \frac{1}{n+1}}(\frac{1}{n})$ contains no point of A other than x. If $y \in B_{\frac{1}{n} - \frac{1}{n+1}}(\frac{1}{n})$, then

$$\frac{1}{n} - \frac{1}{n} + \frac{1}{n+1} < y < \frac{1}{n} + \frac{1}{n} - \frac{1}{n+1} \qquad (*)$$

It is enough to show that $\frac{1}{n+1} < y < \frac{1}{n-1}$ for every $n > 1$ and $\dfrac{1}{2} < y$ for $n = 1$. The latter is clear, by substituting $n = 1$ to $(*)$. To prove the former, it is enough to show $\frac{1}{n} + \frac{1}{n} - \frac{1}{n+1} < \frac{1}{n-1}$, which is true if and only if $\frac{2}{n} < \frac{2n}{n^2-1}$, which holds if and only if $n^2 - 1 < n^2$. This means $1/n$ is not a limit point of A.

Suppose $\frac{1}{n+1} < x < \frac{1}{n}$. Then, the ball of radius $r = \min(x - \frac{1}{n+1}, \frac{1}{n} - x)$ centered at x contains no point of A. If $y \in B_r(x)$, then $|y - x| < r$, and thus $x - r < y < x + r$, or $y < x + \frac{1}{n} - x = \frac{1}{n}$. Similarly $y > x - r > x - (x - \frac{1}{n+1}) = \frac{1}{n+1}$. This means $y \notin A$, and thus x is not a limit point of A. $\square$

Example 6.11. Using the definition, find each limit or show it does not exist.

(a) $\displaystyle\lim_{(x,y)\to(1,0)} \frac{x^2 + 2y}{x + y}$.

(b) $\displaystyle\lim_{(x,y)\to(0,0)} (x + y) \sin\left(\frac{1}{x^2 + y^2}\right)$.

(c) $\displaystyle\lim_{(x,y)\to(0,0)} \frac{y^2}{x^2 + y^2}$.

(d) $\displaystyle\lim_{(x,y)\to(2,1)} xy - x^2 + y$.

(e) $\displaystyle\lim_{(x,y)\to(0,0)} \frac{x^2+2xy}{x^2+y^2}$.

Scratch. For the first part, we know the limit should be 1, since both numerator and denominator are continuous. We will need to prove the following:

$$\forall\ \epsilon > 0\ \exists\ \delta > 0 \ \text{ such that } 0 < ||(x,y)-(1,0)|| < \delta \Rightarrow \left|\frac{x^2+2y}{x+y} - 1\right| < \epsilon \qquad (*)$$

The first inequality in $(*)$ can be written as $\sqrt{(x-1)^2+y^2} < \delta$, which implies $|x-1| < \delta$ and $|y| < \delta$. The latter inequality in $(*)$ can be simplified as

$$\left|\frac{x^2+2y}{x+y} - 1\right| < \epsilon \Longleftrightarrow \left|\frac{x^2+y-x}{x+y}\right| < \epsilon.$$

Since we know $x \approx 1$ and $y \approx 0$, for the numerator we have

$$x^2 - x = x(x-1) \approx 0, \text{ and } y \approx 0.$$

We already know $|y| < \delta$. We can then bound $|x|$ by making sure $\delta \le 1$, which implies

$$|x-1| < 1 \Rightarrow -1 < x-1 < 1 \Rightarrow 0 < x < 2 \Rightarrow |x(x-1)| < 2\delta.$$

On the other hand, the denominator $|x+y|$ is approximately 1. We will guarantee this quantity remains away from zero by choosing an appropriate delta. By assuming $\delta \le 1$ we obtain $-1 < y < 1$ and $0 < x < 2$, which implies $-1 < x+y < 3$, but that is not good enough, since in this range, $x+y$ could be very close to zero. We will make δ even smaller. Letting $\delta \le 1/3$ we will get $2/3 < x < 4/3$ and $-1/3 < y < 1/3$, which implies $1/3 < x+y < 5/3$. Therefore, $1/|x+y| < 3$. Putting what we have so far together we obtain the following:

$$\left|\frac{x^2+2y}{x+y} - 1\right| = \left|\frac{x^2+y-x}{x+y}\right| \le \left|\frac{x^2-x}{x+y}\right| + \left|\frac{y}{x+y}\right| < 3 \times 2\delta + 3\delta = 9\delta.$$

Therefore, we need to make sure $\delta \le 1, \delta \le 1/3$, and $\delta \le \epsilon/9$.

For part (b) we know

$$\left|(x+y)\sin\left(\frac{1}{x^2+y^2}\right)\right| \le |x+y| \le |x| + |y| < \delta + \delta = 2\delta, \text{ if } \sqrt{x^2+y^2} < \delta.$$

Thus, we need to make sure $2\delta \le \epsilon$.

For part (c) we will approach the origin along the line $y = mx$ to obtain $\dfrac{m^2x^2}{x^2+m^2x^2} = \dfrac{m^2}{1+m^2}$. Since this depends on m the limit does not exist. We will show that using proof by contradiction and taking two different values of m that yield different limits, e.g. $m = 0$ and $m = 1$. If the limit were b, then b must be close to both $\dfrac{0^2}{1+0^2} = 0$ and $\dfrac{1^2}{1+1^2} = 1/2$. This is not possible and can be shown by taking $\epsilon = 1/4$, which is half of the distance between 0 and 1/2.

Solution. (a) We will show that the limit is 1. For every $\epsilon > 0$ let $\delta = \min(1/3, \epsilon/9)$. Note that since 1/3 and $\epsilon/9$ are positive, δ is positive as well. If $||(x,y)-(0,0)|| < \delta$, then

$$\sqrt{(x-1)^2+y^2} < \delta \Rightarrow |x-1| < \delta, \text{ and } |y| < \delta \Rightarrow |x-1| < 1/3, \text{ and } |y| < 1/3 \Rightarrow 2/3 < x < 4/3.$$

This also implies

$$1/3 < x+y < 4/3 + 1/3 = 5/3 \Rightarrow |x+y| > 3 \Rightarrow \frac{1}{|x+y|} < 3.$$

This yields the following:

$$\left|\frac{x^2+2y}{x+y} - 1\right| = \left|\frac{x^2+y-x}{x+y}\right| \le \left|\frac{x^2-x}{x+y}\right| + \left|\frac{y}{x+y}\right| = \left|\frac{x(x-1)}{x+y}\right| + \left|\frac{y}{x+y}\right| < \frac{4}{3} \times 3 \times \delta + 3\delta = 7\delta \le \frac{7\epsilon}{9} < \epsilon.$$

(b) We will show the limit is 0. For every $\epsilon > 0$ let $\delta = \epsilon/2$. Suppose $||(x,y) - (0,0)|| < \delta$. We have $\sqrt{x^2+y^2} < \delta$, and thus $|x|, |y| < \delta$. Therefore,

$$\left|(x+y)\sin\left(\frac{1}{x^2+y^2}\right)\right| \le |x+y| \le |x| + |y| < \delta + \delta = 2\delta = \epsilon.$$

This completes the proof of the claim.

(c) We will show this limit does not exist. Assume on the contrary that the limit is a real number b. In the definition of limit let $\epsilon = 1/4$. There is $\delta > 0$ for which

$$0 < \sqrt{x^2+y^2} < \delta \Rightarrow \left|\frac{y^2}{x^2+y^2} - b\right| < \frac{1}{4}.$$

Letting $x = \delta/2$ and $y = 0$ we have $\sqrt{x^2+y^2} = \delta/2 < \delta$. Therefore,

$$|0 - b| < \frac{1}{4} \Rightarrow -\frac{1}{4} < b < \frac{1}{4} \qquad (*)$$

Letting $x = y = \delta/2$ we have $\sqrt{x^2+y^2} = \sqrt{2\delta^2/4} = \delta/\sqrt{2} < \delta$. Therefore, we must have

$$|1/2 - b| < \frac{1}{4} \Rightarrow \frac{1}{4} < b < \frac{3}{4}.$$

This contradicts $(*)$. Thus, the limit does not exist.

(d) We will prove the limit is -1. Let $\epsilon > 0$ and set $\delta = \min(1, \epsilon/8)$. Since $\delta \le 1$, if $||(x,y) - (2,1)|| < \delta$, then $\sqrt{(x-2)^2 + (y-1)^2} < \delta$, which implies $|x-2| < 1$ and $|y-1| < \delta$. Thus, $1 < x < 3$. This means $|x+1| < 4$. Using the Triangle Inequality we have the following:

$$\begin{aligned}
|xy - x^2 + y - (-1)| &= |x(y-1) + x - x^2 + y - 1 + 2| \\
&\le |(y-1)(x+1)| + |x - x^2 + 2| \\
&\le \delta(|x+1|) + |x-2||x+1| \\
&< \delta(4) + \delta(4) \\
&= 8\delta \le \epsilon.
\end{aligned}$$

This completes the proof.

(e) We prove that the limit does not exist. On the contrary, assume the limit exists and is equal to L. Let $\epsilon = 1/2$ in the definition of limit. There must exist $\delta > 0$ for which, if $0 < \sqrt{x^2+y^2} < \delta$, then $\left|\frac{x^2+2xy}{x^2+y^2} - L\right| < 1/2$. We see that $\sqrt{(\delta/2)^2 + 0^2} = \delta/2 < \delta$ and $\sqrt{0^2 + (\delta/2)^2} = \delta/2 < \delta$. Therefore,

$$\left|\frac{(\delta/2)^2 + 2\delta/2 \cdot 0}{(\delta/2)^2 + 0^2} - L\right| < 1/2, \text{ and } \left|\frac{0^2 + 2 \cdot 0 \cdot \delta/2}{0^2 + (\delta/2)^2} - L\right| < 1/2.$$

Therefore, $|1-L| < 1/2$ and $|L| < 1/2$. The first inequality implies $1/2 < L < 3/2$ and the second implies $-1/2 < L < 1/2$, which is a contradiction. $\square$

Example 6.12. Find each limit or show it does not exists. You may use any method.

(a) $\displaystyle\lim_{(x,y,z)\to(1,\pi,0)} (x^2 + \sin(xy) - x\cos y + xz)$.

(b) $\displaystyle\lim_{(x,y)\to(0,0)} \frac{x}{\sqrt{x^2+y^2}}$.

(c) $\displaystyle\lim_{(x,y)\to(0,0)} \frac{x^5+y^5}{x^4+y^4}$.

Solution. (a) Since the projection functions, polynomials, trigonometric functions are all continuous functions, the function $x^2 + \sin(xy) - x\cos y + xz$ is continuous. Therefore, the limit is

$$1 + \sin(\pi) - \cos(\pi) + 0 = 2.$$

(b) Letting $y = x$ we obtain

$$\frac{x}{\sqrt{x^2+y^2}} = \frac{x}{\sqrt{2}|x|} = \pm\frac{1}{\sqrt{2}}.$$

Letting $x \to 0^+$ we obtain $\frac{1}{\sqrt{2}}$, and letting $x \to 0^-$ we obtain $-\frac{1}{\sqrt{2}}$. Since these two values are different, by Theorem 6.1 the limit does not exist.

(c) Note that along all lines $y = mx$ and $x = 0$ the limit is zero, so we suspect the limit might be zero. We can write the following chain of inequalities:

$$\left|\frac{x^5+y^5}{x^4+y^4}\right| \le \left|\frac{x^5}{x^4+y^4}\right| + \left|\frac{y^5}{x^4+y^4}\right| = \left|\frac{x^4}{x^4+y^4}\right||x| + \left|\frac{y^4}{x^4+y^4}\right||y| \le |x| + |y|.$$

This gives us the following inequalities:

$$-|x| - |y| \le \frac{x^5+y^5}{x^4+y^4} \le |x| + |y|.$$

Since $|x|$ and $|y|$ are continuous, we have

$$\lim_{(x,y)\to(0,0)} |x| + |y| = \lim_{(x,y)\to(0,0)} -|x| - |y| = 0.$$

By the Squeeze Theorem, the answer is zero. $\square$

Example 6.13. Prove that the set $\mathbb{Z}$ does not have any limit points in $\mathbb{R}$.

Solution. Suppose on the contrary $a \in \mathbb{R}$ is a limit point of $\mathbb{Z}$. By definition, there would have to be an integer $n \neq a$ that lines in $(a - 0.5, a + .05)$. Since the length of this interval is 1, there cannot be more than one integer in this interval. Let $r = \min(0.5, |a-n|)$. Note that since $a \neq n$, the number r is positive. Since n is the only integer in $(a - 0.5, a + 0.5)$, the open ball $(a - r, a + r)$ contains no integer points. (Why?) This contradiction shows $\mathbb{Z}$ has no limit points. $\square$

More examples can be found on Colley's Vector Calculus: pages 100-109 Examples 4, 8, 9, 10, 14, 16.

6.4 Exercises

Exercise 6.1. *Using the definition of limit, find each of the following limits or show they do not exist:*

(a) $\lim\limits_{(x,y)\to(2,1)} xy - x^2 + y$.

(b) $\lim\limits_{(x,y)\to(0,0)} \dfrac{x^2y}{x^2+y^2}$.

(c) $\lim\limits_{(x,y)\to(2,1)} \dfrac{xy}{x+y}$.

Exercise 6.2. *Prove that every linear transformation $L : \mathbb{R}^n \to \mathbb{R}^m$ is continuous.*

Hint: Use the matrix M_L.

Definition 6.7. A subset U of $\mathbb{R}^n$ is called **convex** if for every $\mathbf{x}, \mathbf{y} \in U$ and every $t \in [0,1]$, we have $t\,\mathbf{x} + (1-t)\,\mathbf{y} \in U$. (Recall that the set consisting of points of the form $t\,\mathbf{x} + (1-t)\,\mathbf{y}$ with $t \in [0,1]$ is called the line segment between $\mathbf{x}$ and $\mathbf{y}$.)

Exercise 6.3. *Prove that for every $\mathbf{a} \in \mathbb{R}^n$ and every $r > 0$, the open ball $B_r(\mathbf{a})$ is convex.*

Exercise 6.4. *Evaluate the limit or show it does not exist:*

$$\lim_{(x,y)\to(0,0)} (3x^3 + y\cos(x+y))\sin\left(\frac{1}{x^2+y^4}\right).$$

Hint: Use the Squeeze Theorem.

Exercise 6.5. *Prove that every real number is a limit point of $\mathbb{Q}$.*

Hint: Given a real number r you need to show there is a rational number in $(r-\epsilon, r+\epsilon)$ that is not r. Choose a positive integer n for which $n > \dfrac{1}{\epsilon}$. Use that to show there is an integer between nr and $nr + n\epsilon$.

Exercise 6.6. *Evaluate each of the following or show they do not exist, once using the $\epsilon - \delta$ definition of limit, and once using an appropriate theorem.*

(a) $\lim\limits_{(x,y,z)\to(1,2,3)} x + 2y - 3z$.

(b) $\lim\limits_{(x,y)\to(0,0)} \dfrac{x^2+y^3}{x^4+y^2}$.

Exercise 6.7. *Evaluate each limit or show it does not exist.*

(a) $\lim\limits_{(x,y)\to(0,0)} \dfrac{xy^4}{x^4+y^4}$.

(b) $\lim\limits_{(x,y)\to(0,0)} \dfrac{\sin(xy)}{x^2+y^2}$.

(c) $\lim\limits_{(x,y)\to(\pi,0)} \dfrac{\sin(x+y)}{y}$.

Exercise 6.8. *Show that the following function is not continuous at* $(0,0)$.

$$f(x,y)=\begin{cases}\dfrac{x^2-y^2}{x^2+y^2} & \textit{if } (x,y)\neq(0,0)\\ 0 & \textit{if } x=y=0\end{cases}$$

Exercise 6.9. *Consider the function* $f(x,y)=\dfrac{xy}{x^2+y^2}$ *defined over* $\mathbb{R}^2-\{(0,0)\}$. *Show that*

(a) $\lim\limits_{(x,y)\to(0,0)} f(x,y)$ *does not exist, but*

(b) $\lim\limits_{x\to 0}\lim\limits_{y\to 0} f(x,y)$, *and* $\lim\limits_{y\to 0}\lim\limits_{x\to 0} f(x,y)$ *both exist and are equal.*

Exercise 6.10. *Show the following function is continuous over* $\mathbb{R}^2$.

$$f(x,y)=\begin{cases}\dfrac{x^5-x^2y^3}{x^4+y^4} & \textit{if } (x,y)\neq(0,0)\\ 0 & \textit{if } (x,y)=(0,0)\end{cases}$$

Exercise 6.11. *Consider the function* $f(x,y)=\dfrac{xy^2}{x^2+y^4}$ *defined over* $\mathbb{R}^2-\{(0,0)\}$. *Prove that every limit of* $f(x,y)$ *along a line through the origin is zero, but* $\lim\limits_{(x,y)\to(0,0)} f(x,y)$ *does not exist.*

Hint: Approach the origin along the parabola $x=y^2$.

Exercise 6.12. *Consider the function* $f:\mathbb{R}^2-\{(0,0)\}\to\mathbb{R}$ *defined by* $f(x,y)=\dfrac{x^2}{x^2+y^2}$. *Prove:*

(a) $\lim\limits_{(x,y)\to(0,0)} f(x,y)$ *does not exist,*

(b) $\lim\limits_{x\to 0}\lim\limits_{y\to 0} f(x,y)$, *and* $\lim\limits_{y\to 0}\lim\limits_{x\to 0} f(x,y)$ *both exist, but they are not equal.*

6.5 Challenge Problems

Exercise 6.13. *Let* a,b *be two positive integers. Prove that* $\lim\limits_{(x,y)\to(0,0)}\dfrac{x^ay^b}{x^2+y^2}$ *exists if and only if* $a+b>2$.

Exercise 6.14. *Find the limit or show it does not exist:* $\lim\limits_{(x,y)\to(0,0)}\dfrac{xy}{x^2+(y\ln(x^2))^2}$.

Exercise 6.15. *Suppose* $f:\mathbb{R}^2\to\mathbb{R}$ *is a function for which* $\lim\limits_{(x,y)\to(0,0)} f(x,y)=L_1$ *and* $\lim\limits_{x\to 0} f(x,y)=g(y)$ *and* $\lim\limits_{y\to 0} g(y)=L_2$ *all exist. Prove* $L_1=L_2$.

6.6 Summary

- To prove $\lim\limits_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x})=\mathbf{b}$ using the definition:
  - Start with writing down the definition.
  - The objective is to find δ in terms of ϵ.
  - Simplify both $||\mathbf{x}-\mathbf{a}||<\delta$ and $||f(\mathbf{x})-\mathbf{b}||<\epsilon$.

  - You may need to break up the inequality $||f(\mathbf{x}) - \mathbf{b}|| < \epsilon$ into portions that tend to zero, then use $||\mathbf{x} - \mathbf{a}|| < \delta$ to find an inequality for each piece in terms of δ.
  - After you find δ you need to re-write the work as a full solution. Start with "Let $\epsilon > 0$ and set $\delta = \cdots$".

- To find the limit of a function $f = (f_1, \ldots, f_n)$ we find the limit of each of the component functions f_i.
- To find the limit of a function $f : D \to \mathbb{R}$ at $\mathbf{a}$:
  - Find the limit of $f(\mathbf{x})$ as $\mathbf{x}$ approaches $\mathbf{a}$ along different paths.
  - If two of these limits are different or if any of the limits does not exist, then the original limit does not exist.
  - If all limits are the same value b, then we suspect the limit might in fact be b.
  - Then follow the process above and prove the limit is b.
  - Sometimes the Squeeze Theorem could help. In order to create appropriate inequalities polar coordinates may be used.

