Lecture notes

Course: Preparatory course in Mathematics
Teaching hours: 15h
Instructor: Guillaume Pommey
Last updated: August 27, 2025

1. Preliminary notions

1.1. Set theory

Definition. One of the most elementary notion in mathematics is that of a **set.** A **set** is defined as a collection of distinct *elements.* These elements are typically mathematical objects such as numbers, symbols, functions, or even other sets.

A set is defined by the elements that compose it either by listing them or by describing their properties.

Example 1.1 *If the set A contains three elements, a, b, and c, we write:* $A = \{a, b, c\}$.

By convention, the curly brackets indicate that we are defining a set and each element is separated by a comma. We also usually prefer to use capital letters to denote a set.

Example 1.2 *If the set A contains all the natural numbers greater or equal than 8 we write* $A = \{x \in \mathbb{N} \mid x \geqslant 8\}$. *This notation reads: "among all x belonging to* $\mathbb{N}$ *take only those satisfying the property* $x \geqslant 8$*".*

We use the symbol $\in$ to say "belongs to". For instance, in the first example we can say that $a \in A$. We sometimes also the symbol $\notin$ to say that an element does not belong to a set.

A set can also contain other sets.

Example 1.3 *Define two sets* $A = \{a, b, c\}$ *and* $B = \{d, e\}$. *We can define* $C = \{A, B\} = \{\{a, b, c\}, \{d, e\}\}$. *Be careful, while A contains 3 elements and B 2 elements, the set C contains only two elements, namely* $\{a, b, c\}$ *and* $\{d, e\}$. *The elements of C are sets themselves and cannot be further "broken down".*

The elements of a set do not have to be "homogeneous", that is, elements of the same *nature.*

Example 1.4 *For instance* $C = \{\{a, b, c\}, d, e\}$. *In that case C contains the set* $A = \{a, b, c\}$ *and the elements d and e.*

Importantly, the order of the elements in the list describing a set is irrelevant. The set $A = \{a, b, c\}$ and the set $B = \{b, c, a\}$ are the same.

Subsets. A subset is a set of which all the elements are contained in another set B.

Definition 1.1 *Let A and B be two distinct sets. If every element in A is also in B we can state that A is a subset of B, denoted by* $A \subseteq B$*. Formally,*

$$\forall a \in A : a \in A \Rightarrow a \in B \Leftrightarrow A \subseteq B,$$

where $\subseteq$ *reads as "is included in".*

Example 1.5 *Let* $A = \{a,b,c\}$ *and* $B = \{a,b,c,d,e\}$*. Each element of A is also an element of B so that A is a subset of B.*

Example 1.6 *Let* $A = \{a,b,c,y\}$ *and* $B = \{a,b,c,z\}$*. Even if a, b, and c are both in A and B, it is clear that* $y \notin B$ *and so that A cannot be a subset of B. The same applies for B and* $z \notin A$*.*

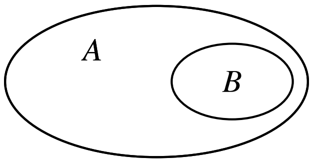


Figure 1: Euler diagram illustrating $A \subseteq B$.

Definition 1.2 *Two sets A and B are equal, if and only if A is a subset of B* ***and*** *B is a subset of A. Formally,* $A = B \Leftrightarrow A \subseteq B$ *and* $B \subseteq A$*.*

Definition 1.3 *We call the empty set* $\varnothing$*, the unique set which contains no element at all. By convention* $\varnothing \in A$ *for any set S.*

Example 1.7 *Define* $A = \{x \in \mathbb{R} \mid x^2 + 1 = 0\}$*. As it is clear that the equation* $x^2 + 1 = 0$ *has no solution in* $\mathbb{R}$*, then* $A = \varnothing$*.*

Complement. The *complement* of a set is a set that contains all the elements that are not in this set.

Definition 1.4 *Let X be a set and* $A \subseteq X$*. We define* A^c *the complement of set A as follows:*

$$A^c := \{x \in X \mid x \notin A\}.$$

Cardinality. Informally, the cardinality of a set is the number of elements contained in the set. The cardinality of a set A is usually denoted by $|A|$. In the case of a set A that contains a finite number of elements, we say that A is a *finite set* and $|A|$ is simply equal to the number of elements in A.

Example 1.8 *Let* $A = \{a,b,c,d,e\}$*, then* $|A| = 5$*.*

It is however possible that a set A contains an infinite number of elements like for instance if A is the set of all even numbers. In that case we say that A is an *infinite set.* The cardinality of an infinite set is a well-defined object but its investigation is beyond the level of this course.

1.2. Sets of numbers

Some *famous* sets are sets of numbers. Each of the following set describes a family of numbers that are regularly used.

- Set of **natural numbers**: $\mathbb{N} = \{0, 1, 2, 4, ...\}$.
- Set of **integers**: $\mathbb{Z} = \{0, \pm 1, \pm 2, \pm 4, ...\}$.
- Set of **rational numbers**: $\mathbb{Q} = \{a/b \mid a, b \in \mathbf{Z}, b \neq 0\}$.
- Set of **real numbers** $\mathbb{R}$. This set includes all rational numbers, together with all irrational numbers.

1.3. Set operations.

We now turn to defining basic operations on sets. Throughout this section, let X be a set, and A and B subsets of X.

Definition 1.5 *The **union** of a collection of sets is the set of all elements in the collection. The union between A and B is denoted by $A \cup B$ and satisfies*

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

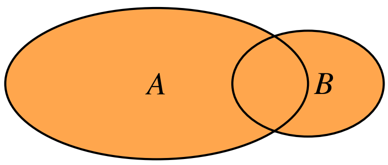


Figure 2: The orange region is the union of A and B.

Example 1.9 *Let $A = \{a, b, g\}$ and $B = \{a, c, g\}$. Then $A \cup B = \{a, b, c, g\}$.*

Definition 1.6 *The **intersection** of a collection of sets is the set of all elements that are common to all sets of the collection. The intersection between A and B is by $A \cap B$ and satisfies*

$$A \cup B = \{x \mid x \in A \text{ and } x \in B\}.$$

Example 1.10 *Let $A = \{a, b, g\}$ and $B = \{a, c, g\}$. Then $A \cap B = \{a, g\}$.*

Definition 1.7 *Two sets are said to be **disjoint** if*

$$A \cap B = \varnothing.$$

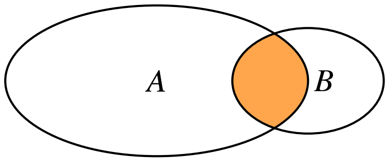


Figure 3: The orange region is the intersection between A and B.

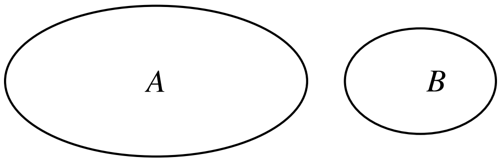


Figure 4: Example of two disjoint sets.

The operations of union and intersection have some basic properties. They are both *associative* as $A \cup (B \cup C) = (A \cup B) \cup C$ and $A \cap (B \cap C) = (A \cap B) \cap C$. They are both *commutative* as $A \cup B = B \cup A$ and $A \cap B = B \cap A$. Taking the union between a set and the empty set gives the set itself, $A \cup \varnothing = A$ while taking the intersection gives the empty set, $A \cap \varnothing = \varnothing$. Intersection distributes over union,

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

and union distributes over intersection,

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

Definition 1.8 *The* ***difference*** *between set A and set B is the set containing all the elements present in A but not in B. It is denoted by $A \backslash B$ and satisfies*

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

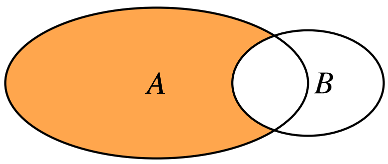


Figure 5: The orange region is the difference between A and B.

Example 1.11 *Let $A = \{a, b, g\}$ and $B = \{a, c, g\} \Rightarrow A \backslash B = \{b\}$.*

1.4. Some other useful definitions

Convexity. We define convexity for subsets of $\mathbb{R}$. Roughly speaking, a set $A \subseteq \mathbb{R}$ is *convex* if any weighted average of any two elements in A is also an element in A.

Definition 1.9 *A set $A \subseteq \mathbb{R}$ is convex if for all x, $y \in A$ and $\alpha \in [0,1]$ we have that $z = \alpha x + (1-\alpha)y \in A$.*

Cartesian product. In many cases, we need to pick several elements for various sets. If for instance $A = \{\text{milk, tea, coffee}\}$ is the set of available drinks and $B = \{\text{cereals, fruit, eggs}\}$ is the set of available items, we may want to define a breakfast as the pair composed by choosing one element in A and one element in B.

