# Lecture Notes for the Math Camp

Max Levine and Luis Linhares

# Introduction

Welcome to the PhD! The purpose of this math camp is to help you make a smooth transition into the life of a doctoral student. Throughout the course, we will cover key mathematical topics essential for graduate-level economics, including real analysis, linear algebra, probability theory, and optimization. These notes should serve as a quick reference to the topics covered in class and are best used in conjunction with other reference materials. If you encounter any errors or typos, please let us know.

Our aim is not to overwhelm, but to provide you with a foundation and ease you into the life of a PhD student in an environment with lower stakes (there are no exams!). Students may come from a wide range of backgrounds, and this course is meant to bridge those gaps. Most of the results and theorems presented in these notes are well-known, and if a proof is not provided, it can be easily found online, in case you are curious.

Do not worry if some of the material is unfamiliar to you — that's exactly what this course is for. Think of it as an opportunity to strengthen your mathematical intuition, get comfortable with writing mathematical proofs, and begin forming a toolkit that will serve you throughout the program. With this out of the way, we can begin.

# 1 Logic

In this section, we will go over some concepts of logic that form the basis of mathematical statements and proofs.

**Definition 1.1.** *A **statement** is a declarative sentence that can be either True (T) or False (F).*

The following are examples of statements:

- The sky is blue.
- I love economics.
- 1+1 = 0.

We represent a statement using letters A, B, P, Q. . . . . For instance, we can say "P is true", or "Q is false." With a slight abuse of language, we may say "assume P" when in fact we mean "assume P is true."

We are often interested in how statements are related to each other: "if P is true, then Q is false," and so on. For instance:

<u>If</u> $\{x_n\}$ is a bounded sequence in $\mathbb{R}^{\mathbb{N}}$, <u>then</u> $\{x_n\}$ has a convergent subsequence.

## 1.1 Logical Operations

We can connect statements to create new ones through logical operations, which are defined using truth tables.

<u>**Negation** (not)</u>

A statement modified by the word "not" or the symbol ¬ is the **negation** of the original statement.

| P | $\neg P$ |
|---|---|
| T | F |
| F | T |

Table 1: Truth table for Negation.

<u>**Conjunction** (and)</u>

A **conjunction** is a new statement formed by joining two original statements using the word "and" or the symbol ∧.

| P | Q | $P \wedge Q$ |
|---|---|---|
| T | T | T |
| T | F | F |
| F | T | F |
| F | F | F |

Table 2: Truth table for Conjunction.

**Disjunction** (or)

A **disjunction** is a new statement formed by joining two original statements using the word "or" or the symbol $\vee$.

| P | Q | $P \vee Q$ |
|---|---|---|
| T | T | T |
| T | F | T |
| F | T | T |
| F | F | F |

Table 3: Truth table for Disjunction.

NOTE: In some contexts, mainly computer sciences, it is also useful to define the "exclusive or", denoted by $XOR$ or $\oplus$, which is only true when only one of the statements is true.

| P | Q | $P \oplus Q$ |
|---|---|---|
| T | T | F |
| T | F | T |
| F | T | T |
| F | F | F |

Table 4: Truth table for Exclusive Or.

**Conditional** (If ..., then ...)

A **conditional statement** is a new statement constructed in the form of "If P, then Q". The statement following "if" is the antecedent, and the statement following "then" is the consequent. The conditional statement "If P, then Q" is also denoted by $P \implies Q$.

The understanding of the truth table for the conditional statement might require a bit more explanation:

- Falsification: The statement $P \implies Q$ is false if [$P$ is true and $Q$ is false] is a result that can be tested (falsifiable). If, however, $P$ is false, then there is not much that can be said about $Q$. For instance, look at the following examples of what happens when $P$ is false and:

| P | Q | $P \implies Q$ |
|---|---|---|
| T | T | T |
| T | F | F |
| F | T | T |
| F | F | T |

Table 5: Truth table for Conditional.

- Q is true: “If it is raining hamburgers (P), then I am mortal (Q).”
- Q is false: “If it is raining hamburgers (P), then I am immortal (Q).”

- This definition of $P \implies Q$ provides a useful and intuitive definition of logical equivalence.

Other ways of saying “If P, then Q”:
- “P implies Q”,
- “P only if Q”,
- “Q is necessary for P”,
- “Q if P”,
- “P is sufficient for Q”.

NOTE: The statement $P \implies Q$ is the same as $(\neg P) \vee Q$, which might be easier to think about when building the truth table.

**Biconditional** (...if, and only if, ...)

A **biconditional statement** is a new statement constructed in the form of “P if, and only if, Q”.

**Definition 1.2.** *We say that $P$ and $Q$ are **(logically) equivalent**, written as $P \iff Q$, if $P \implies Q$ and $Q \implies P$.*

This gives us the following truth table.

| P | Q | $P \iff Q$ |
|---|---|---|
| T | T | T |
| T | F | F |
| F | T | F |
| F | F | T |

Table 6: Truth table of Biconditional.

Other ways of saying “P if, and only if, Q”:
- “P and Q are equivalent”,

- "P is necessary and sufficient for Q",
- "Q is necessary and sufficient for P".

**Definition 1.3.** *Given a statement* $P \implies Q$*,*

*i) we call the* ***converse*** *the statement* $Q \implies P$*;*

*ii) we call the* ***contrapositive*** *the statement* $\neg Q \implies \neg P$*.*

## 1.2 Logical Quantifiers

**Universal Quantifier** ($\forall$)

The **universal quantifier** is the phrase "for every" or "for all". Taking an arbitrary $x$, it is denoted by $\forall x$. Examples:

- $x$ is even $\forall x \in \mathbb{N}$ (false);[1]
- $x + 0 = x \ \ \forall x \in \mathbb{R}$ (true).

Formally, for an arbitrary statement $P(x)$ defined for any $x$, we have

| P(a) | P(b) | P(c) | ... | $\forall x, P(x)$ |
|---|---|---|---|---|
| T | T | T | ... | T |
| F | T | T | ... | F |
| T | F | T | ... | F |
| ... | ... | ... | ... | |

Table 7: Truth table for Universal Quantifier.

**Existencial Quantifier** ($\exists$).

The **existencial quantifier** is the phrase "there exists". Taking an arbitrary $x$, it is denoted by $\exists x$. Examples:

- $\exists x \in \mathbb{N}$ such that $x$ is prime and even (true);
- $\exists x \in \mathbb{N}$ such that $x + 1 = x$ (false).

For an arbitrary statement $P(x)$ defined for any $x$, the corresponding truth table for the existential quantifier is shown in Table 8.

NOTE: The two quantifiers are related by negation. For instance, "not all prime numbers are odd if, and only if, there exists an even prime number". More formally,

$$[\neg(\forall x, P(x))] \iff [\exists x, \neg P(x)].$$

[1] We have not formally defined natural or real numbers, but we assume the reader is already familiar with them.

| P(x) | P(b) | P(c) | ... | $\exists x, P(x)$ |
|---|---|---|---|---|
| F | F | F | ... | F |
| T | F | F | ... | T |
| F | T | F | ... | T |
| ... | ... | ... | ... | |

Table 8: Truth table for Existencial Quantifier.

## 1.3 Proof Techniques

Starting from a set of well-defined assumptions, we want to derive useful conclusions that are true, and the way we reason that something is true from a given set of assumptions is through mathematical proofs. They not only provide logical evidence that the reasoning is correct, but they also highlight the steps and reasoning that lead from the assumptions to the conclusions. Here we provide a few techniques commonly used to prove mathematical results.

### 1.3.1 Direct Proof

To prove $P \implies Q$, assume P is true and deduce Q is true (why is this enough?). It may be done by first proving a series of intermediate statements

$$P \implies P_1 \implies P_2 \implies \cdots \implies Q.$$

Example:

Consider that an integer[2] $n \in \mathbb{Z}$ is even if $\exists m \in \mathbb{Z}$ such that $n = 2m$. We want to prove the following result.

*For any $n \in \mathbb{Z}$, $n$ is even $\implies n^2$ is even.*

*Proof.*

1. Assume $n \in \mathbb{Z}$ is even.
2. Then, $\exists m \in \mathbb{Z}$ s.t. $n = 2m$. (definition)
3. $n^2 = (2m)^2 = 4m^2 = 2(2m^2)$.
4. Since $m \in \mathbb{Z}$, $2m^2 \in \mathbb{Z}$. ($\mathbb{Z}$ is closed under multiplication).
5. Therefore, $n^2$ is even.

$\square$

We do not need bullet points to write the proof. It is possible to make it more concise.

[2] Formally, we have not defined what an integer is yet. We have added this example here for clarity since we assume the reader has had contact with integer numbers before.

*Proof.* Assume $n \in \mathbb{Z}$ is even. Then, $\exists m \in \mathbb{Z}$ s.t. $n = 2m$. Now, note that $n^2 = (2m)^2 = 4m^2 = 2(2m^2)$. But since $m \in \mathbb{Z}$, $2m^2 \in \mathbb{Z}$, which implies that $n^2$ is even. $\square$

NOTE: We show that a proof is over with a $QED$ or with the symbol $\square$.

### 1.3.2 Contrapositive

When we want to prove $P \implies Q$, sometimes it is easier to prove the contrapositive, i.e., that $\neg Q \implies \neg P$. This is sufficient for the proof because $[\neg Q \implies \neg P] \iff [P \implies Q]$. (Can you prove this equivalence?)

### 1.3.3 Proof by Contradiction.

We want to prove that $P$ is true, but the direct approach might be too complicated. In a proof by contradiction, we assume instead that $\neg P$ is true, and use this to derive a contradiction, usually that $\neg P \implies [Q \wedge \neg Q]$ for another statement $Q$, which is impossible, allowing us to reject our initial (wrong) assumption.

Example: Prove $\sqrt{2}$ is irrational.

1. Assume, **by way of contradiction,** that $\sqrt{2}$ is a rational number. [assume $\neg P$]
2. Then, there exist $a, b \in \mathbb{Z}$ such that $\sqrt{2} = \frac{a}{b}, b \neq 0$ and $a, b$ are relative primes (have no common divisor other than 1). $[\neg P \implies Q]$
3. This means that $2 = \frac{a^2}{b^2} \implies 2b^2 = a^2$.
4. We then deduce that $a^2$ is even, and therefore $a$ is also even (prove this statement as an exercise).
5. Since $a$ is even, there exists $m \in \mathbb{Z}$ such that $a = 2m \implies a^2 = 4m^2$ and $a^2$ is divisible by 4.
6. Rearranging, $b^2 = \frac{a^2}{2} = \frac{4m^2}{2} = 2m^2$ and $b^2$ is thus even.
7. But, as we have seen before, $b^2$ is even $\implies b$ is even.
8. Since $a, b$ are both even, they have a common divisor of 2. Contradiction. $[\neg P \implies \neg Q]$
9. We then conclude that $\sqrt{2}$ is irrational.

In which step would there be no contradiction if we were to use this logic to try to (wrongly) prove that $\sqrt{4}$ is not rational?

### 1.3.4 Proof by Mathematical Induction

This is a useful method for proving statements about the natural numbers.

**Theorem 1.1.** ***(The Principle of Mathematical Induction)*** *Let $P_1, P_2, \dots$ be a series of statements indexed by the real numbers. If $P_1$ is True, and ($P_m$ is True $\implies$ $P_{m+1}$ is True) $\forall m \in \mathbb{N}$, then $P_n$ is True $\forall n \in \mathbb{N}$.*

In a proof by induction, we first prove that the statement is True for $n = 1$. Then, we assume that the statement is true for an arbitrary $n$ and prove that this implies that it is True for $n + 1$. By the Principle of Mathematical induction, this means that it is True for all natural numbers.

Example: Prove that the sum of the first $n$ natural numbers is $\frac{(n+1)n}{2}$ for all $n \in \mathbb{N}$.

*Proof.* These are the steps:

1. Check for $n = 1$:
   $\frac{(1+1)(1)}{2} = \frac{2}{2} = 1$, which is the sum of the first $n = 1$ numbers, so it is True.
2. Assume it is True for an arbitrary number $n$, so $1 + 2 + \cdots + n = \frac{(n+1)n}{2}$. Check if this implies it is also true for $n + 1$:
   $1 + 2 + \cdots + n + (n + 1) = \frac{(n+1)n}{2} + (n + 1) = \frac{(n+1)n}{2} + \frac{2(n+1)}{2} = \frac{(n+2)(n+1)}{2}$, so it is True for $n + 1$.

□

# 2 Introduction to Set Theory

Sets form the basis of mathematics.

**Definition 2.1.** *A* ***Set*** *is a collection of distinct elements.*

The notation $x \in S$ (or $S \ni x$) means "$x$ is an element of the set $S$." The negation of this statement is given by $x \notin S$, which means "$x$ is not an element of $S$". A set is defined by listing its elements between curly brackets:

$$S = \{2, 7, \sqrt{43}, \triangle, \text{dog}\}.$$

Instead of listing the elements, we can alternatively define a set by describing the properties satisfied by its elements. For instance,

$$\begin{cases} S \text{ is the set of the first 4 odd natural numbers;} \\ S = \{1, 3, 5, 7\}; \\ S = \{n \in \mathbb{N} : (n < 5) \wedge (n/2 \notin \mathbb{N})\}. \end{cases}$$

A set is completely defined by its elements, or members. Therefore, $\{1, 2, 3\}, \{3, 1, 2\}$, and $\{1, 1, 2, 2, 2, 3\}$ are all the same set. Elements can be anything, including other sets. For instance, $\{\{1\}\}$ is the set containing the set $\{1\}$. Note that $\{\{1\}\} \neq \{1\}$.

The set containing no elements is called the **empty set**, denoted by $\emptyset$. A set with only one element is called a **singleton**.

A set $A$ is a **subset** of a set $B$ (denoted by $A \subseteq B$) if $x \in A \implies x \in B$. For instance, $\{1, 3\} \subseteq \{1, 2, 3\}$. Sets $A$ and $B$ are **equal** (denoted $A = B$) if $A \subseteq B$ and $B \subseteq A$. If $A \subseteq B$ but $A \neq B$, we say that $A$ is a **strict subset** of $B$.

For a non-empty set $S$, we denote by $2^S$ (or sometimes by $P(S)$) the **power set** of $S$, the set of all subsets of $S$,

$$2^S \coloneqq \{T : T \subseteq S\}.$$

The empty set is a subset of all other sets, and a set is always a subset of itself. Therefore, for any arbitrary set $S$,

$$\emptyset \in 2^S \text{ and } S \in 2^S.$$

The notation $2^S$ comes from the fact that if $S$ has $n$ elements, the power set of $S$ is $2^n$.

Example:
Let $S = \{1, 2\}$, and we have $2^S = \{\emptyset, \{1\}, \{2\}, S\}$.

The **cardinality** of a set $S$ is the number of elements of the set

## 2.1 Set Operations

We define the following set operations.

- Set Union: $A \cup B \equiv \{x : x \in A \text{ or } x \in B\}$;
- Set Intersection: $A \cap B \equiv \{x : x \in A \text{ and } x \in B\}$;
- Complement: $A^{\mathsf{c}} = \overline{A} \equiv \{x : x \notin A\}$.
- Relative Complement: $A \setminus B = A - B \equiv \{x : x \in A \text{ and } x \notin B\}$.

**Definition 2.2.** *We say that two sets, $A, B$, are* ***disjoint*** *if $A \cap B = \emptyset$.*

We can expand the definition of unions and intersections to a collection of sets. Given an **index set** I, we can define a collection of sets $\{A_i : i \in I\}$, and we get

$$\bigcup_{i \in I} A_i = \{x : \exists i \in I \text{ s.t. } x \in A_i\};$$
$$\bigcap_{i \in I} A_i = \{x : \forall i \in I, x \in A_i\};$$

**Definition 2.3.** *A family of sets $\{A_i\}_{i \in I}$, indexed by a set $I$, is* ***pairwise disjoint*** *if $A_i \cap A_j = \emptyset\ \forall i, j \in I$ with $i \neq j$,.*

**Theorem 2.1.** ***(De Morgan's Laws)*** *Let $A$ and $B$ be sets. Then,*

*i)* $(A \cup B)^{\mathsf{c}} = A^{\mathsf{c}} \cap B^{\mathsf{c}}$;
*ii)* $(A \cap B)^{\mathsf{c}} = A^{\mathsf{c}} \cup B^{\mathsf{c}}$.

## 2.2 Relations

An **ordered pair** is an ordered list $(a, b)$ of two objects $a$ and $b$. It is ordered because the order matters, i.e.,

$$(a, b) = (a', b') \iff [a = a' \wedge b = b'].$$

The **Cartesian product** of two non-empty sets $A$ and $B$, denoted $A \times B$, is the set of all ordered pairs $(a, b)$ where $a \in A$ and $B \in B$, i.e.,

$$A \times B \coloneqq \{(a, b) : a \in A \text{ and } b \in B\}.$$

Note that $\{a\} \times \{b\} = \{(a,b)\}$. Similarly, for any $n \in \mathbb{N}$, we can define an ordered list $(a_1, \dots, a_n)$. The Cartesian product for $n$ sets, is given by

$$\Pi_{i=1}^n A_i = A_1 \times \cdots \times A_n \coloneqq \{(a_1, \dots, a_n) : \forall i \in \{1, \dots, n\}, a_i \in A_i\}.$$

**Definition 2.4.** *A* ***binary relation*** *from $X$ to $Y$ is a subset $R \subseteq X \times Y$. When $Y = X$, we say that $R$ is a relation on $X$.*

If $(x,y) \in R$, we can alternatively write $xRy$.

Example:

i) The Equality relation: $R = \{(a,a) : a \in \mathbb{R}\}$, also denoted $=$.

ii) Greater or Equal relation: $R = \{(a,b) \in \mathbb{R} \times \mathbb{R} : a \geqslant b\}$, also denoted $\geqslant$.

We usually denote relations with the known symbols like $=, \geqslant, >, \succeq$ .

**Definition 2.5.** *A relation $R$ on $X$ is said to be*

*i)* ***reflexive*** *if $xRx$ for all $x \in X$;*

*ii)* ***complete*** *if, for any $x, y \in X$, we have either $xRy$ or $yRx$;*

*iii)* ***symmetric*** *if, for any $x, y \in X$, $xRy \implies yRx$;*

*iv)* ***antisymmetric*** *if, for any $x, y \in X$, $[xRy \wedge yRx] \implies x = y$;*

*v)* ***transitive*** *if, for any $x, y, z \in X$, we have $[xRy \wedge yRz] \implies xRz$.*

*vi)* ***irreflexive*** *if $(x,x) \notin R$ for all $x \in X$.*

NOTE: A **weak order** is a relation that is complete and transitive. It will be very important for microeconomics.

## 2.3 Order Relations

Fix a relation $\succeq$ on a non-empty set X, which gives us a list $(X, \succeq)$. If $\succeq$ is

- transitive and reflexive, we call it a **pre-order** on $X$, and $(X, \succeq)$ is a pre-ordered set;
- an antisymmetric pre-order on X, we call it a **partial order**, and $(X, \succeq)$ is a poset (partially ordered set);
- a complete partial order, we call it a **linear order**, and $(X, \succeq)$ is a **loset** (linear ordered set) or a chain.

Example:

$(\mathbb{R}, \geqslant)$ is a loset and $\geqslant$ is a linear order. However, $(\mathbb{R}^2, \geqslant)$ where

$$(x_1, x_2) \geqslant (y_1, y_2) \iff [x_1 \geqslant y_1 \wedge x_2 \geqslant y_2],$$

is a poset and $\geqslant$ is a partial order. Indeed, how do you order $(1,2)$ and $(2,1)$ according to $\geqslant$?