Week 7

7.1 Topology of $\mathbb{R}^n$

YouTube Video: https://youtu.be/benGbWqYyAQ

Definition 7.1. A subset A of $\mathbb{R}^n$ is called **open** if given any point $\mathbf{a} \in A$, there exists an open ball $B_r(\mathbf{a})$ (with $r > 0$) that is completely contained in A.

Example 7.1. For any positive real number r and any $\mathbf{a} \in \mathbb{R}^n$ the ball $B_r(\mathbf{a})$ is an open subset of $\mathbb{R}^n$.

Theorem 7.1. *Open sets in $\mathbb{R}^n$ satisfy the following properties:*

(a) $\emptyset$ and $\mathbb{R}^n$ are open.

(b) The union of any collection of open sets is open.

(c) The intersection of any finite number of open sets is open.

Example 7.2. By an example show that the intersection of a collection of open subsets of $\mathbb{R}^n$ may not be open.

Definition 7.2. A subset A of $\mathbb{R}^n$ is said to be **closed** if $\mathbb{R}^n - A$ is open.

Example 7.3. Prove that the intervals $[a, b], [a, \infty), (-\infty, b]$ are all closed subsets of $\mathbb{R}$.

Theorem 7.2. *A subset A of $\mathbb{R}^n$ is closed if and only if it contains all of its limit points.*

Theorem 7.3. *Closed subsets of $\mathbb{R}^n$ satisfy the following properties:*

(a) $\emptyset$ and $\mathbb{R}^n$ are closed.

(b) The union of any finite number of closed sets is closed.

(c) The intersection of any collection of closed sets is closed.

7.2 Continuity, Open and Closed Subsets

YouTube Video: https://youtu.be/GUi0DC84Ymc

Theorem 7.4. *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a function.*

(a) f is continuous if and only if given any open subset U of $\mathbb{R}^m$, the inverse image $f^{-1}(U)$ is an open subset of $\mathbb{R}^n$.

(b) f is continuous if and only if given any closed subset C of $\mathbb{R}^m$ the inverse image $f^{-1}(C)$ is a closed subset of $\mathbb{R}^n$.

Example 7.4. Prove that the circle $x^2 + y^2 = 1$ is a closed subset of $\mathbb{R}^2$.

Example 7.5. Prove that every closed ball in $\mathbb{R}^n$ is a closed subset of $\mathbb{R}^n$.

7.3 Compact Subsets of $\mathbb{R}^n$

YouTube Video: https://youtu.be/asmIQiFv2yo

Definition 7.3. A subset A of $\mathbb{R}^n$ is called **compact** if every infinite subset of A has a limit point which lies in A.

Example 7.6. Prove that $\mathbb{R}$, and $(0, 1)$ are not compact.

Definition 7.4. A subset A of $\mathbb{R}^n$ is said to be **bounded** if it lies inside some open ball (of positive radius).

Example 7.7. Prove that a subset of $\mathbb{R}^n$ is bounded if and only if it is inside an open ball centered at the origin.

Theorem 7.5 (Bolzano–Weierstrass Theorem)**.** *A subset of $\mathbb{R}^n$ is compact if and only if it is bounded and closed.*

Theorem 7.6 (The Extreme Value Theorem)**.** *Suppose A is a compact subset of $\mathbb{R}^n$ and $f : A \to \mathbb{R}$ is continuous. Then, f attains its maximum and minimum values. In other words, there exist $\mathbf{x}_0, \mathbf{y}_0 \in A$ for which $f(\mathbf{x}_0) \le f(\mathbf{x}) \le f(\mathbf{y}_0)$ for all $\mathbf{x} \in A$.*

7.4 Curves in $\mathbb{R}^n$

Recall that for a function $f : \mathbb{R} \to \mathbb{R}$, we define its derivative at a by

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

This can also be written as

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

In other words, the value of $f(a + h) - f(a)$ is very close to $f'(a)h$, when h is small. Note that $f'(a)h$ is a linear function in terms of h.

Definition 7.5. Given a function $f : I \to \mathbb{R}^n$, where $I \subseteq \mathbb{R}$ is an open interval, the **derivative** of f at point $a \in I$ is given by

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

If this limit does not exists we say f is not **differentiable** at a. The n-th derivative of f at a, denoted by $f^{(n)}(a)$, is recursively defined as the derivative of $f^{(n-1)}$ at a. Note that for the n-th derivative of f to exist at a, the $(n-1)$-st derivative of f must exist on an open interval centered at a.

Theorem 7.7. *Suppose $f = (f_1, \dots, f_n) : I \to \mathbb{R}^n$ is a function, where $I \subseteq \mathbb{R}$ is an open interval. Then, f is differentiable at a point $a \in I$ if and only if f_j is differentiable at a for all j, $j = 1, \dots, n$. Furthermore, if f is differentiable at a, then $f'(a) = (f_1'(a), \dots, f_n'(a))$.*

Proof. Follows from Theorem 6.2. □

Theorem 7.8 (Properties of Derivatives)**.** *Let a be a number is an open interval I. Suppose $f, g : I \to \mathbb{R}^n$, and $\varphi : I \to \mathbb{R}$ are differentiable at a. Then,*

(a) $(f + g)'(a) = f'(a) + g'(a)$.

(b) $(f \cdot g)'(a) = f'(a) \cdot g(a) + f(a) \cdot g'(a)$. *[Recall that "$\cdot$" denotes the dot product of $\mathbb{R}^n$.]*

(c) $(\varphi f)'(a) = \varphi'(a)f(a) + \varphi(a)f'(a)$.

Theorem 7.9 (The Chain Rule)**.** *Suppose I and J are open intervals, $\varphi : I \to J$ is differentiable at $a \in I$, and $f : J \to \mathbb{R}^n$ is differentiable at $\varphi(a)$. Then $(f \circ \varphi)'(a) = \varphi'(a)f'(\varphi(a))$.*

Definition 7.6. Let I be an open interval, and $f : I \to \mathbb{R}^n$ is a function that is differentiable at a point $a \in I$. The linear function $L : \mathbb{R} \to \mathbb{R}^n$ defined by $L(h) = f'(a)h$ is denoted by df_a, and is called the **differential** of f at a.

Theorem 7.10. *The function $f : I \to \mathbb{R}^n$ is differentiable at some $a \in I$, where I is an open interval, if and only if there exists a linear transformation $L : \mathbb{R} \to \mathbb{R}^n$ such that*

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

Furthermore, when such a linear transformation exists, it is unique and $L(h) = f'(a)h$.

Example 7.8. Evaluate the derivative and the differential of $f(x) = (\sin x, x^2, x + \cos x)$.

Consider the identity function $x : \mathbb{R} \to \mathbb{R}$. We have $dx_a(h) = 1h = h$. If $\varphi : I \to \mathbb{R}$ is differentiable at a point $a \in I$, then $d\varphi_a(h) = \varphi'(a)h$, which means $d\varphi_a(h) = \varphi'(a)dx_a(h)$, or $d\varphi_a = \varphi'(a)dx_a$. This is quite similar to the notation $\varphi'(x) = \dfrac{d\varphi}{dx}$.

If $f : \mathbb{R} \to \mathbb{R}^n$ is differentiable at $\varphi(a)$, then

$$d(f \circ \varphi)_a(h) = (f \circ \varphi)'(a)h = f'(\varphi(a))\varphi'(a)h = df_{\varphi(a)}(d\varphi_a(h)) = df_{\varphi(a)} \circ d\varphi_a(h).$$

Therefore,

$$d(f \circ \varphi)_a = df_{\varphi(a)} \circ d\varphi_a.$$

Definition 7.7. Let $I \subseteq \mathbb{R}$ be an interval. A curve $f : I \to \mathbb{R}^n$ is called k **times continuously differentiable** or C^k if it satisfies all of the following:

(a) f has k derivatives at all points inside I except at its endpoints, if any,

(b) $f^{(k)}$ is continuous, and

(c) f is continuous on I.

We say f is **smooth** or C^∞ if it is C^k for every $k \in \mathbb{Z}^+$. We say f is **piecewise smooth** if I can be partitioned into finitely many intervals $I_1, I_2, \ldots, I_m$, where the restriction of f to each I_j is a smooth curve. A curve $f : [a, b] \to \mathbb{R}^n$ is called **closed** if $f(a) = f(b)$. A curve $f : I \to \mathbb{R}^n$ is called **simple** if it is one-to-one except possibly at its endpoints. In other words, if $f(s) = f(t)$ for some $s, t \in I$ then, $s = t$ or s, t are endpoints of I.

Example 7.9. Show $r(t) = (\cos t, \sin t, 1), t \in \mathbb{R}$ is smooth.

Theorem 7.11. *Suppose $f : [a, b] \to \mathbb{R}^n$ is a simple, C^1 curve. The arc length of this curve is given by* $\displaystyle\int_a^b ||f'(t)||\ dt$.

Example 7.10. Find the length of the curve $f : \mathbb{R} \to \mathbb{R}^3$ given by $f(t) = (e^t, e^{-t}, \sqrt{2}\,t)$.

7.5 More Examples

Example 7.11. Prove that every open ball in $\mathbb{R}$ is an open interval and every open interval is an open ball.

Solution. Consider an open ball $B_r(a)$ with $r > 0$ and $a \in R$. By definition $x \in B_r(a)$ if and only if $|x - a| < r$, which is equivalent to $-r < x - a < r$ which is equivalent to $a - r < x < r + a$ which is equivalent to $x \in (a - r, a + r)$, by definition of an open interval. Therefore, $B_r(a) = (a - r, a + r)$.

Now, consider an open interval (a, b). We will show $(a, b) = B_r(c)$, where $c = \dfrac{a+b}{2}$ and $r = \dfrac{b-a}{2}$. Note that $r > 0$ since $b > a$.

By definition of an open interval $x \in (a, b)$ if and only if $a < x < b$. This is equivalent to

$$a - \frac{a+b}{2} < x - \frac{a+b}{2} < b - \frac{a+b}{2} \iff -\frac{b-a}{2} < x - \frac{a+b}{2} < \frac{b-a}{2} \iff |x - c| < r \iff x \in B_r(c).$$

Therefore, $(a, b) = B_r(c)$, as desired. $\square$

Example 7.12. Prove that every finite subset of $\mathbb{R}^n$ is closed.

Solution. Since every finite subset of $\mathbb{R}^n$ is a finite union of sets of the form $\{\mathbf{x}\}$, by Theorem 7.3 it is enough to show $\{\mathbf{x}\}$ is closed for every $\mathbf{x} \in \mathbb{R}^n$. We will show its complement is open. If $\mathbf{y} \neq \mathbf{x}$, then let $r = ||\mathbf{y} - \mathbf{x}||$. By positivity of norm, We have $r > 0$. We will show that $\mathbf{x} \notin B_r(\mathbf{y})$. Otherwise $||\mathbf{y} - \mathbf{x}|| < r = ||\mathbf{y} - \mathbf{x}||$, which is a contradiction. Therefore, $B_r(\mathbf{y}) \subseteq \mathbb{R}^n - \{\mathbf{x}\}$. By definition of open sets, $\mathbb{R}^n - \{\mathbf{x}\}$ is open. By definition of closet sets $\{\mathbf{x}\}$ is closed, as desired. $\square$

Example 7.13. Prove that every ellipse $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = 1$ is a compact subset of $\mathbb{R}^2$.

Solution. We need to show this curve is closed and bounded. By definition of the inverse image, such an ellipse is $f^{-1}(\{1\})$, where $f : \mathbb{R}^2 \to \mathbb{R}$ is given by $f(x, y) = \dfrac{x^2}{a^2} + \dfrac{y^2}{b^2}$. Since f is a polynomial, it is continuous. Since $\{1\}$ is a finite set, it is closed. Therefore, by Theorem 7.4, this inverse image is closed.

Now, note that Since $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = 1$ we have $x^2 \le a^2$ and $y^2 \le b^2$, and thus $\sqrt{x^2 + y^2} \le \sqrt{a^2 + b^2}$, which means the ellipse lies inside an open ball of radius $\sqrt{a^2 + b^2} + 1$ centered at the origin, which means it is bounded. Thus, the ellipse is compact. $\square$

Example 7.14. Prove that the only subspace of $\mathbb{R}^n$ that is open is $\mathbb{R}^n$ itself.

Sketch. We know every subspace V contains the origin. Since the subspace is open it must contain a ball around the origin, but a ball contains all directions, e.g. some multiple of $\mathbf{e}_1$must be in the ball. Since V is a subspace, it must contain $\mathbf{e}_1$. Similarly V contains all $\mathbf{e}_i$'s. Since V is a subspace, it must be $\mathbb{R}^n$.

Solution. Suppose V is a subspace of $\mathbb{R}^n$ that is open. We know $\mathbf{0} \in V$, since it is a subspace. Therefore, there is $r > 0$ for which $B_r(\mathbf{0}) \subseteq V$. We see that for every i we have $||\frac{r}{2}e_i|| = \frac{r}{2}||e_i|| = \frac{r}{2} < r$, and thus $\frac{r}{2}e_i \in V$. Since V is closed under scalar multiplication we have $e_i \in V$. Therefore, V contains the span of $e_1, \ldots, e_n$, which is $\mathbb{R}^n$. Thus, $V = \mathbb{R}^n$. $\square$

Example 7.15. Prove that the graph of a continuous function $f : \mathbb{R} \to \mathbb{R}$ defined by

$$\Gamma_f = \{(x, f(x)) \mid x \in \mathbb{R}\}$$

is a closed subset of $\mathbb{R}^2$.

Solution. Consider the function $g : \mathbb{R}^2 \to \mathbb{R}$ defined by $g(x, y) = y - f(x)$. Note that $(x, y) \in \Gamma_f$ if and only if $y = f(x)$ if and only if $g(x, y) = 0$. Therefore, $\Gamma_f = g^{-1}(\{0\})$. By Example 7.12, the set $\{0\}$ is a closed subset of $\mathbb{R}$. So, if we show g is continuous, then by Theorem 7.4 the set $f^{-1}(\{0\})$ is closed. Note that g is a difference of two continuous functions ($y = \pi_2(x, y)$ and $f(x) = f(\pi_1(x, y))$. Therefore, it is continuous. This completes the proof. $\square$

Example 7.16. Suppose $f, g : \mathbb{R}^n \to \mathbb{R}^m$ are continuous functions. Prove the following set is a closed subset of $\mathbb{R}^n$:

$$\{\mathbf{x} \in \mathbb{R}^n \mid f(\mathbf{x}) = g(\mathbf{x})\}.$$

Solution. Let S be the set given above. We note that $\mathbf{x} \in S$ if and only if $f(\mathbf{x}) = g(\mathbf{x})$ if and only if $f(\mathbf{x}) - g(\mathbf{x}) = \mathbf{0}$. Therefore, $S = (f-g)^{-1}(\{0\})$. Since f and g are continuous, their difference $f - g$ is also continuous. We know $\{0\}$ is closed by Example 7.12. Thus, by Theorem 7.4 the set $(f-g)^{-1}(\{0\})$ is closed. Therefore, S is closed, as desired. $\square$

Example 7.17. Suppose $\mathbf{x}_0$ is a point in $\mathbb{R}^n$ and D is a nonempty compact subset of $\mathbb{R}^n$. Prove that there exists a closest point $\mathbf{y}_0 \in D$ to x_0. In other words $||\mathbf{x}_0 - \mathbf{y}_0|| \leq ||\mathbf{x}_0 - \mathbf{y}||$ for all $\mathbf{y} \in D$.

Solution. Define $f : \mathbb{R}^n \to \mathbb{R}$ by $f(\mathbf{x}) = ||\mathbf{x}_0 - \mathbf{x}||$. Let $\mathbf{x}_0 = (a_1, \ldots, a_n)$. Note that this function is

$$f(x_1, \ldots, x_n) = \sqrt{(a_1 - x_1)^2 + \cdots + (a_n - x_n)^2},$$

which is a composition of a polynomial and the square root function, and thus it is continuous. By a theorem, $f(D)$ must have a minimum value. Suppose this minimum value is $f(\mathbf{y}_0)$. This means for all $\mathbf{y} \in D$, we have $f(\mathbf{y}_0) \leq f(\mathbf{y})$. This is the same as $||\mathbf{x}_0 - \mathbf{y}_0|| \leq ||\mathbf{x}_0 - \mathbf{y}||$, as desired. $\square$

Example 7.18. Prove that the union and intersection of any finite number of compact sets is compact.

Solution. Suppose $A_1, \ldots, A_m$ are compact subset of $\mathbb{R}^n$. By Theorem 7.5 each A_j is bounded and closed. We need to show $\bigcap_{j=1}^{m} A_j$ and $\bigcup_{j=1}^{m} A_j$ are both closed and bounded.

By Theorem 7.3 both sets $\bigcap_{j=1}^{m} A_j$ and $\bigcup_{j=1}^{m} A_j$ are closed.

Suppose for every j the set A_j is contained in the ball $B_{r_j}(\mathbf{0})$. Consider $r = \max(r_1, \ldots, r_m)$. Thus, for all j we have $A_j \subseteq B_{r_j}(\mathbf{0}) \subseteq B_r(\mathbf{0})$. Therefore, both the union and intersection of A_j's are in $B_r(\mathbf{0})$. Thus, $\bigcap_{j=1}^{m} A_j$ and $\bigcup_{j=1}^{m} A_j$ are both bounded. $\square$

Example 7.19. Prove that the intersection of a closed subset of $\mathbb{R}^n$ and a compact subset of $\mathbb{R}^n$ is compact.

Solution. Let A be a closed and B be a compact subset of $\mathbb{R}^n$. By the Bolzano-Weierstrass Theorem, B is closed. Therefore, by Theorem 7.3, the set $A \cap B$ is also closed. Since B is compact, it is bounded and thus there is a balls $B_r(\mathbf{p})$ that contains B. Since $A \cap B$ is a subset of B, the ball $B_r(\mathbf{p})$ contains $A \cap B$. Therefore, $A \cap B$ is bounded. This implies $A \cap B$ is both bounded and closed. Thus, by the Bolzano-Weirestrass Theorem $A \cap B$ is compact. $\square$

Example 7.20. Prove that the function $f(x, y) = x^4 + 3xy + y^4$ attains its maximum and minimum values over the circle $x^2 + y^2 = 1$.

Solution. By the Extreme Value Theorem, it is enough to show f is continuous and the circle $x^2 + y^2 = 1$ is compact. Note that f, as a polynomial, is continuous. The given circle lies in the open ball $B_2(0, 0)$, since every point on the circle satisfies $x^2 + y^2 < 4$. Also, the circle can be describes as $g^{-1}(\{1\})$, where $g(x, y) = x^2 + y^2$. Note that g is continuous, since it is a polynomial, and the set $\{1\}$ is closed, since it is finite. Thus, by Theorem 7.4 the given circle is closed. Therefore, the given circle is compact. $\square$

Example 7.21. Prove that the function

$$f(x, y, z) = \sin(x + 2y + 3z) + \cos(z) + \sin(x - y) + \cos(x + y)$$

attains its maximum and minimum values over $\mathbb{R}^3$.