Definition 1.10 *The* ***Cartesian product*** *of two sets A and B, denoted $A \times B$, is the set of all the* ***ordered*** *pairs (a,b) where $a \in A$ and $b \in B$.*

Example 1.12 *Let $A = \{a_1, a_2\}$ and $B = \{b_1, b_2\}$, we have that*

$$A \times B = \{(a_1,b_1),(a_1,b_2),(a_2,b_1),(a_2,b_2)\}.$$

2. One Variable Calculus

2.1. Functions: definitions and properties

One of the main goals of Economics is to understand mechanisms, interactions and relationships between different variables. In many cases these relations can be described by **functions**.

Definition 2.1 *Let X and Y be two (nonempty) sets. A function f from a set X to a set Y is a correspondence associating to each element $x \in X$* ***at most*** *one element $y \in Y$.*

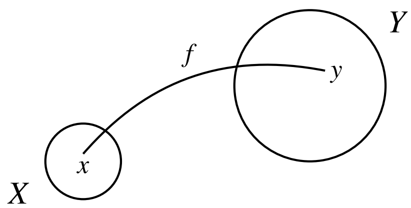


Figure 6: Representation of a function mapping and element of X into an element of Y.

By convention, we define a function with the notation $f : X \to Y$. The set of elements $x \in X$ to which f assigns an element in Y is called the **domain**, while the elements $y \in Y$ associated to x are called **images**. The set of all the images is called **range**. We sometimes denote the domain of f by dom f and the image img f.

Example 2.1 *Define $f : \mathbb{R} \to \mathbb{R}$ such that $f(x) = x^2 + 3$. We can evaluate the image of $x = 5$ under the function f as $f(5) = 5^2 + 3 = 28$.*

Example 2.2 *Let $f : \mathbb{R} \to \mathbb{R}$ where $f(x) = \frac{1}{x-2}$. Notice that the function is not well-defined at $x = 2$. the domain of f is therefore given by all the elements in $\mathbb{R} \backslash \{2\}$.*

The graph of $f : X \to Y$ is the set of ordered pairs (x,y) where $y = f(x)$. It can written as $G(f) := \{(x, f(x)) : x \in \mathrm{dom} f\}$.

Some functions can be classified as even or odd.

Definition 2.2 *The function f is **even** if*

$$f(-x) = f(x),\ \forall x \in X.$$

Example 2.3 *$f(x) = x^2 - 3$, since $(-x)^2 = x^2$, then $f(x) = f(-x)$.*

Definition 2.3 *The function f is **odd** if*

$$f(-x) = -f(x),\ \forall x \in X.$$

Example 2.4 *$f(x) = 3x^3$, since $(-x)^3 = -(x^3)$ then $f(-x) = -f(x)$.*

Some of the basic geometric properties of a function are whether it is increasing, decreasing, or constant. This notion can be true locally or globally (that is, for the entire domain).

Definition 2.4 *Let X, $Y \subseteq \mathbb{R}$. The function $f : X \to Y$ is increasing over the interval $S \subseteq X$ if for all $x_1, x_2 \in S$ such that $x_1 < x_2$ we have $f(x_1) \leqslant f(x_2)$.*

Similarly, we say that f is decreasing over the interval S if for all $x_1, x_2 \in S$ such that $x_1 < x_2$ we have $f(x_1) \geqslant f(x_2)$. Notice also that if f is increasing over S then $-f$ is decreasing over S.

We say f is an increasing (resp. decreasing) function if it is increasing (resp. decreasing) over its entire domain.

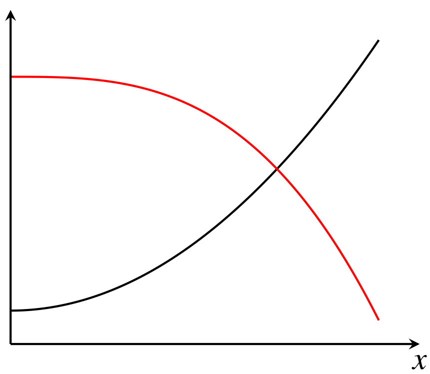


Figure 7: Example of an increasing function (black) and a decreasing function (red).

We can compose a function with another to produce a new function.

Definition 2.5 *Let $f : X \to Y$ and $g : Z \to X$. We can define a new function $h := f \circ g$ defined as $h(z) = f(g(z))$, where $z \in Z$. The new function h is a function $h : Z \to Y$.*

Example 2.5 *Let $f : \mathbb{R} \to \mathbb{R}$ defined as $f(x) = x + 1$ and $g : \mathbb{R} \to \mathbb{R}$ defined as $g(x) = x^2$. Define $h(x) = f(g(x)) = g(x) + 1 = x^2 + 1$. We can also define $k(x) = g(f(x)) = f(x)^2 = (x+1)^2$.*

The way a function maps elements from X into Y can be described more precisely by the notions of bijection, injection and surjection.

Definition 2.6 *A function $f : X \to Y$ is* ***injective****, also called a* ***one-to-one function****, if it maps distinct elements of its domain to distinct elements of its image. Formally, f is injective if for any x_1, $x_2 \in dom\ f$ such that $x_1 \neq x_2$ then $f(x_1) \neq f(x_2)$.*

Example 2.6 *Let $f : \mathbb{N} \to \mathbb{N}$ defined as $f(x) = x+1$. Clearly f is injective as for any every x_1, $x_2 \in \mathbb{N}$ where $x_1 \neq x_2$ we obviously have $x_1 + 1 \neq x_2 + 1$.*

Definition 2.7 *A function $f : X \to Y$ is* ***surjective****, also called an* ***onto function****, if for every element of its codomain there exists at least one element in the function's domain such that $f(x) = y$. Formally, f is surjective if for any $y \in img\ f$, there exists at least one $x \in dom\ f$ such that $f(x) = y$.*

Example 2.7 *Let $f : \mathbb{R} \to \mathbb{R}$ defined as $f(x) = |x|$ where $|\cdot|$ is the absolute value operator. We can say that f is surjective as for any $y \in \mathbb{R}$ we have that both $x_1 = y$ and $x_2 = -y$ give $f(x_1) = f(x_2) = y$.*

A function can be injective but not surjective and vice versa. For instance in the case of $f(x) = x+1$ where $X = Y = \mathbb{N}$, there is no $x \in X$ that yields $f(x) = 1$ $y = 1$

Notice however, that when f is either injective or surjective, there can be elements in the domain or in the image that are not associated to anything. In the case of $f(x) = x+1$ where $X = Y = \mathbb{N}$ it is clear that $y = 0$ is in $\mathbb{N}$ but it cannot be obtained from any $x \in \mathbb{N}$ under function f. Hence this f is injective but not surjective. Instead, $f(x) = |x|$ is surjective but not injective as every element of its image can be obtained from two distinct elements of its domain.

When a function is both injective and surjective, we say that it is bijective.

Definition 2.8 *A function $f : X \to Y$ is* ***bijective****, also called a* ***one-to-one correspondence****, if and only if it is both injective and surjective.*

A bijective function is a function that maps every element of its domain to exactly one element of its image, and vice versa.

Example 2.8 *Let $f : \mathbb{R}_+ \to \mathbb{R}$ defined as $f(x) = x^2$. Given that we are only considering nonnegative values of x, it is easy to see that each x is uniquely associated with an x^2 and reciprocally. Therefore this function is bijective. However, if we were to consider the same function $f(x) = x^2$ but defined over $X = Y = \mathbb{R}$ it would not be bijective anymore as it would fail to be injective.*

A useful construct is that of an *inverse function.*

Definition 2.9 *Let $f : X \to Y$ and assume that f is bijective. The inverse of the function f is denoted by f^{-1}. The inverse is a function $f^{-1} : Y \to X$ whose image $f^{-1}(y)$ returns the unique $x \in X$ such that $f(x) = y$.*

Intuitively, f^{-1} undoes what f does, that is applying the inverse function to the function yields the identity. Indeed, $(f^{-1} \circ f)(y) = f^{-1}(f(x)) = x$.

Example 2.9 *Let $f(x) = 2x$. Its inverse function is $f^{-1}(y) = y/2$. Indeed, starting from $x = 4$, we get $f(x) = 8$. In order to find which x produced 8 under the function we use the inverse $f^{-1}(8) = 8/2 = 4$.*

2.2. Common types of functions

Polynomial function. A polynomial is a map of the form $P(x) = a_n x^n + ... + a_1 x + a_0$ where n is the degree of polynomial and $a_0, \ldots, a_n$ are the polynomial coefficients.

Linear function. Polynomials of degree 1 are interesting functions, they are also called **linear functions**. All linear functions can be written as

$$f(x) = ax + b,$$

where $a, b \in \mathbb{R}$.