Fix a relation $\succeq$ on a non-empty set X. The asymmetric part of $\succeq$ is the relation $\succ$, defined by

$$x \succ y \iff [x \succeq y \wedge \neg(y \succeq x)].$$

The symmetric part of $\succeq$ is the relation $\sim$, defined by

$$x \sim y \iff [x \succeq y \wedge y \succeq x].$$

Let $(X, \succeq)$ be a pre-ordered set. Now, fix $Y \subseteq X$ such that $Y$ is non-empty. For any $x \in Y$, we then say that $x$ is

i) $\succeq$-**maximal** in $Y$ if there is no $y \in Y$ such that $y \succ x$;
ii) $\succeq$-**minimal** in $Y$ if there is no $y \in Y$ such that $x \succ y$;
iii) the $\succeq$-**maximum** of $Y$ if $x \succeq y$ for all $y \in Y$;
iv) the $\succeq$-**minimum** of $Y$ if $y \succeq x$ for all $y \in Y$.

Example:

Let $A \subseteq \mathbb{N}$ be endowed with the divisibility relation. That is, for $a, b \in A$, we say that $a$ "divides" $b$. For instance, if $A = \{2, 3, 5, 6, 12, 15, 30\}$ is the set of divisors of 30 larger than 1, the maximal elements are $2, 3$, and 5, as there are no other elements in the set that divide them. Since there is more than one maximal element, there is no maximum (i.e., a number that divides all numbers in the set). However, 30 is the minimum (and consequently the minimal) as all numbers in the set divide 30.

Example:

In an economy, consider a set of allocations and the partial order of Pareto improvement. Then, the Pareto optimal allocations are maximal elements.

Now, take an element $x \in X$ (it may be in $Y$ or not). We say that $x \in X$ is

i) an $\succeq$-**upper bound** for $Y$ if $x \succeq y$ for all $y \in Y$;
ii) an $\succeq$-**lower bound** for $Y$ if $y \succeq x$ for all $y \in Y$;
iii) the $\succeq$-**supremum** of $Y$ if it is the $\succeq$-minimum of the set of all $\succeq$-upper bounds;
iv) the $\succeq$-**infimum** of $Y$ if it is the $\succeq$-maximum of the set of all $\succeq$-lower bounds;

In other words, the difference between a maximum and a supremum is: a maximum of a set always belongs to the set, but a supremum does not need to belong to the set.

We usually do not include the $\succeq$- symbol when it is clear what relation we are working with when talking about maximum, upper bound, etc...

Example: Consider the linear order $\geqslant$ on $\mathbb{R}$. The maximum and supremum of the closed interval $[0,1]$ is 1. The supremum of the open interval $[0,1)$ is 1, but there is no maximum; you can always get a higher number that gets closer and closer to 1 without leaving the interval $[0,1)$.

## 2.4 Functions

A **function** $f$ that maps $X$ into $Y$, denoted by

$$f : X \longrightarrow Y,$$

is a relation $f \subseteq X \times Y$ such that, for every $x \in X$, the set $\{y \in Y : xfy\}$ is a singleton. This means that

i) for every $x \in X$, there exists $y \in Y$ such that $xfy$;

ii) for every $y, z \in Y$, $[xfy \wedge xfz] \implies y = z$.

We use the notation $f(x) = y$ when $xfy$. When the sets $X, Y$ are implicit, we can also denote $f$ by

$$f : x \mapsto f(x).$$

Finally, $Y^X$ denotes the set of all functions that map $X$ into $Y$. For any function $f \in Y^X$, we say that $X$ is the **domain** of $f$ and $Y$ is the **codomain**. The **range** or **image** of $f$ is

$$f(X) \coloneqq \{y \in Y : \exists x \in X \text{ s.t. } f(x) = y\} = \{f(x) : x \in X\}.$$

In addition, for any $A \subseteq X$, the **image** of $A$ under $f$ is the set

$$f(A) \coloneqq \{y \in Y : \exists x \in A \text{ s.t. } f(x) = y\}.$$

Correspondingly, the **inverse image** of a set $B \subseteq Y$ is given by

$$f^{-1}(B) \coloneqq \{x \in X : f(x) \in B\}.$$

Note that $f^{-1}(B)$ is a set. A function $f \in Y^X$ is **invertible** if $f^{-1}(\{y\})$ is a singleton for all $y \in Y$. If $f$ is invertible, we can define $g : Y \longrightarrow X$ as the **inverse** of $f$, by solving

$$\{g(x)\} = f^{-1}(\{y\})\ \forall x \in X.$$

The **graph** of a function $f$ is defined by

$$gph(f) \coloneqq \{(x, y) \in X \times Y : y = f(x)\} = \{(x, f(x)) : x \in X\}.$$

A function is said to be a

- **surjection** when $f(X) = Y$;
- **injection** if $x \neq x' \implies f(x) \neq f(x')$ for all $x, x' \in X$;
- **bijection** if it is both surjective and injective.

Example:

- An injective, but not surjective function: $f : \mathbb{R}_+ \to \mathbb{R},\ f(x) = \sqrt{x}$.
- A surjective, but not injective function: $f : \mathbb{R} \to \mathbb{R}_+,\ f(x) = x^2$.
  - If, instead, we had $f : \mathbb{R} \to \mathbb{R},\ f(x) = x^2$, it would not be surjective anymore.
- A bijective function: $f : \mathbb{R} \to \mathbb{R},\ f(x) = x + 1$.

An arbitrary set $X$ is said to be **countably infinite** if there exists a bijection that maps $X$ into the natural numbers. The set $X$ is **countable** if it is finite or countably infinite. Otherwise, it is called **uncountable**.

A few important functions:

- The function $f \in Y^X$ is **constant** if there exists $\overline{y} \in Y$ such that $f(x) = \overline{y}$ for all $x \in X$.
- The **identity function** on the set $X$ is defined by $Id_X(x) = x$ for all $x \in X$.
- The **indicator function** (or characteristic function) for a set $S \subseteq X$, $\mathbb{1}_S : X \longrightarrow \{0, 1\}$, is given by

$$\begin{cases} \mathbb{1}_S(x) = 1 \text{ if } x \in S; \\ \mathbb{1}_S(x) = 0 \text{ if } x \notin S. \end{cases}$$

- For a function $f \in X^{X \times Y}$, we say that $f$ is a **projection of** $X \times Y$ **onto** $X$ if $f$ is defined by

$$f(x, y) = x \text{ for all } x \in X.$$

- Given $f : X \longrightarrow Y$ and $g : Y \longrightarrow Z$, the **composition** of $f$ and $g$ is defined as the new function $g \circ f : X \longrightarrow Z$, given by

$$(g \circ f)(x) \coloneqq g(f(x)) \text{ for all } x \in X.$$

## 2.5 Sequences

Sequences are one of the most important topics that we will see. They are used to formalize many of the more advanced concepts we will discuss in the future. They also show up in numerous economic applications, like in discrete-time models, and asymptotic analysis, just to name a few.

We define $\mathbb{N} = \{1, 2, 3, \dots\}$ as the set of real numbers.[3] A **sequence** on a non-empty set $X$ is a function $f : \mathbb{N} \longrightarrow X$. We denote a sequence by the ordered list $(x_1, x_2, \dots)$, where $x_i \coloneqq f(i)$ for all $i \in \mathbb{N}$. It may be helpful to think of sequences as an infinite ordered list of elements of $X$ that is indexed by $\mathbb{N}$. We can also denote a sequence by

$$\{x_i\}_{i=1}^{\infty},\ \{x_i\}_i,\ \text{or } \{x_i\}.$$

The set of all sequences on $X$ is $X^{\mathbb{N}}$ (sometimes also denoted as $X^{\infty}$). A **subsequence** of a sequence $f$ on $X$ is a function $g \in X^{\mathbb{N}}$ such that $g(i) \coloneqq f \circ \varphi(i)$, where $\varphi : \mathbb{N} \longrightarrow \mathbb{N}$ is a strictly increasing function. The subsequence can be represented by $\{x_{i_k}\}_{k=1}^{\infty}$, where for all $k \in \mathbb{N}$,

$$i_k \coloneqq \varphi(k).$$

Example:

Let us consider the alternating sequence $(-1, 1, -1, 1, \dots)$ where $x_i = (-1)^i$. We can find a subsequence $(1, 1, 1, \dots)$, where $\varphi(i) = 2i$.

[3] In an alternate definition, we could also consider 0 a real number, and $\mathbb{N}$ would be the set of all positive integers.

# 3 The Real Number System

We denote the real numbers by $\mathbb{R}$.

## 3.1 Ordering

Important subsets of the real numbers:

- The natural numbers $\mathbb{N}$;
- The integers $\mathbb{Z}$;
- The rational numbers $\mathbb{Q}$;
- The irrational numbers $\mathbb{R} \setminus \mathbb{Q}$.

Let $A \subseteq \mathbb{R}$.

- The number $m \in \mathbb{R}$ is an **upper bound** (**lower bound**) of A if $m \geqslant a$ $(m \leqslant a)$ $\forall a \in A$.
- If $A$ has both an upper bound ($\overline{m} \in \mathbb{R}$) and a lower bound ($\overline{m} \in \mathbb{R}$), then we say that A is bounded.
- Let $m \in \mathbb{R}$ be a lower bound of $A$ such that $N > m \implies N$ is not a lower bound of $A$. Then, we say that $m$ is the **infimum**, or **greatest lower bound** of A. If $\inf A \in A$, we say that $m$ is the **minimum** of A.
- Let $m \in \mathbb{R}$ be an upper bound of $A$ such that $N < m \implies N$ is not an upper bound of $A$. Then, we say that $m$ is the **supremum**, or **least upper bound** of A. If $\sup A \in A$, we say that $m$ is the **maximum** of A.

**Proposition 3.1.** *For any $A \subseteq \mathbb{R}$, if $\sup A$ $(\inf A)$ exists, it is unique.*

*Proof.* Let $a$ and $b$ be suprema of $A$. Then, $b$ is an upper bound of A by definition. Since $a$ is a supremum, $a \leqslant b$. Similarly, $a$ is an upper bound of $A$, and since $b$ is a supremum, $b \leqslant a$. Therefore, $a = b$. The proof for the uniqueness of the infimum is analogous. $\square$

The same holds for the maximum and the minimum of $A$. If $A$ has no upper bound, $\sup A = \infty$. Similarly, if $A$ has no lower bound, $\inf A = -\infty$. We also use $\sup \emptyset = -\infty$ and $\inf \emptyset = \infty$.

**Theorem 3.1.** *Let $A \subseteq R$. If $A$ has an upper (lower) bound, $\sup A$ ($\inf A$) exists in $\mathbb{R}$.*

*Proof.* Let $\overline{m} \in \mathbb{R}$ be an upper bound of $A$. Then, $A$ is non-empty and $\sup A \leqslant \overline{m} \implies \sup A \in \mathbb{R}$. The same logic applies for the inf. $\square$

## 3.2 Intervals

Fix any two real numbers $a < b$. The **open interval** $(a, b)$ is defined as

$$(a, b) \coloneqq \{x \in \mathbb{R} : a < x < b\}.$$

Correspondingly, the semi-open intervals are defined by

$$[a, b) \coloneqq (a, b) \cup \{a\} \text{ and } (a, b)] \coloneqq (a, b) \cup \{b\}.$$

Finally, the **closed interval** $[a, b]$ is defined by

$$[a, b] \coloneqq \{x \in \mathbb{R} : a \leqslant x \leqslant b\}.$$

The unbounded intervals are

$$\begin{aligned}&(a, \infty) \coloneqq \{x \in \mathbb{R} : x > a\}, \text{ and } (-\infty, b) \coloneqq \{x \in \ x < b\};\\ &[a, \infty) \coloneqq (a, \infty) \cup \{a\} \text{ and } (-\infty, b] \coloneqq (-\infty, b) \cup \{b\}.\end{aligned}$$

The set of **extended real numbers** is given by

$$\overline{\mathbb{R}} \coloneqq \mathbb{R} \cup \{-\infty, \infty\}.$$

# 4 Sequences of Real Numbers

A (real) sequence is a function

$$\begin{aligned} f :&\mathbb{N} \longrightarrow \mathbb{R} \\ &n \mapsto x_n. \end{aligned}$$

The set of all real sequences is represented by $\mathbb{R}^{\mathbb{N}}$ or $\mathbb{R}^{\infty}$.

A sequence $\{x_n\}$ is said to be (**strictly**) **increasing** if $x_n$ $(<) \leqslant x_{n+1}$ for all $n \geqslant 1$. Similarly, it is said to be (**strictly**) **decreasing** if $x_n$ $(>) \geqslant x_{n+1}$ for all $n \geqslant 1$. The sequence is **monotone** if it is either increasing or decreasing. If $\{x_n\}$ is an increasing (decreasing) sequence, we denote it by $\{x_n\} \nearrow$ ($\{x_n\} \searrow$).

## 4.1 Convergence

We usually care about the behavior of the sequence as it grows arbitrarily large, or $n \to \infty$. Many things can happen: the sequence may become constant, or may settle around a particular value, or may explode (go to infinity), or it may oscillate forever, etc ...

We say that a sequence $\{x_n\} \in \mathbb{R}^{\mathbb{N}}$ (so we are talking about a sequence over real numbers) converges to some $x \in \mathbb{R}$ if,

$$\forall \varepsilon > 0,\ \exists N \in \mathbb{N} \text{ s.t. } n \geqslant \mathbb{N} \implies |x_n - x| < \varepsilon.$$

The number $x$ is called the **limit** of the sequence. We denote that a sequence converges to $x$ by

$$\lim_{n\to\infty} x_n = x,\ \lim x_n = x,\ \text{or } x_n \to x.$$

We say that a sequence $\{x_n\}$ is **convergent** if there exists $x \in \mathbb{R}$ such that $x_n \to x$. If $x_n \to x$ and $\{x_n\}$ is increasing (decreasing), we write $x_n \nearrow x$ ($x_n \searrow x$).

<u>Example:</u>

Take the sequence $(\frac{1}{2}, \frac{3}{4}, \frac{7}{8}, \frac{15}{16}, \dots)$, which can be thought of as $x_n = \frac{2^n-1}{2^n}$. It converges to 1.

**Theorem 4.1.** *If $\{x_n\}$ converges to $x$, then any subsequence of $\{x_n\}$ also converges to $x$.*

We say that a sequence **diverges** to $\infty$ when

$$\forall x \in \mathbb{R},\ \exists N \in \mathbb{N} \text{ s.t. } \forall n \geqslant N,\ x_n \geqslant x,$$

and we say that a sequence **diverges** to $-\infty$ when

$$\forall x \in \mathbb{R},\ \exists N \in \mathbb{N} \text{ s.t. } \forall n \geqslant N,\ x_n \leqslant x.$$

A sequence that diverges is called **divergent**. When $\{x_n\}$ diverges to $\infty$ we write $\lim x_n = \infty$, and when it diverges to $-\infty$ we write $\lim x_n = -\infty$.

Note that some sequences do not have a defined limit in the extended real numbers $\overline{\mathbb{R}}$, for instance, the sequence given by $x_n = (-1)^n$.

**Lemma 4.1.** ***(Monotonicity)*** *For any two real sequences $\{x_n\}$ and $\{y_n\}$ that converge in $\overline{\mathbb{R}}$, if $\exists N \in \mathbb{N}$ such that $\forall n \geqslant N$, $x_n \leqslant y_n$, then $\lim x_n \leqslant \lim y_n$.*

**Lemma 4.2.** ***(Sandwich)*** *Let $\{x_n\}, \{y_n\}, \{z_n\}$ be three real sequences. If $x_n \leqslant y_n \leqslant z_n$ for all $n$ large enough and $\lim x_n = \lim z_n = l$, then $\lim y_n = l$.*

A real sequence $\{x_n\}$ is said to be **bounded from above** if the set $\{x_1, x_2, \dots\}$ is bounded from above, or equivalently

$$\sup_n x_n := \sup\{x_n : n \in \mathbb{N}\} < \infty.$$

Similarly, the sequence is said to be **bounded from below** if the set $\{x_1, x_2, \dots\}$ is bounded from below, or equivalently

$$\inf_n x_n := \inf\{x_n : n \in \mathbb{N}\} > -\infty.$$

A real sequence is **bounded** if it is bounded from above and from below.

**Proposition 4.1.** *Every real sequence that converges in $\mathbb{R}$ is bounded.*

We also have the following results.

**Proposition 4.2.** *Every increasing (decreasing) real sequence that is bounded from above (below) converges.*

*Proof.* Let $\{x_n\}$ be an increasing sequence bounded from above. Without loss, assume that it is increasing. Since $\{x_n\}$ is bounded, $\sup x_n < \infty$, so let $x := \sup x_n$. We claim that $x_n \to x$. Indeed, fix $\varepsilon > 0$, and we get that $x - \varepsilon < \sup x_n$. Therefore, there exists $N \in \mathbb{N}$ such that

$x_N > x - \varepsilon$. Since the sequence is increasing, $x_n \geqslant x_N > x - \varepsilon$ for all $n \geqslant N$. But since $x$ is the supremum, $x_n \leqslant x < x + \varepsilon$, and so $|x_n - x| < \varepsilon$ for all $n \geqslant N$. $\square$

**Theorem 4.2.** ***(Monotone Sequence Theorem)*** *Every real sequence has a monotone subsequence.*

*Proof.* Let $\{x_n\}$ be a real sequence. Call $x_n$ a peak if $x_n \geqslant x_m$ for all $m \geqslant n$. We then have two cases.

Case 1: the sequence has infinitely many peaks. For all $k \in \mathbb{N}$, let $x_{n_k}$ be defined as the $k - th$ peak.

Case 2: There are finitely many peaks. Denote by $\tilde{n} \in \mathbb{N}$ the position of the last peak. Take $x_{n_1} > x_{\tilde{n}}$. Since $x_{n_1}$ is not a peak, there exists $n_2 > n_1$ such that $x_{n_2} > x_{n_1}$. Similarly, since $x_{n_2}$ is not a peak, there exists $n_3 > n_2$ such that $x_{n_3} > x_{n_2}$. Proceed by induction: since we have no more peaks, for all $k \in \{3, 4, 5, \dots\}$, there exists $l > k$ such that $x_{n_l} > x_{n_k}$. The resulting sequence is increasing. $\square$

**Theorem 4.3.** ***(Bolzano-Weierstrass)*** *Every bounded real sequence has a convergent subsequence.*

*Proof.* Let $\{x_n\}$ be a bounded real sequence. From the Monotone Sequence theorem, we know that it admits a monotone subsequence. Since this monotone subsequence must also be bounded, by Proposition 4.2 it must converge. $\square$

Example:

The sequence given by $x_n = (-1)^n$ is bounded, but does not converge. However, we can take the subsequence $x_{n_k} = x_{2k}$ which gives us the constant sequence $(1, 1, 1, \dots)$ which clearly converges to 1.

The following result is relevant later on.