Solution. First, note that

$$f(x + 2\pi k, y + 2\pi\ell, z + 2\pi m) = f(x, y, z), \qquad \forall k, \ell, m \in \mathbb{Z}.$$

This means all functional values can be obtained by assuming $x, y, z \in [0, 2\pi]$. Therefore, we can consider the function f over the cube C given by $0 \le x, y, z \le 2\pi$. This cube is bounded since every point in the cube satisfies $x^2 + y^2 + z^2 \le 3 \times 4\pi^2 < 13\pi^2$. This cube is the intersection of the sets

$$M = \{(x, y, z) \in \mathbb{R}^3 \mid 0 \le x \le 2\pi\}, N = \{(x, y, z) \in \mathbb{R}^3 \mid 0 \le y \le 2\pi\}, \text{ and } P = \{(x, y, z) \in \mathbb{R}^3 \mid 0 \le z \le 2\pi\}.$$

The inequality $0 \le x \le 2\pi$ can be described by $0 \le \pi_1(x, y, z) \le 2\pi$, where $\pi_1 : \mathbb{R}^3 \to \mathbb{R}$ is given by $\pi_1(x, y, z) = x$. Hence $M = \pi_1^{-1}([0, 2\pi])$. Since $[0, 2\pi]$ is closed, by Theorem 7.4 the set M is closed. Similarly N and P are also closed subsets of $\mathbb{R}^3$. The intersection of these three sets gives us the desired cube C. Therefore, by Theorem 7.3 the set C is closed. Thus, C is compact by Theorem 7.5. Since f is continuous, f attains its maximum and minimum values by the Extreme Value Theorem. $\square$

Example 7.22. Given a real number a, find the derivative and the differential of each of the following functions at a:

(a) $f(x) = (1 + x, e^x, \sin(2x))$.

(b) $g(x) = (x^2, 3, x)$.

(c) $h(t) = (1 + t^2, 2t - \cos t, \sqrt{1 + t^2})$.

Solution. (a) The derivative of f is $f'(a) = (1, e^a, 2\cos(2a))$. Its differential is the function $df_a : \mathbb{R} \to \mathbb{R}^3$ given by $df_a(h) = (h, e^a h, 2h\cos(2a))$.

(b) The derivative of g is $g'(a) = (2a, 0, 1)$. Its differential is the function $dg_a : \mathbb{R} \to \mathbb{R}^3$ given by $dg_a(h) = (2ah, 0, h)$.

(c) The derivative of h is $h'(a) = (2a, 2 + \sin a, 1/2 \cdot (1 + a^2)^{-1/2}(2a)) = (2a, 2 + \sin a, a/\sqrt{1 + a^2})$. The differential is a function $df_a : \mathbb{R} \to \mathbb{R}^3$ given by $df_a(h) = (2ah, 2h + h\sin a, ah/\sqrt{1 + a^2})$. $\square$

Example 7.23. Prove that the curve $f : \mathbb{R} \to \mathbb{R}^2$ given by

$$f(t) = \begin{cases} (t^2 \sin(1/t), \sin t) & \text{if } t \neq 0 \\ (0, 0) & \text{otherwise} \end{cases}$$

is differentiable, but not continuously differentiable.

Solution. First, note that $\sin t$ is differentiable on $\mathbb{R}$ and $t^2 \sin(1/t)$ is differentiable everywhere except possibly at $t = 0$. We need will now find the derivative of the first component of f (call it f_1) at $t = 0$.

$$f_1'(0) = \lim_{h \to 0} \frac{f(h) - f(0)}{h} = \lim_{h \to 0} h \sin(1/h).$$

We see

$$|h \sin(1/h)| \leq |h| \Rightarrow -|h| \leq h \sin(1/h) \leq |h|.$$

Since $|h|$ and $-|h|$ are continuous at $h = 0$, we have $\lim_{h \to 0} |h| = \lim_{h \to 0} -|h| = 0$. By the Squeeze Theorem, $\lim_{h \to 0} h \sin(1/h) = 0$. Therefore, $f_1'(0) = 0$. This means f' is differentiable on $\mathbb{R}$.

We will show f' is not continuous at $t = 0$. By Theorem7.7, for every $t \neq 0$ we have

$$f'(t) = (2t \sin(1/t) - \cos(1/t), \cos t).$$

Setting $t_n = \dfrac{1}{\pi n}$ for every positive integer n we see $f_1'(t_n) = -\cos(\pi n)$. This is -1, when n is even and 1 when n is odd. Therefore, $\lim_{t \to 0} f_1'(t)$ does not exist, and hence f' is not continuous. This means f is not continuously differentiable. $\square$

Example 7.24. Prove that curves $f, g : \mathbb{R} \to \mathbb{R}^2$ defined by $f(t) = (t, t)$ and $g(t) = (\sqrt[3]{t}, \sqrt[3]{t})$ have the same trace (i.e. image), they are both simple, even though f is smooth and g is not.

Solution. Any point in the trace of f is of the form $f(t) = (t, t) = g(t^3)$. Thus trace of f is a subset of trace of g. Any point in the trace of g is of the form $g(t) = (\sqrt[3]{t}, \sqrt[3]{t}) = f(\sqrt[3]{t})$. Therefore, trace of g is a subset of trace of f. This means, trace of these two curves are the same. Note also that both f and g are one-to-one.

t as a polynomial is C^∞. Thus f is C^∞, i.e. smooth. $\sqrt[3]{t}$ is not differentiable at $t = 0$. Thus g is not smooth. $\square$

Example 7.25. Find the arc length of the curve given by $f : [0, 1] \to \mathbb{R}^4$,

$$f(t) = (t^{3/2}, \cos t, \sin t, t)$$

Solution. First, note that functions given by $\cos t, \sin t, t$ are all C^1. Also, $t^{3/2}$ is continuous on $[0, \infty)]$ and its derivative $\dfrac{3}{2} t^{1/2}$ is continuous on $(0, \infty)$. So f is C^1. Note also that $f(a) = f(b)$ implies $a = b$, so f is simple. Therefore, the arc length of f is given by

$$\int_0^1 ||f(t)||\ dt = \int_0^1 \sqrt{\frac{9}{4} t + \sin^2 t + \cos^2 t + 1}\ dt = \frac{1}{2} \int_0^1 \sqrt{9t + 8}\ dt = \frac{1}{27} (9t + 8)^{3/2}]_{t=0}^{t=1}.$$

So, the answer is $\dfrac{17\sqrt{17}-16\sqrt{2}}{27}$. □

Example 7.26. Prove that the following set is a compact subset of $\mathbb{R}^4$:

$$\{(x_1, x_2, x_3, x_4) \in \mathbb{R}^4 \mid x_1^2 + 4x_2^6 + \sin(x_1 + x_2x_4) + x_3^2 + x_4^4 \leq 7\}.$$

Solution. Call the given set S. Consider the function $f(x_1, x_2, x_3, x_4) = x_1^2 + 4x_2^6 + \sin(x_1 + x_2x_4) + x_3^2 + x_4^4$. This function is continuous as it is a composition of continuous functions. The set S, by definition, is the same as $f^{-1}((-\infty, 7])$. By Theorem 7.4, since $(-\infty, 7]$ is a closed subset of $\mathbb{R}$, the set S is closed. On the other hand, since $-1 \leq \sin(x_1 + x_2x_4)$, for every $(x_1, x_2, x_3, x_4) \in S$ we have

$$x_1^2 + 4x_2^6 - 1 + x_3^2 + x_4^4 \leq 7 \Rightarrow x_1^2 + 4x_2^6 + x_3^2 + x_4^4 \leq 8.$$

Since all the terms on the left side of the latter inequality are nonnegative, $x_1^2 \leq 8$, $x_2^2 \leq \sqrt[3]{2} < 2$, $x_3^2 \leq 8$, and $x_4^2 \leq \sqrt{8} < 3$. Therefore, $x_1^2 + x_2^2 + x_3^2 + x_4^2 < 21$. Therefore, $S \subseteq B_{\sqrt{21}}(\mathbf{0})$. Hence, S is bounded. Therefore, by the Bolzano-Weierstrass Theorem, S is compact. □

7.6 Exercises

Exercise 7.1. *Determine if each of the following sets is closed, open or neither.*

(a) $\{(x, y) \in \mathbb{R}^2 \mid 1 \leq x^2 + y^2 \leq 2\}$.

(b) $\mathbb{Q}^2$ *as a subset of* $\mathbb{R}^2$.

(c) $\{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid x_i \leq 0 \text{ for some } i\}$.

(d) $\{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid \text{none of } x_1, \ldots, x_n \text{ is an integer}\}$.

Exercise 7.2. *Consider the set*

$$A = \{(x, y) \in \mathbb{R}^2 \mid x \geq 0, \text{ and } y > 0\}.$$

(a) Geometrically sketch this set and explain if it is closed, open or neither closed nor open.

(b) Carefully prove your claim in part (a).

Exercise 7.3. *Prove that the intersection of a closed subset of* $\mathbb{R}^n$ *and a compact subset of* $\mathbb{R}^n$ *is compact.*

Exercise 7.4. *Suppose* $f : \mathbb{R} \to \mathbb{R}^n$ *is differentiable. Prove that* $||f(t)||$ *is constant, if and only if* $f(t)$ *and* $f'(t)$ *are orthogonal for every* $t \in \mathbb{R}$.

Hint: $||f(t)||^2 = f(t) \cdot f(t)$.

Exercise 7.5. *(a) Prove that every nonempty open subset of* $\mathbb{R}^n$ *is a union of a collection of balls; all of which have a rational radius.*

(b) *Prove that every nonempty open subset of* $\mathbb{R}^n$ *is a union of a collection of balls; all of which have an irrational radius.*

Exercise 7.6. *Prove part (b) of Theorem 7.4: Let* $f : \mathbb{R}^n \to \mathbb{R}^m$ *be a function. Then,* f *is continuous if and only if given any closed subset* C *of* $\mathbb{R}^m$ *the inverse image* $f^{-1}(C)$ *is a closed subset of* $\mathbb{R}^n$.

Exercise 7.7. *Let* D *be a nonempty compact subset of* $\mathbb{R}^n$. *For every* $\mathbf{x} \in \mathbb{R}^n$ *let* $f(\mathbf{x})$ *be the minimum distance between* $\mathbf{x}$ *and points of* D. *(See Example 7.17). Prove that* $f : \mathbb{R}^n \to \mathbb{R}$ *is continuous.*

Hint: Use the $\epsilon - \delta$ definition of limit.

Exercise 7.8. *Prove that every subspace of* $\mathbb{R}^n$ *is a closed subset of* $\mathbb{R}^n$.

Hint: Write down a linear transformation whose kernel is the given subspace of $\mathbb{R}^n$.

Definition 7.8. Let A be a subset of $\mathbb{R}^n$. The point $\mathbf{a}$ is said to be a **boundary point** of a set A if every open ball centered at $\mathbf{a}$ contains at least one point that is in A and at least one point that is outside of A. The set of boundary points of A is denoted by ∂A and is called the **boundary** of A.

Exercise 7.9. *Prove that for every subset* A *of* $\mathbb{R}^n$, *the boundary of* A *and the boundary of its complement* A^c *are the same.*

Exercise 7.10. *Prove that the boundary of every subset of* $\mathbb{R}^n$ *is a closed subset of* $\mathbb{R}^n$.

Exercise 7.11. *Suppose* $A = \{\mathbf{a}_1, \mathbf{a_2}, \ldots\}$ *is a subset of* $\mathbb{R}^m$ *for which* $||\mathbf{a}_n|| \geq n$ *for all* $n \geq 1$. *Prove that* A *has no limit points.*

Exercise 7.12. *Let* A, B *be two nonempty subset of* $\mathbb{R}^n$, *and* $\mathbf{x} \in \mathbb{R}^n$. *Define*

$$\mathbf{x} + A = \{\mathbf{x} + \mathbf{a} \mid \mathbf{a} \in A\}, \text{ and } A + B = \{\mathbf{a} + \mathbf{b} \mid \mathbf{a} \in A, \text{ and } \mathbf{b} \in B\}.$$

(a) *Prove that if* A *is open, then so is* $\mathbf{x} + A$.

(b) *Prove that if* A *is open, then so is* $A + B$.

(c) *Prove that if* A *is closed, then so is* $\mathbf{x} + A$.

(d) *Prove that if* A *is closed and* B *is finite, then* $A + B$ *is also closed.*

(e) *Prove that if* A *and* B *are bounded, then so is* $A + B$.

(f) *With an example show that it is possible that both* A *and* B *are closed but* $A + B$ *is not.*

Definition 7.9. Let $\mathbf{a} \in \mathbb{R}^n$ and r be a positive real number. A **sphere** of radius r centered at $\mathbf{a}$, denoted by $S_r(\mathbf{a})$, is given by

$$S_r(\mathbf{a}) = \{\mathbf{x} \in \mathbb{R}^n \mid ||\mathbf{x} - \mathbf{a}|| = r\}.$$

Exercise 7.13. *Prove that for every* $r > 0$ *and every* $\mathbf{a} \in \mathbb{R}^n$ *we have* $S_r(\mathbf{a}) = \overline{B}_r(\mathbf{a}) - B_r(\mathbf{a})$. *Deduce that every sphere is closed.*

Exercise 7.14. *Find all constants a, b, c for which the following represents a sphere in $\mathbb{R}^3$:*

$$ax^2 + (2a - b)y^2 + z^2 + 2ax + 2y + c = 0.$$

Exercise 7.15. *Consider the point $A = (1, 2, 0) \in \mathbb{R}^3$. Find all constants λ for which the set of points P whose distance to the origin is λ times their distance to A is a sphere.*

Exercise 7.16. *Prove Theorem 7.8.*

Exercise 7.17. *Find the derivative and differential of each function given below over its domain.*

(a) $f(t) = (t^3, \tan t, \sqrt{1+t^2})$.

(b) $g(t) = (t, t^2 + 1, 2t)$.

(c) $h(t) = (1 - t, t^2)$.

Exercise 7.18. *Suppose $f, g : \mathbb{R} \to \mathbb{R}^n$ are two differentiable curves and $(s_0, t_0) \in \mathbb{R}^2$ is a point for which the points $f(t_0)$ and $g(s_0)$ are closer than any other points on the two curves. Prove that $f(t_0) - g(s_0)$ is orthogonal to both $f'(t_0)$ and $g'(s_0)$. Use this fact to find the closest distance between lines $f(t) = (t+1, t, t-1)$ and $g(s) = (2s, s-1, 2s+1)$. You may assume this minimum distance exists.*

Hint: Show that t_0 must be a critical point of $||f(t) - g(s_0)||^2$.

Exercise 7.19. *Suppose $A_1, \ldots, A_m$ are open subsets of $\mathbb{R}^n$ and C is a closed subset of $\mathbb{R}^n$. Prove that $(\bigcup_{i=1}^{m} A_i) - C$ is an open subset of $\mathbb{R}^n$.*

Exercise 7.20. *Give an example of a continuous function $f : \mathbb{R} \to \mathbb{R}$, a closed subset C of $\mathbb{R}$ and an open subset U of $\mathbb{R}$ for which $f(C)$ is not closed, and $f(U)$ is not open.*

Exercise 7.21. *Give examples of continuous functions $f : \mathbb{R}^2 \to \mathbb{R}$ and subsets A of $\mathbb{R}$ that each of the following holds:*

(a) A is not open, but $f^{-1}(A)$ is open.

(b) A is not closed, but $f^{-1}(A)$ is closed.

(c) A is compact, but $f^{-1}(A)$ is not compact.

Exercise 7.22. *Suppose U is a nonempty open subset of $\mathbb{R}^n$ for which there is a positive real number r such that $B_r(\mathbf{x}) \subseteq U$ for all $\mathbf{x} \in U$. Prove that $U = \mathbb{R}^n$.*

Exercise 7.23. *Find the following limit or show it does not exist:*

$$\lim_{(x,y)\to(0,0)} (x^3 + y^4)\frac{e^{-x^2-y^2}}{x^2 + y^2}.$$

Exercise 7.24. *Consider the following function*

$$f(x,y) = \begin{cases} \dfrac{|y|e^{-|y|/(x^2+y^2)}}{x^2+y^2} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{otherwise} \end{cases}$$

Prove the limit of $f(x,y)$ as $(x,y) \to (0,0)$ along any line through the origin is zero, however $\lim\limits_{(x,y)\to(0,0)} f(x,y)$ does not exist.

Exercise 7.25. *Suppose A is a subset of $\mathbb{R}^n$. Prove:*

(a) If $A \cap U$ is open for every open subset U of $\mathbb{R}^n$, then A is open.

(b) If $A \cup U$ is open for every open subset U of $\mathbb{R}^n$, then A is open.

Prove the analogous of the above statements when "open" is replaced by "closed".

Exercise 7.26. *Determine if each of the following limits exist:*

(a) $\lim\limits_{(x,y)\to(0,0)} \dfrac{x^2+y^2}{x^2+|y|}$.

(b) $\lim\limits_{(x,y)\to(0,0)} \dfrac{\tan(x^2+y^2)}{x^2+y^2}$.

Exercise 7.27. *Evaluate* $\lim\limits_{(x,y)\to(0,0)} \dfrac{y^4}{x^2+y^4}$ *or show it does not exist.*

Exercise 7.28. *Suppose $f : \mathbb{R} \to \mathbb{R}$ is a periodic continuous function. Prove the image of f is bounded.*

7.7 Challenge Problems

Exercise 7.29. *Suppose $A \subseteq \mathbb{R}^n$ and $B \subseteq \mathbb{R}^m$. Considering $A \times B$ as a subset of $\mathbb{R}^{m+n}$, prove the following:*

(a) If A and B are both open, then $A \times B$ is open.

(b) If A and B are both closed, then $A \times B$ is closed.

(c) If A and B are both compact, then $A \times B$ is compact.

Exercise 7.30. *In this exercise we will prove that $\mathbb{R}^n$ has no nonempty, proper subset that is both open and closed. On the contrary, assume $\emptyset \neq U \neq \mathbb{R}^n$ is both open and closed and set $V = \mathbb{R}^n - U$.*

(a) Let $\mathbf{u} \in U$, $\mathbf{v} \in V$, and r be a real number with $\max(||\mathbf{u}||, ||\mathbf{v}||) < r$. Prove that $U \cap \overline{B}_r(\mathbf{0})$ and $V \cap \overline{B}_r(\mathbf{0})$ are both compact and nonempty.

(b) Define $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ by $f(\mathbf{x}, \mathbf{y}) = ||\mathbf{x} - \mathbf{y}||$. Show that f is continuous.

(c) Deduce there are points $\mathbf{x} \in U \cap \overline{B}_r(\mathbf{0})$ and $\mathbf{y} \in V \cap \overline{B}_r(\mathbf{0})$ that are closest among all points of the two subsets.

(d) Using the fact that $\dfrac{\mathbf{x}+\mathbf{y}}{2}$ must be either in U or in V obtain a contradiction.

Exercise 7.31. *Suppose $A \in M_{m\times n}(\mathbb{R})$. Prove there is a real number C for which both of the following happens:*

(a) $||A\mathbf{x}|| \leq C||\mathbf{x}||$ for all $\mathbf{x} \in \mathbb{R}^n$, and

(b) There is some nonzero point $\mathbf{a} \in \mathbb{R}^n$ for which $||A\mathbf{a}|| = C||\mathbf{a}||$.