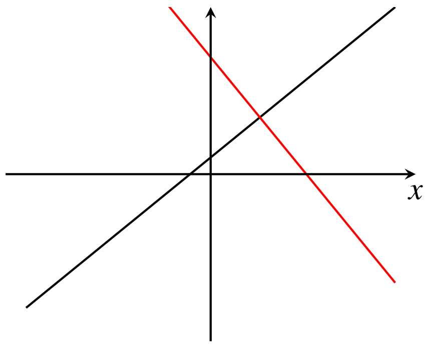


Figure 8: Examples of linear functions.

One of the main features which distinguishes two different lines is the **slope** (steepness) that is given by a. This function is increasing as $a > 0$ and decreasing if $a < 0$; if $a = 0$ the function degenerates to the constant function $f(x) = b$. The slope is given by the ratio of the growth in y ($y_2 - y_1$) and the growth of x ($x_2 - x_1$), that is $a = \frac{y_2 - y_1}{x_2 - x_1}$.

Quadratic function. A quadratic function is a polynomial of degree 2. We write quadratic functions as

$$f(x) = ax^2 + bx + c,$$

where $a, b, c \in \mathbb{R}$. The graph of a univariate quadratic function is a parabola whose axis of symmetry is parallel to the y-axis.

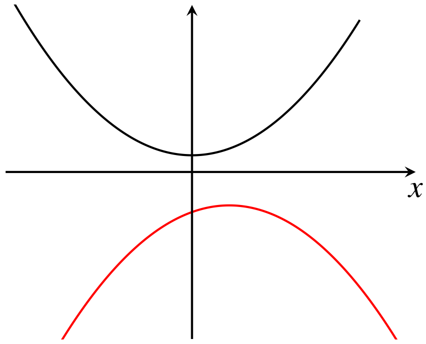


Figure 9: Examples of quadratic functions.

Exponential function. An **exponential function** is a function of the form

$$f(x) = a^x,$$

where $a > 0$.

Logarithmic function. The logarithm is the inverse of the exponential function.

$$f(x) = \log_a x.$$

The logarithm of a given number x is the exponent to which another fixed number, the base a, must be raised, to produce that number x. A particular case of the logarithm is the **natural logarithm** which has the number e (that is $e \approx 2.718$) as the base.

3. Limit and continuity

3.1. Limit

Roughly speaking, a function is said to have a limit l at point c if $f(x)$ gets closer and closer to l as x gets close to c. Before introducing the formal definition, we establish some notation.

Let $X, Y \subseteq \mathbb{R}$ and $f : X \to Y$. If f has a limit l at point c we write it as

$$\lim_{x \to c} f(x) = l.$$

In the most simplest cases, we can find the limit of a function at a point simply by evaluating the function at this point.

Example 3.1 *For instance,* $\lim_{x\to 2}(3x - x^2) = 3 \cdot 2 - 2^2 = 2$.

Example 3.2 *Another example is:* $\lim_{x\to 2} \ln(x-1) = \ln(2-1) = \ln(1) = 0$.

A function does not always have a limit, that is, converges to a finite number. In that cases, we use the notation $+\infty$ or $-\infty$ to denote *infinity*.

Example 3.3 *Evaluating* $\lim_{x \to 0} \frac{1}{x} = +\infty$ *shows that a function does not always have a limit.*

Sometimes we are interested in the limit of a function when x goes to $+\infty$ or $-\infty$.

Example 3.4 *For instance,* $\lim_{x \to +\infty} e^x = +\infty$.

We now introduce a formal definition of the limit of a function. This is not the unique definition, there exists other more general definitions but we the one we introduce will be enough for the purpose of this course.

Definition 3.1 *A function f has limit l as x approaches c if for all $\varepsilon > 0$, there exists a real $\delta > 0$ such that for all x, $|x - c| < \delta$ implies that $|f(x) - l| < \varepsilon$.*

In words, saying that $f(x)$ converges to l at c means that if we take x close enough to c we can always make $f(x)$ arbitrarily close to l.

So far, when we said that we were taking x closer and closer to c we did not specify *from which side.* Indeed, x could approach c by the *left* or by the *right*.

Definition 3.2 *The **right limit** of a function at point c is writes as*

$$\lim_{x \to c^+} f(x) = l,$$

and the left limit writes as

$$\lim_{x \to c^-} f(x) = l.$$

The $\lim_{x \to c} f(x) = l$ exists only if the right limit and the left limit exist and they are equal.

Example 3.5 *Consider the function $f(x) = \frac{1}{x-1}$ and we are interested in the behavior of f when x is close to 1. If we approach 1 from the right, that is $x \to 1^+$ then $x - 1 \to 0^+$, where 0^+ means "arbitrarily close to 0 and positive". If we instead approach 1 from the left, that $x \to 1^-$ then $x - 1 \to 0^-$, where 0^+ means "arbitrarily close to 0 and negative".*

It immediately follows that

$$\lim_{x \to 1^+} f(x) = +\infty \quad \textit{and} \quad \lim_{x \to 1^-} f(x) = -\infty.$$

As the right and left limit differ, we can conclude that f does not have a limit at 1.

Properties of limits. The limit of the sum of functions is equivalent to taking the sum of the limits.

$$\lim_{x \to c} [f(x) \pm g(x)] = \lim_{x \to c} f(x) \pm \lim_{x \to c} g(x)$$

The limit of the product of function is the product of the limits.

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

The limit of the ratio of function is the ratio of the limits.

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

Indeterminate forms. We may end up in a **indeterminate form**, that is, when we evaluate the limit of a function we encounter one of this four cases:

$$\frac{0}{0}, \ \frac{\infty}{\infty}, \ 0\times\infty, \ \infty-\infty.$$

In these cases, we have to work on the expression of the limit to try and remove the indeterminacy.

Example 3.6 *Consider evaluating* $f(x) = \frac{2x^2-3x^5}{x+2x^2}$ *when x goes to 0. It is clear that both the numerator and the denominator are equal to 0 when evaluated at* $x = 0$*. We therefore face the case* $0/0$*. Notice that we can rewrite the limit as follows:*

$$\lim_{x\to 0} \frac{2x^2-3x^5}{x+2x^2} = \lim_{x\to 0} \frac{x^2(2-3x^3)}{x(1+2x)} = \lim_{x\to 0} \frac{x(2-3x^3)}{1+2x}.$$

The numerator is still going to 0 but now the denominator converges to 1. Hence, this removes the indeterminacy and we can conclude that $\lim_{x\to 0} f(x) = 0$.

Example 3.7 *Consider now* $f(x) = \frac{2x^3+5x^2-x+7}{4x^3-x^2+x-3}$*. We want to evaluate the limit at* $x \to +\infty$*. As before we can factorize the function such that:*

$$\lim_{x\to+\infty} \frac{2x^3+5x^2-x+7}{4x^3-x^2+x-3} = \frac{2x^3(1+\frac{5}{2x}-\frac{1}{2x^2}+\frac{7}{2x^3})}{4x^3(1-\frac{1}{4x}+\frac{1}{4x^2}-\frac{3}{4x^3})} = \frac{1}{2}.$$

Example 3.8 *Finally consider evaluating* $f(x) = e^x - \sqrt{x}$ *at* $x \to +\infty$*. This yields the indeterminate case* $\infty - \infty$*. To provide an answer here, we need to compare whether* e^x *grows faster than x or not. We do not provide a proof but here this limit is:*

$$\lim_{x\to+\infty} e^x - \sqrt{x} = +\infty.$$

3.2. Continuity

Intuitively, a function is said to be continuous when slightly changing the input x induces only *small* variation of the output $f(x)$. In other words there are no *jumps*. Once again, several definitions exists but we will rely on the simplest one in terms of limits.

Definition 3.3 *A function $f : X \to Y$ is continuous at point $c \in X$ if*

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

3.3. Differentiation

Under some conditions, a function can be *differentiated* to obtain its *derivative.* The derivative of a function at a given point describes at which rate the function changes around this point.

Consider a function $f : X \to Y$, where $X, Y \subseteq \mathbb{R}$. We denote the absolute change of the function at x when it increases by an amount Δx as follows

$$\Delta f(x) = f(x_0 + \Delta x) - f(x_0).$$

To compute the rate of change at x we are interested in the ratio:

$$\frac{\Delta f(x)}{\Delta x} = \frac{f(x+\Delta x) - f(x)}{\Delta x}.$$

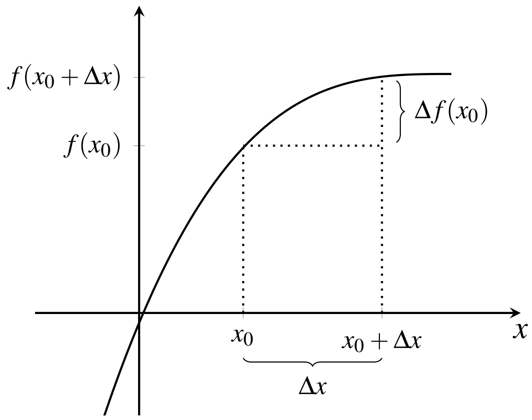


Figure 10: Graphical representation of the rate of change.