**Lemma 4.3.** *If $S$ is a non-empty subset in $\mathbb{R}$, then*

*i) there exists an increasing sequence $\{x_n\} \in S$ such that $x_n \nearrow \sup S$,*

*ii) there exists a decreasing sequence $\{x_n\} \in S$ such that $x_n \searrow \inf S$,*

*Proof.* The proof only covers the first case when $\sup S$ is a real number. The other cases follow similar arguments. Let $s = \sup S$. If $s$ is actually a maximum, i.e., $s \in S$, then we take $x_n = s$ for every $n$ and we are done. Now, suppose $s$ is not a maximum of $S$. Then, the real number $s - 1$ is not an upper bound of $S$. Therefore, there exists $x_1 \in S$ such that $x_1 \geqslant s - 1$. Since $s$ is not a maximum, $x_1 < s$, and we can define $\varepsilon_1 = (s - x_1)/2 < 1/2$. In particular, $s - \varepsilon_1$ is not an upper bound, and there exists $x_2 \in S$ such that $x_2 \geqslant s - \varepsilon_1$. By construction, $x_2 > x_1$ and $s - x_2 \leqslant \varepsilon_1$. Since $s$ is not a maximum, $x_2 < s$, and we can define $\varepsilon_2 = (s - x_2)/2 < 1/4$. In particular, $s - \varepsilon_2$ is not an upper bound of $S$, and it follows that there exists $x_3 \in S$ such that $x_3 \geqslant s - \varepsilon_2$, with $x_3 > x_2$ and $s - x_3 \leqslant \varepsilon_2$. By induction, we can construct a sequence $\{x_n\}$ in $S$ such that $x_n < x_{n+1}$ and $s - x_{n+1} = \varepsilon_n \leqslant 1/2^n$. This implies that $x_n$ is increasing and converges to $s$. □

## 4.2 Cauchy Sequences

A real sequence $\{x_n\}$ is **Cauchy** if

$$\forall \varepsilon > 0,\ \exists N \in \mathbb{N} \text{ s.t. } \forall n, m \geqslant N,\ |x_m - x_n| < \varepsilon.$$

Any convergent sequence is Cauchy. For real sequences, the converse is also true: every real Cauchy sequence converges.[4]

## 4.3 Series

Series are sums of real numbers, where we usually index each term in the sum by a natural number. Formally, $\{x_n\}$ be a sequence of real numbers. For any integer $1 < k < n$, we define

$$\sum_{i=1}^{n} x_i := x_1 + \cdots + x_n \text{ and } \sum_{i=k}^{n} x_i := x_k + \cdots + x_n.$$

For any $S \subseteq \mathbb{N}$, we write $\sum_{i \in S} x_i$ to denote the sum of the terms of $\{x_n\}$ such that their index belongs to $S$. For instance, if $S = \{3, 43, 117\}$,

$$\sum_{i \in S} x_i = x_3 + x_{43} + x_{117}.$$

An **infinite series** is a sequence of partial sums $\{\sum_{i=1}^{n} x_i\}$ of some real sequence $\{x_n\}$. When the limit of the infinite series exists, we say

$$\sum_{i=1}^{\infty} x_i := \lim_{n \to \infty} \sum_{i=1}^{n} x_i.$$

We say that the infinite series is **convergent** if it attains a limit in $\mathbb{R}$. It is possible for an infinite series to not have a limit. Take, for instance, $\sum_{i=1}^{\infty} (-1)^i$.

[4]This is not always true for sequences over arbitrary sets. Whenever we are in a space in which every Cauchy sequence converges, we say that we are in a complete metric space. It turns out that $\mathbb{R}^n$ is complete with the usual Euclidean distance.

Infinite series appear often in economic models. For instance, the total utility of an infinitely lived agent is the sum of the utility the agent experiences in each period.

**Lemma 4.4.** *If an infinite series* $\sum_{i=1}^{n} x_i$ *converges, then* $\lim_{i \to 0} x_i = 0$.

The converse does not necessarily hold. A good example is the Harmonic series ($\sum_{i=1}^{n} \frac{1}{i}$), in which we have

$$\lim_{i \to \infty} \frac{1}{i} = 0, \text{ but}$$
$$\sum_{i=1}^{\infty} \frac{1}{i} = \infty.$$

**Lemma 4.5.** *If an infinite series* $\sum_{i=1}^{n} x_i$ *converges, then for every* $k \in \mathbb{N}$ *the sequence* $\sum_{i=k}^{\infty} x_i$ *also converges and* $\lim_{k \to n} \sum_{i=k}^{\infty} x_i = 0$.

Example:

1) The infinite series

$$\sum_{i=1}^{\infty} \frac{1}{i^{\alpha}}$$

converges if, and only if, $\alpha > 1$.

2) The infinite geometric series is given by

$$\sum_{i=1}^{\infty} \delta^{i-1} a,$$

for $a \in \mathbb{R}$. If $|\delta| \geqslant 1$, the series does not converge. When $\delta < 1$ it converges to

$$\frac{a}{1-\delta}.$$

# 5 Real Functions

A **real function** (or real-valued function) on a non-empty set $X$ is an element of $\mathbb{R}^X$, or simply

$$f : X \longrightarrow \mathbb{R}.$$

Given two real functions $f, g \in \mathbb{R}^X$, we endow $\mathbb{R}^X$ with the partial order "$\geqslant$", defined by

$$f \geqslant g \Longleftrightarrow f(x) \geqslant g(x) \ \forall x \in X.$$

We then say,

1. $f > g \Longleftrightarrow [f \geqslant g \wedge \neg(g \geqslant f)]$;
2. $f \gg g \Longleftrightarrow f(x) > g(x) \ \forall x \in X$.

Now, consider a relation $\succeq$ on $X$ such that $(X, \succeq)$ is a poset. A real function $f$ is **increasing** if, for any $x, y \in X$,

$$x \succeq y \implies f(x) \geqslant f(y).$$

Moreover, $f$ is **strictly increasing** if, for any $x, y \in X$,

$$x \succ y \implies f(x) > f(y).$$

Similarly, we say that $f$ is (**strictly**) **decreasing** if $-f$ is (strictly) increasing. Finally, $f$ is monotonic if it is either increasing or decreasing.

We call $f$ a **bounded** function if there exists $M \in \mathbb{R}$ such that

$$|f(x)| \leqslant M, \ \forall x \in X.$$

The following results are of interest:

i) Given $a, b \in \mathbb{R}$, with $-\infty < a \leqslant b < \infty$, every monotonic funtion $f : [a, b] \longrightarrow \mathbb{R}$ is bounded.

ii) If $X \subseteq \mathbb{R}$, then any function $f : X \longrightarrow \mathbb{R}$ that is strictly increasing is injective.

## 5.1 Limits of Real Functions

For this section, fix $X \subseteq \mathbb{R}$, $X$ non-empty. Now, consider a function $f : X \longrightarrow \mathbb{R}$. We say that $y \in \overline{\mathbb{R}}$ is the **right-limit** of $f$ at $x$ if

$$\lim_{n\to\infty} f(x_n) = y$$

for all sequences $\{x_n\}$ in $X \setminus \{x\}$, with $x_n \searrow x$. We can also denote the right-limit $y$ as $f(x^+)$, or $\lim_{t\to x^+} f(t)$. Similarly, we say that $y \in \overline{\mathbb{R}}$ is the **left-limit** of $f$ at $x$ if

$$\lim_{n\to\infty} f(x_n) = y$$

for all sequences $\{x_n\}$ in $X \setminus \{x\}$, with $x_n \nearrow x$. We can also denote the left-limit $y$ as $f(x^-)$, or $\lim_{t\to x^-} f(t)$.

NOTE: The definition of right or left side limit of a function does not require the function itself to be defined at the limit. For instance, take $f : \mathbb{R}_{++} \longrightarrow \mathbb{R}_+$ defined by $f(x) = 1/x$. Then, $f(0^+) = \infty$, even though $f(x)$ is itself undefined at $x = 0$.

Finally, we say that $y \in \overline{\mathbb{R}}$ is the **limit** of $f$ at $x$ if

$$\lim_{n\to\infty} f(x_n) = y$$

for every sequence $\{x_n\}$ in $X \setminus \{x\}$ with $x_n \to x$.

If a limit $y$ of $f$ at $x$ exists, we write

$$\lim_{t\to x} f(t) = y.$$

NOTE: For $x \in X$, it is possible that $\lim_{t\to x} f(t) \neq f(x)$.

Example: Take

$$f(x) = \begin{cases} 0, & \text{if } x \neq 0, \\ 1, & \text{if } x = 0. \end{cases}$$

Then $\lim_{t\to x} f(t) = 0$ but $f(0) = 1$.

**Proposition 5.1.** *If $x, y \in \mathbb{R}$. We have $\lim_{t\to x} f(t) = y$ if, and only if,*

$$\forall \varepsilon > 0,\ \exists \delta > 0 \ s.t.\ \forall x \in X,\ |t - x| < \delta \implies |f(t) - y| < \varepsilon.$$

## 5.2 Continuity

For any $x \in X$, we say that $f$ is **continuous at** $x$ if either

i) $x$ is an isolated point of X, i.e, there is no sequence $\{x_n\}$ in $X \setminus \{x\}$ with $x_n \to x$;

ii) **or** there exists $\{x_n\}$ in $X \setminus \{x\}$ with $x_n \to x$ and $\lim_{t\to x} f(t) = f(x)$ (the limit exists).

For any $S \subseteq X$, we say that $f$ is **continuous on** $S$ if $f$ is continuous at all $x \in S$. Furthermore, if $S = X$, we simply say that $f$ is **continuous**.

We also have the $\varepsilon - \delta$ **definition of continuity**. A function $f : X \longrightarrow \mathbb{R}$ is continuous at $x' \in X$ if, and only if,

$$\forall \varepsilon > 0,\ \exists \delta > 0 \text{ s.t. } \|x - x'\| < \delta \implies |f(x) - f(x')| < \varepsilon.$$

We denote by $C(X)$ the set of all continuous functions on $X$. In economics, we sometimes employ stronger definitions of continuity.

**Definition 5.1.** *A function $f \in \mathbb{R}^T$ is **uniformly continuous** on $S \subseteq T$ if*

$$\forall \varepsilon > 0, \exists \delta > 0,\ \textit{s.t.}\ \forall x, t \in S, |x - t| < \delta \implies |f(x) - f(t)| < \varepsilon.$$

*If $S = T$, we say that $f$ is **uniformly continuous**.*

We are now ready to present one of the more important results for this course.

**Theorem 5.1.** ***(Weierstrass or Extreme Value Theorem)*** *For all $f \in C[a, b]$, there exists $x, x' \in [a, b]$ such that*

$$f(x') \leqslant f(t) \leqslant f(x)\ \forall x \in [a, b].$$

The proof of EVT relies on Lemma 4.3. We then get the following conclusion.

**Corollary 5.1.** *The function $f$ attains its maximum and minimum in $[a, b]$.*

NOTE: This result is later expanded by the Maximum Theorem.

**Theorem 5.2.** ***(Intermediate Value Theorem - IVT)*** *If $f \in C[a, b]$, then $f([a, b])$ is an interval, i.e.,*

$$\forall x, x' \in [a, b] \textit{ and } \forall y \in \mathbb{R},\ f(x') < y < f(x') \implies \exists x'' \in [a, b] \textit{ with } f(x'') = y.$$

<u>Example:</u>

To show that $6x - e^x = 0$ has a solution on $[0, 1]$, let $f(x) = 6x - e^x$. We then have that $f(0) = -1$, $f(1) = 6 - e > 0$. By the IVT, there must exists $x \in [0, 1]$ such that $f(x) = 0$.

## 5.3 Multivariable Functions

An element $x \in \mathbb{R}^n$ is the vector $x = (x_1, \dots, x_n)'$., where $'$ denotes the transpose ($x$ is a column vector).[5] Now we work with functions $f : \mathbb{R}^n \to \mathbb{R}^m$, where $n, m \geqslant 1$. We then get

$$f(x) = \begin{pmatrix} f_1(x) \\ \vdots \\ f_n(x) \end{pmatrix}$$

so that $f(x)$ is an $m \times 1$ vector. Multivariable functions can appear in two different ways:

1. Domain: $f : \mathbb{R}^n \to \mathbb{R}$ (more common);
2. Codomain: $f : \mathbb{R} \to \mathbb{R}^m$ (more complicated).

Of course, both generalizations can happen at the same time as well. This family of functions are common in economics. Think of an equilibrium parameter $y^* = (y_1^*, \dots, y_m^*)$ that is a function of exogenous parameters in the model.

Example:

A classic example is a utility function over a bundle of goods:

$$\begin{aligned} u : \mathbb{R}^n &\longrightarrow \mathbb{R}, \\ (x_1, \dots, x_n) &\mapsto u(x_1, \dots, x_n). \end{aligned}$$

[5] Some authors represent vectors in a bold font, $\mathbf{x} \in \mathbb{R}^n$.

# 6 Calculus

In this section, we will look into formal definitions for the concepts of derivative and integral of a function.

## 6.1 Differentiation

Let $I$ be a non-degenerate interval, and fix $f : I \to \mathbb{R}$. The **growth rate map** $G_{f,x} : I \setminus \{x\} \to \mathbb{R}$ as

$$G_{f,x}(t) \coloneqq \frac{f(t) - f(x)}{t - x}.$$

We are interested in what happens to this growth rate as $t \to x$. We say that $f$ is **right-differentiable at x** if $G_{f,x}(x^+)$ exists in $\mathbb{R}$. In this case, we denote the **right-derivative** of $f$ at $x$ by $f'_+(x)$ and define it by,

$$f'_+(x) \coloneqq G_{f,x}(x^+) = \lim_{t \to x^+} G_{f,x}(x).$$

Similarly, we say that $f$ is **left-differentiable at x** if $G_{f,x}(x^-)$ exists in $\mathbb{R}$. The **left-derivative** is defined by

$$f'_-(x) \coloneqq G_{f,x}(x^-) = \lim_{t \to x^-} G_{f,x}(x).$$

**Definition 6.1.** *Given the interval $I$, we say that $f$ is* ***differentiable at*** *$x$ if one of the following properties is satisfied:*

*i) $x$ is the left end point of $I$ and $f'_+(x)$ exists;*

*ii) $x$ is the right end point of $I$ and $f'_-(x)$ exists;*

*iii) $x$ is not an end point of $I$ and $f$ is right and left-differentiable at $x$, with $f'_-(x) = f'_+(x)$.*

The idea is that the function will not be differentiable at a point $x$ if the right and left derivatives are different. This would lead to confusion as to what value should be considered as the derivative at the point $x$. When one of these 3 properties is satisfied, we denote the derivative of $f$ at $x$ by $f'(x)$, and its value equals the limit of the growth rate map at $x$.

Given Definition 6.1, we can equivalently say that $f$ is **differentiable** at $x$ if, and only if,

$$\lim_{t \to x} \frac{f(t) - f(x)}{t - x}$$

exists in $\mathbb{R}$. This limit is the **derivative** of $f$ at $x$ and it is denoted by $f'(x)$. We can also use Leibniz's notation, which keeps in mind the idea that we are taking a ratio of differences,

$$\frac{df(x)}{dx} = f'(x).$$

Let $J \subseteq I$. If $f$ is differentiable at $x$ for all $x \in J$, we say that $f$ is **differentiable on** J. If $J = I$, we simply say that $f$ is **differentiable**. When $f$ is differentiable, the **derivative of** $f$ is the function

$$\begin{aligned} f' :& I \to \mathbb{R} \\ & x \mapsto f'(x). \end{aligned}$$

Similarly, if $f'$ is in itself differentiable, we say that $f$ is **twice differentiable**, and the second derivative of $f$ is denoted by $f''$. This can be extended to any higher-order derivatives. For the $n - th$ order derivative, we denote it by $f^{(n)}$, as long as it is well defined. In addition, if $f'$ exists and it is continuous, we say that $f$ is **continuously differentiable**. If $f''$ exist and it is continuous, we say that $f$ is **twice continuously differentiable**, and so on.

**Definition 6.2.** *We say that $x$ is a **critical point** of $f$ if $f$ is not differentiable at $x$, or if $f'(x) = 0$.*

As seen in Calculus classes, we can interpret the derivative at $x$ as the slope, or tangent line, at that point. For a given function evaluated at two points $x_1, x_2$, we consider the growth rate as the slope of the line that passes through the points $(x_1), f(x_1))$ and $(x_2, f(x_2))$. This slope is given by the formula

$$\frac{f(x_2) - f(x_1)}{x_2 - x_1}.$$

Let $h = x_2 - x_1$, and we can rewrite the slope as

$$\frac{f(x_1 + h) - f(x_1)}{x_1 + h - x_1} = \frac{f(x_1 + h) - f(x_1)}{h}.$$

Note then that the derivative is the value of the above formula as $h \to 0$, given the interpretation of the derivative as the tangent line.

Below, we list a few useful differentiation rules. They can be obtained by applying the definition of the derivative to each case.

1. $(f + g)' = f' + g'$;
2. $(fg)' = f'g + fg'$;
3. $(f/g)' = (f'g - fg')/g^2$;

4. (Chain Rule) $\frac{d}{dx}f(g(x)) = f'(g(x))g'(x)$.

Common derivatives:

1. $\frac{d}{dx}x^n = nx^{(n-1)}$ for any $n \in \mathbb{Q}$;
2. $\frac{d}{dx}a^x = \ln(a)a^x$;
3. $\frac{d}{dx}ln(x) = 1/x$;
4. $\frac{d}{dx}sin(x) = cos(x)$;
5. $\frac{d}{dx}cosx) = -sin(x)$.

Example:
$f(x) = g_1(x)g_2(x)\ldots g_n(x)$. To find $f'(x)$, we go

$$ln(f(x)) = \ln(g_1(x)) + ln(g_2(x)) + \ldots ln(g_n(x)) = \sum_{i=1}^{n} ln(g_i(x)).$$

Taking the derivative on both sides, $\frac{f'(x)}{f(x)} = \frac{g_1'(x)}{g(x)} + \frac{g_2'(x)}{g(x)} + \cdots + \frac{g_n'(x)}{g(x)}$.

We also have the following result.

**Proposition 6.1.** *If $f$ is differentiable at $x$, then $f$ is continuous at $x$.*

Given the interpretation as a tangent line, the derivative has a lot of implications for solving maximization problems.

**Definition 6.3.** *For any $f : X \to \mathbb{R}$, an element $c$ in the interior of $X$ is said to be a* ***local extremum*** *of $f$ if there exists $\varepsilon > 0$ such that the sign of $f(x) - f(c)$ is constant for any $x$ in the interval $(c - \varepsilon, c + \varepsilon)$.*

The following is one of the most useful results for economics.

**Proposition 6.2.** ***(First Order Condition.)*** *If $c$ is a local extremum of $f$ and $f$ is differentiable at $c$, then $f'(c) = 0$.*

**Theorem 6.1.** ***(Rolle)*** *Let $[a, b]$ be a non-degenerate interval, and fix $f \in C^1[a, b]$ (that is $f$ is differentiable on (a,b)). Then,*

$$f(a) = f(b) \implies \exists c \in (a, b) \ s.t. \ f'(c) = 0.$$

**Corollary 6.1.** ***(Mean value Theorem - MVT)*** *Let $[a, b]$ be a nondegenerate interval, and fix $f \in C[a, b]$ that is differentiable on (a,b). There exists $c \in (a, b)$ such that*

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

From the previous results, the following properties can be obtained:

1. if $f' = 0$ (meaning $f'(x) = 0$ for every $x$ in the domain), then $f$ is a constant function;
2. if $f' \geqslant (\leqslant)\ 0$, then $f$ is increasing (decreasing);
3. if $f' > (<)\ 0$, then $f$ is strictly increasing (decreasing).

**Theorem 6.2.** ***(Inverse Function Theorem)*** *For $X \subseteq \mathbb{R}$, let $f : X \to \mathbb{R}$ be continuously differentiable. Then, for every $x \in X$ where $f'(x) \neq 0$, there exists $a_x$, $b_x \in \mathbb{R}$ such that $x \in (a_x, b_x)$ and $f(x)$ is invertible on $(a_x, b_x)$ with inverse $g(y)$ and*

$$g'(y) = \frac{1}{f'(y)}.$$

The above can be generalized to higher dimensions.

Example:

Take $f : \mathbb{R}_{++} \to \mathbb{R}$, where $f(x) = \sqrt{x}$. Then, $f'(x) = \frac{1}{2\sqrt{x}}$. Since $f'(x) \neq 0$ for any $x \in \mathbb{R}_{++}$, we can deduce the existence of $g(y) = f^{-1}(y) = y^2$. Note that $g'(y) = 2y$.

**Theorem 6.3.** ***(L'Hôpital's Rule)*** *Let $f(x), g(x)$ be a differentiable on the interval $(a, b)$, with $g'(x) \neq 0$. If, for some $z \in (a, b)$, when $x \to z$ either $f(x) \to 0$ and $g(x) \to 0$, or $f(x) \to \pm\infty$ and $g(x) \to \pm\infty$, then,*

$$\lim_{x \to z} \frac{f(x)}{g(x)} = \lim_{x \to z} \frac{f'(x)}{g'(x)}.$$

### 6.1.1 Taylor Approximation

The following is a very useful result that allows us to approximate functions by polynomials, which are much easier to work with than many of the functions that one may find in economics.