7.8 Summary

- To prove A is open start with an arbitrary $\mathbf{a} \in A$ and show there is an open ball $B_r(\mathbf{a})$ that completely lies in A.
- $B_r(\mathbf{a})$ is open.
- To prove A is closed, either show its complement is open or show all limit points of A belong to A.
- To show A is bounded prove there is r for which $||\mathbf{a}|| < r$ for all $\mathbf{a} \in A$.
- To show A is compact, show it is closed and bounded.
- To show a function $f : A \to \mathbb{R}$ attains its maximum and minimum values:
  - Show A is closed and bounded, i.e. compact.
  - Show f is continuous.
  - Invoke the Extreme Value Theorem to conclude f attains its maximum and minimum values.
- The derivative of a function $f : \mathbb{R} \to \mathbb{R}^n$ is given by differentiating each coordinate function of f.
- The differential of f at $\mathbf{a}$ is a linear transformation $df_{\mathbf{a}} : \mathbb{R} \to \mathbb{R}^n$ given by $df_{\mathbf{a}}(h) = f'(\mathbf{a})h$. This linear transformation is the only linear transformation L that satisfies the following:
$$\lim_{h\to 0} \frac{f(a+h) - f(a) - L(h)}{h} = 0.$$

Week 8

8.1 Directional Derivatives

Definition 8.1. Let $\mathbf{a}$ be a point in an open subset U of $\mathbb{R}^n$ and $\mathbf{F} : U \to \mathbb{R}^m$ be a function, and $\mathbf{0} \neq \mathbf{v} \in \mathbb{R}^n$. The **directional derivative of F with respect to $\mathbf{v}$ at $\mathbf{a}$** is defined by:

$$D_{\mathbf{v}}\mathbf{F}(\mathbf{a}) = \lim_{h \to 0} \frac{\mathbf{F}(\mathbf{a} + h\mathbf{v}) - \mathbf{F}(\mathbf{a})}{h}.$$

When $\mathbf{v} = \mathbf{e}_i$, this directional derivative is denoted by

$$D_{\mathbf{e}_i}\mathbf{F}(\mathbf{a}) = D_i\mathbf{F}(\mathbf{a}) = \frac{\partial \mathbf{F}}{\partial x_i}(\mathbf{a}) = \mathbf{F}_{x_i}(\mathbf{a}).$$

This is called the i-th partial derivative of $\mathbf{F}$ at $\mathbf{a}$.

Example 8.1. Evaluate the partial derivatives of $x^2 + xy - y^3$.

Remark. To evaluate the partial derivative of a function with respect to x_i, we fix all other variables and differentiate with respect to x_i.

Example 8.2. Evaluate the directional derivative of the following function with respect to the vector $(1, 2)$ at the origin:

$$\mathbf{F}(x, y) = \begin{cases} \dfrac{x^3}{x^2 + y^2} & \text{if } (x, y) \neq (0, 0) \\ 0 & \text{otherwise} \end{cases}$$

Theorem 8.1. *Let U be an open subset of $\mathbb{R}^n$. Given a function $\mathbf{F} : U \to \mathbb{R}^m$, a vector $\mathbf{0} \neq \mathbf{v} \in \mathbb{R}^n$, a point $\mathbf{a} \in U$, and $0 \neq c \in \mathbb{R}$, we have $D_{c\mathbf{v}}\mathbf{F}(\mathbf{a}) = cD_{\mathbf{v}}\mathbf{F}(\mathbf{a})$.*

8.2 Derivative and Differential

We know from the definition of directional derivative that

$$\lim_{h \to 0} \frac{\mathbf{F}(\mathbf{a} + h\mathbf{v}) - \mathbf{F}(\mathbf{a}) - hD_{\mathbf{v}}\mathbf{F}(\mathbf{a})}{h} = \mathbf{0}.$$

This brings us to the following definition:

Definition 8.2. Let $\mathbf{a}$ be a point in an open subset U of $\mathbb{R}^n$. We say $\mathbf{F} : U \to \mathbb{R}^m$ is **differentiable at a** if there exists a linear transformation $L : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{\mathbf{F}(\mathbf{a}+\mathbf{h}) - \mathbf{F}(\mathbf{a}) - L(\mathbf{h})}{||\mathbf{h}||} = \mathbf{0}.$$

Theorem 8.2. *If a linear transformation L in the previous definition exists, then it is unique.*

Definition 8.3. The linear transformation in the above theorem is called the **differential of F at a**, and is denoted by $d\mathbf{F}_{\mathbf{a}}$. Its matrix is called the **derivative of F at a**, and is denoted by $\mathbf{F}'(\mathbf{a})$.

Remark. Suppose U is an open subset of $\mathbb{R}^n$. Let $\mathbf{F} : U \to \mathbb{R}^m$ be a function that is differentiable at some $\mathbf{a} \in U$. Then $d\mathbf{F}_{\mathbf{a}} : \mathbb{R}^n \to \mathbb{R}^m$ is linear and its matrix $\mathbf{F}'(\mathbf{a})$ is an $m \times n$ matrix for which $d\mathbf{F}_{\mathbf{a}}(\mathbf{h}) = \mathbf{F}'(\mathbf{a})\mathbf{h}$, where $\mathbf{h}$ is a column vector in $\mathbb{R}^n$.

Theorem 8.3. *Suppose $\mathbf{a}$ is a point in an open subset U of $\mathbb{R}^n$ and $c \in \mathbb{R}$ is a scalar. Assume $\mathbf{F}, \mathbf{G} : U \to \mathbb{R}^m$ are both differentiable at $\mathbf{a}$. Then, $\mathbf{F} + \mathbf{G}$ and $c\mathbf{F}$ are both differentiable at $\mathbf{a}$. Furthermore,*

(a) $(\mathbf{F} + \mathbf{G})'(\mathbf{a}) = \mathbf{F}'(\mathbf{a}) + \mathbf{G}'(\mathbf{a})$*, and*

(b) $(c\mathbf{F})'(\mathbf{a}) = c\mathbf{F}'(\mathbf{a})$.

Theorem 8.4. *Let $\mathbf{a}$ be a point in an open subset U of $\mathbb{R}^n$. If $\mathbf{F} = (F_1, \ldots, F_m) : U \to \mathbb{R}^m$ is differentiable at $\mathbf{a}$, then*

$$D_{\mathbf{v}}\mathbf{F}(\mathbf{a}) = d\mathbf{F}_{\mathbf{a}}(\mathbf{v}) = \mathbf{F}'(\mathbf{a})\mathbf{v}.$$

Furthermore, the (i, j) entry of $\mathbf{F}'(\mathbf{a})$ is $\dfrac{\partial F_i}{\partial x_j}(\mathbf{a})$. In other words,

$$\mathbf{F}'(\mathbf{a}) = \begin{pmatrix} \dfrac{\partial F_1}{\partial x_1}(\mathbf{a}) & \dfrac{\partial F_1}{\partial x_2}(\mathbf{a}) & \cdots & \dfrac{\partial F_1}{\partial x_n}(\mathbf{a}) \\ \\ \dfrac{\partial F_2}{\partial x_1}(\mathbf{a}) & \dfrac{\partial F_2}{\partial x_2}(\mathbf{a}) & \cdots & \dfrac{\partial F_2}{\partial x_n}(\mathbf{a}) \\ \\ \vdots & \vdots & \vdots & \vdots \\ \\ \dfrac{\partial F_m}{\partial x_1}(\mathbf{a}) & \dfrac{\partial F_m}{\partial x_2}(\mathbf{a}) & \cdots & \dfrac{\partial F_m}{\partial x_n}(\mathbf{a}) \end{pmatrix}.$$

Definition 8.4. The matrix in the previous Theorem is called the **Jacobian matrix** of $\mathbf{F}$ at $\mathbf{a}$.

Example 8.3. Assume we know $\mathbf{F} : \mathbb{R}^2 \to \mathbb{R}^3$ defined by $\mathbf{F}(x, y) = (x^2 + y, x - 1, y^2)$ is differentiable everywhere. Find its derivative $\mathbf{F}'(1, 2)$ and its differential $d\mathbf{F}_{(1,2)}$. Use that to find the directional derivative $D_{(2,3)}\mathbf{F}(1, 2)$.

As a consequence of this theorem we obtain the following:

Corollary 8.1. Suppose U is an open subset of $\mathbb{R}^n$ and $f : U \to \mathbb{R}$ is a function that is differentiable at some $\mathbf{a} \in U$. Then, for every $\mathbf{0} \neq \mathbf{v} \in \mathbb{R}^n$ we have

$$D_{\mathbf{v}}f(\mathbf{a}) = \mathbf{v} \cdot (D_1 f(\mathbf{a}), \ldots, D_n f(\mathbf{a})).$$

Definition 8.5. The **gradient** of a function $f : U \to \mathbb{R}$, where U is an open subset of $\mathbb{R}^n$ is the function $\nabla f : U \to \mathbb{R}^n$ defined by $\nabla f(\mathbf{a}) = (D_1 f(\mathbf{a}), \ldots, D_n f(\mathbf{a}))$.

Definition 8.6. A **direction** is a unit vector, i.e. a vector with length 1. The **directional derivative of a function F in the direction of a nonzero vector $\mathbf{v}$ at point $\mathbf{a}$** is $D_{\mathbf{u}} F(\mathbf{a})$, where $\mathbf{u} = \mathbf{v}/||\mathbf{v}||$.

Theorem 8.5. *Let $\mathbf{a} \in U$, where U is an open subset of $\mathbb{R}^n$. Suppose $f : U \to \mathbb{R}$ is differentiable at $\mathbf{a}$ and that $\nabla f(\mathbf{a}) \neq \mathbf{0}$. Then, the maximum directional derivative of f at $\mathbf{a}$ is in the direction of $\nabla f(\mathbf{a})$, and this maximum directional derivative is equal to $||\nabla f(\mathbf{a})||$. Similarly, the minimum directional derivative of f at $\mathbf{a}$ is in the direction of $-\nabla f$, and this minimum directional derivative is equal to $-||\nabla f(\mathbf{a})||$.*

Example 8.4. Find the maximum and minimum directional derivative of the function $f : \mathbb{R}^3 \to \mathbb{R}$ given by $f(x, y, z) = \sin(xyz) + x^2 + yz$ at $(2\pi, 1, 3)$. Assume f is differentiable on $\mathbb{R}^3$.

Definition 8.7. Let $\mathbf{a}$ be a point in an open subset U of $\mathbb{R}^n$. A function $F : U \to \mathbb{R}^m$ is said to be **continuously differentiable** at $\mathbf{a}$ if all partial derivatives $D_1 F, \ldots, D_n F$ exist on U and they are all continuous at $\mathbf{a}$.

Theorem 8.6. *Suppose U is an open subset of $\mathbb{R}^n$. If $\mathbf{F} : U \to \mathbb{R}^m$ is continuously differentiable at a point $\mathbf{a} \in U$, then $\mathbf{F}$ is differentiable at $\mathbf{a}$.*

Example 8.5. Prove that $\mathbf{F} : \mathbb{R}^2 \to \mathbb{R}^3$ defined by $\mathbf{F}(x, y) = (x^2 + y, 2xy, y^2 - x)$ is differentiable everywhere.

Definition 8.8. Suppose U is an open subset of $\mathbb{R}^n$, and $f : U \to \mathbb{R}$ is differentiable. A point $\mathbf{a} \in U$ is called a **critical point** of f, iff $\nabla f(\mathbf{a}) = \mathbf{0}$.

Definition 8.9. Let $f : U \to \mathbb{R}$ be a function, where U is an open subset of $\mathbb{R}^n$. We say f attains a **local minimum** (resp., a local maximum) at $\mathbf{a}$, if there is an open subset V of U for which $f(\mathbf{a}) \leq f(\mathbf{x})$ (resp., $f(\mathbf{a}) \geq f(\mathbf{x})$) for all $\mathbf{x} \in V$. If f has a local maximum or a local minimum at $\mathbf{a}$ we say f has a local extremum at $\mathbf{a}$.

Theorem 8.7. *Suppose $f : U \to \mathbb{R}$ is differentiable, where U is an open subset of $\mathbb{R}^n$. If f attains a local extremum at a point $\mathbf{a} \in U$, then $\mathbf{a}$ is a critical point of f.*

Definition 8.10. Let U be an open subset of $\mathbb{R}^n$, and $\mathbf{F} : U \to \mathbb{R}^m$ be differentiable. Suppose $\mathbf{a} \in U$. Then, the approximation

$$\mathbf{F}(\mathbf{x}) \approx \mathbf{F}(\mathbf{a}) + d\mathbf{F}_{\mathbf{a}}(\mathbf{x} - \mathbf{a})$$

is called the **tangent plane approximation** of $\mathbf{F}$ near $\mathbf{a}$.

Example 8.6. Approximate $\sqrt{1.95 \times 2.01 \times 4.01}$ using tangent plane approximation.

Given a function $f : U \to \mathbb{R}$, where U is an open subset of $\mathbb{R}^n$, and $\mathbf{a} \in U$, we have the following:

$$df(\mathbf{h}) = D_{\mathbf{h}} f = \nabla f \cdot \mathbf{h} = \sum D_i f h_i = D_i f dx_i(\mathbf{h}).$$

Therefore, we can write $df = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} dx_i$.

Definition 8.11. Let $f : U \to \mathbb{R}$ be a differentiable function, where U is an open subset of $\mathbb{R}^n$. The function L given by $L(\mathbf{a}) = df_{\mathbf{a}}$ which assigns to any point $\mathbf{a}$ the linear transformation $df_{\mathbf{a}} : \mathbb{R}^n \to \mathbb{R}$ is called a **differential form**.

8.3 The Chain Rule

Theorem 8.8 (The Chain Rule)**.** *Suppose U and V are open subsets of $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. Suppose $\mathbf{F} : U \to \mathbb{R}^m$ and $\mathbf{G} : V \to \mathbb{R}^k$ are differentiable at points $\mathbf{a} \in U$, and $\mathbf{F}(\mathbf{a}) \in V$, respectively. Assume $\mathbf{F}(U) \subseteq V$. Then, the composition $\mathbf{H} = \mathbf{G} \circ \mathbf{F} : U \to \mathbb{R}^k$ is differentiable at $\mathbf{a}$ and $d\mathbf{H}_{\mathbf{a}} = d\mathbf{G}_{\mathbf{F}(\mathbf{a})} \circ d\mathbf{F}_{\mathbf{a}}$. Furthermore, $\mathbf{H}'(\mathbf{a}) = \mathbf{G}'(\mathbf{F}(\mathbf{a}))\mathbf{F}'(\mathbf{a})$.*

Example 8.7. Write down the Chain Rule for functions $f = (f_1, \ldots, f_m) : \mathbb{R} \to \mathbb{R}^m$, and $g : \mathbb{R}^m \to \mathbb{R}$.

Example 8.8. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the usual polar coordinate function defined by $T(r, \theta) = (r\cos\theta, r\sin\theta)$. For a function $f(x, y)$ from the cartesian plane $\mathbb{R}^2$ to $\mathbb{R}$. Find the partial derivatives of the function $f(r\cos\theta, r\sin\theta)$ with respect to r and θ.

Definition 8.12. Given two points $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$, the **segment** L from $\mathbf{a}$ to $\mathbf{b}$ is the set given by

$$L = \{\mathbf{c} \in \mathbb{R}^n \mid \mathbf{c} = t\mathbf{b} + (1-t)\mathbf{a}, \text{ where } 0 \le t \le 1\}.$$

Definition 8.13. A subset E of $\mathbb{R}^n$ is called **connected** if for every $\mathbf{a}, \mathbf{b} \in E$ there is a continuous function $\varphi : [0, 1] \to E$ whose restriction to $(0, 1)$ is differentiable, such that $\varphi(0) = \mathbf{a}$, and $\varphi(1) = \mathbf{b}$.

Theorem 8.9 (Intermediate Value Theorem)**.** *Suppose E is a connected subset of $\mathbb{R}^n$, and let $f : E \to \mathbb{R}$ be a continuous function. Suppose $\mathbf{a}, \mathbf{b} \in E$ are two points and r is a real number between $f(\mathbf{a})$ and $f(\mathbf{b})$. Then, there is a $\mathbf{c} \in E$ for which $f(\mathbf{c}) = r$.*

Definition 8.14. A function $\mathbf{F} : U \to \mathbb{R}^m$ is called **constant** if there is some $\mathbf{c} \in \mathbb{R}^m$ for which $F(\mathbf{x}) = \mathbf{c}$ for all $\mathbf{x} \in U$.

Theorem 8.10. *Let U be an open and connected subset of $\mathbb{R}^n$. A differentiable function $\mathbf{F} : U \to \mathbb{R}^m$ is constant if and only if $\mathbf{F}'(\mathbf{x}) = \mathbf{0}$ for all $\mathbf{x} \in U$.*

Theorem 8.11 (Mean Value Theorem)**.** *Suppose U is an open subset of $\mathbb{R}^n$, and $\mathbf{a}, \mathbf{b}$ are two points in U such that U contains the line segment L from $\mathbf{a}$ to $\mathbf{b}$. If $f : U \to \mathbb{R}$ is differenatible, then there is a point $\mathbf{c} \in L$ for which*

$$f(\mathbf{b}) - f(\mathbf{a}) = f'(\mathbf{c})(\mathbf{b} - \mathbf{a}) = \nabla f(\mathbf{c}) \cdot (\mathbf{b} - \mathbf{a}).$$

Example 8.9. Find all second partial derivatives of $f(x, y) = x^2y + xy\ln x$.

Theorem 8.12 (Clairaut's Theorem or Mixed-Partial Theorem)**.** *Suppose U is an open subset of $\mathbb{R}^n$. Suppose $f : U \to \mathbb{R}$ has continuous first and second partial derivatives. Then for every i, j we have $D_jD_if(\mathbf{a}) = D_iD_jf(\mathbf{a})$ for all $\mathbf{a} \in U$.*

Example 8.10. Let $f(x,y)$, with $f : \mathbb{R}^2 \to \mathbb{R}$, be a function with continuous first and second partial derivatives, and let $g(u,v) = f(Au+Bv, Cu+Dv)$, where A, B, C, D are constants. Prove that

$$\frac{\partial^2 g}{\partial u \partial v} = AB\frac{\partial^2 f}{\partial x^2} + CD\frac{\partial^2 f}{\partial y^2} + (AD+BC)\frac{\partial^2 f}{\partial x \partial y}.$$

8.4 More Examples

Example 8.11. Find all directional derivatives of each function f below at the given point $\mathbf{a}$.

(a) $f(x,y) = x^3 + 3xy$ with $\mathbf{a} = (0,1)$.

(b) $f(x,y,z) = \begin{cases} \dfrac{x^4+y^2+z^3}{x^2+y^2+z^2} & \text{if } (x,y,z) \neq (0,0,0) \\ 0 & \text{otherwise} \end{cases}$ with $\mathbf{a} = \mathbf{0}$.

(c) $f(x,y,z) = \dfrac{\sin(x+y)}{x^2+1}$ with $\mathbf{a} = \mathbf{0}$.