The derivative of f at point x is defined as the rate of change at x when Δx is arbitrarily close to zero. That is, we want to compute how the function changes when we increase the input by an infinitesimal amount.

Definition 3.4 *The derivative of $f : X \to Y$, $X = Y = \mathbb{R}$ at point x is denoted by $f'(x)$, or equivalently by, $\frac{\partial}{\partial x} f(x)$, and is defined by*

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

Naturally, the existence of the derivative relies on the existence of the limit of the function. If the limit of the function does not exist at a point x_0, we say that f is not differentiable at x_0. As we also defined continuity at x as the existence of the limit at x it means that continuity is a *necessary* condition for a function to be differentiable.

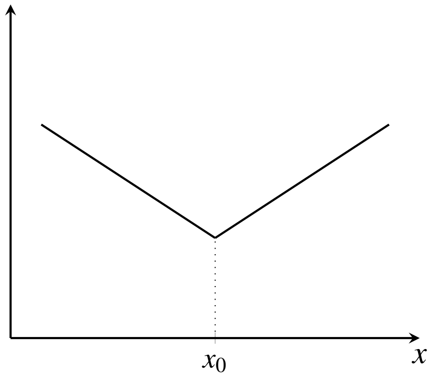


Figure 11: Continuous function not differentiable at x_0.

However, continuity is only a necessary condition and is not sufficient for differentiability. For instance, the function in Figure 11 is continuous at x_0 but is not differentiable. This happens when the *right* and *left* derivatives (defined similarly as the right and left limits) do not coincide.

In short "differentiable at x" implies "continuous at x" but not the converse.

Example 3.9 *Consider $f(x) = 2x$. We can compute its derivative f' by computing*

$$\lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x} = \lim_{\Delta x \to 0} \frac{2(x+\Delta x) - 2x}{\Delta x} = \lim_{\Delta x \to 0} \frac{2\Delta x}{\Delta x} = 2.$$

Hence the derivative is constant, as for any x we have $f'(x) = 2$.

Example 3.10 *Consider $f(x) = 3x^2 + 1$. We can compute its derivative f' by computing*

$$\begin{aligned}\lim_{\Delta x \to 0} \frac{[3(x+\Delta x)^2 + 1] - [3x^2 + 1]}{\Delta x} &= \lim_{\Delta x \to 0} \frac{3x^2 + 6x\Delta x + 2(\Delta x)^2 + 1 - 3x^2 - 1}{\Delta x} \\ &= \lim_{\Delta x \to 0} \frac{2\Delta x(\Delta x + 3x)}{\Delta x} = \lim_{\Delta x \to 0} 2(\Delta x + 3x) = 6x.\end{aligned}$$

In that case, the derivative has different values according to where it is evaluated.

Table of derivatives. It is useful to differentiate a function using a table of derivatives that lists the most common functions and their derivatives. Figure 12 lists some of the most usual ones.

Higher order derivatives In many cases, we are interested in differentiating a function more than one time. That is, we may want to differentiate its derivative or differentiate the derivative of the derivative. Starting with a function f, the *first-order derivative* f' is what we called the derivative so far. The *second order derivative* is simply $f'' := \frac{\partial}{\partial x} f'$, the *third order derivative* is $f''' := \frac{\partial}{\partial x} f''$, and so on. Generally, we denote the *nth order derivative* by $f^{(n)}(x)$.

We can also define the nth order derivative of f with the notation $\frac{\partial^n}{\partial x^n} f$. Hence we have that $f' = \frac{\partial^1}{\partial x^1} f$, $f'' = \frac{\partial^2}{\partial x^2} f$, and so on. We usually omit the 1 for the first order derivative for convenience.

Function f	Derivative f'	Function f	Derivative f'
a	0	uv	$u'v+uv'$
ax	a	$\frac{u}{v}$	$\frac{u'v-uv'}{v^2}$
x^a	ax^{a-1}	u^a	au'^{a-1}
a^x	$a^x \ln(a)$	$u(v(x))$	$v'u'(v)$
e^{kx}	ke^x	e^u	$u'e^u$
$\ln(ax)$	$\frac{a}{x}$	$\ln u$	$\frac{u'}{u}$

Figure 12: Table of common derivatives. We assume a is a constant, u and v are functions.

Example 3.11 *Differentiate once* $f(x) = \frac{4x-x^2}{x-5}$ *on* $\mathbb{R}\backslash\{5\}$. *We can use the formula to differentiate* u/v *where* $u(x) = 4x - x^2$ *and* $v(x) = x - 5$. *The first order derivative is therefore* $f'(x) = \frac{(4-2x)(x-5)-(4x-x^2)}{(x-5)^2} = \frac{-x^2+10x-20}{(x-5)^2}$.

Example 3.12 *Find all nth order derivatives of* $f(x) = e^{3x+1}$. *Let us first compute the first, second and third derivative:* $f'(x) = 3e^{3x+1}$, $f''(x) = 9e^{3x+1}$, *and* $f'''(x) = 27e^{3x+1}$. *Hence notice that* $f'(x) = 3f(x)$ *and therefore* $f''(x) = 3f'(x) = 3*3f(x)$ *and* $f'''(x) = 3*3f'(x) = 3*3*3f(x)$. *We can easily generalize* $\frac{\partial^n}{\partial x^n} f(x) = 3^n f(x)$.

Example 3.13 *Let* $f(x) = e^x$ *and* $g(x) = \frac{1-x}{x^2}$. *Find the first order derivative of the function* $h := f \circ g$. *We compute* $h(x) = e^{\frac{1-x}{x^2}}$ *and now we have that*

$$h'(x) = \frac{-x^2 - 2x(1-x)}{x^4} e^{\frac{1-x}{x^2}} = \frac{x^2 - 2x}{x^4} e^{\frac{1-x}{x^2}} = \frac{x-2}{x^3} e^{\frac{1-x}{x^2}}.$$

'Graphical' representation of derivatives. To better understand what the derivative of a function means, we can try to visualize it. Formally, it can be proved that the derivative of a function f at point x is precisely the slope of the tangent line to the curve passing through x.

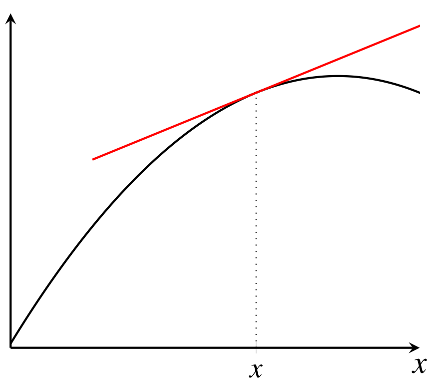


Figure 13: The straight red line is the tangent line to the black curve at point x.

The tangent line in Figure 13 that passes through x has slope $f'(x)$.

3.4. Usefulness of derivatives

Derivatives of a function provide information about the function behavior such as how it varies, how fast it varies, and how its variations vary.

First order derivative. The first order derivative directly informs us whether the function is increasing, decreasing, constant. A differentiable function $f : X \to Y$ is strictly increasing (resp. decreasing) on an interval $S \subseteq \mathbb{R}$ if it's first order derivative is strictly positive (resp. negative) on that interval. Similarly, f is constant on S if its derivative is null on S. Intuitively, a positive first derivative means that the rate of change of the function is positive, that is, when slightly increasing the input the function is also increasing.

Example 3.14 *Let $f(x) = 3x+1$. It is straightforward to see by inspection that f is an increasing function over all $\mathbb{R}$. Computing the derivative confirms this as $f'(x) = 3 > 0$.*

Example 3.15 *Let $f(x) = x^2$. Notice here that f is decreasing on $\mathbb{R}_-$ and increasing on $\mathbb{R}_+$. Indeed, the derivative is $f'(x) = 2x$ and is such that $f'(x) < 0$ for $x \in \mathbb{R}_-$ and $f'(x) > 0$ for $x \in \mathbb{R}_+$.*

In the two previous examples, we do not really need to compute the derivative to understand how the function varies. In some more complicated cases it can instead helps a lot.