**Theorem 6.4.** ***(Taylor's Theorem)*** *Fix $n \in \mathbb{N}$, $a \in \mathbb{R}$, and let $f : \mathbb{R} \in \mathbb{R}$ be continuously differentiable at least $n$ times, and differentiable at least $n+1$ times. Then, for all $x \in \mathbb{R}$, there exists $c < |x - a|$ such that*

$$f(x) = f(a) + f'(a)(x-a) + \frac{f^{(2)}(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n + \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1}$$

The last term is like a remainder. Without it, the function is an $n$-th order Taylor Approximation of $f(x)$ around $a$, as the two sides are not equal anymore. Note that as $n$ increases, $(n+1)!$ grows very quickly, so the last term tends to vanish as well, and the better the approximation tends to become. In addition, the closer $x$ and $a$ are, the better the approximation tends to be as well. Note also that, for $n = 1$, we have the Mean Value Theorem.

Example:

One important application of the Taylor approximation is in log-linearization. Let $f : (a, b) \to \mathbb{R}$ be defined as $f(x) = \ln(x)$. Now, fix $x^* \in (a, b)$. A first-order approximation of $f(x)$ around $x^*$ is given by:

$$\begin{aligned} f(x) &\approx f(x^*) + f'(x^*)(x - x^*) \\ \ln(x) &\approx \ln(x^*) + \frac{x - x^*}{x^*} \\ \implies \ln(x) - \ln(x^*) &\approx \frac{x - x^*}{x^*} \text{ (\% variation)} \end{aligned}$$

We can interpret $x^*$ as the steady state of the system, and $x$ is a small perturbation around this steady state.

## 6.2 Multivariable Calculus

We begin with a generalization of the differentiation that we saw before to functions defined over higher dimensions of the Euclidean space.

**Definition 6.4.** *Let $X \subseteq \mathbb{R}^n$. The partial derivative of $f : X \to \mathbb{R}$ at $x \in X$ with respect to $x_i$ is*

$$\frac{\partial f(x)}{\partial x_i} = \lim_{h \to 0} \frac{f(x_1, \ldots, x_i + h, \ldots, x_n) - f(x)}{h}.$$

The partial derivative sees the rate of change of the function when one of the variables changes, keeping all others constant.

Example:

Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by $f(x, y) = x^2 y$. Then,

$$\frac{\partial f}{\partial x} = 2xy, \text{ and } \frac{\partial f}{\partial y} = x^2.$$

A differentiable[6] $f : \mathbb{R}^n \to \mathbb{R}$ has $n$ partial derivatives. We then have the following definition.

[6]We direct the interested reader to other sources for a more formal definition of what it means for a function $f : \mathbb{R}^n \to \mathbb{R}^m$ to be differentiable. For instance, see Rudin (2021).

**Definition 6.5.** *The* ***gradient*** *of a differentiable $f : \mathbb{R}^n \to \mathbb{R}$, denoted by $\nabla f$, is the colunm vector*

$$\nabla f := \left(\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n}\right)'.$$

The gradient of a function at a point shows the direction of maximum change of the function.

**Definition 6.6.** *The* ***total differential*** *of $f : \mathbb{R}^n \to \mathbb{R}$ at a point $x \in \mathbb{R}^n$ is*

$$df \equiv \frac{\partial f}{\partial x_1}dx_1 + \frac{\partial f}{\partial x_2}dx_2 + \cdots + \frac{\partial f}{\partial x_n}dx_n = \nabla f' \cdot dx'.$$

We can interpret the total differentiation as the sum of the change of $f$ in each direction.

NOTE: We simplify here the definition of total differentiation.

Example:

Let the vector $x$ vary continuously on time, so that we have $x(t) = (x_1(t), \dots, x_n(t))'$. Then, by total differentiation,

$$\frac{df}{dt} = \frac{\partial f}{\partial x_1}\frac{dx_1}{dt} + \cdots + \frac{\partial f}{\partial x_n}\frac{dx_n}{dt}.$$

Example:

Let $Y(t) = K(t)^\alpha L(t)^{1-\alpha}$, with $K(t) = \ln t$ and $L(t) = \beta - \frac{1}{t}$. Then,

$$\dot{Y} := \frac{dY}{dt} = \frac{\partial Y}{\partial K}\frac{dK}{dt} + \frac{\partial Y}{\partial L}\frac{dL}{dt} \implies \dot{Y} = \frac{\alpha k^{\alpha-1}L^{1-\alpha}}{t} + \frac{(1-\alpha)K^\alpha L^{-\alpha}}{t^2}.$$

**Definition 6.7.** *For a twice-differentiable $f : \mathbb{R}^n \to \mathbb{R}$, the* ***Hessian*** *of $f$ is the $n \times n$ matrix given by*

$$Hf := \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n} & \frac{\partial^2 f}{\partial x_2 \partial x_n} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}.$$

**Theorem 6.5.** *Let all the second partial derivaties of $f$ be continuous. Then, $Hf$ is symmetric.*

We may also denote by $f_{x_i}$ de partial derivative of $f$ in relation to the $x_i$ variable.

Example:

Take $f(x, y) = x^2 y + y \sin x$. Then, we have:

$$f_x(x, y) = 2xy + y \cos x \implies f_{xy} = 2x + \cos x;$$
$$f_y(x, y) = x^2 + \sin x \implies f_{yx} = 2x + \cos x.$$

**Definition 6.8.** *The* ***Jacobian*** *of $f$ is the $m \times n$ matrix given by*

$$J_x(f) = \begin{pmatrix} \nabla f_1' \\ \vdots \\ \nabla f_n' \end{pmatrix} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_m} \end{pmatrix}.$$

### 6.2.1 Implicit functions

It is very common for a variable be interpreted as a function of another. For instance, when we write $y = f(x)$ for $f : \mathbb{R} \to \mathbb{R}$. We can reinterpret the relationship between $y$ and $x$ as

$$F(x, y) = f(x) - y.$$

Note then that we can think of points in $\mathbb{R}^2$ that satisfy $y = f(x)$ as the same points that satisfy $F(x, y) = 0$. This way of interpreting this relationship can be useful to analyze $f(x)$ by looking only at $F(x, y)$.

Example:

Let $U$ be a utility function over $\mathbb{R}^2$ given by $U(x_1, x_2) = x_1^{\alpha} x_2^{1-\alpha}$. For a given utility level $c$, we can find the locus of the bundles that give exactly $c$ utility by

$$c = x_1^{\alpha} x_2^{1-\alpha} \implies x_2 = c^{\frac{1}{1-\alpha}} x_1^{\frac{-\alpha}{1-\alpha}}.$$

Why is it helpful to think of $x_2(x_1)$ like this? One important application is that now we can then find the Marginal Rate of Substitution (MRS) by looking at

$$\frac{dx_2}{dx_1} = c^{\frac{1}{1-\alpha}} \frac{\alpha}{1-\alpha} x_1^{\frac{-1}{1-\alpha}} = \frac{\alpha}{\alpha - 1} \frac{x_2}{x_1}.$$

In general, we assume that $y = y(x)$ and we set $F(x, y(x)) = 0$. The total differentiation of $F$ would be

$$dF = F_x d_x + F_y d_y = 0 \implies \frac{dy}{dx} = -\frac{F_x}{F_y}.$$

Example:

$$\begin{aligned} U(x_1, x_2) &= x_1^{\alpha} x_2^{1-\alpha} \\ &\implies U_1 = \alpha x_1^{\alpha-1} x_2^{1-\alpha} \text{ and } U_2 = (1-\alpha) x_1^{\alpha} x_2^{1-\alpha}; \\ &\implies \frac{dx_2}{dx_1} = -\frac{U_1}{U_2}. \end{aligned}$$

However, given $F$, how can we be sure that a function $y = y(x)$ that satisfies $F(x, y(x)) = 0$ even exists?

**Definition 6.9.** ***(Implicit Function Theorem)*** *Let $F(x, y)$ be $C^1$ on an open ball centered at $(x^*, y^*)$, with $F(x^*, y^*) = 0$. Then, if $\frac{\partial F}{\partial y}(x^*, y^*) \neq 0$, $\exists y(x) \in C^1$ defined on an open interval $(a, b)$ such that $x^* \in (a, b)$ and*

*i)* $F(x, y(x)) = 0$ *for all* $x \in (a, b)$*;*

*ii)* $y(x^*) = y^*$*;*

*iii)*

$$\frac{dy(x^*)}{dx} = -\frac{\frac{\partial F(x^*, y^*)}{dx}}{\frac{\partial F(x^*, y^*)}{dy}}.$$

### 6.2.2 Homogeneous Functions

**Definition 6.10.** *Let $k \in \mathbb{Z}$. We say that $f : \mathbb{R}^n \to \mathbb{R}$ is homogeneous of degree $k$ if $f(tx) = t^k f(x)$ for all $t > 0$.*

Example:

1) Take a utility function $u(x_1, x_2) = x_1^{\alpha} x_2^{\beta}$. Then,

$$u(tx_1, tx_2) = (tx_1)^{\alpha} (tx_2)^{\beta} = t^{\alpha+\beta} x_1^{\alpha} x_2^{\beta} = t^{\alpha+\beta} u(x_1, x_2).$$

2) Take a demand function $x_i(p, \omega) = \omega / p_i$. Then, it is homogeneous of degree zero: $x_i(tp, t\omega) = t\omega / tp_i = \omega / p_i$.

Why is this concept important? Functions that are homogeneous of degree 1 exhibit constant returns to scale. We then get the following result.[7]

**Theorem 6.6.** ***(Euler)*** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable and homogeneous of degree $k$. Then, $\nabla f(x) \cdot x = k f(x)$.*

[7] To see a definition of the inner product, go to the section on Linear Spaces.

*Proof.* We have that $f(tx) = t^k f(x)$ for all $t > 0$. Differentiating with respect to $t$ yields $\nabla f(tx) \cdot x = kt^{k-1} f(x)$. Applying this relation to $t = 1$ yields the desired result. $\square$

**Definition 6.11.** *A function $f : \mathbb{R}^n \to \mathbb{R}$ is homothetic if $f(x) = g(h(x))$, where $h : \mathbb{R}^n \to \mathbb{R}$ is homogeneous of degree 1 and $g : \mathbb{R} \to \mathbb{R}$ is strictly increasing.*

Homotheticity is a crucial property, widely used in utility theory as it relates to ordinality.

## 6.3 Integration

We fix an interval $[a, b]$, with $a \leqslant b$, and we work with bounded functions in $\mathbb{R}^{[a,b]}$. Our goal for this section is to provide the foundation for a formal definition of an Integral operation.

**Definition 6.12.** *For any $n \in \mathbb{N}$, a* ***dissection*** *of an interval $[a, b]$ is a set of intervals*

$$[a_0, \ldots, a_n] \coloneqq \{[a_0, a_1], [a_1, a_2], \ldots, [a_{n-1}, a_n]\},$$

*where $a = a_0 < a_1 < a_2 < \cdots < a_{n-1} < a_n = b$.*

The class of all dissections on $[a, b]$ is denoted by $\mathcal{D}[a, b]$. We aim to define the creation of increasingly finer dissections of the interval. For any two dissections $\mathbf{d}^1 = [d_0^1, \ldots, d_n^1]$ and $\mathbf{d}^2 = [d_0^2, \ldots, d_n^2]$, we say that dissection $\mathbf{d}_1$ is finer than $\mathbf{d}_2$ if

$$\{d_0^2, \ldots, d_n^2\} \subseteq \{d_0^1, \ldots, d_n^1\}.$$

We can then define a relation $\succeq$ on $\mathcal{D}[a, b]$ where $\mathbf{d}^1 \succeq \mathbf{d}^2$ if $\mathbf{d}^1$ is finer than $\mathbf{d}^2$. We can create a new dissection, finer than both, using the join of two dissections, denoted by $\mathbf{d}^1 \vee \mathbf{d}^2$, so that

$$\mathbf{d}^1 \vee \mathbf{d}^2 = \{d_0^1, \ldots, d_n^1\} \cup \{d_0^2, \ldots, d_n^2\}.$$

We can now highlight the role of the join of two dissections in our analysis. Note that $\mathbf{d}^1 \vee \mathbf{d}^2 = \sup\{\mathbf{d}^1, \mathbf{d}^2\}$ in terms of the $\succeq$ relation we have just defined. Indeed, $\mathbf{d}^1 \vee \mathbf{d}^2$ is finer than both $\mathbf{d}^1$ and $\mathbf{d}^2$, and any other dissection finer than both $\mathbf{d}^1$ and $\mathbf{d}^2$ must also be finer than $\mathbf{d}^1 \vee \mathbf{d}^2$. Finally, note that $\mathbf{d}^1 \succeq \mathbf{d}^2$ if, and only if, $\mathbf{d}^1 \vee \mathbf{d}^2 = d^1$.

For a bounded $f \in \mathbb{R}^{[a,b]}$, we we fix a dissection $d \coloneqq [d_0, \ldots, d_n] \in \mathcal{D}[a, b]$.

**Definition 6.13.** *The $d$-**upper Riemann sum** of $f$ is the number*

$$\overline{R}_d(f) \coloneqq \sum_{i=1}^{n} [\sup f([d_{i-1}, d_i])][(d_i - d_{i-1})],$$

*and the* $d$*-**lower Riemann sum** of* $f$ *is the number*

$$\underline{R}_d(f) := \sum_{i=1}^{n}[\inf f([d_{i-1}, d_i])][(d_i - d_{i-1})].$$

The term $\sup f([d_{i-1}, d_i])$ represents the highest value that $f$ attains in the interval $[d_{i-1}, d_i]$. Similarly, $\inf f([d_{i-1}, d_i])$ represents the lowest value in the same interval. By definition,

$$\overline{R}_d(f) \geqslant \underline{R}_d(f).$$

Note that the map $\mathbf{d} \mapsto \overline{R}_d(f)$ is decreasing as

$$\mathbf{d}^1 \succeq \mathbf{d}^2 \implies \overline{R}_{d^1}(f) \leqslant \overline{R}_{d^2}(f).$$

Similarly, the map $\mathbf{d} \mapsto \underline{R}_d(f)$ is increasing as

$$\mathbf{d}^1 \succeq \mathbf{d}^2 \implies \underline{R}_{d^1}(f) \geqslant \underline{R}_{d^2}(f).$$

**Definition 6.14.** *The **upper Riemann integral** is the real number* $\overline{R}(f)$ *defined by*

$$\overline{R}(f) := \inf\{\overline{R}_d(f) : \boldsymbol{d} \in \mathcal{D}[a, b]\}.$$

*Similarly, the **lower Riemann integral** is the real number* $\underline{R}(f)$ *defined by*

$$\underline{R}(f) := \sup\{\underline{R}_d(f) : \boldsymbol{d} \in \mathcal{D}[a, b]\}.$$

Since $f$ is bounded, $\underline{R}(f)$ and $\overline{R}(f)$ are real numbers with $\underline{R}(f) \leqslant \overline{R}(f)$.

**Definition 6.15.** *Let* $f \in \mathbb{R}^{[a,b]}$ *be a bounded function. If* $\overline{R}(f) = \underline{R}(f)$*, then* $f$ *is said to be **Riemann integrable**, and the number*

$$\int_a^b f(t)dt := \overline{R}(f) = \underline{R}(f)$$

*is called the **Riemann integral** of* $f$*. When* $\overline{R}(f) > \underline{R}(f)$*, we say that* $f$ *is not Riemann integrable.*

For these notes, when we say that a function is integrable, we mean Riemann integrable.

Then, for an integrable $f$, we also define

$$\int_b^a f(t)dt \coloneqq -\overline{R}(f) = -\underline{R}(f).$$

When we write $\int_b^a f(t)dt$, the variable $t$ is considered a "dummy variable" as $\int_b^a f(t)dt = \int_b^a f(x)dx$, for any other variable $x$.

**Definition 6.16.** *Let $f \in \mathbb{R}^{[a,\infty]}$ be a bounded function such that $f$ is integrable at any interval $[a,b]$ for any $b > a$. If the limit*

$$\lim_{b\to\infty} \int_a^b f(t)dt$$

*exists in $\overline{\mathbb{R}}$, then $f$ admits an improper Riemann integral denoted*

$$\int_a^\infty f(t)dt \coloneqq \lim_{b\to\infty} \int_a^b f(t)dt.$$

We can similarly define an improper Rieman integral for $-\infty$.

**Proposition 6.3.** *Any continuous function $f \in C[a,b]$ is integrable.*

**Proposition 6.4.** *If $f, g \in \mathbb{R}^{[a,b]}$ are bounded and integrable, then*

$$f \geqslant g \implies \int_a^b f(t)dt \geqslant \int_a^b g(t)dt.$$

Common integral properties for any $f, g$ integrable on $[a,b]$:

i) $(\alpha f + g)$ is integrable for any $\alpha \in \mathbb{R}$, and

$$\int_a^b (\alpha f + g)(t)dt = \alpha \int_a^b f(t)dt + \int_a^b g(t)dt;$$

ii) for any $c \in [a,b]$,

$$\int_a^b f(t)dt = \int_a^c f(t)dt + \int_c^b f(t)dt;$$

**Theorem 6.7.** ***(Fundamental Theorem of Calculus)***

1. *Let* $f : [a, b] \to \mathbb{R}$ *be Riemann Integrable. Define*

$$F(x) = \int_a^x f(t)d(t).$$

*Then* $F(x)$ *is continuous on* $[a, b]$. *If* $f$ *is continuous at* $z \in (a, b)$, *then* $F$ *is differentiable at* $z$, *and* $F'(z) = f(z)$.

2. *Let* $F$ *and* $f$ *be real-valued functions defined on* $[a, b]$, $F$ *continuous on* $[a, b]$ *and* $F' = f$, $f$ *integrable on* $[a, b]$. *Then*

$$\int_a^b f(x)dx = F(b) - F(a).$$

Example:

$$\int_0^2 x^2 dx = \frac{1}{3}x^3 \mid_0^2 = \frac{8}{3}.$$

There are a few useful methods of integration that make the calculation easier.

**Proposition 6.5.** ***(Integration by Parts*** *If* $f, g \in C^1[a, b]$, *then*

$$\int_a^b f(t)g'(t)dt = f(b)g(b) - f(a)g(a) - \int_a^b f'(t)g(t)dt.$$

**Proposition 6.6.** ***(Integration by Substitution*** *Fix* $f \in C[a, b]$ *integrable and fix* $u \in C^1[a, b]$. *Then,*

$$\int_a^b f(u(t))u'(t)dt = \int_a^b f(u)du.$$

By the Fundamental Theorem of Calculus, if $F$ is the antiderivative of $f$, $\int_a^b f(u(t))u'(t)dt = F(u(b)) - F(u(a))$. The integration by substitution works as if it were an inverse of the chain rule.