Solution. (a) Note that $f_x(x,y) = 3x^2 + 3y$ and $f_y(x,y) = 3x$ are both polynomials and thus continuous. Therefore, f is continuously differentiable. By Theorem 8.6 its directional derivative at $(0,1)$ with respect to a nonzero vector $\mathbf{v} = (a,b)$ is given by:

$$D_{\mathbf{v}} f(0,1) = (a,b) \cdot (f_x(0,1), f_y(0,1)) = (a,b) \cdot (3,0) = 3a.$$

(b) We will have to use the limit definition of directional derivatives. Let $\mathbf{v} = (a,b,c)$ be a nonzero vector.

$$\begin{aligned} D_{\mathbf{v}} f(0,0,0) &= \lim_{h\to 0} \frac{f((0,0,0)+h(a,b,c)) - f(0,0,0)}{h} \\ &= \lim_{h\to 0} \frac{f(ha,hb,hc)}{h} \\ &= \lim_{h\to 0} \frac{h^4a^4 + h^2b^2 + h^3c^3}{h^3(a^2+b^2+c^2)} \\ &= \lim_{h\to 0} \frac{h^2a^4 + b^2 + hc^3}{h(a^2+b^2+c^2)} \end{aligned}$$

The denominator approaches zero while the numerator approaches b^2 as $h \to 0$. Thus, if $b \neq 0$ the limit does not exist.

If $b = 0$, then the limit is as follows:

$$D_{\mathbf{v}} f(0,0,0) = \lim_{h\to 0} \frac{ha^4 + c^3}{a^2+b^2+c^2} = \frac{c^3}{a^2+b^2+c^2}.$$

(c) Similar to part (a), this function is continuously differentiable, since

$$\frac{\partial f}{\partial x} = \frac{\cos(x+y)(x^2+1) - 2x\sin(x+y)}{(x^2+1)^2}, \; \frac{\partial f}{\partial y} = \frac{\cos(x+y)}{x^2+1}.$$

Therefore, by Theorem 8.6, f is differentiable and $D_{\mathbf{v}}f = \mathbf{v} \cdot \nabla f$. We see $f_x(0,0) = f_y(0,0) = 1$. Letting $\mathbf{v} = (a,b) \in \mathbb{R}^2$ we obtain

$$D_{\mathbf{v}}f(0,0) = \mathbf{v} \cdot \nabla f(0,0) = a + b.$$

□

Example 8.12. Evaluate $D_1D_2f(x,y)$ at all points for each of the following functions:

(a) $f(x,y) = x^2 + xy$.

(b) $f(x,y) = \begin{cases} \dfrac{xy}{x^2+y^2} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{otherwise} \end{cases}$

Solution. (a) We see that $D_2f(x,y) = x$ and thus $D_1D_2f(x,y) = 1$.
(b) Note that the function is given by $\dfrac{xy}{x^2+y^2}$ on the open set $\mathbb{R}^2 - \{(0,0)\}$. (Recall that finite sets are closed. See Example 7.12.) Thus, for every $(x,y) \neq (0,0)$ we can find the answer by applying the Quotient Rule:

$$D_2f = \frac{x(x^2+y^2) - 2y(xy)}{(x^2+y^2)^2} = \frac{x^3 - xy^2}{(x^2+y^2)^2} \qquad (*)$$

We can now apply the Quotient Rule again to find D_1D_2f at points other than the origin:

$$D_1D_2f = \frac{(3x^2 - y^2)(x^2+y^2)^2 - 2(x^2+y^2)(2x)(x^3 - xy^2)}{(x^2+y^2)^4}.$$

For the origin this can be done using the definition of directional derivatives:

$$D_2f(0,0) = \lim_{h\to 0} \frac{f(0,h) - f(0,0)}{h} = \lim_{h\to 0} \frac{0}{h} = 0.$$

Similarly we have

$$D_1D_2f(0,0) = \lim_{h\to 0} \frac{D_2f(h,0) - D_2f(0,0)}{h}$$

Note that by $(*)$ we have $D_2f(h,0) = \dfrac{h^3 - h0^2}{(h^2+0^2)^2} = \dfrac{1}{h}$. Therefore,

$$D_1D_2f(0,0) = \lim_{h\to 0} \frac{1/h}{h}.$$

This limit is not a real number. Therefore, $D_1D_2f(0,0)$ does not exist. □

Example 8.13. Suppose $F : U \to \mathbb{R}^m$ is differentiable, where U is an open subset of $\mathbb{R}^n$ with $m < n$. Prove that for every $\mathbf{a} \in U$, there is a nonzero vector $\mathbf{v} \in \mathbb{R}^n$ for which $D_{\mathbf{v}}F(\mathbf{a}) = \mathbf{0}$.

Solution. By Theorem 8.4, $D_{\mathbf{v}}F(\mathbf{a}) = F'(\mathbf{a})\mathbf{v}$. We know $F'(\mathbf{a})$ is an $m \times n$ matrix. Since there are n columns (with $n > m$), and these columns are all in $\mathbb{R}^m$, by a theorem the columns of $F'(\mathbf{a})$ are linearly dependent. Therefore, there is a vector $\mathbf{v} \in \mathbb{R}^n$ for which $F'(\mathbf{a})\mathbf{v} = \mathbf{0}$. Therefore, $D_{\mathbf{v}}F(\mathbf{a}) = \mathbf{0}$. □

Example 8.14. Consider the function given by

$$f(x,y) = \begin{cases} \dfrac{x^2y - y^3}{x^2+y^2} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{otherwise} \end{cases}$$

(a) Show that $f(x,y)$ is continuous on $\mathbb{R}^2$.

(b) Find $D_{\mathbf{u}}f(0,0)$ for every nonzero vector $\mathbf{u} = (a,b)$.

(c) Show that f is not differentiable at $(0,0)$.

Solution. (a) First, note that $\dfrac{x^2y - y^3}{x^2+y^2}$ is a composition of continuous function and thus it is continuous at any point (x,y) that satisfies $x^2 + y^2 \neq 0$. Thus, f is continuous everywhere except possibly at the origin. In order to show f is continuous at $(0,0)$ we need to show $\lim_{(x,y)\to(0,0)} f(x,y) = f(0,0)$. We know $f(0,0) = 0$. Therefore, we need to show $\lim_{(x,y)\to(0,0)} f(x,y) = 0$.

By the Triangle Inequality and the fact that $0 < x^2 + y^2 \leq x^2$ and $0 < x^2 + y^2 \leq y^2$ we obtain the following chain of inequalities:

$$\left|\frac{x^2y - y^3}{x^2+y^2}\right| \leq \left|\frac{x^2y}{x^2+y^2}\right| + \left|\frac{y^3}{x^2+y^2}\right| \leq |y| + |y| = 2|y|.$$

Using properties of absolute value, we can rewritten this as

$$-2|y| \leq \frac{x^2y - y^3}{x^2+y^2} \leq 2|y|.$$

Note that since $|y|$ is continuous, $\pm 2|y| \to 0$ as $(x,y) \to (0,0)$. Thus, by the Squeeze Theorem $f(x,y) \to 0$ as $(x,y) \to (0,0)$. Therefore, f is continuous everywhere.

(b) We will use the definition of directional derivatives:

$$D_{\mathbf{u}}f(0,0) = \lim_{h\to 0} \frac{f((0,0) + h(a,b)) - f(0,0)}{h} = \lim_{h\to 0} \frac{f(ha,hb) - 0}{h} = \lim_{h\to 0} \frac{\dfrac{h^3a^2b - h^3b^3}{h^2a^2 + h^2b^2}}{h} = \frac{a^2b - b^3}{a^2+b^2}.$$

(c) Assume on the contrary f is differentiable at $(0,0)$. By Corollary 8.1,

$$D_{\mathbf{u}}f(0,0) = \nabla f(0,0) \cdot \mathbf{u}.$$

By part (b) we have $D_1 f(0,0) = \dfrac{1^2 \times 0 - 0^3}{1^2 + 0^2} = 0$, and $D_2 f(0,0) = \dfrac{0^2 \times 1 - 1^3}{0^2 + 1^2} = -1$. Therefore, $\nabla f(0,0) = (0,-1)$. Therefore,

$$D_{\mathbf{u}}f(0,0) = (0,-1) \cdot (a,b) = -b.$$

This contradicts the formula that we found in part (b) for $D_{\mathbf{u}}f(0,0)$, e.g. for $\mathbf{u} = (1,1)$. $\square$

Example 8.15. Find the maximum and minimum directional derivatives of the function $f(x,y) = x^3 \sin y + xe^y$ at the origin.

Solution. Partial derivatives of this function are

$$f_x = 3x^2 \sin y + e^y, \text{and } f_y = x^3 \cos y + xe^y.$$

Since both f_x, and f_y are continuous, f is continuously differentiable. By Theorem 8.6, f is differentiable. Therefore, by Theorem 8.5 the maximum and minimum directional derivatives of the function is $||\nabla f(0,0)|| = \sqrt{1^2+0^2} = 1$ and $-||\nabla f(0,0)|| = -1$, respectively. $\square$

Example 8.16. Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ given by

$$f(x,y) = \begin{cases} \dfrac{xy^2}{x^2+y^2} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{otherwise} \end{cases}$$

Prove that $D_{\mathbf{u}} f(0,0)$ exists for all nonzero vectors $\mathbf{u} \in \mathbb{R}^2$, but f is not differentiable at $(0,0)$

Solution. Let $\mathbf{u} = (a,b)$. We have

$$D_{\mathbf{u}} f(0,0) = \lim_{h\to 0} \frac{f((0,0)+h(a,b)) - f(0,0)}{h}.$$

This fraction simplifies to

$$\frac{f((0,0)+h(a,b)) - f(0,0)}{h} = \frac{f(ha,hb) - 0}{h} = \frac{\frac{hah^2b^2}{h^2a^2+h^2b^2}}{h} = \frac{ab^2}{a^2+b^2}.$$

Since this is independent of h we obtain

$$D_{\mathbf{u}} f(0,0) = \frac{ab^2}{a^2+b^2} \qquad (*)$$

On the contrary assume f were differentiable at $(0,0)$. By a theorem $D_{\mathbf{u}} f(0,0) = \nabla f(0,0) \cdot \mathbf{u}$. We have the following:

$$f_x(0,0) = D_{\mathbf{e}_1} f(0,0) = \frac{1 \cdot 0^2}{1^2+0^2} = 0, \text{ and } f_y(0,0) = D_{\mathbf{e}_2} f(0,0) = \frac{0 \cdot 1^2}{0^2+1^2} = 0.$$

Therefore, $\nabla f(0,0) = (0,0)$. Thus, $D_{\mathbf{u}} f(0,0) = 0$ for every vector $\mathbf{u}$. This contradicts $(*)$, e.g. for $a = b = 1$. Which means f is not differentiable at $(0,0)$. $\square$

Example 8.17. Approximate $\sqrt{(3.1)^2 + (3.99)^2}$ using tangent plane approximation.

Solution. We see that $f : \mathbb{R}^2 \to \mathbb{R}$ defined by $f(x,y) = \sqrt{x^2+y^2}$ has partial derivatives

$$\frac{\partial f}{\partial x} = x(x^2+y^2)^{-1/2}, \text{ and } \frac{\partial f}{\partial y} = y(x^2+y^2)^{-1/2}$$

which are both continuous on an open disk about $(3,4)$. The partial derivative of f at $(3,4)$ is $(3/5, 4/5)$. Therefore, $f(x,y) \approx f(3,4) + (3/5, 4/5) \cdot (0.1, -0.01) = 5 + 0.3/5 - 0.04/5 = 5.052$. $\square$

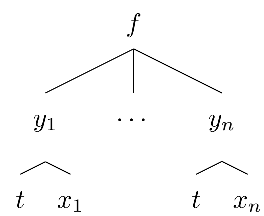
Definition 8.15. A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be **homogeneous** of degree m, where m is a positive integer, if

$$f(tx_1, \ldots, tx_n) = t^m f(x_1, \ldots, x_n), \text{ for all } t, x_1, \ldots, x_n \in \mathbb{R}.$$

Example 8.18. Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable and homogeneous of degree m. Prove that

$$x_1 \frac{\partial f}{\partial x_1} + \cdots + x_n \frac{\partial f}{\partial x_n} = mf.$$

Solution. Consider the function $f(y_1, \ldots, y_n)$ with $y_j = tx_j$ for $j = 1, \ldots, n$ and assume $y_j = tx_j$. This gives the following tree:



Using the Chain Rule we obtain the following:

$$\frac{\partial f}{\partial t} = \sum_{k=1}^{n} D_k f(y_1, \ldots, y_n) \frac{\partial y_k}{\partial t} = \sum_{k=1}^{n} D_k f(y_1, \ldots, y_n) x_k \qquad (*)$$

By assumption $f(tx_1, \ldots, tx_n) = t^m f(x_1, \ldots, x_n)$ and thus

$$\frac{\partial f}{\partial t} = mt^{m-1} f(x_1, \ldots, x_n).$$

Substituting this into $(*)$ and setting $t = 1$ and using the fact that $y_k = tx_k$ we obtain the result. $\square$

Example 8.19. Consider the function

$$f(x, y) = \begin{cases} \dfrac{x^2y^2}{x^2 + y^2} & \text{if } (x, y) \neq (0, 0) \\ 0 & \text{if } x = y = 0 \end{cases}$$

Prove that $f_{xy} = f_{yx}$ everywhere, even though f_{xy} is not continuous at $(0, 0)$. Compare this with Clairaut's Theorem.

Solution. For $(x, y) \neq (0, 0)$ we have

$$f_x = \frac{2xy^2(x^2 + y^2) - 2x(x^2y^2)}{(x^2 + y^2)^2} = \frac{2xy^4}{(x^2 + y^2)^2},$$

and

$$f_{xy} = \frac{8xy^3(x^2 + y^2)^2 - 2(x^2 + y^2)2y(2xy^4)}{(x^2 + y^2)^4} = \frac{8xy^3(x^2 + y^2) - 8xy^5}{(x^2 + y^2)^3} = \frac{8x^3y^3}{(x^2 + y^2)^3}.$$

By symmetry f_{yx} would be the same at points that are not the origin.

At $(0, 0)$ we have

$$f_x(0, 0) = \lim_{x \to 0} \frac{f(x, 0) - f(0, 0)}{x} = \lim_{x \to 0} \frac{0 - 0}{x} = 0.$$

Using this we obtain

$$f_{xy}(0,0) = \lim_{y \to 0} \frac{f_x(0,y) - f_x(0,0)}{y} = \lim_{y \to 0} \frac{0-0}{y} = 0.$$

By symmetry we have $f_{yx}(0,0) = 0$. This shows $f_{xy}(0,0) = f_{yx}(0,0)$.

Approaching $(0,0)$ along the lines of the form $y = mx$ yields

$$f_{xy}(x, mx) = \frac{x^3(mx)^3}{(x^2 + m^2x^2)^3} = \frac{m^3}{(1+m^2)^3}.$$

Since this value depends on m, by Theorem 6.1 the limit does not exist.

This example shows that the converse of Clairaut's Theorem is not valid. □

Example 8.20. Prove that the graph of a differentiable function $f : \mathbb{R} \to \mathbb{R}^n$ defined by

$$\Gamma_f = \{(x, f(x)) \mid x \in \mathbb{R}\}$$

is a connected subset of $\mathbb{R}^{n+1}$.

Solution. Let $(a, f(a)), (b, f(b))$ be two points in Γ_f, i.e. $a, b \in \mathbb{R}$, with $a < b$. Define $\varphi : [0,1] \to [a,b]$ by $\varphi(t) = a(1-t) + bt$. We see that φ is differentiable on $(0,1)$ and continuous on $[0,1]$. Furthermore, $\varphi(0) = a$ and $\varphi(1) = b$. By Theorem 6.5 and the Chain Rule $f \circ \varphi : [0,1] \to \mathbb{R}^n$ is differentiable on $(0,1)$ and continuous on $[0,1]$. Furtheremore, we see $f \circ \varphi(0) = f(a)$ and $f \circ \varphi(1) = f(b)$. Since φ and $f \circ \varphi$ are differentiable, the function $g : [0,1] \to \Gamma_f$ given by $g(t) = (\varphi(t), f \circ \varphi(t))$ is differentiable on $(0,1)$ and continuous on $[0,1]$. Note that $g(0) = (a, f(a))$ and $g(1) = (b, f(b))$. Thus, Γ_f is connected. □

Check pages 66-69, examples 1-4 of Advanced Calculus of Several Variables by Edwards.

8.5 Exercises

Exercise 8.1. *Using the definition of directional derivative prove the following:*

(a) $D_{\mathbf{v}}(f+g)(\mathbf{a}) = D_{\mathbf{v}}f(\mathbf{a}) + D_{\mathbf{v}}g(\mathbf{a})$.

(b) $D_{\mathbf{v}}(cf)(\mathbf{a}) = cD_{\mathbf{v}}f(\mathbf{a})$.

Exercise 8.2. *Prove that the function* $F : \mathbb{R}^3 \to \mathbb{R}^2$ *defined by* $F(x,y,z) = (xyz, x^2+y+z^3)$ *is differentiable everywhere, find its derivative, and its differential at* $(1,2,-1)$*. Use that to find the derictional derivative of this function in the direction* $(1,-2,2)$*. (Note that directional derivative in a direction should not depend on the length of the vector.)*

Exercise 8.3. *The position of a particle in* $\mathbb{R}^3$ *is given by*

$$\mathbf{r}(t) = (\cos(t), \sin(t), t).$$

(a) Show that this particle is always located on the cylinder $x^2 + y^2 = 1$. Use that to sketch the trajectory of this particle.

(b) Show the speed of this particle is constant, even though its velocity is not. (Recall that speed is the norm of velocity.)

(c) Show that the velocity always makes a constant nonzero angle with the z-axis.

(d) Letting $t_1 = 0$, and $t_2 = 2\pi$, show that $\mathbf{r}(t_2) - \mathbf{r}(t_1)$ is vertical.

(e) Conclude that there cannot be any $c \in (0, 2\pi)$ for which $\mathbf{r}(t_2) - \mathbf{r}(t_1) = \mathbf{r}'(c)(t_2 - t_1)$. Explain why this does not contradict the Mean Value Theorem.

Exercise 8.4. *Suppose $L : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation with matrix A.*

(a) Using the definition of derivatives, show that the differential of L is itself. Deduce the derivative of L is A.

(b) Conversely, prove that if the derivative of a function $F : \mathbb{R}^n \to \mathbb{R}^m$, satisfying $F(\mathbf{0}) = \mathbf{0}$, is a constant matrix A, then F is linear.

Exercise 8.5. *Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ given by $f(x, y) = \sqrt[3]{x^3 + y^3}$. Prove that $D_{\mathbf{u}} f(0, 0)$ exists for all nonzero vectors $\mathbf{u} \in \mathbb{R}^2$, but f is not differentiable at $(0, 0)$.*

Exercise 8.6. *Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ given by*

$$f(x,y) = \begin{cases} (x^2+y^2)\sin\left(\dfrac{1}{x^2+y^2}\right) & \text{if } (x,y) \neq (0,0) \\ 0 & \text{if } (x,y) = (0,0) \end{cases}$$

(a) Prove that $f_x(0, 0) = f_y(0, 0) = 0$.

(b) Prove f is differentiable at $(0, 0)$.

(c) Prove f_x and f_y are not continuous at $(0, 0)$.

Exercise 8.7. *Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ given by*

$$f(x,y) = \begin{cases} x^2 \sin(1/x) + y^2 & \text{if } x \neq 0 \\ y^2 & if x = 0 \end{cases}$$

(a) Prove that f_x and f_y exist everywhere.

(b) Prove that f_x is not continuous at $(0, 0)$, however f_y is continuous everywhere.

(c) Prove that f is differentiable at $(0, 0)$.