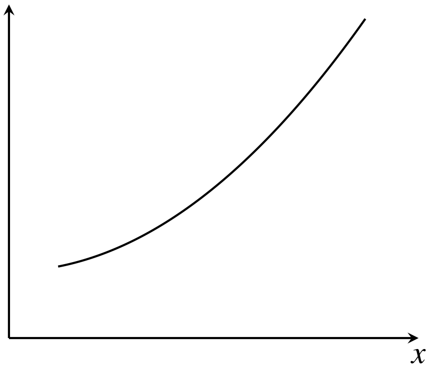
Example 3.16 *Let $f(x) = 3x^2 - 4x$. Here, it is a bit less obvious how f behaves. Take its derivative $f'(x) = 6x-4$. Then, we can simply notice that $f'(x) < 0$ whenever $6x < 4 \Leftarrow x < 2/3$ and $f'(x) \geqslant 0$ whenever $6x \geqslant 4 \Leftarrow x \geqslant 2/3$. We can therefore conclude that f is decreasing on $(-\infty, 2/3]$ and increasing on $[2/3, +\infty)$.*

The value of the derivative at a point also informs us about how *fast* the function varies. Take two points x and z, a function f whose derivative is $f'(x) = 2$ at x and $f'(z) = 1$ at z, is increasing both at both point but is increasing *at a faster rate* at x than at z.

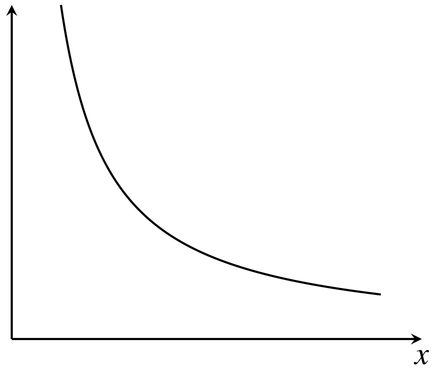
Second order derivative. By definition the second order derivative of a function f is *the derivative of the derivative* of f. Formally, if we let g be $g := f'$ then $g' := f''$, that is, we can say that the second order derivative of f is the first order derivative of f'.

We can therefore make use of the second order derivative to deduce whether the first order derivative is increasing, decreasing or constant using the same rules that we saw previously. However, we have to be careful as a proper interpretation of the second order derivative depends on whether the function is increasing or decreasing.

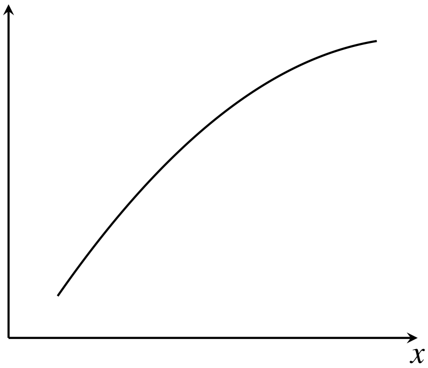
We have essentially four possible combinations for the sign of (f', f''). Let us denote the four possible cases $(+,+)$, $(+,-)$, $(-,+)$ and $(-,-)$, where for instance $(-,+)$ means that f' is negative and f'' is positive. In the case $(+,+)$, the function is increasing and it is increasing at an increasing rate, that is, it is increasing and increasing faster and faster. Instead, in the case $(+,-)$ the function is also increasing but at a decreasing rate, that is, increasing but less and less fast. Be



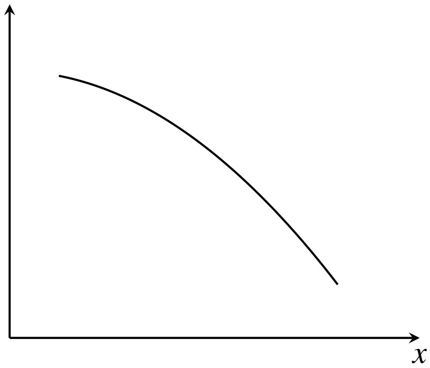
(a) $f' > 0$, $f'' > 0$.



(b) $f' < 0$, $f'' > 0$.



(c) $f' > 0$, $f'' < 0$.



(d) $f' < 0$, $f'' < 0$.

Figure 14: Examples of increasing and decreasing convex and concave functions.

careful, for decreasing functions we have to interpret the sign of the second order derivative in the opposite way. Indeed, $(-,+)$ corresponds to a decreasing function that decreases less and less fast while $(-,-)$ corresponds to a decreasing function decreasing faster and faster.

Figure 14 illustrates all four cases.

Convexity and concavity. We have already seen the notion of convexity in the case of sets. In the case of function, convexity (and concavity) is a notion of *curvature* of the function. The definition of a convex function is as follows.

Definition 3.5 *A function $f : \mathbb{R} \to \mathbb{R}$ is convex on an interval $S \subseteq \mathbb{R}$ if for all x, $y \in S$ and all $\alpha \in [0,1]$ we have that*

$$f(\alpha x + (1-\alpha)y) \leqslant \alpha f(x) + (1-\alpha) f(y).$$

Intuitively, a function is convex when the image of the weighted average $\alpha x + (1-\alpha)y$ is lower than the weighted average of the images of x and y.

Definition 3.6 *A function $f : \mathbb{R} \to \mathbb{R}$ is concave on an interval $S \subseteq \mathbb{R}$ if for all x, $y \in S$ and all $\alpha \in [0,1]$ we have that*

$$f(\alpha x + (1-\alpha)y) \geqslant \alpha f(x) + (1-\alpha) f(y).$$

Concavity can be understood as the *opposite* of convexity. Notice that a function that is both convex and concave is linear.

Sometimes, the above definition may prove difficult (or long) to apply. When a function f is differentiable, the second order derivative fully characterizes convexity and concavity.

If a twice-differentiable function f has a positive second order derivative then is it convex. If it has a negative second order derivative it is concave. And if the second order derivative is null, the function is linear.

Example 3.17 *Consider $f(x) = x^2$ for $x \in \mathbb{R}$. It is easy to compute $f'(x) = 2x$ and $f''(x) = 2 > 0$. Hence the function is convex on $\mathbb{R}$.*

Example 3.18 *Consider $f(x) = x^3 - 2x^2$ on $S = \mathbb{R}_+$. We have $f'(x) = 3x^2 - 4x$ and $f''(x) = 6x - 4$. We can immediately conclude that f is concave on $[0, 2/3]$ and convex on $[2/3, +\infty)$.*

Example 3.19 *Consider $f(x) = 5x + 20$ on $\mathbb{R}$. We have $f'(x) = 5$ and $f''(x) = 0$. Hence the function is linear.*

3.5. Maximum and minimum of a function

The maximum and the minimum of a function are called the extrema of the function. Throughout this section we consider functions of the type $f : X \to \mathbb{R}$, where X can be any finite or infinite set.

The maximum and minimum of a function, if they exist, can be either *local* or *global.* A function has a global maximum at a point x^* when the value of the function at this point is greater than the value of the function at any other point of its domain. Instead, a function has a local maximum at a point $\tilde{x}$ when the value of the function at this point is greater than the value of the function at any other point in a *neighborhood* of point $\tilde{x}$.

Global maximum and minimum are formally defined as follows.

Definition 3.7 *Let $f : X \to \mathbb{R}$. The function f has a* ***global maximum*** *at $x^* \in X$ if*

$$f(x^*) \geqslant f(x), \text{ for all } x \in X.$$

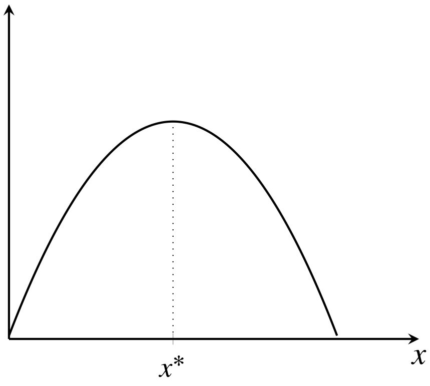
Definition 3.8 *Let $f : X \to \mathbb{R}$. The function f has a* ***global minimum*** *at $x^* \in X$ if*

$$f(x^*) \leqslant f(x), \text{ for all } x \in X.$$

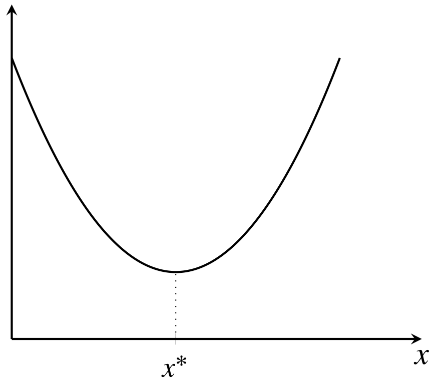
Local maximum and minimum are defined as follows.