Example:
Find the value of

$$\int_0^{\pi/2} e^{\cos(x)} \sin(x)dx.$$

Let $u(x) = \cos(x)$ and $f(x) = e^x \implies F(x) = e^x$. By substitution,

$$-\int_0^{\pi/2} f(u(x))u'(x)dx = -(F(u(\pi/2)) - F(u(0))) = -e^{\cos(\pi/2)} + e^{\cos(0)} = -e^0 + e^1 = e - 1.$$

**Theorem 6.8.** ***(Leibniz's Rule)*** *Let $a(x), b(x) \in C^1[x_0, x_1]$ be two differentiable functions. Let $f(x,t)$ be a function such that $f$ and its partial derivative $f_x$ are continuous for any $x \in [x_0, x_1]$ and $t \in [a(x), b(x)]$. Then, for any $x \in [x_0, x_1]$,*

$$\frac{d}{dx}\left(\int_{a(x)}^{b(x)} f(x,t)dt\right) = f(x, b(x))\frac{db(x)}{dx} - f(x, a(x))\frac{da(x)}{dx} + \int_{a(x)}^{b(x)} \frac{\partial}{\partial x} f(x,t)dt.$$

**Corollary 6.2.** *If $a(x) = a$ and $b(x) = b$ for any $x \in [x_0, x_1]$, then*

$$\frac{d}{dx}\left(\int_a^b f(x,t)dt\right) = \int_a^b \frac{\partial}{\partial x} f(x,t)dt.$$

Example:

Let $f(t)$ be a continuous income stream over the time interval $[0, T]$. We want to find out how the present value of this income stream varies with the instantaneous interest rate of $r$. We have

$$PV = \int_0^T e^{-rt} f(t)dt,$$

so that

$$\begin{aligned}\frac{dPV}{dr} = \frac{\partial}{\partial r}\left(e^{-rt}f(t)dt\right) &= \int_a^b \frac{\partial}{\partial r} e^{-rt} f(t)dt \\ &= \int_0^T -te^{-rt} f(t)dt.\end{aligned}$$

## 6.4 Multiple Integrals

Let $f : \mathbb{R}^n \to \mathbb{R}$. Then, the **multiple integral** of $f$ over the set $A = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]$ is given by

$$\int_a^b f(\mathbf{x})d\mathbf{x} = \int_{a_1}^{b_1} \int_{a_2}^{b_2} \cdots \int_{a_n}^{b_n} f(x_1, \ldots, x_n)dx_n \ldots dx_2 dx_n,$$

When $f$ is continuous, the order of integration does not matter.

# 7 Metric and Topological Spaces

We start by defining a non-empty set $X$. However, in most of our applications, we can think of $X$ as being the usual Euclidean space, or $\mathbb{R}^n$.

**Definition 7.1.** *A function $d : X \times X \to \mathbb{R}_+$ is a **metric**, or distance, on $X$ if, for any $x, y, z \in X$*

*i) $d(x, y) = 0 \iff x = y$;*

*ii) (Symmetry) $d(x, y) = d(y, x)$;*

*iii) (Triangle Inequality) $d(x, z) \leqslant d(x, y) + d(y, z)$.*

We then say that $(X, d)$ is a **metric space**.

Example:

(i) Euclidean Metric. For any $n \in \mathbb{N}$ and $x, y \in \mathbb{R}^n$, we have

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}.$$

(ii) Discrete Metric. For a set $X$ and $x, y \in X$,

$$d(x, y) = \begin{cases} 1 \text{ if } x \neq y; \\ 0 \text{ if } x = y. \end{cases}$$

In what follows, we will denote the Euclidean distance between $x, y$ by $\|x - y\|$.[8]

## 7.1 Topological Spaces

A topological space is with the minimal structure necessary to define limits and continuity. It can be very general and abstract, but here we will keep things simple and build our topology on $\mathbb{R}^n$, which can be done using the metrics we have already defined.

**Definition 7.2.** *Given $x \in \mathbb{R}^n$ and $\varepsilon > 0$, the **open ball** centered at $x$ with radius $\varepsilon$ is the set $B_\varepsilon(x) = \{z \in \mathbb{R}^n : d(x, z) < \varepsilon\}$.*

**Definition 7.3.** *For any $S \subseteq \mathbb{R}^n$, we say that*

*i) $x \in \mathbb{R}^n$ is an **interior point** of $S$ if $\exists \varepsilon > 0$ such that $B_\varepsilon(x) \subseteq S$;*

*ii) $x \in \mathbb{R}^n$ is a **boundary point** of $S$ if $\forall \varepsilon > 0$, $B_\varepsilon(x) \cap S \neq \emptyset$ and $B_\varepsilon(x) \cap S^c \neq \emptyset$;*

[8]As we will see, $\|\cdot\|$ is the Euclidean norm on $\mathbb{R}^n$, but it can be used to define the Euclidean distance.

*iii) $S$ is an **open set** if it only has interior points, i.e., $\forall x \in S$, $\exists \varepsilon > 0$ such that $B_\varepsilon(x) \subseteq S$;*

*iv) $S$ is a **closed set** if $S^c$ is open;*

**Definition 7.4.** *For any $S \subseteq \mathbb{R}^n$,*

*i) the **interior** of $S$ is the set $int(S) = \{x \in S : x \text{ is an interior point of } S\}$;*

*ii) the **boundary** of $S$ is the set $bd(S) = \{x \in \mathbb{R}^n : x \text{ is a boundary point of } S\}$;*

*iii) the **closure** of $S$ is the set $cl(S) = S \cup bd(S)$.*

Example:

In $\mathbb{R}$, the set $[0,1]$ is closed, the set $(0,1)$ is open, and the set $[0,1)$ is neither open nor closed.

**Proposition 7.1.** *A set $S \subseteq \mathbb{R}^n$ is open if, and only if, $S = int(S)$.*

**Definition 7.5.** *The **topology** associated with a metric $d$ on a set $X$ is the collection of all open subsets of $X$.*

Recall that a topology gives enough structure to a space that it allows one to define limits. We will now see that we can use sequences and limits to give an alternate definition of closed sets. The following result is more general than for the Euclidean space.

**Proposition 7.2.** *A set $S \in \mathbb{R}^n$ is closed if, and only if, every sequence in $S$ that converges in $\mathbb{R}^n$ has its limit point in $S$.*

*Proof.* $\Rightarrow$ Take the contrapositive. Assume that there exists a sequence $\{x_n\}$ in $S$ that converges to $x \notin S$. Since $x_n \to x$, $\forall \varepsilon > 0$, there exists $N_\varepsilon \in \mathbb{N}$ such that, for all $n > N_\varepsilon$, $\|x_n - x\| < \varepsilon$. Fix $\varepsilon > 0$ and the corresponding $N_\varepsilon$. Then, let $l > N_\varepsilon$, and we get that $x_l \in B_\varepsilon(x) \implies B_\varepsilon(x) \cap S \neq \emptyset$. Therefore, $B_\varepsilon(x) \nsubseteq S^c$. Since this is true for any $\varepsilon > 0$ and $x \in S^c$, $x$ is not an interior point of $S^c$, and thus $S^c$ is not open. Therefore, $S$ is not closed.

$\Leftarrow$. Assume, by way of contradiction, that every sequence in $S$ that converges, converges in $S$, and that $S$ is not closed. Then, $S^c$ is not open, and there exists $x \in S^c$ such that $x \in bd(S^c)$, and so for all $\varepsilon > 0$, $B_\varepsilon(x) \cap S \neq \emptyset$. Let $\varepsilon_n := \frac{1}{n}$, and define a sequence $\{x_n\}$ in $S$ such that $x_n \in B_{\varepsilon_n}(x) \cap S$. By construction, $x_n \to x \in S^c$. Contradiction. □

**Definition 7.6.** *Let $S \subseteq \mathbb{R}^n$. We say that $S$ is **bounded** if $\exists M > 0$ and $x \in S$ such that $S \subseteq B_M(x)$.*

An equivalent definition is if there exists $M > 0$ such that $\sup\{\|x - y\| : x, y \in S\} < M$.

**Definition 7.7.** *A set $S \in \mathbb{R}^n$ is **compact** if every sequence in $S$ has a subsequence that converges to a point in $S$.*

**Theorem 7.1.** ***(Heine-Borel)*** *A subset of* $\mathbb{R}^n$ *(with the Euclidean metric) is compact if, and only if, it is closed and bounded.*

*Proof.* $\Rightarrow$ Let $S \subseteq \mathbb{R}^n$ be compact.

<u>Closed</u>: take a sequence $\{x_n\}$ in $S$ that converges to $x$. We must show that $x \in S$. Since $S$ is compact, $\{x_n\}$ admits a subsequence $\{x_{n_k}\}$ that converges to a point in $x' \in S$. Since every subsequence of a convergent sequence must converge to the same point, $x' = x$ and $x \in S$.

<u>Bounded</u>: Assume, by way of contradiction, that $S$ compact and unbounded. Pick $x_1 \in S$. Then, since $S$ is unbounded, there exists $x_2 \in S$ such that $\|x_2 - x_1\| > 1$. Inductively, given $\{x_i\}_{i=1}^{k-1}$ such that $\|x_j - x_m\| > 1$ for all $j \neq m$, pick $x_k$ such that $\|x_k - x_i\| > 1$ for all $i \in \{1, 2, \ldots, k-1\}$. By construction, $\{x_n\}$ has no convergent subsequence.

$\Leftarrow$. Let $S \subseteq \mathbb{R}^n$ be closed and bounded. Fix a sequence $\{x_n\}$ in $S$. Since $S$ is bounded, $\{x_n\}$ is bounded. Therefore, it has a convergent subsequence that converges to some $x \in \mathbb{R}^n$. But since $S$ is closed, $x \in S$, and therefore $S$ is compact. $\square$

Outside of Euclidean space, the theorem may not hold.

<u>Example:</u>

i) Take the sequence in $\mathbb{R}$ defined by $x_n = n$. It has no convergent subsequence, and therefore $\mathbb{R}$ is not compact.
ii) Take $S = (0, 1)$ and the sequence with $x_n = \frac{1}{n+1}$, which is in $S$. Note that $x_n \to 0$, which is not in $S$, and so $S$ is not compact.

## 7.2 Continuity in Topology

We will now expand on the definition of continuity of functions to a topological space. We will repeat the definition we saw before, but add to it. Take any $X \subseteq \mathbb{R}^n$.

**Proposition 7.3.** *The following are equivalent definitions of continuity.*

*i) (Sequence) We say that* $f : X \to \mathbb{R}$ *is* ***continuous*** *if, for all* $x \in X$,
   *(a)* $x$ *is an isolated point of X;*
   *(b)* ***or*** *there exists* $\{x_n\}$ *in* $X \setminus \{x\}$ *with* $x_n \to x$ *and* $\lim_{t \to x} f(t) = f(x)$.

*ii)* $(\varepsilon{-}\delta)$ *A function* $f : X \to \mathbb{R}$ *is continuous at* $x' \in X$ *if* $\forall \varepsilon > 0$, $\exists \delta > 0$ *such that* $d(x, x') < \delta \implies |f(x) - f(x')| < \varepsilon$. *We say that* $f$ *is continuous if it is continuous at all* $x' \in X$.

*iii) (open sets) We say that* $f : X \to \mathbb{R}$ *is continuous if for every open* $O \subseteq \mathbb{R}$, $f^{-1}(O)$ *is open in* $X$.

**Theorem 7.2.** *Let $f : A \to B$ be continuous. If $X \subseteq A$ is compact, then $f(X)$ is compact.*

*Proof.* Take $\{y_n\}$ in $f(X)$. We want to show that $\{y_n\}$ has a subsequence convergent in $f(X)$. By definition of $f(X)$, for each $n$, there exists $x_n \in X$ such that $f(x_n) = y_n$. This defines a sequence $\{x_n\}$ in $X$. Since $X$ is compact, it has a convergent subsequence $\{x_{n_k}\}$ that converges to $x \in X$. Since $f$ is continuous, $y_{n_k} = f(x_{n_k}) \to f(x) \in f(X)$, and $f(X)$ is therefore compact. $\square$

# 8 Linear Spaces

**Definition 8.1.** *A **linear space** is a set $X$ for which the following are satisfied:*

*i)* $x, y \in X \implies x + y \in X$.

*ii) For any* $\alpha \in \mathbb{R}$, $x \in X \implies \alpha x \in X$.

Elements of $X$ are called **vectors**. The identity element $\mathbf{0}$ such that $x + \mathbf{0} = x$ for all $x$ is called the origin, or zero. A clear example of a linear space is $\mathbb{R}^n$.

**Definition 8.2.** *A **linear subspace** of $X$ is a non-empty subset of $X$ that is itself a linear space.*

Example:

$S = \{(a, a) : a \in \mathbb{R}\}$ is a linear subspace of $R^2$. However, $T = \{(a, a) + (0, 2) : a \in \mathbb{R}\}$ is not a linear subspace as $(0, 0) \notin T$.

**Definition 8.3.** *The **inner product**, or dot product, of two vectors $x, y \in \mathbb{R}^n$ is*

$$x \cdot y = \sum_{i=1}^{n} x_i y_i.$$

*It is also sometimes denoted as $\langle x, y \rangle$.*

Example

Let $x = (x_1, \ldots, x_n)' \in \mathbb{R}^n$ be a bundle of goods, and $p = (p_1, \ldots, p_n)' \in \mathbb{R}^n$ be a vector of prices. Then, the total cost of the bundle $x$ is

$$p \cdot x = p_1 x_1 + \cdots + p_n x_n.$$

## 8.1 Norms

Let $X$ be an arbitrary non-empty set.

**Definition 8.4.** *A **norm** is a real-valued function $\| \cdot \| : X \to \mathbb{R}_+$ that, for all $x, y \in X$ and $\lambda \in \mathbb{R}$, satisfies*

*i) (Definite)* $\|x\| = 0$ *if, and only if,* $x = 0$;

*ii) (Absolute Homogeneity)* $\|\lambda x\| = |\lambda| \|x\|$;

*iii) (Subadditivity)* $\|x + y\| \leqslant \|x\| + \|y\|$.

**Lemma 8.1.** *Fix $X$, and endow it with norm $\|\cdot\|$. Then, $\|x\| \geqslant 0$ for all $x \in X$.*

*Proof.* We have that $0 = \|x - x\| \leqslant \|x\| + \|-x\| = 2\|x\|$. ☐

Note that $\|\cdot\|$ is a norm, then

$$d(x, y) \coloneqq \|x - y\|$$

is the distance associated with the norm $\|\cdot\|$.

<u>Example:</u>

The **Euclidean Norm** on $\mathbb{R}^n$ is defined as

$$\|x\| = \sqrt{\sum_{i=1}^{n} x_i^2},$$

and the associated **Euclidean Distance** between two vectors $x, y \in \mathbb{R}^n$ is given by

$$\|x - y\| = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}.$$

For any arbitrary set $X$ and a norm $\|\cdot\|$ on $X$, the pair $(X, \|\cdot\|)$ is a normed space.

**Definition 8.5.** *$x, y \in \mathbb{R}^n$ are **orthogonal** if $x \cdot y = 0$.*

**Theorem 8.1.** ***(Cauchy-Schwarz Inequality)*** *For any $x, y \in \mathbb{R}^n$, $|x \cdot y| \leqslant \|x\|\|y\|$.*

*Proof.* We begin by proving the following claims:

*Claim 1:* For any $x, y \in \mathbb{R}^n$, if $x \perp y$ (i.e., the angle between $x$ and $y$ is $90°$ or $\frac{\pi}{2}$ rad), then $x \cdot y = 0$.

*Proof of Claim 1:* Since $x \perp y$, Pythagoras implies that

$$\begin{aligned}
\|x\|^2 + \|y\|^2 = \|y - x\|^2 &\implies \left(\sqrt{\sum_i x_i^2}\right)^2 + \left(\sqrt{\sum_i y_i^2}\right)^2 = \left(\sqrt{\sum_i (y_i - x_i)^2}\right)^2 \\
&\implies x \cdot x + y \cdot y = y \cdot y - 2y \cdot x + x \cdot x \\
&\implies -2x \cdot y = 0 \implies x \cdot y = 0.
\end{aligned}$$

*Claim 2:* $x \cdot y = \|x\|\|y\| \cos\theta$, where $\theta$ is the angle between $x$ and $y$.

*Proof of Claim 2:* Decompose $x$ into $x_1, x_2 \in \mathbb{R}^n$ such that

i) $x_1 + x_2 = x$,

ii) $x_1$ is parallel to $y$,

iii) $x_2 \perp y$.

By definition, $\cos\theta = \frac{\|x_1\|}{\|x\|}$. We then have

$$x \cdot y = (x_1 + x_2) \cdot y = x_1 \cdot y + x_2 \cdot y.$$

Since $x_2 \perp y$, $x_2 \cdot y = 0$. Moreover, since $x_1$ and $y$ are parallel, we can deduce the existence of $\alpha \in \mathbb{R}$ such that $x_1 = \alpha y \implies \|x_1\| = |\alpha|\|y\|$. Then,

$$x \cdot y = x_1 \cdot y = \sum_i x_{1,i} y_i = \alpha \sum_i y_i y_i = |\alpha|\|y\|\|y\| = \|x_1\|\|y\|.$$

Substituting $\|x_1\| = \|x\| \cos\theta$ achieves the desired result.

With both claims proved, we are ready to show the final result. We have that $x \cdot y = \|x\|\|y\| \cos\theta$. Since $-1 \leqslant \cos\theta \leqslant 1$,

$$-\|x\|\|y\| \leqslant x \cdot y \leqslant \|x\|\|y\| \implies |x \cdot y| \leqslant \|x\|\|y\|.$$

□

## 8.2 Linear Combination and Dependence

Let $V = \{v^1, \dots, v^n\}$ be a set of $n$ vectors in $X$, so each $v^i \in X$, and let $\lambda = (\lambda_1, \dots, \lambda_n) \in \mathbb{R}^n$ be a vector of $n$ scalars. We then have,

1. The vector $\sum_{i=1}^n \lambda_i v^i$ is called a **linear combination** of the vectors $v^1, \dots, v^n$. The scalars $\lambda_1, \dots, \lambda_n$ are the **coefficients** of the linear combination;
2. If $\sum_{i=1}^n \lambda_i = 1$ then the vector $\sum_{i=1}^n \lambda_i v^i$ is called an **affine combination** of the vectors $v^1, \dots, v^n$;
3. If $\sum_{i=1}^n \lambda_i = 1$ and $\lambda_i \geqslant 0$ for each $i$, then the vector $\sum_{i=1}^n \lambda_i v^i$ is called a **convex combination** of the vectors $v^1, \dots, v^n$.

**Definition 8.6.** *The **span** of a subset $S$ of a linear space $X$ is the set of all linear combinations of finitely many members of $S$, i.e.,*

$$\mathit{span}(S) := \left\{ \sum_{i=1}^n \lambda_i v^i : n \in \mathbb{N} \text{ and } (\lambda_i, v^i) \in \mathbb{R} \times S,\ i = 1, \dots, n \right\}$$

By convention, $span(\emptyset) = \{\mathbf{0}\}$. Note that $span(X) = X$ for any linear space $X$, and that $span(\{x\}) = \{\lambda x : \lambda \in \mathbb{R}\}$.

**Definition 8.7.** *The* ***affine hull*** *of a subset $S$ of a linear space $X$, denoted by aff$(S)$, is the set of all affine combinations of finitely many members of $S$, i.e.,*

$$aff(S) := \Big\{\sum_{i=1}^{n} \lambda_i v^i : n \in \mathbb{N} \ and \ (\lambda_i, v^i) \in \mathbb{R} \times S, \ i = 1, \ldots, n \ with \ \sum_{i=1}^{n} \lambda_i = 1\Big\}.$$

**Definition 8.8.** *We say that $V$ is* ***linearly dependent*** *(LD) if there exists scalars $\lambda_i \in \mathbb{R}$, $i = 1, \ldots, n$, such that $\sum_{i=1} \lambda_i v^i = \mathbf{0}$ where $\lambda_i \neq 0$ for at least one $i$.*[9] *$V$ is* ***linearly independent*** *(LI) if $\sum_{i=1} \lambda_i v^i = 0 \implies \lambda_i = 0$ for all $i$.*

Example:
Let $X = \mathbb{R}^2$ and $V = \{(1,2), (2,4)\}$. Then, take $c_1 = 2$ and $c_2 = -1$ and $2(1,2) - 1(2,4) = \mathbf{0}$. In fact, in $R^2$, any two parallel lines are LD. Any set of three vectors in $\mathbb{R}^2$ is also LD.

**Definition 8.9.** *A* ***basis*** *for a linear space $X$ is a finite set of LI vectors $V$ such that $span(V) = X$. The* ***dimension*** *of $X$ is then defined as $dim(X) = |V|$.*

Example: For $i = 1, \ldots, n$, let the vector $e_i \in \mathbb{R}^n$ have 1 in the $i-$coordinate, and 0 everywhere else, so $e_1 = (1, 0, \ldots, 0)'$, $e_2 = (0, 1, 0, \ldots, 0)'$, and so on. Then, is easy to see that $\{e_1, e_2, \ldots, e_n\}$ forms a basis for $\mathbb{R}^n$, and $dim(\mathbb{R}^n) = n$.