Exercise 8.8. *Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a function with continuous second partials. Define a function g by $g(r, \theta) = f(r\cos\theta, r\sin\theta)$. Prove that*

$$||\nabla f||^2 = \left(\frac{\partial g}{\partial r}\right)^2 + \frac{1}{r^2}\left(\frac{\partial g}{\partial \theta}\right)^2.$$

Exercise 8.9. *Consider three differentiable functions $\varphi : \mathbb{R} \to \mathbb{R}^n, f : \mathbb{R}^n \to \mathbb{R}^m$, and $g : \mathbb{R}^m \to \mathbb{R}$. If $h = g \circ f \circ \varphi$, prove that $h'(t) = \nabla g(f(\varphi(t))) \cdot D_{\varphi'(t)} f(\varphi(t))$ for every $t \in \mathbb{R}$.*

Exercise 8.10. *Suppose $f(x), g(x)$ are functions defined over open intervals I, J, and are differentiable at x_0, y_0, respectively, Let a, p be two functions defined over $I \times J$ by $a(x,y) = f(x)+g(y)$ and $p(x,y) = f(x)g(y)$. Prove the Clairaut's Theorem for a and p, at (x_0, y_0). In other words, show $a_{xy}(x_0,y_0) = a_{yx}(x_0,y_0)$ and $p_{xy}(x_0,y_0) = p_{yx}(x_0,y_0)$.*

Exercise 8.11. *Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ given by*

$$f(x,y) = \begin{cases} \dfrac{x^3y - xy^3}{x^2+y^2} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{otherwise} \end{cases}$$

(a) Find $D_1 f$ and $D_2 f$ at all points.

(b) Find $D_1 D_2 f$ and $D_2 D_1 f$ at all points.

(c) Show that $D_1 D_2 f(0,0) \neq D_2 D_1 f(0,0)$. How do you reconcile this with the Clairaut's Theorem?

Exercise 8.12. *Suppose U is an open subset of $\mathbb{R}^n$ and let $\mathbf{a} \in U$. Assume $F : U \to \mathbb{R}^m$ is differentiable at $\mathbf{a}$. Prove that F is continuous at $\mathbf{a}$.*

Exercise 8.13. *Suppose U is an open subset of $\mathbb{R}^n$ and $f, g : U \to \mathbb{R}$ are differentiable. Prove the following:*

(a) $\nabla(f+g) = \nabla f + \nabla g$.

(b) $\nabla(fg) = f\nabla g + g\nabla f$.

(c) $\nabla(f^n) = nf^{n-1}\nabla f$, for every positive integer n.

Exercise 8.14. *Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ given by*

$$f(x,y) = \begin{cases} \dfrac{x^3}{x^2+y^2} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{if } x = y = 0 \end{cases}$$

(a) Find f_x and f_y at all points in $\mathbb{R}^2$.

(b) Show that $f_{xx}(0,0)$ and $f_{xy}(0,0)$ both exist but f_x is not continuous at the origin.

(c) Show $f_{yy}(0,0)$ and $f_{yx}(0,0)$ both exist but f_y is not continuous at the origin.

(d) Show f_{xy} is not continuous at the origin, however $f_{xy}(0,0) = f_{yx}(0,0)$. How do you reconcile this with the Clairaut's Theorem?

(e) Prove that f is not differentiable at the origin.

Exercise 8.15. *Consider the function*

$$f(x,y) = \begin{cases} \dfrac{x^2 y}{x^2+y^2} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{otherwise} \end{cases}$$

Prove that:

(a) f *is continuous on* $\mathbb{R}^2$.

(b) $\dfrac{\mathrm{d}}{\mathrm{d}x}\left(\lim_{y\to 0} f(x,y)\right)$ *exists for every* $x \in \mathbb{R}$.

(c) $\lim_{(x,y)\to(0,0)} \left(\dfrac{\partial}{\partial x}(f(x,y))\right)$ *does not exist.*

Exercise 8.16. *Suppose* $f : \mathbb{R}^n \to \mathbb{R}$ *is a differentiable function for which* $f'(\mathbf{a}) = \mathbf{a}^T A$, *for all* $\mathbf{a} \in \mathbb{R}^n$ *and a fixed matrix* $A \in M_n(\mathbb{R})$. *Prove* A *must be symmetric.*

Exercise 8.17. *Determine if each of the following can be the Jacobian matrix of some function* $F : \mathbb{R}^n \to \mathbb{R}^m$ *for appropriate* n, m. *If it can be provide* <u>*one*</u> *such function.*

(a) $F'(x_1,x_2,x_3) = \begin{pmatrix} x_1 & x_2+x_3 & x_3+x_3^2+x_2 \\ x_2 & x_1 & x_3 \end{pmatrix}$ *for all* $x_1, x_2, x_3 \in \mathbb{R}$.

(b) $F'(x_1,x_2) = \begin{pmatrix} x_1+x_2 & x_1+x_2^2 \\ e^{x_2} & x_1 \end{pmatrix}$ *for all* $x_1, x_2 \in \mathbb{R}$.

8.6 Challenge Problems

Exercise 8.18. *Let* n *be a positive integer. Identify all vectors of* $\mathbb{R}^{n^2}$ *with* $n \times n$ *matrices by placing components of these vectors in the entries of rows of the matrix starting from the upper left corner and moving to the right and down. Let* $f : \mathbb{R}^{n^2} \to \mathbb{R}^{n^2}$ *be a function defined by* $f(A) = A^2$. *Find the differential of this function.*

Exercise 8.19. *Consider the function* f *defined by*

$$f(x,y) = \begin{cases} \dfrac{xy^2\sqrt{x^2+y^2}}{x^2+y^4} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{if } (x,y) = (0,0) \end{cases}$$

(a) Prove $D_{\mathbf{v}} f(0,0)$ *exists for every nonzero* $\mathbf{v} \in \mathbb{R}^2$, *and that* $D_{\mathbf{v}} f(0,0) = \nabla f(0,0) \cdot \mathbf{v}$.

(b) Show f *is differentiable everywhere on its domain, except at* $(0,0)$.

(c) Is f *continuous at* $(0,0)$?

Exercise 8.20. *Suppose* U *is a connected subset of* $\mathbb{R}^n$. *Suppose* $U \subseteq A \cup B$, *where* A *and* B *are open and disjoint subsets of* $\mathbb{R}^n$. *Prove* $U \subseteq A$ *or* $U \subseteq B$.

Exercise 8.21. *Suppose U is an open and connected subset of $\mathbb{R}^n$, and $f : U \to \mathbb{R}$ attains a local minimum at every point in U. Prove f is constant.*

Exercise 8.22. *Consider the function:*

$$f(x,y) = \begin{cases} (x^2+y^2)\cos\left(\dfrac{4xy^2}{x^2+y^4}\right) & \text{if } (x,y) \neq (0,0) \\ 0 & \text{if } (x,y) = (0,0) \end{cases}$$

(a) Prove f is differentiable (and hence continuous) on $\mathbb{R}^2$.

(b) Prove f attains a local minimum value at $(0,0)$ along every line that passes through $(0,0)$.

(c) Prove f does not attain a local minimum at $(0,0)$.

8.7 Summary

- Partial derivative of a function with respect to x can be found by fixing all variables and differentiating with respect to x.
- When a function has different rules at different values you need to use the limit definition to find its directional derivatives:
$$D_{\mathbf{v}}F(\mathbf{a}) = \lim_{h\to 0}\frac{F(\mathbf{a}+h\mathbf{v}) - F(\mathbf{a})}{h}.$$
- The (i,j) entry of the derivative of $(F_1,\ldots,F_m)$ is the partial of the F_i with respect to x_j.
- To show a function is differentiable we could find all partials of its component functions and show they are all continuous. Note that if these conditions are satisfied then the function is differentiable, but the converse is not true.
- If a function is differentiable, then $D_{\mathbf{u}}f(\mathbf{a}) = F'(\mathbf{a})\mathbf{v}$.
- To show a function is not differentiable:
  - Find all partial derivatives of the component functions.
  - Form the Jacobian Matrix.
  - Show that this Jacobian matrix fails to satisfy either the limit definition of differentials or the equality $D_{\mathbf{u}}f(\mathbf{a}) = F'(\mathbf{a})\mathbf{v}$.
- For a function $f : U \to \mathbb{R}$, where $U \subseteq \mathbb{R}^n$ is open, the differential is often called the gradient and is denoted by $\nabla f = (f_{x_1}, f_{x_2}, \ldots, f_{x_n})$.
- When finding directional derivatives, i.e. rate of change, we need to first normalize the vector.
- The maximum directional derivative of a function $f : U \to \mathbb{R}$ is $||\nabla f||$ and is obtained in the direction of gradient. The minimum is obtained in the direction of $-\nabla f$.

- To evaluate the derivative $\dfrac{\partial f}{\partial t}$:
  - Draw a tree diagram with f as its top vertex (called the root).
  - Place all variables that f depend on in the next row.
  - Draw edges from f to the variables that f depend on.
  - Repeat this process for all variables in the second row of the tree. Continue until you end up with the dependent variables.
  - For each path starting with f and ending at t write a product of derivatives along that path.
  - Add up all the products formed in the previous step. That is equal to $\dfrac{\partial f}{dt}$.
- If the derivative of a function over an open and connected set is zero, then the function is constant.
- The Mean Value Theorem also holds for functions $f : U \to \mathbb{R}$:

$$f(\mathbf{b}) - f(\mathbf{a}) = \nabla f(\mathbf{c}) \cdot (\mathbf{b} - \mathbf{a}).$$

- Caliraut's Theorem states that when dealing with partial derivatives, the order does not matter as long as all partials are continuous. For example $D_1 D_2 f = D_2 D_1 f$, if they are both continuous.

Week 9

9.1 Critical Points in Two Dimensions

In this section we would like to classify critical points of a function $f : \mathbb{R}^2 \to \mathbb{R}$. Recall that if a point is a local extremum, then it must be a critical point. Let's first look at a simple case when

$$f(x, y) = ax^2 + 2bxy + cy^2, \text{ where } a, b, c \text{ are real constants.}$$

Such a function is called a quadratic form.

We note that $(0, 0)$ is a critical point of this function, and $f(0, 0) = 0$. So, the question is: Under what conditions on a, b, c can we guarantee that $f(x, y) \geq 0$ for points (x, y) near the origin?

Completing the square we obtain the following

$$f(x, y) = \frac{(ax + by)^2 + (ac - b^2)y^2}{a}.$$

This gives the following:

- If $a > 0$, and $ac - b^2 \geq 0$, then $f(x, y)$ has a local (and absolute) minimum at $(0, 0)$.
- If $a < 0$, and $ac - b^2 \geq 0$, then $f(x, y)$ has a local (and absolute) maximum at $(0, 0)$.
- If $ac - b^2 < 0$, then $f(x, y)$ has neither a local minimum nor a local maximum at $(0, 0)$.

Definition 9.1. A **quadratic form** is a function $f : \mathbb{R}^n \to \mathbb{R}$ given by

$$f(x_1, \ldots, x_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j, \text{ where } a_{ij} \in \mathbb{R} \text{ is a constant.}$$

Definition 9.2. A quadratic form f is called **positive definite** (resp., **positive semidefinite**) if $f(\mathbf{x}) > 0$ (resp., $f(\mathbf{x}) \geq 0$) for all $\mathbf{0} \neq \mathbf{x} \in \mathbb{R}^n$. It is called **nondefinite** if f attains both positive and negative values. Similarly we say f is **negative definite** (resp., **negative semidefinite**) if $f(\mathbf{x}) < 0$ (resp., $f(\mathbf{x}) \leq 0$) for all $\mathbf{0} \neq \mathbf{x} \in \mathbb{R}^n$.

The above discussion gives us the following theorem:

Theorem 9.1. *The quadratic form $f(x,y) = ax^2 + 2bxy + cy^2$ is*

- *positive definite if $a > 0$, and $ac - b^2 > 0$.*
- *negative definite if $a < 0$, and $ac - b^2 > 0$.*
- *nondefinite if $ac - b^2 < 0$.*

Example 9.1. Determine and classify all critical points of $f(x,y) = x^2 - y^2$.

Definition 9.3. Let U be an open subset of $\mathbb{R}^n$. A critical point $\mathbf{a}$ of a function $f : U \to \mathbb{R}$ is called a **saddle point** if every open ball containing $\mathbf{a}$ contains points $\mathbf{x}, \mathbf{y} \in U$ for which $f(\mathbf{x}) < f(\mathbf{a}) < f(\mathbf{y})$.

Theorem 9.2 (Second Partials Test)**.** *Let $f : U \to \mathbb{R}$ be twice continuously differentiable, where U is an open subset of $\mathbb{R}^2$. Suppose $\mathbf{a} \in U$ is a critical point of f. Let*

$$\Delta = \frac{\partial^2 f}{\partial x^2}(\mathbf{a}) \cdot \frac{\partial^2 f}{\partial y^2}(\mathbf{a}) - \left(\frac{\partial^2 f}{\partial x \partial y}(\mathbf{a})\right)^2$$

Then f has

- *a local minimum at $\mathbf{a}$ if $\Delta > 0$, and $\dfrac{\partial^2 f}{\partial x^2}(\mathbf{a}) > 0$.*
- *a local maximum at $\mathbf{a}$ if $\Delta > 0$, and $\dfrac{\partial^2 f}{\partial x^2}(\mathbf{a}) < 0$.*
- *a saddle point at $\mathbf{a}$ if $\Delta < 0$.*

Note that if $\Delta = 0$, the above test is inconclusive.

Example 9.2. Show that the Second Partial Tests for each of the following functions is inconclusive, however $(0,0)$ is a local minimum, a local maximum, a saddle point for each of the following functions, respectively.

$$f(x,y) = x^4 + y^4, g(x,y) = -x^4 - y^4, h(x,y) = x^4 - y^4.$$

Example 9.3. Classify all critical points of $f(x,y) = xy + 2x - y$.

To understand quadratic forms on n variables, note that for a quadratic form $f(x_1, \ldots, x_n)$ we have

$$f(cx_1, \ldots, cx_n) = c^2 f(x_1, \ldots, x_n).$$

Thus, in order to understand if the origin is a local maximum or minimum we need to understand f over the unit sphere $x_1^2 + \cdots + x_n^2 = 1$.

9.2 Lagrange Multipliers

Definition 9.4. Suppose S is a subset of $\mathbb{R}^n$, and $f : U \to \mathbb{R}$ is function defined over an open subset U of $\mathbb{R}^n$ containing S. We say the **restriction of f to S attains a local minimum (resp., a local maximum) at a point $\mathbf{a} \in S$** if there is an open ball $B_r(\mathbf{a})$ for which $f(\mathbf{x}) \geq f(\mathbf{a})$ (resp., $f(\mathbf{x}) \leq f(\mathbf{a})$) for every $\mathbf{x} \in B_r(\mathbf{a}) \cap S$. If the restriction of f to S attains a local minimum or a local maximum at $\mathbf{a}$, we say f attains a local extremum at $\mathbf{a}$.

Theorem 9.3. *Let S be a subset of $\mathbb{R}^n$. Assume f is a differentiable real-valued function defined on some open set containing S, and the restriction of f to S attains a local extremum at* $\mathbf{a}$*, then the gradient vector $\nabla f(\mathbf{a})$ is orthogonal to all tangent lines to all curves on S that pass through* $\mathbf{a}$*. In other words, if $\varphi : \mathbb{R} \to S$ is a differentiable curve with $\varphi(0) = \mathbf{a}$ then $\nabla f(\mathbf{a})$ is orthogonal to $\varphi'(0)$.*

Example 9.4. Find the maximum and minimum values of $f(x, y) = xy$ subject to the constraint $x^2+y^2 = 1$.

Example 9.5. Find the equation of the plane tangent to the surface $x^2 + 2y^2 + 3z^2 = 6$ at $(1, -1, 1)$.

Definition 9.5. A **k-dimensional manifold** (or a **k-manifold**) M is a subset of $\mathbb{R}^n$ for which for every point $\mathbf{a} \in M$ there is an open subset U of $\mathbb{R}^n$ containing $\mathbf{a}$ for which $U \cap M$ "looks like" the k-dimensional space $\mathbb{R}^k$. (Yes, this is not a rigorous definition!)

Example 9.6. A sphere in $\mathbb{R}^3$ is a 2-dimensional manifold.

Theorem 9.4. *If M is a k-dimensional manifold in $\mathbb{R}^n$ and $\mathbf{a} \in M$, then M has a k-dimensional tangent plane at* $\mathbf{a}$*. In other words all lines tangent to curves on M at* $\mathbf{a}$ *that pass through* $\mathbf{a}$ *form the translation of a k-dimensional subspace of $\mathbb{R}^n$.*

Theorem 9.5. *Suppose $g : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable. If M is the set of all points* $\mathbf{x}$ *with both $g(\mathbf{x}) = \mathbf{0}$ and $\nabla g(\mathbf{x}) \neq \mathbf{0}$, then M is an $(n-1)$-manifold. Given $\mathbf{a} \in M$, the gradient vector $\nabla g(\mathbf{a})$ is orthogonal to the tangent plane to M at* $\mathbf{a}$*.*

Theorem 9.6 (Lagrange Multipliers Theorem, Simplified Version)**.** *Suppose $g : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable, and let M be the set of all points $\mathbf{x} \in \mathbb{R}^n$ that both $g(\mathbf{x}) = \mathbf{0}$, and $\nabla g(\mathbf{x}) \neq \mathbf{0}$. Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable. Assume the restriction of f to M attains a local extremum at a point $\mathbf{a} \in M$. Then $\nabla f(\mathbf{a}) = \lambda \nabla g(\mathbf{a})$ for some scalar λ.*

Example 9.7. Find the maximum and minimum values of $f(x, y, z) = x + 3y + z$ under the constraint $x^2 + y^2 + z^2 = 1$.

Theorem 9.7 (Lagrange Multipliers Theorem)**.** *Suppose $G = (G_1, \ldots, G_m) : \mathbb{R}^n \to \mathbb{R}^m$ is continuously differentiable, and denote by M the set of all points $\mathbf{x} \in \mathbb{R}^n$ such that $G(\mathbf{x}) = \mathbf{0}$, and also the gradient vectors $\nabla G_1(\mathbf{a}), \ldots, \nabla G_m(\mathbf{a})$ are linearly independent. Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable. Assume the restriction of f to M attains a local extremum at some point $\mathbf{a} \in M$. Then $\nabla f(\mathbf{a})$ is a linear combination of $\nabla G_1(\mathbf{a}), \ldots, \nabla G_m(\mathbf{a})$.*

Example 9.8. Find the highest and lowest points, if they exist, of the ellipse of intersection of the cylinder $x^2 + y^2 = 1$ and the plane $x + y + z = 1$.

9.3 More Examples

Example 9.9. Find all critical points of $f(x, y, z) = x^3 + y^2 + z^2 + 3xyz$.

Solution. The critical points satisfy the system below:

$$\begin{cases} f_x = 3x^2 + 3yz = 0 \Rightarrow x^2 + yz = 0 \\ f_y = 2y + 3xz = 0 \Rightarrow y = -3xz/2 \\ f_z = 2z + 3xy = 0 \end{cases}$$

Substituting $y = -3xz/2$ into the last equation we obtain $2z + 3x(-3xz/2) = 0$, which implies $z = 0$ or $x = \pm 2/3$. We will take three cases.

Case I. $z = 0$. The first and second equations yield $x = y = 0$. This gives the point $(0, 0, 0)$.

Case II. $x = 2/3$. Substituting into the second equation we obtain $2y + 2z = 0$, which implies $z = -y$. The first equation yields $4/9 - y^2 = 0$, which gives $y = \pm 2/3$. Therefore, we obtain the critical points $(2/3, 2/3, -2/3)$ and $(2/3, -2/3, 2/3)$.