Definition 3.9 *Let $f : X \to \mathbb{R}$. The function f has a* ***local maximum*** *at $\tilde{x} \in X$ if there exists an $\varepsilon > 0$ such that $f(\tilde{x}) \geqslant f(x)$ whenever $|x - \tilde{x}| < \varepsilon$.*

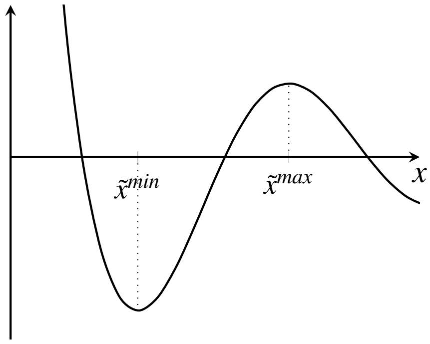
Definition 3.10 *Let $f : X \to \mathbb{R}$. The function f has a* ***local minimum*** *at $\tilde{x} \in X$ if there exists an $\varepsilon > 0$ such that $f(\tilde{x}) \leqslant f(x)$ whenever $|x - \tilde{x}| < \varepsilon$.*



(a) Global maximum



(b) Global minimum



(c) Local maximum and local minimum.

Figure 15: Examples of global and local extrema.

Notice that a global maximum (resp minimum) is also a local maximum (resp. minimum). The converse is not true however.

First-order conditions. When a function is differentiable there is a very simple way to find its extrema. Assume $f : X \to \mathbb{R}$ is differentiable, then at any local or global maximum or minimum x^* we must have that $f'(x^*) = 0$. This condition is often called the *first-order condition* for an extreme point. Be careful, we only state that if a point x^* is an extreme then $f'(x^*) = 0$. Therefore the first-order condition is a *necessary* condition for an extreme point but not a *sufficient* one.

Figure 16 illustrates the case of a global maximum at point x^*. As we have seen earlier, the derivative of a function in an point corresponds to the slope of the tangent line at this point. Hence, the first-order condition $f'(x^*) = 0$ corresponds to the case in which the tangent line is "flat", that is, has slope zero.

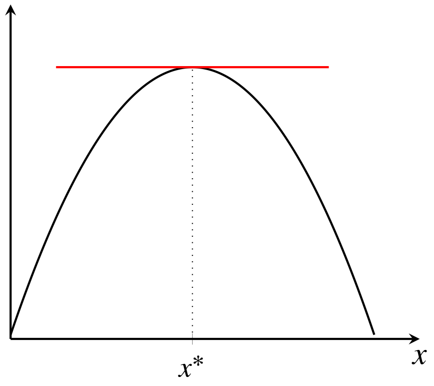


Figure 16: Graphical representation of the first-order condition.

(Advanced) Intuitive proof. It is easy to provide an intuitive proof of the statement that for any differentiable function, its derivative must be null at any type of extreme point. We show that it is the case of a local maximum – hence we implicitly show that it is also true for a global maximum. The case of a local minima can be easily found by symmetry.

Assume that x^* is a local maximum of f. By definition, this implies that there exists an $\varepsilon > 0$ such that $f(x^*) \geqslant f(x)$ for all $x \in X$ such that $|x - x^*| < \varepsilon$. Now assume that $f'(x^*) > 0$. This means that moving from x^* to $x^* + \Delta x$, where $\Delta x \to 0^+$ is making the function greater, that is $f(x^* + \Delta x) > f(x^*)$. As we can make Δx arbitrarily small, we can always find one such that $|x^* + \Delta x - x^*| < \varepsilon$, therefore contradicting the assumption that x^* is a local maximum. Assume instead that $f'(x^*) < 0$. Now if we were to take $x^* - \Delta x$ with $\Delta x \to 0^+$ we would also be able to find a greater value at $f(x^* - \Delta x)$ than at $f(x^*)$, contradicting once again the assumption that x^* is a local maximum. Therefore, we must have that $f'(x^*) = 0$.

Second-order conditions. As pointed before, first-order conditions are only necessary conditions to identify an extremum. Figure 17 illustrates the case in which the first-order condition is satisfied although the point is not an extrema of the function. Indeed at $\tilde{x}$ notice that we have that $f'(\tilde{x}) = 0$ (slope of the tangent line is null) so that the first-order condition is satisfied at $\tilde{x}$. However we can clearly see that this point is neither a maximum nor a minimum, we say that it is a *saddle point*.

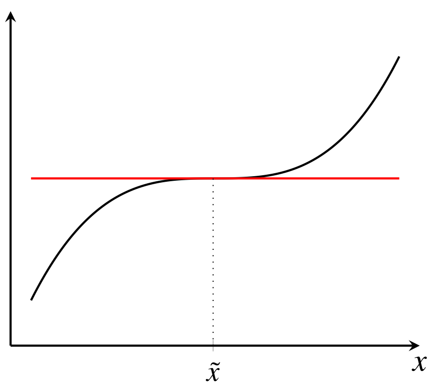


Figure 17: Graphical representation of the first-order condition.

When the function is twice differentiable, we can make use of the so-called *second-order conditions.* They are *sufficient* conditions that can help identify whether a point satisfying the first-order condition is indeed a maximum or a minimum, or if it is a saddle point.

Type	First-order condition	Second-order condition
Local maximum	$f'(x^*) = 0$	$f''(x^*) > 0$
Local minimum	$f'(x^*) = 0$	$f''(x^*) < 0$
Saddle point	$f'(x^*) = 0$	$f''(x^*) = 0$
Global maximum	$f'(x^*) = 0$	$f''(x) < 0$, for all x
Global minimum	$f'(x^*) = 0$	$f''(x^*) > 0$, for all x

Recall that the sign of the second-order derivative indicates whether the function is (locally/-globally) convex or concave and that these properties help to identify the nature of our candidate extremal points.

Example 3.20 *Let $f(x) = 3-(x-2)^2$. It is easy to compute that $f'(x) = -2(x-2)$ so that the only solution to the first-order condition is $x^* = 2$. We also have that $f''(x) = -2 < 0$ meaning that our function is globally concave. We can conclude that $x^* = 2$ is a global maximum of f.*

Example 3.21 *Let $f(x) = (x-1)^3 + 1$. The first and second order derivative respectively write as $f'(x) = 3(x-1)^2$ and $f''(x) = 6(x-1)$. The only point satisfying the first-order condition $f'(x^*) = 0$ is $x^* = 1$. Notice also that $f''(x^*) = 0$. Hence, x^* is a saddle point.*

4. Integral

We have seen how to capture the rate of change of a function thanks to its derivative. We now proceed to the "inverse" operation, the antiderivative.

Definition 4.1 *A function F is an antiderivative of the function f if for all x in the domain of f we have that*

$$F'(x) = f(x).$$

In other words, differentiating F gives f. Notice that the antiderivative F is not unique as any other function $G(x) := F(x) + c$, where $c \in \mathbb{R}$ is also an antiderivative of $f(x)$, given that differentiating a constant term always yields zero.

Example 4.1 *The function $F(x) = x^2$ is an antiderivative of the function $f(x) = 2x$ as we obviously have that $F'(x) = f(x)$. The function $G(x) = x^2 + 3$ and $H(x) = x^2 + 12$ also are antiderivatives of f.*

Definition 4.2 *Let f be a function with domain (a,b). The class of functions whose derivative is f is denoted by*

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

We call $\int f(x)dx$ the *indefinite integral* of f, where $\int$ is the *integral sign* and f is said to be the *integrand.* The symbol dx indicates with respect to which variable integration is performed. When there is no ambiguity about the variable of integration we sometimes simply write the integral $\int f$.

Linearity. The integral is a **linear operator**. Take two functions f and g whose antiderivatives are denoted by F and G, respectively, we have that:

$$\begin{aligned}\int(\alpha f(x)+\beta g(x))dx &= \alpha\int f(x)dx+\beta\int g(x)dx\\ &= \alpha(F(x)+c_1)+\beta(G(x)+c_2)\\ &= \alpha F(x)+\beta G(x)+c,\ \ c\in\mathbb{R}.\end{aligned}$$

In words, the integral of the weighted sum of f and g is equal to the weighted sum of their respective antiderivatives modulo a constant term $c \in \mathbb{R}$.

Common antiderivatives. As for derivatives, it is useful to know the most common antiderivatives summarized in Figure 18.

Indefinite integral	Antiderivative
$\int x^\alpha dx$	$\frac{x^{\alpha+1}}{\alpha+1},\ \alpha\neq-1$
$\int \frac{1}{x}dx$	$\ln\lvert x\rvert+c$
$\int[f(x)]^\alpha f'(x)dx$	$\frac{[f(x)]^{\alpha+1}}{\alpha+1}+c,\ \alpha\neq-1$
$\int\frac{f'(x)}{f(x)}dx$	$\ln\lvert f(x)\rvert+c$
$\int e^{f(x)}f'(x)dx$	$e^{f(x)}+c$
$\int e^x dx$	e^x+c

Figure 18: Table of common antiderivatives. We assume that c and α are constants in $\mathbb{R}$.