One may ask if the dimension of linear space $X$ is well-defined. What if two different bases have different cardinalities? The following result shows we need not be concerned.

**Theorem 8.2.** *Let $\{v^1, \ldots, v^n\}$ be a basis for $X$, and $\{y_1, \ldots, y_k\}$ a set of LI vectors in $X$. Then, $k \leqslant n$.*

*Proof.* Assume, by way of contradiction that $k > n$. Since $\{v^1, \ldots, v^n\}$ is a basis for $X$, then there exist constants $c_1, \ldots, c_n$ such that $y_1 = c_1 v_1 + \cdots + c_n v_n$. Therefore,

$$\begin{aligned}\frac{y_1}{c_i} &= \frac{c_1}{c_i} v_1 + \frac{c_2}{c_1} v_2 + \cdots + v_i + \cdots + \frac{c_n}{c_i} v_n \\ \implies v_i &= \frac{y_1}{c_i} - \frac{c_1}{c_i} v_1 - \frac{c_2}{c_1} v_2 - \cdots - \frac{c_n}{c_i} v_n,\end{aligned}$$

[9] Remember that each $v^i = (v_1^i, \ldots, v_n^i)'$ is a vector and each $\lambda_i$ is a real number.

and therefore $span\{v_1, \dots, v_{i-1}, y_i, v_{i+1}, \dots, v_n\} = X$. Since $\{y_1, \dots, y_k\}$ is linearly independent, and this holds for any $i$, we can write $span(y_1, \dots, y_n) = X$. But as $y_{n+1} \in X$, there exists $d_1, \dots, d_n$ such that $y_{n+1} = \sum_{i=1}^{n} d_i y_i$, and so $\{y_1, \dots, y_k\}$ is not linearly independent. Contradiction. □

The proof is by contradiction, as $k > n \implies \{y_1, \dots, y_k\}$ is LD. A corollary of this result is that the dimension of $X$ is well-defined. A basis is a maximal set of $LI$ vectors. Any set with $k < n$ cannot be a basis.

**Definition 8.10.** *Two linear spaces $X, Y$ are* ***orthogonal*** *if $\forall x \in X$ and $\forall y \in Y$, $x \cdot y = 0$.*

Example:
Take $X = \{(x, y) \in \mathbb{R}^2 : x = y\}$ and $Y = \{(x, y) \in \mathbb{R}^2 : x = -y\}$.

**Definition 8.11.** *For a linear space $X$, the* ***orthogonal complement*** *of $X$ is $X^{\perp} := \{y \in \mathbb{R}^n : y \cdot x = 0 \ \forall x \in X\}$.*

**Theorem 8.3.** *For a linear space $X \subseteq \mathbb{R}^n$, $dim(X) + dim(X^{\perp}) = n$.*

## 8.3 Linear Operators and functionals

**Definition 8.12.** *A function between linear spaces $L : X \to Y$ is linear if*

$$L(\alpha x + y) = \alpha L(x) + L(y) \ \forall x, y \in X, \ \forall \alpha \in \mathbb{R}.$$

Let $\mathcal{L}(X, Y)$ denote the set of all linear operators from $X$ to $Y$.

**Definition 8.13.** *The* ***Kernel*** *(or null space) of a linear function $L : \mathbb{R}^n \to \mathbb{R}^m$ is $K(L) := \{x \in \mathbb{R}^n : L(x) = \mathbf{0}\}$.*

Example:
Let $D : C^1[0, 1] \to C[0, 1]$ be defined by $D(f) := f'$. The derivative operator $D$ is a linear function. $K(D)$ is the set of all constant functions on $[0, 1]$.

**Proposition 8.1.** *The following holds.*

*i) If $f$ is linear, then $K(f)$ is a linear subspace;*
*ii) If $f$ is linear, $K(f) = \{\mathbf{0}\} \iff f$ is injective;*

**Theorem 8.4.** ***(Fundamental Theorem of Linear Algebra)*** *For any linear operator $L \in \mathcal{L}(X, Y)$, we have*

$$dim(K(L)) + dim(span(L)) = dim(X).$$

## 8.4 Matrices

A $m \times n$ **matrix** in a non-empty set $X$ is a function

$$f : \{1, \dots, m\} \times \{1, \dots, n\} \to X.$$

We can represent this function by $A = [a_{ij}]_{m \times n}$, where $a_{ij} \coloneqq f(i, j)$. We can equivalently represent this matrix as an array with $m$ rows and $n$ columns:

$$\begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix}.$$

We say that $m \times n$ are the dimensions of the matrix. When $m = n$, we say that $A$ is a **square matrix**. When $n = 1$, we have a colunm vector, which we usually denote using lowercase letters.

The sum of two matrices $A = [a_{ij}], B = [b_{ij}]$ with the same dimensions is defined as

$$A + B = [a_{ij} + b_{ij}]_{m \times n}.$$

Example:

$$\begin{pmatrix} 1 & 0 \\ 2 & 4 \end{pmatrix} + \begin{pmatrix} 3 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 4 & 0 \\ 3 & 5 \end{pmatrix}.$$

The multiplication of the matrix $A_{m \times n}$ by a scalar $\alpha$ is given by $\alpha A_{m \times n} = [\alpha a_{ij}]_{m \times n}$.

Example:

$$2 \begin{pmatrix} 1 & 0 \\ 2 & 4 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 4 & 8 \end{pmatrix}$$

.

**Definition 8.14.** *The* ***trace*** *of a matrix $A$, denoted $trace(A)$, is the sum of its main diagonal elements, i.e.,*

$$trace(A) \coloneqq a_{11} + \dots + A_{nn}.$$

Let $A_{m\times n}$ and $B_{n\times k}$ be two matrices. The **matrix product** (or matrix multiplication) $C = A \times B$ is the $m$-by-$k$ matrix with elements $c_{ij} = \sum_{t=1}^{n} a_{it}b_{tj}$. Note that matrix multiplication is not commutative: $AB \neq BA$ in general (in fact, this only makes sense when $A$ and $B$ are both square matrices).

**Definition 8.15.** *The **Identity matrix** of dimension $n$ is the $n-$by$-n$ matrix with main diagonal entries 1 and 0 everywhere else:*

$$I_n := \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & \vdots \\ 0 & 0 & \ddots & \dots & \vdots \\ \vdots & \vdots & \dots & \dots & 0 \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix}_{n\times n}.$$

**Proposition 8.2.** *Let $A_{n\times n}$. If $Ax = x$ for all $x \in \mathbb{R}^n$, then $A = I_n$.*

Assuming that the dimensional requirements are satisfied, the following properties of matrix operations hold:

1) $A + B = B + A$;
2) $(A + B) + C = A + (B + C)$;
3) $A(BC) = (AB)C$;
4) $AI_n = I_nA = A$;
5) $(A + B)C = AC + BC$.

Note then that matrices are linear operators. Indeed, we can go a bit further as the next result shows.

**Theorem 8.5.** ***(Matrix Representation)*** *For any linear $f : \mathbb{R}^n \to \mathbb{R}^m$, there exists a matrix $A_{m\times n}$ such that $f(x) = Ax$ for all $x \in \mathbb{R}^n$.*

Therefore, we can think of any function $L : \mathbb{R}^n \to \mathbb{R}^m$ as a linear operator given by:

$$L(x) := Ax.$$

**Definition 8.16.** *The **transpose** of a matrix $A = [a_{ij}]_{m\times n}$ is denoted by $A'$ or $A^T$, and is given by*

$$A' = [b_{ij}]_{n\times m}, \text{ where } b_{ij} = a_{ji}.$$

**Definition 8.17.** *We say that a square matrix $A$ is **symmetric** if $A = A'$.*

Properties of transpose matrices (again assuming that the dimensional requirements are satisfied):

1) $(A')' = A$;
2) $(A+B)' = A' + B'$;
3) $(AB)' = B'A'$;
4) $(\alpha A)' = \alpha A'$ for any $\alpha \in \mathbb{R}$.

**Definition 8.18.** *An n-by-n square matrix A is called* ***invertible*** *if there exists an n-byn square matrix B such that* $AB = BA = I_n$*. If this is the case, B is denoted by* $A^{-1}$ *and it is called the* ***inverse*** *of A.*

**Theorem 8.6.** *Fix an invertible matrix A. Then,*

*i)* $(\alpha A)^{-1} = \alpha^{-1}A^{-1}$;
*ii)* $(AB)^{-1} = B^{-1}A^{-1}$;
*iii)* $(A^{-1})' = (A')^{-1}$.

**Definition 8.19.** *The* ***column (row) rank*** *of matrix A is the number of linearly independent columns (rows) of A.*

**Theorem 8.7.** *For any matrix A, the row rank and column rank are equal. Therefore, we simply write* $rank(A)$.

**Definition 8.20.** *We say that* $A_{n\times n}$ *has full rank if* $rank(A) = n$.

**Theorem 8.8.** *A square matrix A is invertible if, and only if, it has full rank.*

*Proof.* $\Leftarrow$. Let $A$ have full rank. Then $span\{A_1, \ldots, A_n\}$, where $A_i$ is the vector formed by the $i$-th column of $A$, has dimension $n$. Define a linear operator $L : \mathbb{R}^n \to \mathbb{R}^n$ by $L(x) = Ax$. Then, by the Fundamental Theorem of Linear Algebra, $dim(span(L)) + dim(K(L)) = n$. Since $A$ has full rank it is surjective and $dim(span(L)) = n$. Then, $dim(K(L)) = 0$ and so $K(L) = \{0\}$. By a previous result, this implies that $L$ is injective. Therefore $L$ is bijective, and so $L^{-1}$ exists. Let $B$ be a matrix such that $f^{-1}(x) = Bx$. We then get

$$
\begin{aligned}
& BAx = f^{-1} \circ f(x) = x \\
\implies & ABx = f \circ f^{-1}(x) = x \\
\implies & B = A^{-1}.
\end{aligned}
$$

$\Rightarrow$. Note that the arguments above are if, and only if. □

The following properties hold (again assuming that all dimension requirements are satisfied):

1. $rank(A) = rank(A') = rank(AA')$;
2. $rank(A_{m\times n}) \leqslant \min\{m, n\}$;
3. $rank(A) = rank(PA) = rank(AQ) = rank(PAQ)$, where $P, Q$ are square matrices of full rank.
4. The rank of a matrix is unchanged by:
   i) multiplying a row or column by a scalar;
   ii) interchanging any two rows or columns;
   iii) adding a row to a different row or a column to a different column.

### 8.4.1 Determinant

For an $n$-by-$n$ matrix $A$, let $A_{ij}$ denote the $(n-1)\times(n-1)$ matrix formed by removing the $i$-th row and the $j$-th column of $A$.

Example:

$$A = \begin{pmatrix} 1 & 0 & 2 \\ 3 & 1 & 1 \\ 2 & 1 & 1 \end{pmatrix}, A_{13} = \begin{pmatrix} 3 & 1 \\ 2 & 1 \end{pmatrix}.$$

Let the determinant of a 1-by-1 matrix $A_{1\times 1}$ be given by $det(A_{1\times 1}) = a_{11}$. The **determinant** of a $n$-by-$n$ square matrix $A$, denoted by $|A|$ or $det(A)$, is the number

$$|A| := \sum_{j=1}^{n} (-1)^{i+j} a_{ij} det(A_{ij}),$$

for a fixed $i$, and where $det(A_{1\times 1}) = a_{11}$. It is a recursive definition.

**Proposition 8.3.** *A square matrix $A$ is invertible $\iff det(A) \neq 0 \iff$ it has full rank.*

A corollary of this result is that $det(A) \neq 0$ if, and only if, $A$ has two or more LD columns (or rows).

Let $A = (a_1, \ldots, a_n)$, where $a_i$ is a $n \times 1$ column vector for each $i$. The following properties of the determinant hold:

1. $det(a_1, \ldots, \alpha a_i, \ldots, a_n) = \alpha det(A)$ for any $i$;
2. $det(a_1, \ldots, a_i, \ldots, a_j, \ldots, a_n) = -det(a_1, \ldots, a_j, \ldots, a_i, \ldots, a_n)$ (you change two columns);
3. for a $n \times 1$ column vector $b$, $det(a_1, \ldots, a_i + b, \ldots, a_n) = det(a_1, \ldots, a_i, \ldots, a_n) + det(a_1, \ldots, b, \ldots, a_n)$;
4. $det(AB) = det(A)det(B)$; $det(A^{-1}) = [det(A)]^{-1}$.

**Proposition 8.4.** *Let $A$ be an invertible $2 \times 2$ matrix given by*

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}. \textit{ Then, } A^{-1} = \frac{1}{det(A)} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

### 8.4.2 Quadratic Forms

**Definition 8.21.** *Let $A$ be a symmetric $n$-by-$n$ matrix. The* ***quadratic form*** *of $A$ is a function $Q_A : \mathbb{R}^n \to \mathbb{R}$ defined as $Q_A(x) = x'Ax$ for all vectors $x \in \mathbb{R}^n$.*

Example:

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ and } x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}. \text{ Then, } Q_A(x) = x_1^2 + x_2^2.$$

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 4 \end{pmatrix} \text{ and } x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}. \text{ Then, } Q_A(x) = 2x_1^2 + 2x_1x_2 + 4x_2^2.$$

**Definition 8.22.** *Let $A_{n \times n}$ be a symmetric matrix. We say that $A$ is*

- ***positive definite*** *if $Q_A(x) > 0$ for all $x \in \mathbb{R}^n \setminus \{0\}$;*
- ***positive semi-definite*** *if $Q_A(x) \geqslant 0$ for all $x \in \mathbb{R}^n \setminus \{0\}$;*
- ***negative definite*** *if $Q_A(x) < 0$ for all $x \in \mathbb{R}^n \setminus \{0\}$;*
- ***negative semi-definite*** *if $Q_A(x) \leqslant 0$ for all $x \in \mathbb{R}^n \setminus \{0\}$.*

*Otherwise, $A$ is* ***indefinite****.*

This is a very useful definition that extends the notion of positive and negative to matrices.

### 8.4.3 Eigenvalues and Eigenvectors

Consider a $n$-by-$n$ matrix $A$.

**Definition 8.23.** *A scalar $\lambda \in \mathbb{R}$ is called an* ***eigenvalue*** *of $A$ when there exists a vector $x \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ such that $Ax = \lambda x$. The vector $x$ is called an* ***eigenvector*** *of $A$ associated with the eigenvalue $\lambda$.*[10]

If $x$ is an eigenvector of $A$ associated with the eigenvalue $\lambda$ we have:

$$Ax = \lambda x \implies Ax - \lambda x = \mathbf{0} \implies (A - \lambda I_n)x = \mathbf{0}.$$

[10] This definition applies for eigenvalues/eigenvectors of any linear operator $L : X \to X$.

Since $x \neq \mathbf{0}$, the columns of the matrix $(A - \lambda I_n)$ are LD. Therefore,

$$det(A - \lambda I_n) = 0. \tag{1}$$

Equation (1) is the **characteristic equation** of matrix $A$, which is a polynomial equation of degree $n$ in $\lambda$. Therefore, $A$ has (possibly repeated) $n$ eigenvalues. The characteristic equation lets us find the eigenvalues and eigenvectors.

<u>Example:</u>

$$A = \begin{pmatrix} 0 & 1 \\ -2 & -3 \end{pmatrix} \implies A - \lambda I_n = \begin{pmatrix} -\lambda & 1 \\ -2 & -\lambda - 3 \end{pmatrix} \implies det(A - \lambda I_n) = -\lambda(-3 - \lambda) + 2.$$

The characteristic equation is then

$$-\lambda(-3 - \lambda) + 2 = 0 \implies \lambda^2 + 3\lambda + 2 = 0 \implies \lambda_1 = -1, \lambda_2 = -2.$$

To find the corresponding eigenvectors, we solve for $\lambda_1 = -1$,

$$\begin{pmatrix} 0 & 1 \\ -2 & -3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -x_1 \\ -x_2 \end{pmatrix},$$

and we have that $x = (1, -1)'$ is the eigenvector associated with $\lambda_1$ (in fact any vector $(a, a)$, $a \in \mathbb{R}$ is an eigenvector). Similarly, for $\lambda_2$ we have

$$\begin{pmatrix} 0 & 1 \\ -2 & -3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -2x_1 \\ -2x_2 \end{pmatrix},$$

and we have that $x = (1, -2)'$ is the eigenvector associated with $\lambda_2$.

Bogt eigenvalues and eigenvectors are very useful for matrix algebra. They also show up when you want to prove that some dynamic systems converge to an equilibrium.

**Theorem 8.9.** *A square matrix $A$ is not invertible if, and only if, $0$ is an eigenvalue of $A$.*

*Proof.* We have that 0 is an eigenvalue of $A \iff Ax = 0 \iff A$ has LD columns $\iff A$ is not invertible. $\square$

### 8.4.4 Diagonalization

**Definition 8.24.** *We say that a matrix $D = [d_{ij}]$ is diagonal if $i \neq j \implies d_{ij} = 0$.*

Diagonal matrices are very easy to calculate with.

**Definition 8.25.** *A square matrix $A$ is **diagonalizable** if there exists a full rank matrix $P$ and a diagonal matrix $D$ such that $P^{-1}AP = D$.*

**Theorem 8.10.** *A square matrix is diagonalizable if, and only if, it has linearly independent eigenvectors.*

**Theorem 8.11.** *Let $A$ be a square matrix with (possibly repeated) eigenvalues $\lambda_1, \ldots, \lambda_n$. Then,*

*i) $det(A) = \Pi_{i=1}^{n} \lambda_i$;*

*ii) $trace(A) = \sum_{i=1}^{n} \lambda_i$.*

**Theorem 8.12.** *Let $A_{n \times n}$ be a symmetric matrix. Then,*

- *$A$ is positive definite $\iff$ all eigenvalues of $A$ are strictly positive;*
- *$A$ is positive semi-definite $\iff$ all eigenvalues of $A$ are non-negative;*
- *$A$ is negative definite $\iff$ all eigenvalues of $A$ are strictly negative;*
- *$A$ is negative semi-definite $\iff$ all eigenvalues of $A$ are non-positive.*

# 9 Convexity and Concavity

**Definition 9.1.** *A subset $S$ of a linear space $X$ is **convex** if*

$$\forall x, y \in S,\ \forall \lambda \in (0,1),\ \lambda x + (1-\lambda) y \in S.$$

The definition of a convex set implies that any convex combination of elements of the set is also in the set.

Example:

1) The interval $[a, b]$ in $\mathbb{R}$ is convex. In Euclidean space, lines, planes, hyperplanes, are all convex.

2) The simplex of a set, $\triangle(T) := \{\mu \in \mathbb{R}_+^T : \sum_{x \in T} \mu(x) = 1\}$, also known as the set of all probability functions defined on $T$, is convex.

**Definition 9.2.** *Let $S$ be a subset of a linear space $X$. The **convex hull** of $S$, denoted by $co(S)$, is the smallest convex set containing $S$.*

**Proposition 9.1.** *Let $X$ be a linear space and fix $S \subseteq X$. Then, $co(S)$ coincides with the set*

$$\left\{ \sum_{x \in T} \lambda(x) x : T \in P(S) \text{ and } \lambda \in \mathbb{R}^{T_+} \text{ with } \sum_{x \in T} \lambda(x) = 1 \right\}.$$

In what follows, we will work in Euclidean space. so let $T \subseteq \mathbb{R}^n$ be a convex set.