Case III. $x = -2/3$. The second equation gives us $y = z$, and the first equation yields $4/9 + y^2 = 0$, which is impossible. □

Example 9.10. Find and classify all critical points of each function:

(a) $f(x, y) = x^2 + y^2 + xy + 2x - 2y$.

(b) $f(x, y) = x^4 + x^2 + y^4$.

Solution. (a) First, we will find all critical points: $f_x = 2x + y + 2, f_y = 2y + x - 2$. This gives the following system of equations:

$$\begin{cases} 2x + y + 2 = 0 \\ 2y + x - 2 = 0 \end{cases}$$

This yields $x = -2, y = 2$. We will now use the Second Partials Test. $f_{xx} = 2, f_{xy} = 1, f_{yy} = 2$. This gives $\Delta = 4 - 1^2 = 3$ which is positive. Since $f_{xx} = 2$ is also positive, $(-2, 2)$ is a local minimum.

(b) $f_x = 4x^3 + 2x, f_y = 4y^3$. The critical points satisfy the system

$$\begin{cases} 4x^3 + 2x = 0 \Rightarrow x(4x^2 + 2) = 0 \Rightarrow x = 0. \\ 4y^3 = 0 \Rightarrow y = 0. \end{cases}$$

The Second Partials Test gives $f_{xx} = 12x^2 + 2, f_{xy} = 0, f_{yy} = 12y^2$. This gives us $\Delta(0, 0) = 2 \times 0 - 0^2 = 0$. Therefore, the Second Partials Test is inconclusive.

Note that $f(0, 0) = 0$ and $f(x, y) = x^4 + x^2 + y^4 \geq 0$ since perfect squares are nonnegative. Therefore, $(0, 0)$ is a local (and absolute) minimum. □

Example 9.11. Find the plane or hyper-plane tangent to each manifold at the given point. Assume the given set is a manifold.

(a) $x_1^2 + 3x_2^2 + x_3^2 = 2$ at $(1, 0, -1)$.

(b) $x_1^4 + 4x_2 \sin(x_1 x_3) + x_3^2 + 3x_4^2 = 4$ at $(0, 0, 1, -1)$.

Solution. First, note that both functions on the left side of eqaulities are continuously differentiable.

(a) The vector orthogonal to the tangent plane at the point $(1, 0, -1)$ is the gradient vector if it is not zero:

$$(2x_1, 6x_2, 2x_3) = (2, 0, -2).$$

Thus, if (x_1, x_2, x_3) is on this plane, then

$$(x_1 - 1, x_2 - 0, x_3 + 1) \cdot (2, 0, -2) = 0.$$

Thus, the equation of the plane is $x_1 - 1 - x_3 - 1 = 0$.
(b) Similarly the orthogonal vector to the hyperplane tangent to this is the gradient vector

$$(4x_1^3 + 4x_2 x_3 \cos(x_1 x_3), 4\sin(x_1 x_3), 4x_2 x_1 \cos(x_1 x_3) + 2x_3, 6x_4) = (0, 0, 2, -6).$$

Thus, the equation of the hyperplane is

$$(x_1, x_2, x_3 - 1, x_4 + 1) \cdot (0, 0, 2, -6) = 0.$$

The equation simplifies to $x_3 - 3x_4 - 4 = 0$. □

Example 9.12. In each case below, find the maximum and minimum values of the given function subject to the given constraint or show they do not exist:

(a) $f(x, y) = x^3 + 2y^2$ given that $x^2 + 3y^2 = 1$.

(b) $f(x, y) = 3x^4 + 4y^4$ with the constraint $x^2 + y^2 = 1$.

(c) $f(x, y, z) = \sin x + \sin y + \sin z$ subject to $x + y + z = \pi$.

(d) $f(x, y, z) = x^2 + 2y^2 + z^2$ given $3x + 2y + z = 1$.

Solution. (a) The function f is a polynomial and thus it is continuous. The constraint gives us an ellipse which is closed and bounded (See Example 7.13.) Thus, by the Extreme Value Theorem, f attains its maximum and minimum values given the constraint. By the Lagrange Multiplier's Theorem these extreme points must satisfy either of the following:

$$(3x^2, 4y) = \lambda(2x, 6y), \text{ or } (2x, 6y) = (0, 0).$$

The second equality can not hold, since otherwise we will have $x = y = 0$ which does not lie on the ellipse $x^2 + 3y^2 = 1$.

The first equality gives us the following system:

$$\begin{cases} 3x^2 = 2\lambda x \\ 4y = 6\lambda y \\ x^2 + 3y^2 = 1 \end{cases}$$

The first equation can be written as $x(3x - 2\lambda) = 0$. Thus, $x = 0$ or $3x = 2\lambda$. We will take two cases:

Case I: $x = 0$. The third equation yields $3y^2 = 1$ or $y = \pm 1/\sqrt{3}$. This gives us $f(0, \pm 1/\sqrt{3}) = 2/3$.

Case II: $3x = 2\lambda$. Substituting this into the second equation we obtain $4y = 9xy$. Thus, $y = 0$ or $x = 4/9$. These give us the following four points:

$$(\pm 1, 0), \text{ and } (4/9, \pm\sqrt{65/243}).$$

We not evaluate the function f at these points

$$f(\pm 1, 0) = \pm 1, \text{ and } f(4/9, \pm\sqrt{65/243}) = 64/729 + 130/243 = 322/726.$$

Comparing these values we see the maximum is 1 and the minimum is -1.

(b) Similar to above, the contraint gives us a circle that is closed and bounded. Therefore, by the Extreme Value Theorem, the maximum and minimum values exist. We will now use the Lagrange Multipliers Theorem. At an extreme point we have one of the following:

$$(12x^3, 16y^3) = \lambda(2x, 2y) \text{ or } (2x, 2y) = (0, 0).$$

The second equality does not hold, since otherwise, we will obtain $x = y = 0$ which does not lie on the circle $x^2 + y^2 = 1$.

The first equality yileds the following system:

$$\begin{cases} 12x^3 = 2\lambda x \\ 16y^3 = 2\lambda y \\ x^2 + y^2 = 1 \end{cases}$$

If $x = 0$ or $y = 0$ then we obtain the points $(0, \pm 1)$ and $(\pm 1, 0)$. The functional values at these points are

$$f(0, \pm 1) = 4, \text{ and } f(\pm 1, 0) = 3.$$

If neither x nor y is zero, we obtain: $\lambda = 6x^2$ and $\lambda = 8y^2$. Combining this with $x^2 + y^2 = 1$ we conclude

$$\frac{\lambda}{6} + \frac{\lambda}{8} = 1 \Rightarrow \lambda = 24/7.$$

From here we obtain the following four points

$$(\pm 2/\sqrt{7}, \pm\sqrt{3/7}).$$

The functional values for these four points are

$$f(\pm 2/\sqrt{7}, \pm\sqrt{3/7}) = 12/7.$$

Comparing these we conclude that the absolute maximum of this function is 4 and the absolute minimum is 12/7.

(c) Note that the plane $x+y+z=\pi$ is closed, since it is the inverse image $g^{-1}(\{\pi\})$ with $g(x,y,z)=x+y+z$, and g is a continuous function, and $\{\pi\}$ is closed. This plane is not bounded so we cannot simply invoke the Extreme Value Theorem. In order to resolve this issue we will replace x by $x+2\pi n$ for some integer n to make sure $x \in [0, 2\pi]$ and do the same with y. This would change x to $x+2n\pi$ and y to $y+2m\pi$ and z to $z-2n\pi-2m\pi$. This does not change the sum $\cos x+\cos y+\cos z$. In other words, we can assume $x, y \in [0, 2\pi]$. Since $z = \pi - x - y$, we have $-3\pi \leq z \leq \pi$. In other words, we can focus on the rectanglular cube $[0, 2\pi] \times [0, 2\pi] \times [-3\pi, \pi]$. Since this region is bounded and closed, and f is continuous, the function does have absolute maximum and minimum values. Invoking the Lagrange Multipliers Theorem, the extreme points must satisfy one of the following:

$$(\cos x, \cos y, \cos z) = \lambda(1,1,1) \text{ or } (1,1,1) = \mathbf{0}.$$

The second equality is impossible. Therefore, we must have $\cos x = \cos y = \cos z$. This implies $y = x$ or $2\pi - x$. Since $x+y+z=\pi$ we must have $z = \pi - 2x$ or $z = -\pi$. Therefore, we have two possibilities:
Case I: $y = x, z = \pi - 2x$. Since $\cos z = \cos x$ we must have

$$\cos(\pi - 2x) = \cos x \Rightarrow x = 2n\pi \pm (\pi - 2x) \Rightarrow x = \frac{(2n+1)\pi}{3} \text{ or } (1-2n)\pi.$$

This yields, the following:

$$x = y = z = \frac{\pi}{3}, \text{ or } x = y = \pi, z = -\pi, \text{ or } x = y = \frac{5\pi}{3}, z = \frac{-7\pi}{3}$$

The values of f at these points are $\frac{3\sqrt{3}}{2}$, 0, and $-\frac{3\sqrt{3}}{2}$, respectively.
Case II: $y = 2\pi - x$. This yields, $z = -\pi$. In this case we have

$$f(x, 2\pi - x, -\pi) = \sin x + \sin(2\pi - x) + \sin(-\pi) = \sin x - \sin x + 0 = 0.$$

Comparing the values that we found we conclude that the maximum and minimum values are $3\sqrt{3}/2$ and $-3\sqrt{3}/2$, respectively.

(d) Everything is similar to parts (a) and (b), except since $3x+2y+z=1$ does not determine a bounded region we cannot invoke the Extreme Value Theorem. Note that we can make x as large as we would like. For example for every x the point $(x, x, -5x+1)$ lies on the plane $3x+2y+z=1$. However

$$f(x, x, -5x+1) = x^2 + 2x^2 + (-5x+1)^2 \geq 3x^2.$$

This means f does not have a maximum value, as $3x^2$ could be arbitrarily large.

Now, note that $(0,0,1)$ satisfies the constraint and $f(0,0,1) = 1$. If $|x| \geq 1$ or $|y| \geq 1$ or $|z| \geq 1$, then $f(x,y,z) \geq 1 = f(0,0,1)$. This means if there is a minimum for f the minimum must satisfy

$$|x|, |y|, |z| \leq 1.$$

The cube given above is closed and bounded. (Why?) Thus, we may invoke the Extreme Value Theorem for f applied to the intersection of this cube and the plane $3x + 2y + z = 1$. The rest is similar to parts (a), (b) and (c). □

Example 9.13. Find the minimum distance from the origin to the points of the surface $x^2+2x+y^2+3z^2 = 1$ or show no minimum exists.

Solution. The distance from the origin to a point (x, y, z) is $\sqrt{x^2 + y^2 + z^2}$. In order to minimize this, it is enough to minimize $f(x,y,z) = x^2 + y^2 + z^2$ subject to $x^2 + 2x + y^2 + 3z^2 = 1$. The constraint can be written as $(x+1)^2 + y^2 + 3z^2 = 2$. This means $(x+1)^2, y^2, 3z^2 \leq 2$. Thus, $|x| \leq |x+1| + |-1| \leq \sqrt{2}+1 < 3$. Using these we conclude that if (x, y, z) satisfies the given constraint, then

$$x^2 + y^2 + z^2 \leq 9 + 2 + \frac{2}{3} < 12 \Rightarrow (x, y, z) \in B_{\sqrt{12}}(0, 0, 0).$$

Therefore, the constraint gives us a bounded subset of $\mathbb{R}^3$. This subset is also closed, as it is the same as $g^{-1}(\{1\})$, where $g(x, y, z) = x^2 + 2x + y^2 + 3z^2$ is continuous and $\{1\}$ is closed in $\mathbb{R}$. Then we will use the Lagrange Multipliers Theorem to find the minimum distance.

If $\nabla f = (2x, 2y, 2z) = \mathbf{0}$, then $x = y = z = 0$, which does not satisfy the constraint. Therefore, it is always the case that $\nabla f \neq \mathbf{0}$ under the given constraint. By the Lagrange Multipliers Theorem, the minimum must satisfy $\nabla f = \lambda \nabla g$. This yields the following system:

$$\begin{cases} 2x = \lambda(2x + 2) \\ 2y = \lambda(2y) \Rightarrow \lambda = 1 \text{ or } y = 0 \\ 2z = \lambda(6z) \Rightarrow \lambda = 1/3 \text{ or } z = 0 \\ x^2 + 2x + y^2 + 3z^2 = 1 \end{cases}$$

Case I. $\lambda = 1$. Substituting this into the first equation, we obtain $2x = 2x + 2$, which is a contradiction.

Case II. $y = 0$ and $\lambda = 1/3$. The first equation yields, $2x = 2x/3 + 2/3$. Therefore, $x = 1/2$. Substituting into the last equation we obtain $z^2 = -1/4$, which is impossible.

Case III. $y = 0$ and $z = 0$, which yields $x = \pm\sqrt{2} - 1$. Therefore, the shortest distance is $\sqrt{(\sqrt{2} - 1)^2} = \sqrt{2} - 1$. □

Example 9.14. Find the minimum distance from the point $(0,0,1)$ to the points on the surface S given by $z = 2x^2 + y^2$ or show no minimum exists.

Solution. We are trying to minimize $f(x,y,z) = x^2 + y^2 + (z-1)^2$ subject to the constraint $g(x,y,z) = z - 2x^2 + y^2 = 0$. Note that f is continuous and $g(x,y,z) = 0$ is closed as this surface is $g^{-1}(\{0\})$. This surface is not bounded, however. We will show we can ignore points on this surface that are "far away". To do this, note that $(0,0,0)$ is on the given surface and $f(0,0,0) = 1$. If $|x| \geq 1$ or $|y| \geq 1$ or $|z| \geq 2$, then $x^2 \geq 1$ or $y^2 \geq 1$ or $(z-1)^2 \geq 1$. This implies,

$$f(x,y,z) \geq 1 = f(0,0,0).$$

Therefore, if the absolute minimum exists it must satisfy $|x|, |y| \leq 1$ and $|z| \leq 2$. Now, we can invoke the Extreme Value Theorem to show such an absolute minimum exists. Note that the set of all points satisfying $|x| \leq 1$ is closed, because it is the same as $\pi_1^{-1}([0,1])$, where $\pi_1(x,y,z) = x$ is continuous, and $[0,1]$ is closed in $\mathbb{R}$. Similarly, the set of all points satisfying $|y| \leq 1$ and the set of all points satisfying $|z| \leq 2$ are also closed. Since the intersection of finitely many closed sets is closed, the set E consisting of all points satisfying $|x| \leq 1$ and $|y| \leq 1$ and $|z| \leq 2$ is closed. Therefore, $S \cap E$ is closed. By the Extreme Value Theorem, f attains a minimum value over $S \cap E$. Let this minimum value be $f(\mathbf{a})$. Since $\mathbf{0} \in S \cap E$, we have $f(\mathbf{0}) \geq f(\mathbf{a})$. We already proves $f(\mathbf{x}) \geq f(\mathbf{0})$ for all $\mathbf{x} \in E \cap S$. Thus, $f(\mathbf{a})$ is an absolute minimum value. We will now find $\mathbf{a}$ using the Lagrange Multipliers Theorem.

If $\nabla f = \mathbf{0}$, then $x = y = z - 1 = 0$. However, this point is not on the given surface S. Thus, $\nabla f \neq \mathbf{0}$ for all points on S. Therefore, $\nabla f = \lambda \nabla g$. This yields the following system:

$$\begin{cases} 2x = \lambda(-4x) \Rightarrow \lambda = -1/2 \text{ or } x = 0 \\ 2y = \lambda(2y) \Rightarrow \lambda = 1 \text{ or } y = 0 \\ 2(z-1) = \lambda \\ z = 2x^2 + y^2 \end{cases}$$

Case I. $\lambda = -1/2$ and $y = 0$. Substituting into the third equation we obtain $z = 3/4$. The last equation yields $x = \pm\sqrt{3}/\sqrt{8}$. We see that $f(\pm\sqrt{3}/\sqrt{8}, 0, 3/4) = 3/8 + 1/16 = 7/16$.

Case II. $x = 0$ and $\lambda = 1$. The third equation yields $z = 3/2$. The last equation gives us $y = \pm\sqrt{3}/\sqrt{2}$. We see that $f(0, \pm\sqrt{3}/\sqrt{2}, 3/2) = 3/2 + 1/4 = 7/4$.

Case III. $x = y = 0$. The last equation yields $z = 0$. We have $f(0,0,0) = 1$.

Comparing the values found above, we conclude that the minimum of f is $7/16$. Thus, the minimum distance is $\dfrac{\sqrt{7}}{4}$. $\square$

9.4 Exercises

Exercise 9.1. *Find the points on the xy-plane on the ellipse* $x^2/9 + y^2/4 = 1$ *that are closest and farthest to the point* $(1, 0)$*. (Or show no such points exist.)*

Exercise 9.2. *Find the largest and smallest values of* $5x+12y$*, provided* x, y *satisfy* $x^2+y^2-6x+4y+12 = 0$*.*

Exercise 9.3. *Find the global extreme values of* $2x - 3y + 7z$ *given* $2x^2 + 3y^2 + 7z^2 = 12$*.*

Exercise 9.4. *Find the plane tangent to the surface* $x^3 + 2y^2z + \cos(xyz) = 2$ *at point* $(1, -1, 0)$*.*

Exercise 9.5. *Find and classify all critical points of* $f(x, y) = x^3 + 3xy^2 - 3xy$*.*

Exercise 9.6. *Find two points on the line* $x + y = 10$ *and the ellipse* $x^2 + 2y^2 = 1$ *which are closest. You need to show this minimum distance exists.*

Hint: You need to use the Lagrange Multipliers Theorem. Since there are two points on different curves, we need four variables. Thus, this is a problem in $\mathbb{R}^4$. The constraints are $x + y = 10$ and $z^2 + 2t^2 = 1$. We are trying to minimize $f(x, y, z, t) = (x - z)^2 + (y - t)^2$. To show the absolute minimum exists first show that both sets $x + y = 10$ and $z^2 + 2t^2 = 1$ are closed subsets of $\mathbb{R}^4$. These sets are unfortunately not bounded. However you can make them bounded by taking the intersection of these sets with a ball $B_r(\mathbf{0})$ in $\mathbb{R}^4$. That way you can show the minimum exists inside a ball (as long as the radius of the ball is large enough so the ball *does* intersect the constraints.) Then show that if x is "large" or y is "large" (e.g. $|x| \geq 100$ or $|y| \geq 100$), then $f(x, y, z, t)$ is more than $f(5, 5, 1, 0)$.) This means the minimum inside the set satisfying $|x|, |y| < 100$ is the same as the minimum inside $\mathbb{R}^4$. Take a look at Example 9.12 part (d).

Exercise 9.7. *Consider the function* $f(x, y) = x^3 + y^3$*.*

(a) Find all critical points of f*.*

(b) Explain why the Second Partials Test is inconclusive.

(c) Determine if each critical point is a local minimum, a local maximum or a saddle point.

Exercise 9.8. *Consider the surface given by* $x^3 + x^2 + y^2 - 2y + z^2 = 3$*. Find all points on this surface where the tangent plane is parallel to the xy-plane.*

Exercise 9.9. *Show that among all triangles whose perimeters is a fixed positive real number p the equalilateral triangle has the largest area.*

Hint: Use the Heron's Formula from Euclidean Plane Geometry.