Example 4.2 *Let $f(x)=2xe^{x^2}$. It is easy to see that $F(x)=e^{x^2}$ is an antiderivative of f.*

Example 4.3 *Let $f(x)=\frac{4x^3}{x^4}$. We have that $F(x)=\ln x^4$.*

Definite integral. In many cases, we are interested in evaluating an integral over an interval (a,b).

Definition 4.3 *The definite integral of f over interval (a,b) is noted by*

$$\int_a^b f(x)dx.$$

We call a and b the lower and upper limits of integration, respectively. Roughly speaking, $\int_a^b f(x)dx$ can be seen as the 'sum' of all the values f over the interval (a,b).

The following properties are important to remember.

1. $\int_a^b f(x)dx = -\int_b^a f(x)dx$.
2. $\int_a^a f(x)dx = 0$.
3. $\int_a^b f(x)dx = F(b) - F(a)$.

The last property shows that if one knows the antiderivative F of function f then evaluating the integral of f over (a,b) is equivalent to evaluate the difference between $F(b)$ and $F(a)$.

Example 4.4 *Assume $f(x) = 3x$. Let us evaluate $\int_1^2 f(x)dx$. First, we find that an antiderivative is $F(x) = \frac{3}{2}x^2$. Hence we can simply compute that $\int_1^2 f(x)dx = \frac{3}{2}2^2 - \frac{3}{2}1^2 = 4.5$.*

Integration by parts. In some cases, it is not possible to find an antiderivative using the common rules for antiderivatives. Integration by parts is a method can sometimes help to solve the problem.

Definition 4.4 *Take two differentiable functions f and g. The following holds:*

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

Example 4.5 *Consider the function $h(x) = xe^{2x}$. At first sight, there is no obvious way to apply one of the common rule to find an antiderivative. However, notice that we can define $f(x) = x$ and $g'(x) = e^{2x}$ so that evaluating $\int_a^b h(x)dx$ is equivalent to evaluating $\int_a^b f(x)g'(x)dx$.*

It straightforward to compute that $f'(x) = 1$ and that $g(x) = \frac{1}{2}e^{2x}$. Hence we can easily compute that $\int_a^b f'(x)g(x)dx = \int_a^b 1 * \frac{1}{2}e^{2x}dx$. The antiderivative G of g is $G(x) = \frac{1}{4}e^{2x}$ so that $\int_a^b f'(x)g(x)dx = G(b) - G(a) = \frac{1}{4}e^{2b} - \frac{1}{4}e^{2a}$.

Let us evaluate h on $(0,1)$ for instance:

$$\begin{aligned}\int_0^1 h(x)dx = \int_0^1 f(x)g'(x)dx &= f(1)g(1) - f(0)g(0) - \int_0^1 f'(x)g(x)dx \\ &= e^2 - 0 - \frac{1}{4}e^2 - \frac{1}{4} \\ &= \frac{3}{4}e^2 - \frac{1}{4}.\end{aligned}$$

5. Vectors and matrices

5.1. Vectors

In many cases, we need to represent objects that cannot be characterized by a single scalar. For instance, we need three scalars to represent a point in a three-dimensional space or we may need n scalars to represent a system of prices in an economy.

A *vector*, sometimes called a *tuple*, is an ordered sequence of scalars $(x_1, x_2, \ldots, x_n)$ whose *size* is defined by the number of elements it contains (Note: in reality, a vector is not necessarily a sequence of scalars, it can also be a sequence of functions for instance). For instance, $(3,2)$ and $(0,1,7)$ are vectors of size 2 and 3, respectively.

Definition 5.1 *A vector x of size n is defined as $x := (x_1, \ldots, x_n)$ where each x_i corresponds to the i-th element (or coordinate) of the vector.*

We will work with *real vectors*, that is, vectors whose elements are real numbers. Hence, we define a real vector as $x \in \mathbb{R}^n$, that is, $x = (x_1, \ldots, x_n)$ where $x_i \in \mathbb{R}$ for all $i = 1, \ldots, n$.

Vector operations. Vectors can be manipulated with operations such as *equality*, *addition*, *multiplication*.

Definition 5.2 *Two vectors $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$ are said to be equal, $x = y$, if and only if $x_i = y_i$ for all $i = 1, \ldots, n$.*

That is, two vectors are equal if and only if all their elements are equal.

Definition 5.3 *The sum of two vectors $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$, $x + y$, produces a vector $z \in \mathbb{R}^n$ defined as*

$$z = x + y = (x_1 + y_1, \ldots, x_n + y_n).$$

Summing two initial vectors of size n therefore means creating a new vector of size n for which each i-th element is the sum of the i-th elements of each initial vector.

Example 5.1 *The sum of $x = (1,2)$ and $y = (3,4)$ is defined as $z = (1+3, 2+4) = (4,6)$. We can also sum three, four of k vectors at the same time, for instance, $(1,1) + (2,2) + (3,3) = (6,6)$.*

Notice that both for equation or adding vectors we need to take vectors of the same size. We cannot add $(2,3)$ and $(1,1,2)$ nor we can say anything about whether they are equal or not.

Definition 5.4 *Let $x \in \mathbb{R}^n$ be a vector and $a \in \mathbb{R}$ a scalar. Scalar multiplication is defined as ax and produces a vector $z \in \mathbb{R}^n$ such that*

$$z = ax + y = (ax_1, \ldots, ax_n).$$

Hence, multiplying a vector of size n by a scalar amounts to multiply each element of the vector by the scalar.

Example 5.2 *Consider* $x = (1,2,3)$ *and* $a = 5$. *Then we have that* $ax = (5*1, 5*2, 5*3) = (5,10,15)$.

We can also define the **linear combination** as the operation that consists in summing several vectors each of them weighted by a (potentially) different scalar.

Definition 5.5 *Let* $x^{(1)}, \ldots, x^{(k)}$ *a sequence of vectors in* $\mathbb{R}^n$ *and* $a_1, \ldots, a_k$ *scalars in* $\mathbb{R}$. *The linear combination is defined as a vector* $z \in \mathbb{R}^n$ *such that*

$$z = \sum_{i=1}^{k} a_k x^{(k)}.$$

Example 5.3 *Consider* $x = (1,2)$, $y = (3,4)$ *and coefficients* $a_1 = 5$, $a_2 = 6$. *The linear combination of these gives* $a_1 x + a_2 y = (5,10) + (18,24) = (23,34)$.

Definition 5.6 *A sequence of vectors* $(x^{(1)}, x^{(2)}, \ldots, x^{(k)})$ *is said to be linearly **dependent**, if there exist scalars* $a_1, a_2, \ldots, a_k$, *not all zero, such that*

$$a_1 x^1 + a_2 x^2 + \cdots + a_k x^k = 0.$$

*On the other hand, two vectors are said to be linearly **independent** if there is no linear combination which gives as a result the null vector (except for the trivial case* $a_1 = a_2 = \cdots = a_k = 0$*).*

5.2. Matrices

A **matrix** is a rectangular array of numbers. The size of a matrix is indicated by the number of its rows and number of its columns. A matrix with k rows and n columns is called a $k \times n$ matrix. The element in row i and column j is called the (i, j)th entry, and it is often written as a_{ij}. A matrix with the number of columns equal to the number of rows is called **square matrix**.

$$A_{k,n} = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k,1} & a_{k,2} & \cdots & a_{k,n} \end{pmatrix}.$$

Matrices operations. We can also define addition, multiplication, and multiplication by a scalar for matrices.

Definition 5.7 *Let A and B two $k \times n$ matrices. Their sum is the matrix C whose elements are* $c_{ij} = a_{ij} + b_{ij}\ \forall i, j$

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & a_{ij} & \vdots \\ a_{k1} & \cdots & a_{kn} \end{pmatrix} + \begin{pmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & b_{ij} & \vdots \\ b_{k1} & \cdots & b_{kn} \end{pmatrix} = \begin{pmatrix} a_{11}+b_{11} & \cdots & a_{1n}+b_{1n} \\ \vdots & a_{ij}+b_{ij} & \vdots \\ a_{k1}+b_{k1} & \cdots & a_{kn}+b_{kn} \end{pmatrix}.$$

Matrices may be multiplied by scalars. This operation is called **scalar multiplication**. More generally:

Definition 5.8 *The product of the matrix A and the number α, denoted αA, is the matrix whose elements are* $\alpha a_{ij}\ \forall i, j$

$$\alpha \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & a_{ij} & \vdots \\ a_{k1} & \cdots & a_{kn} \end{pmatrix} = \begin{pmatrix} \alpha a_{11} & \cdots & \alpha a_{1n} \\ \vdots & \alpha a_{ij} & \vdots \\ \alpha a_{k1} & \cdots & \alpha a_{kn} \end{pmatrix}.$$

We can define the **matrix product** AB if and only if

number of columns of A = number of rows of B.