**Definition 9.3.** *We say that a function $f : T \to \mathbb{R}$ is **concave** if*

$$f(\lambda x + (1-\lambda) y) \geqslant \lambda f(x) + (1-\lambda) f(y)$$

*for any $x, y \in T$ and $\lambda \in [0,1]$.*

**Definition 9.4.** *We say that a function $f : T \to \mathbb{R}$ is **strictly concave** if*

$$f(\lambda x + (1-\lambda) y) > \lambda f(x) + (1-\lambda) f(y)$$

*for any $x, y \in T$ with $x \neq y$ and $\lambda \in (0,1)$.*

The function $f \in \mathbb{R}^T$ is **(strictly) convex** if $-f$ is (strictly) concave. Can a function be both concave and convex at the same time?

Example:

1) The function $f(x) = x^2$ is convex.
2) The function $f(x) = \ln(x)$ is concave.
3) The function $f(x) = 2x - 3$ is both concave and convex.

A function $f$ is **affine** if it is both concave and convex. It is not possible for a function to be both strictly convex and concave, or strictly concave and convex. The following properties are true.

1. If $f$ and $g$ are concave functions, then $\alpha f + g$ is a concave function for any $\alpha \geqslant 0$.
2. If $S$ is an interval with $f(T) \subseteq S$, and $f \in \mathbb{R}^T$ and $g \in \mathbb{R}^S$ are concave and $g$ is increasing, then $g \circ f$ is concave.
3. Let $\{f_n\}$ be a sequence of concave functions in $\mathbb{R}^T$. If $\lim_n f_n(x)$ exists in $\mathbb{R}$ for all $x \in T$, then the map $x \mapsto \lim_n f_n(x)$ is a concave function in $\mathbb{R}^T$.

**Lemma 9.1.** *Fix any $a \leqslant b \in \mathbb{R}$. If $f : [a, b] \to \mathbb{R}$ is concave (or convex), then*

$$\inf f([a, b]) > -\infty \text{ and } \sup f([a, b]) < \infty.$$

**Proposition 9.2.** *Let $I$ be an open interval and $f : I \to \mathbb{R}$. If $f$ is a concave (or convex) function, then it is continuous.*

**Definition 9.5.** *For any function $f : T \to \mathbb{R}$, its **hypograph** is the set*

$$hypo(f) \coloneqq \{(x, y) \in \mathbb{R}^n \times \mathbb{R} : x \in T \text{ and } y \leqslant f(x)\}.$$

**Theorem 9.1.** *A function $f : T \to \mathbb{R}$ is concave if, and only if, $hypo(f)$ is convex.*

*Proof.* Let $S = hypo(f)$.

$\Rightarrow$. Now, assume $f$ is concave. Then, fix $(x^1, y^1), (x^2, y^2) \in S$ and take $t \in [0, 1]$. Then, by the definition of $S$,

$$ty^1 + (1 - t)y^2 \leqslant tf(x^1) + (1 - t)f(x^2) \leqslant f(tx^1 + (1 - t)x^2),$$

where the second inequality follows from the concavity of $f$.

$\Leftarrow$. Now, assume that $S$ is convex, and take $x^1, x^2 \in T$. We then have that $(x^1, f(x^1))$ and $(x^2, f(x^2))$ are in $S$. Since $S$ is convex, for $t \in [0, 1]$,

$$(tx^1 + (1 - t)x^2, tf(x^1) + (1 - t)f(x^2)) \in S.$$

Then, by definition of $S$, $tf(x^1) + (1 - t)f(x^2) \leqslant f(tx^1 + (1 - t)x^2)$ and $f$ is concave.

$\square$

**Theorem 9.2.** *Suppose that $f \in \mathbb{R}^T$ is differentiable in an open set that contains $T$. Then, $f$ is concave if, and only if, for all $x, y \in T$,*

$$f(y) \leqslant f(x) + \nabla f(x) \cdot (y - x).$$

*Moreover, $f$ is strictly concave if, and only if,*

$$f(y) < f(x) + \nabla f(x) \cdot (y - x)$$

*for all $x, y \in T$ with $x \neq y$.*

**Theorem 9.3.** *Let $f : T \to \mathbb{R}$ be twice differentiable. Then, $f$ is concave if, and only if, $Hf(x)$ is negative semi-definite for all $x \in T$. Moreover, $f$ is strictly concave if, and only if, $Hf(x)$ is negative definite for all $x \in T$.*

Consequently, $f$ is convex if, and only if, $Hf$ is positive semi-definite, and $f$ is strictly convex if, and only if, $Hf$ is positive definite. The following results may be useful as well, and sometimes easier to verify.

**Proposition 9.3.** *Assume that $I$ is an open interval and $f : I \to \mathbb{R}$ is differentiable.*
*i) $f$ is (strictly) concave if, and only if, $f'$ is (strictly) decreasing.*
*ii) $f$ is (strictly) convex if, and only if, $f'$ is (strictly) increasing.*

**Theorem 9.4.** *Let $f(x, y) \in C^2$ and let $S \subseteq \mathbb{R}^2$ be an open convex set. Then,*
*i) $f$ is convex $\iff \frac{\partial^2 f}{\partial x^2} \geqslant 0$; $\frac{\partial^2 f}{\partial y^2} \geqslant 0$, $det(Hf) \geqslant 0$,*
*ii) $f$ is concave $\iff \frac{\partial^2 f}{\partial x^2} \leqslant 0$; $\frac{\partial^2 f}{\partial y^2} \leqslant 0$, $det(Hf) \leqslant 0$.*
*For strict concavity/convexity, the equalities are strict.*

## 9.1 Quasiconcavity

**Definition 9.6.** *A function $f : T \to \mathbb{R}$ is quasi-concave if for all $x, y \in T$ and $\lambda \in [0, 1]$,*

$$f(\lambda x + (1 - \lambda)y) \geqslant \min\{f(x), f(y)\}.$$

It is easy to see that a concave function is quasi-concave, but that the converse is not necessarily true.

**Definition 9.7.** *A function $f : T \to \mathbb{R}$ is strictly quasi-concave if for all $x, y \in T$ with $x \neq y$ and $\lambda \in [0, 1]$,*

$$f(\lambda x + (1 - \lambda)y) > \min\{f(x), f(y)\}.$$

It is straitforward to show that a strictly concave function is strictly quasi-concave. We say that $f$ is **quasi-convex** if $f(\lambda x + (1-\lambda)y) \leqslant \max\{f(x), f(y)\}$ for all $x, y \in T$. We say that $f$ is **strictly quasi-convex** if $f(\lambda x + (1-\lambda)y) > \max\{f(x), f(y)\}$ for all $x, y \in T$ with $x \neq y$. We say $f$ is **quasilinear** if it is both quasi-concave and quasi-convex.

**Definition 9.8.** *Let $f : T \subseteq \mathbb{R}^n \to \mathbb{R}$. For each $c \in \mathbb{R}$, the* ***upper contour*** *set of $f$ at $c$ is the set*

$$U_c(f) \coloneqq \{x \in T : f(x) \geqslant c\}.$$

*Similarly, the* ***lower contour*** *set of $f$ at $c$ is the set*

$$L_c(f) \coloneqq \{x \in T : f(x) \leqslant c\}.$$

**Theorem 9.5.** *Consider $f : T \to \mathbb{R}$.*

*i) The function $f$ is quasi-concave if, and only if, $U_c(f)$ is convex for all $c \in \mathbb{R}$.*

*ii) The function $f$ is quasi-convex if, and only if, $L_c(f)$ is convex for all $c \in \mathbb{R}$.*

*Proof.* We prove $i)$. The proof of $ii)$ is analogous.

$\Leftarrow$. Assume $U_c(f)$ is convex for all $c$. Take $x, y \in X$ and, without loss, assume $f(x) \geqslant f(y)$. Denote $f(y) = a$, and so $f(x) \geqslant a \implies x, y \in U_a(f)$. Since $U_a(f)$ is convex, for any $t \in [0,1]$, $tx + (1-t)y \in U_a$, this implies $f(tx + (1-t)y) \geqslant a = f(y) = \min\{f(x), f(y)\}$.

$\Rightarrow$. Assume $f$ is quasiconcave. Fix $c$, and let $x, y \in U_c$, and so $f(x), f(y) \geqslant c$. Without loss, assume $f(x) \geqslant f(y)$. By quasiconcavity,

$$f(tx + (1-t)y) \geqslant \min\{f(x), f(y)\} \geqslant c,$$

and so $tx + (1-t)y \in U_c(f)$. $\square$

**Theorem 9.6.** *Let $f : T \to \mathbb{R}$ be a quasi-concave function, and let $g : \mathbb{R} \to \mathbb{R}$ be non-decreasing. Then, $g \circ f$ is quasi-concave.*

*Proof.* Take $x, y \in T$.

<u>Case 1:</u> $g \circ f(x) = g \circ f(y)$. Without loss, let $f(x) \geqslant f(y)$. Then, for $t \in [0,1]$

$$\begin{aligned} f \text{ quasiconcave: } & f(tx + (1-t)y) \geqslant f(y) \\ g \text{ non-decreasing: } & g \circ f(tx + (1-t)y) \geqslant f(y) \geqslant g \circ f(y). \end{aligned}$$

Case 2: $g \circ f(x) \neq g \circ f(y)$. Without loss, assume $g \circ f(x) \geqslant g \circ f(y)$. Since $g$ is non-decreasing, $f(x) > f(y)$. By quasiconcavity of $f$, for any $t \in [0, 1]$, $f(tx + (1-t)y \geqslant f(y)$. Since $g$ is non decreasing, $g \circ f(tx + (1-t)y) \geqslant g \circ f(y)$. $\square$

# 10 Optimization Problems

This section discusses the solution to constrained and unconstrained maximization and minimization problems. We detail the conditions for a solution to exist, when we can say the solution is unique, and how we can find it.

Note that solving for $\max_x f(x)$ is the same as solving for $\min_x -f(x)$, so the properties of the problems are very closely related. Therefore, we will focus on maximization problems.

## 10.1 Unconstrained Optimization

A canonical **unconstrained optimization** (or **programming**) **problem** is of the form

$$\max_{x \in X} f(x),$$

where $f : X \to \mathbb{R}$ is the **objective function**. A solution to the problem is $x^* \in X$ such that $f(x) \leqslant f(x^*)$ for all $x \in X$. We may write $\max_x f(x)$ when leaving $X$ out does not cause confusion. We say that the problem has a solution when there exists such $x^* \in X$, otherwise there is no solution to the problem and we can say the problem is not well defined.

Example:
1) The problem $\max_{x \in \mathbb{R}} 2x$ does not have a solution.
2) The problem $\max_{x \in [0,1]} 2x$ has a solution.
3) The problem $\max_{x \in [0,1)} 2x$ does not have a solution.

Optimization problems are everywhere in economics.

1. Utility maximization by a consumer with a budget constraint.
2. Firms maximize profits.
3. Least squares.
4. Etc...

That is why it is important to understand how they work, and how we can solve them. We have seen the following result before, which we state again, but now for functions in $\mathbb{R}^n$.

Consider $f : T \to \mathbb{R}$. We say that $x \in T$ is a **local maximum** of $f$ if there exists an open ball $B$ centered at $x$ such that $f(x) \leqslant f(y)$ for all $y \in B \cap T$. We say that $x$ is a **global maximum** of $f$ if $f(x) \geqslant f(y)$ for all $y \in T$. The next result is very important for solving optimization problems. Solving a problem $\max_{x \in X} f(x)$ is then finding the global maximum in of $f$ in $X$.

**Theorem 10.1.** ***(First Order Condition)*** *Let $f : T \subseteq \mathbb{R}^n \to \mathbb{R}$ be differentiable at a local maximizer $x \in T$. Then, $\nabla f(x) = 0$.*

The FOC is a necessary condition for $x$ to be a local maximizer, but is not sufficient.

Example: Take $f : \mathbb{R} \to \mathbb{R}$ with $f(x) = x^3$. Then, $f'(0) = 0$, but 0 is not a local maximum.

The following will give us a sufficient condition for a critical point $x$ to be a local maximizer.

**Theorem 10.2.** ***(Second Order Condition)*** *Let $f : T \subseteq \mathbb{R}^n \to \mathbb{R}$ be twice differentiable, and let $x \in T$ be such that $\nabla f(x) = 0$. Then,*

*i) (necessary) if $x$ is a local maximizer, then $Hf(x)$ is negative semi-definite.*

*ii) (sufficient) if $Hf(x)$ is negative definite, then $x$ is a local maximizer of $f$.*

What about a global maximizer?

**Theorem 10.3.** *If $f : T \to \mathbb{R}$ is concave everywhere and $\nabla f(x) = 0$, then $x$ is a global maximizer of $f$.*

The following result gives us uniqueness of the solution for the maximization problem.

**Theorem 10.4.** *If a function $f : T \to \mathbb{R}$ is strictly quasi-concave, then $f$ has at most one global maximum.*

## 10.2 Constrained Optimization

A canonical **constrained optimization** (or programming) problem is of the form

$$\begin{aligned} &\max_{x \in X} f(x) \\ &\text{s.t. } \; g_1(x) \geqslant 0, \ldots, g_m(x) \geqslant 0 \\ &\qquad h_1(x) = 0, \ldots, h_n(x) = 0, \end{aligned}$$

where $f : X \to \mathbb{R}$ is again the objective function, $g_i$, $i = 1, \ldots, m$, are the **inequality constraints**, and $h_j, j = 1, \ldots, n$, are the **equality constraints**. Constraints put limits in the domain of the problem. The set $C = \{x \in X : g_i(x) \geqslant 0 \ \forall i \text{ and } h_j(x) = 0 \ \forall j\}$ is the constraint set of the problem, or the set of **feasible points**.

The following may be one of the most useful results for microeconomics. It is a direct consequence of Weierstrass's theorem.

**Theorem 10.5.** *Suppose $C$ is compact and $f$ is continuous. Then, the constrained optimization problem has a solution.*

### 10.2.1 Equality Constraints

These are the simplest constraints. In some cases, it suffices to use substitution: we use the equation $h_j(x) = 0$ to isolate one $x_k$ in the vector $x$, and plug it back into the objective function.

Example:
Consider the problem

$$\max_{x_1,x_2} x_1 x_2 \text{ s.t. } x_1 + 2x_2 = 6.$$

Then, $x_1 = 6 - 2x_2$, and we can rewrite the problem as

$$\max_{x_1,x_2} (6 - 2x_2)x_2.$$

By The FOC,

$$[x_2] : 6 - 4x_2 = 0 \implies x_2 = \frac{3}{2} \text{ and } x_1 = 6 - 2(\frac{3}{2}) = 3.$$

We check the SOC,

$$f''(x) = -4 < 0,$$

and $(x_1, x_2) = (3, \frac{3}{2})$ is the global maximizer.

Sometimes we cannot use substitution, or it is too impractical. We will now describe another method to solve the unconstrained optimization problem. Let us write the problem as

$$\max_x \ f(x) \text{ s.t. } h(x) = 0,$$

where

$$h(x) = \begin{pmatrix} h_1(x) \\ \vdots \\ h_n(x) \end{pmatrix}.$$

**Theorem 10.6.** *If $x^*$ solves the constrained optimization problem with only equality constraints, and the rank of $Jh(x^*)$ is $m$, then there exists unique Lagrange multipliers $\lambda^* \in \mathbb{R}^n$*

*such that*

$$\nabla f(x^*) = (\lambda^*)' Jh(x^*) = \sum_{i=1}^{n} \lambda_i^* \nabla h_i(x^*).$$

This is similar to the idea that $\nabla f(x) = 0$ at a local maximizer from before, but now taking into account the constraints. This result leads us to the following definition.

**Definition 10.1.** *The* ***Lagrangian*** *of a constrained problem is the function*

$$\mathcal{L}(x, \lambda) = f(x) + \lambda \cdot h(x).$$

**Theorem 10.7.** *If $f$ attains a constrained local maximum at $x^*$, then $\exists \lambda^* \in \mathbb{R}^m$ such that $\nabla \mathcal{L}(x^*, \lambda^*) = 0$.*

Therefore, we can solve the dual problem of maximizing $\mathcal{L}(x, \lambda)$. The solution for this dual problem will also be a solution for the original problem. As before, this is only the necessary condition. If $\mathcal{L}$ is concave in $X$, then the condition is also sufficient.

### 10.2.2 Inequality Constraints

Now, we have constraints of the form $g_j(x) \geqslant 0$. This adds an extra complication because at a solution $x^*$, either $g_j(x^*) = 0$ or $g_j(x^*) > 0$, but we may have no way of knowing beforehand which is true.

We say that the constraint $g_j(x) \geqslant 0$ **binds at the solution** $x^*$ if $g_j(x^*) = 0$. Otherwise, we say that the constraint is **slack**. If we knew beforehand which constraints are slack, we could simply disregard them and consider the simpler optimization problem with only the remaining equality constraints.

One way of solving optimization problems with inequality constraints involves deducing which constraints are slack at the optimal and disregarding them. Another way is to "guess" that a constraint is slack, find an optimal for the resulting simplified problem, and go back to check if the guess was correct. A more general approach is given by the result below.

**Theorem 10.8.** ***(Karush-Kuhn-Tucker)*** *Consider the problem*

$$\max_{x \in \mathbb{R}^n} f(x) \ \ s.t.$$

$$g(x) = \begin{pmatrix} g_1(x) \\ \vdots \\ g_m(x) \end{pmatrix} \geqslant 0.$$

*Under some regularity conditions,*[11] *if $x^*$ is a constrained local maximizer of $f$, then there exists $\lambda^* \in \mathbb{R}^m_+$ such that*

1. $g(x^*) \geqslant 0$;
2. $\lambda_j^* \geqslant 0$ *for* $j = 1, \ldots, m$;
3. $\nabla f(x^*) + \sum_{j=1}^m \lambda_j^* \nabla g_j(x^*) = 0$ $(\nabla \mathcal{L}(x^*, \lambda^*) = 0)$;
4. *(complementary slackness)* $\lambda_j^* g_j(x^*) = 0$ *for all* $j = 1, \ldots, m$.

This is again only a necessary condition. Under other regularity conditions, for instance, if $f$ is concave and $g_1, \ldots, g_m$ are quasi-concave, then KKT is also sufficient. Usually we need to check if the conditions are satisfied case-by-case.

### 10.2.3 Envelope Theorem

Let us think of the maximization problem in terms of choice variables $x \in \mathbb{R}^n$ and parameters $p \in \mathbb{R}^l$. We can write the problem as

$$\max_{x \in \mathbb{R}^n} f(x, p) \text{ s.t.}$$
$$\begin{pmatrix} g_1(x, p) \\ \vdots \\ g_m(x, p) \end{pmatrix} \geqslant 0.$$

where $f : \mathbb{R}^{n+l} \to \mathbb{R}$ and $g_j : \mathbb{R}^{n+l} \to \mathbb{R}$ for all $j = 1, \ldots, m$. The corresponding Lagrangian for this problem is $\mathcal{L}(x, \lambda, p) = f(x, p) + \lambda \cdot g(x, p)$.

Assume that the problem has a solution $x^*$ for a any vector of parameters $p \in T \subseteq \mathbb{R}^l$. We can then define a function $x^*(p)$ that gives the solution[12] to the problem when the value of the parameters is $p$.

**Definition 10.2.** *The **value function** $V : T \to \mathbb{R}$ is given by $V(p) = f(x^*(p), p)$.*

We want to understand how the value function changes as we change the parameters of the problem.

Example:

Consider the following utility maximization problem for a consumer with income $\omega$ and

[11] These are the constraint qualifications, which can be a bit ad-hoc. To see a formal discussion, check the mathematical appendix of Mas-Colell et al. (1995).

[12] We are implicitly assuming that the solution is unique for any $p \in T$.

subject to a vector of prices $p$,

$$\max_{x \in \mathbb{R}^2} u(x)$$
$$\text{s.t. } p \cdot x \leqslant \omega.$$

How does the utility of the consumer changes as the prices change?