Exercise 9.10. *Suppose* $x_1, \ldots, x_n, a_1, \ldots, a_n$ *are real numbers for which* $x_1^2 + \cdots + x_n^2 = 1$*. Using the Lagrange Multipliers Theorem prove that*

$$(a_1x_1 + \cdots + a_nx_n)^2 \leq (a_1^2 + \cdots + a_n^2).$$

Using the above prove the Cauchy-Schwarz inequality for the standard inner product on $\mathbb{R}^n$*.*

$$(a_1b_1 + \cdots + a_nb_n)^2 \leq (a_1^2 + \cdots + a_n^2) \cdot (b_1^2 + \cdots + b_n^2).$$

Exercise 9.11. *Find the minimum and maximum values of* $x^2 - y^2$ *on the ellipse* $4x^2 + 9y^2 = 13$*, or show they do not exist.*

Exercise 9.12. *Find and classify all critical points of* $(x^2 + y^2)e^{x^2-y^2}$*.*

Exercise 9.13. *Consider the plane* $x + 2y + 3z = 4$*. Find the point on this plane closest to the origin or show no such point exist. Generalize this: Show the shortest distance from a point* (x_0, y_0, z_0) *to a plane* $ax + by + cz + d = 0$*, where* a, b, c, d *are real constants with at least one of* a, b *or* c *nonzero is given by*

$$\frac{|ax_0 + by_0 + cz_0 + d|}{\sqrt{a^2 + b^2 + c^2}}$$

Exercise 9.14. *Find the points of the ellipsoid* $x^2 + 2y^2 + 3z^2 = 1$ *which are closest to and farthest from the plane* $x + y + z = 10$*, or show no such points exist.*

Exercise 9.15. *Consider a quadratic form* $q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$*, where* $A^T = A$*. Prove the set of all critical points of* q *is precisely* Ker A*.*

Exercise 9.16 (Arithmetic Mean-Geometric Mean Inequality)**.** *Let* $x_1, \ldots, x_n$ *be nonnegative real numbers. In this exercise we will prove the Arithmetic Mean-Geometric Mean Inequality stated below:*

$$\frac{x_1 + x_2 + \cdots + x_n}{n} \geq \sqrt[n]{x_1 x_2 \cdots x_n} \qquad (*)$$

(a) Show that the set S *of all points* $(y_1, \ldots, y_n) \in \mathbb{R}^n$ *satisfying* $\sum_{j=1}^{n} y_j^2 = 1$ *is closed and bounded.*

(b) Find the maximum of $\sqrt[n]{y_1^2 y_2^2 \cdots y_n^2}$ *over* S*.*

(c) By setting $y_j = \sqrt{x_j/A}$*, where* A *is the mean of* $x_1, \ldots, x_n$ *prove* $(*)$*.*

9.5 Summary

- To find the maximum and minimum values of a function f given the constraint $g = 0$, you should invoke the Lagrange Multipliers Theorem as follows:
  - Check f and g are continuously differentiable. (It is enough for f to be differentiable, if that works better.)
  - If you are trying to find absolute maximum and minimum values, then show these values exist using the Extreme Value Theorem.
  - Find all points $\mathbf{a}$ for which $\nabla g(\mathbf{a}) = \mathbf{0}$.
  - Solve the system $\nabla f = \lambda \nabla g$, and $g = 0$.
  - Compare the values of function f at all points found in the previous two steps.
- To find maximum and minimum values of f given multiple constraints $g_1 = \cdots = g_m = 0$ follow the steps above, except you would need to find all points $\mathbf{a}$ for which $\nabla g_1(\mathbf{a}), \ldots, \nabla g_m(\mathbf{a})$ are linearly dependent instead of finding those for which $\nabla g(\mathbf{a}) = \mathbf{0}$. Also, the equation $\nabla f = \lambda \nabla g$ would become $\nabla f = \lambda_1 \nabla g_1 + \cdots + \lambda_m \nabla g_m$.

Week 10

10.1 Classification of Critical Points

The Second Partials Test for a function $f : \mathbb{R}^2 \to \mathbb{R}$ allows us to determine if a critical point is a local minimum, a local maximum, or a saddle point. Now, we will turn our focus to n variable real-valued functions. Similar to what we did in the 2-dimensional case, we start with studying quadratic forms.

Definition 10.1. A matrix A is called **symmetric** if $A^T = A$. In other words, the (i, j) entry of A is the same as its (j, i) entry for all i, j.

Theorem 10.1. *Any quadratic form $q : \mathbb{R}^n \to \mathbb{R}$ can be written as $q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$, where $A \in M_n(\mathbb{R})$ is symmetric and $\mathbf{x} \in \mathbb{R}^n$ is a column vector. Furthermore, given q, this symmetric matrix A is unique.*

Note that in the theorem above if $q(x_1, \ldots, x_n) = \sum_{1\leq i\leq j\leq n} a_{ij}x_ix_j$ then the (i, j) entry of A is $a_{ij}/2$ or $a_{ji}/2$ depending on whether $i < j$ or $j < i$, and the (i, i) entry of A is a_{ii}.

Example 10.1. Write down the quadratic form below in the form $q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$, where A is symmetric.

$$q(x, y, z) = x^2 + 2y^2 - z^2 + 3xy + xz - yz.$$

Definition 10.2. Given a symmetric $n \times n$ matrix A the quadratic form $q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ is called the quadratic form **associated** with A. We also say A is the matrix associated with q.

Note that $\mathbf{0}$ is a critical point of q. Also, for a quadratic form q, a scalar c, and a vector $\mathbf{x}$ we have $q(c\mathbf{x}) = c^2q(\mathbf{x})$. Therefore, to determine if $\mathbf{0}$ is a local minimum, local maximum or a saddle point, we need to determine the maximum and minimum of q over the unit sphere given by $||\mathbf{x}|| = 1$. This can be done using the Lagrange Multipliers Theorem.

Theorem 10.2. *A quadratic form q has a local minimum (resp. local maximum) at $\mathbf{0}$ if and only if q is positive semidefinite (resp. negative semidefinite). Similarly, q has a saddle point at $\mathbf{0}$ if and only if q is nondefinite.*

Theorem 10.3. *Let q be a quadratic form associated with an $n \times n$ symmetric matrix A. If q attains its maximum or minimum value on the unit sphere in $\mathbb{R}^n$ at a point $\mathbf{v}$ (with $||\mathbf{v}|| = 1$), then $A\mathbf{v} = \lambda\mathbf{v}$ for some $\lambda \in \mathbb{R}$.*

Definition 10.3. Given a square matrix A, we say a nonzero vector $\mathbf{v}$ is an **eigenvector** of A if there is $\lambda \in \mathbb{R}$ for which $A\mathbf{v} = \lambda\mathbf{v}$. The number λ is called an **eigenvalue** of A, and the pair $(\lambda, \mathbf{v})$ is called an **eigenpair** of A.

Note that if $(\lambda, \mathbf{v})$ is an eigenpair of a matrix A associated to a quadratic form q, then $q(\mathbf{v}) = \lambda||\mathbf{v}||^2$.

Theorem 10.4. *Given a quadratic form q on $\mathbb{R}^n$, the maximum and minimum value of this quadratic form on the unit sphere $\{\mathbf{x} \in \mathbb{R}^n \mid ||\mathbf{x}|| = 1\}$ are the largest and smallest eigenvalues of the matrix A associated with q.*

Theorem 10.5. *A real number λ is an eigenvalue of a square matrix A if and only if $\det(A - \lambda I) = 0$, where I is the identity matrix.*

Corollary 10.1. Let A be the matrix associated with a quadratic form q. Then, the maximum and minimum values of $q(\mathbf{x})$ where $\mathbf{x}$ is on the unit sphere is the largest and smallest real root λ of the equation $\det(A - \lambda I) = 0$.

Example 10.2. Find the maximum and minimum values of $q(x, y) = 3x^2 + 2y^2 - 2xy$ subject to the condition that $x^2 + y^2 = 1$.

Example 10.3. Consider the quadratic form $q(x, y, z) = 3x^2 + 4xy + y^2 + 2z^2$. Find the maximum and minimum value of this quadratic form over the unit sphere. Determine whether $\mathbf{0}$ is a local maximum, local minimum, or a saddle point.

Notation. Let A be an $n \times n$ matrix. For every $k \leq n$ we denote the determinant of the upper left-hand $k \times k$ submatrix of A is denoted by Δ_k.

Definition 10.4. A symmetric matrix A is called **positive definite** if the quadratic form $q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ associated with A is positive definite. Similarly we define a negative definite, nondefinite, positive semidefinite and negative semidefinite matrix.

Theorem 10.6. *Let $q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ be a quadratic form with $A^T = A$.*

(a) q is positive definite (resp. positive semidefinite) if and only if all eigenvalues of A are positive (resp. nonnegative).

(b) q is negative definite (resp. negative semidefinite) if and only if all eigenvalues of A are negative (resp. nonpositive).

(c) q is nondefinite if and only if it has both a positive and a negative eigenvalue.

Theorem 10.7. *Let $q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ be a quadratic form with $A^T = A$.*

(a) q positive definite if and only if $\Delta_k > 0$ for all k.

(b) q is negative definite if and only if $(-1)^k\Delta_k > 0$ for all k.

(c) Assume $\det A \neq 0$*. The quadratic form* q *is nondefinite if and only if neither of the previous two conditions is satisfied.*

To classify a critical point $\mathbf{a}$ of a function f we approximate the function f with a quadratic form and then determine if this quadratic form is positive-definite, negative-definite, or nondefinite.

Definition 10.5. Let U be an open subset of $\mathbb{R}^n$. Suppose $f : U \to \mathbb{R}$ is a C^2 function. The **Hessian matrix** of f at a point $\mathbf{a} \in U$ is the $n \times n$ matrix whose (i, j) entry is $D_i D_j f(\mathbf{a})$. The Hessian matrix of f at $\mathbf{a}$ is denoted by $Hf(\mathbf{a})$.

Theorem 10.8. *Let* U *be an open subset of* $\mathbb{R}^n$*. Suppose* $f : U \to \mathbb{R}$ *is* C^2*, and let* $\mathbf{a} \in U$ *be a critical point of* f*.*

(a) If $Hf(\mathbf{a})$ *is positive-definite, then* f *has a local minimum at* $\mathbf{a}$*.*

(b) If $Hf(\mathbf{a})$ *is negative-definite, then* f *has a local maximum at* $\mathbf{a}$*.*

(c) If $Hf(\mathbf{a})$ *is nondefinite, then* f *has a saddle point at* $\mathbf{a}$*.*

Example 10.4. Consider the function

$$f(x, y, z) = 2x^2 + 5y^2 + 2z^2 + 2xz + x^4 + \sin(y^4).$$

Prove $(0, 0, 0)$ is a critical point of f, and classify this critical point.

10.2 More Examples

Example 10.5. Prove that for every three numbers x, y, z we have

$$2x^2 + 5y^2 + 10z^2 \geq 4xy + 2xz - 6yz.$$

Solution. The matrix associated to the quadratic form $2x^2 + 5y^2 + 10z^2 - 4xy - 2xz + 6yz$ is

$$\begin{pmatrix} 2 & -2 & -1 \\ -2 & 5 & 3 \\ -1 & 3 & 10 \end{pmatrix}.$$

We will use Theorem 10.7.

$$\Delta_1 = 2, \Delta_2 = \det \begin{pmatrix} 2 & -2 \\ -2 & 5 \end{pmatrix} = 10 - 4 = 6.$$

$$\Delta_3 = \begin{pmatrix} 2 & -2 & -1 \\ -2 & 5 & 3 \\ -1 & 3 & 10 \end{pmatrix} = 2(50 - 9) + 2(-20 + 3) - 1(-6 + 5) = 49.$$

Since $\Delta_1, \Delta_2, \Delta_3$ are all positive, the quadratic form is positive-definite and thus,

$$2x^2 + 5y^2 + 10z^2 - 4xy - 2xz + 6yz \geq 0,$$

for all x, y, z. This completes the proof. □

Example 10.6. Prove that the eigenvalues of an upper triangular matrix is its diagonal entries.

Solution. Consider the upper triangular matrix A whose diagonal entries are $a_1, \ldots, a_n$. The matrix $A - \lambda I$ is also upper triangular with diagonal entries $a_1 - \lambda, \ldots, a_n - \lambda$. By an exercise,

$$\det(A - \lambda I) = (a_1 - \lambda) \cdots (a_n - \lambda).$$

The roots of this polynomial are $a_1, \ldots, a_n$. This completes the proof. □

Example 10.7. Classify $\mathbf{0}$ as a minimum, maximum or a saddle point of each quadratic form:

(a) $f(x, y, z) = x^2 + y^2 + 2z^2 - xy - yz$.

(b) $f(x, y, z) = -x^2 - 2y^2 + z^2 + 4xy + 6zy$.

Solution. (a) The matrix associated to this quadratic form is

$$\begin{pmatrix} 1 & -1/2 & 0 \\ -1/2 & 1 & -1/2 \\ 0 & -1/2 & 2 \end{pmatrix}.$$

We will use Theorem 10.7:

$$\Delta_1 = 1, \Delta_2 = 1 - 1/4 = 3/4, \text{ and } \Delta_3 = 5/4.$$

Since $\Delta_1, \Delta_2, \Delta_3$ are all positive, $\mathbf{0}$ is a local (and absolute) minimum.

(b) The matrix associated to this quadratic form is

$$\begin{pmatrix} -1 & 2 & 0 \\ 2 & -2 & 3 \\ 0 & 3 & 1 \end{pmatrix}.$$

We will again use Theorem 10.7:

$$\Delta_1 = -1, \Delta_2 = -2, \Delta_3 = 7.$$

Since Δ_1 and Δ_2 are both negative, and Δ_3 is nonzero, $\mathbf{0}$ is a saddle point. □

Example 10.8. Find and classify all critical points of the function:

$$f(x, y, z) = x^3 + xy^2 + x^2 + y^2 + 3z^2.$$

Solution. To find the critical points we need to solve the following system:

$$\begin{cases} f_x = 3x^2 + y^2 + 2x = 0 \\ f_y = 2xy + 2y = 0 \Rightarrow 2y(x+1) = 0 \Rightarrow x = -1 \text{ or } y = 0 \\ f_z = 6z = 0 \Rightarrow z = 0 \end{cases}$$

If $x = -1$, the first equation yields $y^2 + 1 = 0$, which has no roots.

If $y = 0$, the first equation yields $3x^2 + 2x = 0$ which implies $x = 0$ or $x = -2/3$. Therefore, we obtain two critical points $(0, 0, 0)$ and $(-2/3, 0, 0)$.
The Hessian matrix is

$$Hf(x, y, z) = \begin{pmatrix} 6x+2 & 2y & 0 \\ 2y & 2x+2 & 0 \\ 0 & 0 & 6 \end{pmatrix}.$$

Evaluating this at $(0, 0, 0)$ gives us the matrix

$$Hf(0, 0, 0) = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 6 \end{pmatrix}.$$

The eigenvalues of this matrix are $2, 2, 6$ which are all positive. Therefore, $\mathbf{0}$ is a local minimum.

At $(-2/3, 0, 0)$ the Hessian matrix becomes

$$Hf(-2/3, 0, 0) = \begin{pmatrix} -2 & 0 & 0 \\ 0 & 2/3 & 0 \\ 0 & 0 & 6 \end{pmatrix}.$$

The eigenvalues are $-2, 2/3, 6$. Since there are both negative and positive eigenvalues, $\mathbf{0}$ is a saddle point. $\square$

Example 10.9. Suppose at least one of the diagonal entries of a symmetric matrix is negative. Prove this matrix has at least one negative eigenvalue.

Solution. Suppose the (i, i) entry of a symmetric matrix A is negative. Since the i-th column of A is the same as $A\mathbf{e}_i$, the i-th entry of $A\mathbf{e}_i$ is negative. Therefore, $\mathbf{e}_i^T A\mathbf{e}_i$ is negative. Thus, the minimum of the quadratic form q given by $q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ on the unit sphere is negative. By Theorem 10.3, this minimum is an eigenvalue of A, which means A has a negative eigenvalue. $\square$

More examples from Edwards:
Pages 145-156, Examples 1-5

10.3 Exercises

Exercise 10.1. *Classify* $\mathbf{0}$ *as a local minimum, local maximum or a saddle point of the following quadratic form, in two ways:*

$$f(x, y, z) = x^2 - y^2 - z^2 + 4xy + 6xz$$

(a) Using an appropriate Theorem.

(b) By evaluating e-values.

Exercise 10.2. *Consider the function*

$$f(x, y, z) = x^2 + 4y^2 + z^2 + 2xz + (x^2 - y^2 + z^2)\cos(xyz).$$

Prove that $(0, 0, 0)$ *is a critical point of* f *and classify this critical point.*

Exercise 10.3. *Prove that for all real numbers* x, y, z *we have*

$$3x^2 + 2y^2 + 6z^2 + 2xy + 2xz + 6yz \geq 0.$$

Exercise 10.4. *Consider the quadratic form*

$$q(x_1, x_2, x_3) = -x_1^2 - x_2^2 + 4x_3^2 + 4x_1x_2 - 2x_1x_3 - 2x_2x_3.$$

Determine if this quadratic form is positive semidefinite, negative semidefinite or nondefinite, once by finding its eigenvalues and once by applying an appropriate theorem.

Exercise 10.5. *Find the derivative of a quadratic form* $q(x) = \mathbf{x}^T A\mathbf{x}$*, where* $A^T = A \in M_n(\mathbb{R})$*.*

10.4 Challenge Problems

Exercise 10.6. *Suppose* $f : \mathbb{R}^n \to \mathbb{R}$ *is a differentiable function for which* $f'(\mathbf{a}) = \mathbf{a}^T A$*, for all* $\mathbf{a} \in \mathbb{R}^n$ *and a fixed symmetric matrix* $A \in M_n(\mathbb{R})$*. Prove if* $f(\mathbf{0}) = 0$*, then* f *is a quadratic form.*

Additional Examples From Edwards' Book: p. 159: 8.6

10.5 Summary

- To determine if a quadratic form is positive-definite, negative-definite, or nondefinite:
  - Form the $n \times n$ matrix associated with the quadratic form.
  - Find all eigenvalues of A.
  - If all eigenvalues are positive (resp. nonnegative), then the quadratic form is positive definite (resp. positive semidefinite). If all eigenvalues are negative (resp. nonpositive), then the quadratic form is negative definite (resp. negative semidefinite). If there are both positive and negative e-values, the form is nondefinite.
  - If finding the e-values is not easy you could also do the following:
    - Evaluate $\Delta_1, \ldots, \Delta_n$, the determinants of upper left submatrices of A of various sizes.
    - If $\Delta_k > 0$ for all k, then the form is positive definite.

    * If $(-1)^k\Delta_k > 0$ for all k, then the form is negative definite.
    * If Δ_n is nonzero and neither of the above is true, then the form is nondefinite.

- To find out if a critical point is a local maximum, local minimum or a saddle point:
  - Find the Hessian matrix at the critical point.
  - Check the determinant of the Hessian matrix is nonzero.
  - Determine if the quadratic form associated with this matrix is positive definite, negative definite or nondefinite.
  - Positive definite implies we have a local minimum.
  - Negative definite implies we have a local maximum.
  - Nondefinite implies there is a saddle point.