**Theorem 10.9.** ***(Envelope Theorem)*** *Let $V$ and $\mathcal{L}$ be continuously differentiable. Then,*

$$\frac{\partial V}{\partial p_k} = \frac{\partial \mathcal{L}(x^*(p), \lambda^*(p), p)}{\partial p_k} \text{ for all } k = 1, \ldots, l.$$

When $p$ changes, there should be two effects on $V$. The direct effect from a change in $p$, and an indirect effect from how $p$ changes $x^*(p)$. The envelope theorem tells us that only the direct effect counts. Indeed, because $x^*(p)$ is a maximum,

$$\frac{\partial \mathcal{L}(x^*(p), \lambda^*(p), p)}{\partial x_i} = 0.$$

# 11 Probability

This may be one of the most important topics for this course, and it will be revisited frequently throughout the PhD. It is a fundamental part of how we model beliefs in economics, and it forms the foundation for statistical inference and econometrics.

The rigorous treatment of probability is covered in measure theory, which defines the necessary mathematical objects. In these notes, however, we provide a formal, yet more elementary introduction to the topic. The interested readers are encouraged to look at other sources.

At a more intuitive level, probability assigns a "likelihood" that a certain event will happen. We start with a frequentist notion, in which we observe the outcome of a large number of trials and count the frequency with which a certain event happens.

There is a universal set of outcomes $\Omega$. An **event** $E$ is a subset of $\Omega$.

Example:

When rolling a die, the set of outcomes is $\Omega = \{1, 2.3.4.5.6\}$. An event $E_1$ can be to "roll an even number," so that $E_1 = \{2, 4, 6\}$. Another event $E_2$ can be to roll a prime number, so $E_2 = \{2, 3, 5\}$.

A **probability** is a function $P : 2^{\Omega} \to [0, 1]$ that satisfies the following properties:

i) $P(A) \in [0, 1]$ for all $A \in 2^{\Omega}$;

ii) $P(\Omega) = 1$;

iii) Countable additivity: if $\{B_i\}_{i=0}^{\infty}$ is a class of pairwise disjoint sets, then

$$P(\bigcup_{i=1}^{\infty} B_i) = \sum_{i=1}^{\infty} P(B_i).$$

The above properties directly imply other important properties of probability functions listed below. The proofs of properties $(iii) - (vi)$ are left as an exercise.

i) $P(\emptyset) = 0$;

Take a class of sets defined by $B_1 = \Omega$ and $B_i = \emptyset$ for all $i > 1$. This gives us a class of pairwise disjoint sets. Then, $P(\bigcup_{i=1}^{\infty} B_i) = \sum_{i=1}^{\infty} P(B_i) = P(\Omega) + \sum_{i=2}^{\infty} P(\emptyset)$. Remember that $P(\emptyset) \geqslant 0$ and $P(\bigcup_{i=1}^{\infty} B_i) \leqslant 1$ as $P(\cdot)$ is a probability function. Since we have an infinite sum, this implies that $P(\emptyset) = 0$.

ii) Finite Additivity: If $A \cap B = \emptyset$, then $P(A) + P(B) = P(A \cup B)$;

Let $B_1 = A$, $B_2 = B$ and $B_i = \emptyset$ for $i \geqslant 3$. Then, $P(\bigcup_i B_i) = P(A) + P(B) + \sum_{i=3}^{\infty} P(\emptyset) = P(A) + P(B)$. Since $\bigcup_{i=0}^{\infty} B_i = A \cup B$ we have the desired result.

iii) Excision (a): $P(A^c) = 1 - P(A)$;

iv) Excision (b): $P(B \setminus A) = P(B) - P(A \cap B)$;

v) Monotonicity: For any class of events (not necessarily pairwise disjoint) $\{B_i\}_i$,

$$P(\bigcup_i B_i) \leqslant \sum_{i=1}^{\infty} P(B_i).$$

vi) General Additivity: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

vii) Continuity:[13] If $B_i \subseteq B_{i+1}$ for all $i$, or if $B_i \supseteq B_{i+1}$ for all $i$, then $\lim_{i\to\infty} P(B_i) = P(\lim_{i\to\infty} B_i)$.

The properties of a probability function are very intuitive, given that we want them to represent the likelihood that an event will happen.

NOTE: As its name implies, countable additivity does not apply over uncountable sets. For instance, take $\Omega = [0, 1]$, where $P((a, b)) = b - a$ for all $0 < a < b < 1$. What is the probability of choosing the number 1/2, i.e, $P(\{1/2\}) =$? By continuity, we have that

$$P(\{\frac{1}{2}\}) = \lim_{n\to\infty} P((\frac{1}{2} - \frac{1}{n}, \frac{1}{2} + \frac{1}{n})) = \lim_{n\to\infty} \frac{2}{n} = 0.$$

Countable additivity implies that, when $\Omega$ is uncountable, any events formed by countable sets have zero probability. Also, note that $1 = P([0, 1]) = P(\bigcup_{x\in[0,1]}\{x\}) \neq \sum_x P(\{x\}) = 0$.

## 11.1 Conditional Probability

Fix a universe set $\Omega$ and let $A, B$ be two events with $P(B) > 0$.

**Definition 11.1.** *The probability of $A$ conditional on $B$ is*

$$P(A|B) := \frac{P(A \cap B)}{P(B)}.$$

The idea behind conditional probability is to give the likelihood of an event happening given that we know that another event happened.

Example:

With a fair six-sided die, what is the probability that we roll an odd number, conditional on rolling a prime number? With a fair six-sided die, each number 1 to 6 is rolled with 1/6

[13] This property requires the definition the limits of sets, which is not covered in these notes, so no need to prove.

probability. The event of rolling a prime number is $E_1 = \{2, 3, 5\}$, and the event of rolling an odd number is $E_2 = \{1, 3, 5\}$. Then,

$$P(E_2|E_1) = \frac{P(E_1 \cap E_2)}{P(E_1)} = \frac{P(\{3,5\})}{P(\{2,3,5\})} = \frac{2/6}{3/6} = \frac{2}{3}.$$

**Theorem 11.1.** ***(Bayes' Rule)*** *Let $A$ and $B$ be events with $P(B) > 0$. Then,*

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}.$$

Bayes' Rule is a direct consequence of the definition of conditional probability as

$$P(A|B)P(B) = P(A \cap B) = P(B|A)P(A).$$

The concept of Baye's Rule is used to model how economic agents update their beliefs based on data and observations.

**Definition 11.2.** *Two events $A$ and $B$ are **independent** if $P(A \cap B) = P(A)P(B)$.*

This definition comes from the fact that, for two independent events $A$ and $B$,

$$\begin{aligned} P(A|B) &= \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A), \\ P(B|A) &= \frac{P(A \cap B)}{P(A)} = \frac{P(A)P(B)}{P(A)} = P(B). \end{aligned}$$

Therefore, the occurrence of one event gives no new information about whether the other event is more or less likely to happen. The following are direct consequences of the definition of independence:

i) if $P(A) = 1$ or $P(A) = 0$ for some event $A$, then $A$ is independent of any other event;
ii) if $A$ and $B$ are independent, then $A$ and $B^c$ are independent, and so are $A^c$ and $B^c$, and $A$ and $B^c$.

NOTE: Independence is **almost always** an assumption as it cannot be directly proven in real life.

**Definition 11.3.** *A class of sets $\{B_i\}_{i \in I}$ is a **partition** of a set $X$ if $\{B_i\}_{i \in I}$ is pairwise disjoint*[14] *and $\bigcup_{i \in I} B_i = X$.*

[14]This means $B_i \cap B_j = \emptyset$ for all $i \neq j$.

**Theorem 11.2.** *(Law of Total Probability) Let $\{B_i\}$ be a partition of $\Omega$, with $I$ countable. Then, for any event $A \subseteq \Omega$,*

$$P(A) = \sum_{i \in I} P(A|B_i)P(B_i).$$

## 11.2 Probability Distributions

When we are talking about uncertain events, we are usually dealing with probability distributions of a random variable.

A **random variable**[15] is a real-valued function $X : \Omega \to \mathbb{R}$. Intuitively, we can think of it in terms of a function whose values cannot be determined precisely prior to observation.

<u>Example:</u> Let $\Omega = \{$⚀, ⚁, ⚂, ⚃, ⚄, ⚅$\}$ represent all possible results of rolling a six-sided die. Let $X$ be a random variable (r.v.) defined as

$X($⚀$) = 1$, $X($⚁$) = 2$, $X($⚂$) = 3$, $X($⚃$) = 4$, $X($⚄$) = 5$, $X($⚅$) = 6$.

Take any $\Omega$, and for any $x \in \mathbb{R}$, with a slight abuse of notation we write $P(X \leqslant x) := P(\{\omega \in \Omega : X(\omega) \leqslant x\})$.

Note that a random variable induces a probability on the real numbers, which will be our probability distribution. Random variables are useful because they allow us to think of probabilities in terms of numbers instead of thinking of arbitrary events that can be anything.

**Definition 11.4.** *The* ***Cumulative Distribution Function (CDF)*** *of a r.v. $X$ is the function*

$$F(x) = P(X \leqslant x) \text{ for all } x \in \mathbb{R}.$$

If $X$ follows a distribution $F$ parametrized by $\theta$, we denote $X \sim F(\theta)$. CDFs have the following properties:

i) weakly increasing;
ii) right-continuous;
iii) $\lim_{x\to-\infty} F(x) = 0$, $\lim_{x\to\infty} F(x) = 1$.

**Definition 11.5.** *A r.v. $X$ is called* ***discrete*** *if it can take only countably many values. Otherwise it is called* ***continuous****.*

[15] The formal definition is that given a probability space $(\Omega, \mathcal{F}, P)$, a random variable is a function $X : \Omega \to \mathbb{R}$ that is measurable, but that is going too deep into measure theory.

**Definition 11.6.** *The* ***probability mass function (PMF)*** *of a discrete r.v. $X$ is a function $g : \mathbb{R} \to [0, 1]$ defined by*

$$g(x) = P(X = x).$$

**Definition 11.7.** *The* ***probability density function (PDF)*** *of a continuous r.v. $X$ is a function $f : \mathbb{R} \to [0, 1]$ such that*

*i)* $f(x) \geqslant 0$ *for all* $x \in \mathbb{R}$*;*
*ii)* $\int_a^b f(x)dx = P(a \leqslant X \leqslant b)$*;*
*iii)* $\int_{-\infty}^{\infty} f(x)dx = 1$*.*

NOTE: PDFs need not exist!

**Definition 11.8.** *A distribution is* ***absolutely continuous*** *if a density function exists.*

However, we usually work with distributions that have defined PDFs. When the PDF exists,

i) $F(x) = \int_{-\infty}^{x} f(s)ds$;
ii) $F'(x) = f(x)$.

Here, we list a few of the most relevant random variables and their distributions.

1) Bernoulli:

Example: What is the probability of Heads in a single coin flip?

We say that $X \sim Be(p)$ when $P(X = 1) = p$ and $P(X = 0) = 1 - p$ for $p \in [0, 1]$. It is a discrete r.v. and the corresponding pmf is

$$P(X = x) = p^x(1 - p)^{1-x}.$$

2) Binomial:

Example: what is the probability of getting 3 heads in 7 independent coin flips?

We say that $X \sim B(n, p)$ when it is a series of $n$ independent $Be(p)$ trials. It is a discrete r.v. and the corresponding pmf for any number $x \in \mathbb{N}$ of successes in $n$ trials is

$$P(X = x) = \binom{n}{x} p^x(1 - p)^{n-x}, \text{ where } \binom{n}{x} = \frac{n!}{(n - x)!x!}.$$

3) Geometric:

Example: flip a coin until you get your first Heads. What is the probability that it takes $n$ flips to get a Head?

We say that $X \sim Ge(p)$ where each trial is a Bernoulli with probability $p$. It is a discrete r.v. and the corresponding pmf for any number $x \in \mathbb{N}$ of tries until the first success is

$$P(X = x) = (1-p)^{x-1}p.$$

4) Poisson:

Example: it models the total number of arrivals, or successes, for a given fixed time period. It is often used in continuous-time models.

We say that $X \sim Poisson(\lambda)$ for a parameter $\lambda \geqslant 0$. It is a discrete r.v. and the corresponding PMF for any number $x \in \mathbb{N}$ of successes in a time period is

$$P(X = x) = e^{-\lambda}\frac{\lambda^x}{x!}$$

5) Uniform:

Example: pick a at random a real number between 0 and 1.

We say that $X \sim U[a, b]$. It is a continuous r.v. and the corresponding PDF is

$$f(x) = \begin{cases} \frac{1}{b-a} \text{ if } x \in [a, b], \\ 0 \text{ otherwise.} \end{cases}$$

6) Normal:

Example: the distribution of height in a population. It is an extremely important distribution.

We say that $X \sim N(\mu, \sigma^2)$ for mean $\mu$ and variance $\sigma^2$. It is a continuous r.v. and the corresponding PDF is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

The standard normal distribution is when $\mu = 0$ and $\sigma^2 = 1$.

**Definition 11.9.** *The* ***support*** *of a random variable $X$ is given by the range $\omega \mapsto X(\omega)$.*

In other words, we can more easily think of the support of a random variable as the values that are reached with positive probability.

Example:

Let $X \sim U[-1,1]$. Then, $\text{supp}(X) = [-1,1]$. Indeed, $3 \notin \text{supp}(X)$ as $f(3) = 0$, where $f$ is the PDF of the uniform $[-1,1]$.

## 11.3 Moments

Commonly, we are interested in some properties of how the distribution is "shaped". This can be given by the moments of the distribution.

The **Expected Value**, or the mean, of a r.v. $X$ is the number

$$\mathbb{E}[x] = \int_{-\infty}^{\infty} x dF(x) := \begin{cases} \sum_x xg(x) \text{ if } X \text{ is discrete;} \\ \int_x xf(x)dx \text{ if } X \text{ is continuous.} \end{cases}$$

It is probability-weighted average, and give information on the "central tendency" of the r.v.

Example:

For $X \sim Be(p)$ we have

$$E[X] = 1g(1) + 0g(0) = 1p + 0(1-p) = p.$$

For $X \sim U[a,b]$, we get

$$E[X] = \int_{-\infty}^{\infty} x \frac{1}{b-a} dx = \int_b^a x \frac{1}{b-a} dx = \frac{1}{b-a} \frac{x^2}{2}\bigg|_b^a = \frac{b+a}{2}.$$

The **variance** of a r.v. $X$ is a measure of sprea/dispersion around the mean, given by

$$Var[X] = \mathbb{E}[(X - \mathbb{E}[X])^2].$$

We can also denote the variance by $\sigma^2(X)$. When interpreting the variance in practice, it may be useful to work instead with the **standard deviation** defined by $\sigma(X) = \sqrt{Var[x]}$.

Example:

For $X \sim Be(p)$ we have

$$Var[X] = (1-p)^2 g(1) + (0-p)^2 g(0) = (1-p)^2 p + (0-p)^2 (1-p) = p(1-p).$$

A very useful result:

$$Var[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

In many cases we may be interested in measuring a co-movement between two different r.v. $X, Y$. The **covariance** between $X, Y$ is the number

$$Cov[X,Y] \coloneqq \mathbb{E}[(X - E[X])(Y - \mathbb{E}[Y])].$$

When $Cov[X,Y] = 0$ we say that $X$ and $Y$ are **uncorrelated**. A very useful result:

$$Cov[X,Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

We can normalize the covariance using the **correlation** between $X, Y$, which is the number

$$\rho(X,Y) \coloneqq \frac{Cov[X,Y]}{\sigma(X)\sigma(Y)}.$$

Note that $-1 \leqslant \rho(X,Y) \leqslant 1$.

Let $a, b \in \mathbb{R}$ be constants and $X, Y$ be two r.v. We then have few useful formulas:

- $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$ (linearity);
- $Var[aX \pm bY] = a^2 Var[X] + b^2 Var[Y] \pm 2ab Cov[X,Y]$;
- $Var[a] = 0$;
- $Cov[X,X] = Var[X]$.

## 11.4 Multivariate Distributions

We now work with random variables

$$\mathbf{X} : \Omega \to \mathbb{R}^K$$

where $K > 1$. For $K = 2$, we can write $\mathbf{X} = (X, Y)$.

**Definition 11.10.** *The **joint** distribution (CDF) of two random variables $X$ and $Y$ is*

$$F_{X,Y}(x,y) = P(X \leqslant x, Y \leqslant y) \text{ for all } (x,y) \in \mathbb{R}^2.$$

If the support of $(X, Y)$ is countable, we have the following definitions.

**Definition 11.11.** *The* ***joint PMF*** *of* $X, Y$ *is*

$$g_{X,Y}(x,y) = P(X = x, Y = y).$$

*The* ***marginal PMF*** *of* $X$ *is*

$$g_X(x) = \sum_y g_{X,Y}(x,y),$$

*while* ***conditional PMF*** *of* $X$ *given* $Y$ *is*

$$g_{X|Y}(x|y) = \frac{g_{X,Y}(x,y)}{g_Y(y)}$$

*and the corresponding* ***conditional CDF*** *is*

$$F_{X,Y}(x|y) = P(X \leqslant x|Y = y).$$

**Definition 11.12.** $X, Y$ *are independent if*

$$P(X = x, Y = y) = P(X = x)P(Y = y).$$

If $(X, Y)$ is absolutely continuous, we have the following definitions.

**Definition 11.13.** *The* ***joint PDF*** *of* $X, Y$ $f_{X,Y}$ *is such that*

$$F_{X,Y}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(x,y)dydx;$$

*and has the following properties:*

1. $f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y}$, *when* $F_{X,Y}$ *is differentiable,*
2. $f_{X,Y}(x,y) \geqslant 0$;
3. $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)dydx = 1$

*The* ***marginal PDF*** *of* $X$ *is*

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy.$$

*The* ***conditional PDF*** *of* $X$ *given* $Y$ *is*

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}, \text{ if } f_Y(y) > 0.$$

*and the corresponding* ***conditional CDF*** *is*

$$F_{X,Y}(x|y) = P(X \leqslant x|Y = y).$$

Note that if absolutely continuous $X, Y$ are independent r.v.s, then $f(x, y) = f_X(x)f_Y(y)$.

NOTE: $X, Y$ independent $\implies$ $X, Y$ are uncorrelated. The converse is not necessarily true, unless $(X, Y)$ follows a bivariate normal distribution.

The conditional distribution $F_{X|Y}$ can be interpreted as what is the distribution of $X$ when we know that $Y$ takes a certain value. It can be interpreted as the information that $Y$ gives about $X$. Note also that $X|Y$ is a random variable itself, with CDF $F_{X|Y}$.

Example:

In a population we have a distribution of *gender* and *height*. Assume that in the general population, we have that $height \sim N(5'7'', 1')$. However, for males, we have $height_{male} \sim N(5'10'', 1')$.

**Definition 11.14.** *The* ***conditional expectation*** *of $X$ given $Y$ is*

$$\mathbb{E}[X|Y = y] = \mathbb{E}[X|y] = \int_y x f_{X|Y}(x|y)dx.$$

The conditional expectation is very important in econometrics, as we are usually interested in the relationship between different economic variables. For instance, what is the expected wage of someone who has a graduate degree?

We denote by $\mathbb{E}[X|Y]$ the random variable

$$y \mapsto \mathbb{E}[X|Y = y].$$

**Theorem 11.3.** ***(Law of Iterated Expectations)*** *For two r.v. defined on the same probability space,*

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|Y]].$$

**Theorem 11.4.** ***(Jensen's Inequality)*** *Let $X$ be a r.v. and $h$ a convex function. Then,*

$$h(\mathbb{E}[X]) \leqslant \mathbb{E}[h(X)].$$

**Theorem 11.5.** ***(Convolution)*** *Let $X, Y$ be two independent r.v.s with PDFs $f_X$ and $f_Y$.*

*Then, the density funtion of $Z = X + Y$ is*

$$f_Z(z) = \int_x f_X(x) f_Y(z-x) dx.$$

# References

Mas-Colell, A., M. D. Whinston, J. R. Green, et al. (1995). *Microeconomic theory*, Volume 1. Oxford university press New York.

Rudin, W. (2021). Principles of mathematical analysis.