To obtain the $(i, j-th)$ entry of AB, multiply the i-th row of A and the j-th column of B as follows

$$\begin{pmatrix} a_{i1} & a_{i2} & \cdots & a_{im} \end{pmatrix} \cdot \begin{pmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{ni} \end{pmatrix} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{im}b_{mj}.$$

In other words, the (i, j)-th entry of the product AB is defined to be

$$\sum_{n-1}^{m} a_{in}b_{hj}.$$

If A is a $k \times m$ and B is $m \times n$, then the product $C = AB$ will be $k \times n$.

Usually, $AB \neq BA$.

The $n \times n$ matrix

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

with $a_{ii} = 1,\ \forall i$ and $a_{ij} = 0,\ \forall i \neq j$, has the following property

$$AI = A.$$

for any $m \times n$ matrix A. I is called **identity matrix**.

Definition 5.9 *The **transpose** of a $k \times n$ matrix A is the $n \times k$ matrix obtained by interchanging the rows and the columns.*

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix}^T = \begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \\ a_{13} & a_{23} \end{pmatrix}.$$

Definition 5.10 *A **Triangular matrix** is a square matrix containing element different from zero only above/below the main diagonal.*

Example 5.4 Upper Triangular

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

Example 5.5 Lower Triangular

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

Definition 5.11 *To any square matrix is associated a **determinant**, $|A|$. From a geometric point of view, it represents the area (volume) of the parallelogram generated by the vectors of the matrix. Let A be a 2×2 matrix,*

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}.$$

the determinant is $|A| = a_{11}a_{22} - a_{12}a_{21}$.

In case of a $n \times n$ matrix with $n > 3$, in order to find the determinant, we can use the **minor** of the matrix.

Definition 5.12 *A minor of a matrix A is the determinant of some smaller square matrix, cut down from A by removing one or more of its rows and columns. Minors obtained by removing just one row and one column from square matrices (first minors) are required for calculating matrix cofactors, which in turn are useful for computing both the determinant and inverse of square matrices.*

The cofactor is:

$$A_{ik} = (-1)^{i+k} M_{ik}.$$

where M_{ik} is a minor of the matrix.

*Therefore, using the **Laplace theorem** we can obtain the determinant of a matrix $n \times n$ as the sum of the product, of any row or column, by their cofactor.*

The **rank** of A is the largest order of any non-zero minor in A.

Definition 5.13 *Let A be a $n \times n$ matrix. The $n \times n$ matrix A^{-1} is an **inverse** for A if $AA^{-1} =$*

$A^{-1}A = I_n$. A matrix can have at most one inverse and not every matrix is invertible. In order to be invertible, we need $|A| \neq 0$.

Definition 5.14 *Any symmetric A is:*

- ***Positive semidefinite*** *if $x'Ax > 0$, $\forall x \neq 0$;*
- ***Positive definite*** *if $x'Ax \geqslant 0$, $\forall x \neq 0$;*
- ***Negative definite*** *if $x'Ax < 0$, $\forall x \neq 0$;*
- ***Negative semidefinite*** *if $x'Ax \leqslant 0$, $\forall x \neq 0$.*

5.3. Linear system of equations

In several instances, one might have to study system of equations. We will focus on linear systems of equations.

Definition 5.15 *Generally, an equation is said to be linear if it has the form*

$$a_1x_1 + a_2x_2 + ... + a_nx_n = b$$

where $a_1, ..., a_n$ are **parameters** and $x_1, ..., x_n$ are **variables**.

A solution to the linear equation is a vector $(x_1, ..., x_n)$ such the above equality holds.

Example 5.6 *Consider the linear equation $5x_1 + 10x_2 = 0$. It is easy to see that $(0, 0)$ (trivially) solves this equation. The vectors $(1, -2)$ and $(2, -4)$ are also solutions to this equation.*

When there is more than one (linear) equation, we say we have to deal with a system of (linear) equations. A solution to the system is a vector $(x_1, ..., x_n)$ such that **all** the equations in the system hold simultaneously.

Example 5.7 *Consider the system of two linear equations $x_1 + x_2 = 5$ **and** $2x_1 - x_2 = 1$. Clearly $(3, 2)$ is a not a solution to this system as it solves only the first equation. Neither is $(4, 7)$ a solution as it solves only the second equation. Instead $(2, 3)$ is a solution to the system as it simultaneously solves the first and the second equations.*

Matrix notation and operations can be conveniently used to solve linear systems of equations.

Example 5.8

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + ...a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + ...a_{2n}x_n = b_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + ...a_{mn}x_n = b_m \end{cases}$$

The solution of the system is given by $(x_1, x_2, ..., x_n)$ which solves all the equations contemporaneously. It can be expressed much more compactly using matrix notations. Let A denote the coefficient matrix of the system:

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & a_{ij} & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}.$$

Also, let

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \quad \text{and} \quad \mathbf{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}.$$

Then, the system of equations can be written as

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & a_{ij} & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}.$$

or simply as

$$A\mathbf{x} = \mathbf{b}.$$

Then, if A is *nonsingular* ($|A| \neq 0$), we can solve the system as $\mathbf{x} = A^{-1}\mathbf{b}$. To solve a linear system of simultaneous equations we can use also the **Cramer's rule**. If the matrix A is nonsingular, the linear system of system of n linear equations and n unknowns. Then the theorem states that in this case the system has a unique solution, whose individual values for the unknowns are given by:

$$x_i = \frac{det(A_i)}{det(A)} \quad i = 1, ..., n.$$

where A_i is the matrix formed by replacing the ith column of A by the column vector b.

Example 5.9 *Solve the following system*

$$\begin{cases} x - 2y + z = 1 \\ 3x + y - 7z = 0 \\ x - z = 1. \end{cases}$$

Since $|A| \neq 0$*, we can use Cramer's rule*

$$x = \frac{det\begin{pmatrix} 1 & -2 & 1 \\ 0 & 1 & -7 \\ 1 & 0 & -1 \end{pmatrix}}{6} = 2$$

$$y = \frac{det\begin{pmatrix} 1 & 1 & 1 \\ 3 & 0 & -7 \\ 1 & 1 & -1 \end{pmatrix}}{6} = 1$$

$$z = \frac{det\begin{pmatrix} 1 & -2 & 1 \\ 3 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}}{6} = 1.$$

In economics, we may be interested in system of the following form

$$Ax = \lambda x,$$

where A is a square matrix. It is equivalent to write that

$$(A - \lambda I)x = 0.$$

Definition 5.16 *The values* λ *that solve* $det(A - \lambda I) = 0$ *are called* ***eigenvalues***.

Definition 5.17 *While the non-trivial vectors,* $\mathbf{x}$*, obtained as the solution of* $(A-\lambda I)x=0$ *is called* ***eigenvector****.*

From a geometric point of view, the eigenvector (x) is the vector which is only scaled by a value λ when we apply to it a transformation A.

6. Multivariable calculus

We have previously investigated function of one variable and many of their properties and applications. In many cases, however, functions may depend on several variables. Formally, we are investigating cases in which a function f is defined as $f:\mathbb{R}^n\to\mathbb{R}$. Such a function is called a multivariate function, and takes as an input a vector $x\in\mathbb{R}^n$ and outputs a scalar in $\mathbb{R}$.

Example 6.1 *Consider the two-dimensional function* $f:\mathbb{R}^2\to\mathbb{R}$ *defined as* $f(x,y)=xy$*. Then we simply have that* $f(1,2)=1*2=2$*, or* $f(4,5)=4*5=20$*.*

6.1. Differentiation

The concept of differentiation can be extended to function of several variables. There are two main ways to differentiate a function of several variables: **partial differentiation** and **total differentiation.** Partial differentiation considers how the function changes with respect to one variable while holding others constant, while total differentiation considers the overall change of a function with respect to a variable, taking into account how all other variables,

Partial differentiation. Partially differentiating a function is very close to the differentiation process of a unidimensional function. There are as many partial derivatives as there are input variables. To compute the partial derivative of a function f with respect the i-th variable at a given point $x^0\in\mathbb{R}^n$ we consider a small change in variable i while keep all other variables constant.

Definition 6.1 *Let* $f:\mathbb{R}^n\to\mathbb{R}$*. The partial derivative of* f *with respect to the i-th variable at point* $x^0\in\mathbb{R}^n$ *in the domain of* f *is defined by*

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

There are several (equivalent) notation for denoting the partial derivative of a function with respect to its i-th variable. The most common way is probably $\frac{\partial f}{\partial x_i}$ but we also find notations such as f_{x_i}, f_i $\partial_{x_i}f$, $\partial_i f$.

In practice, computing a partial derivative is very similar to computing the derivative of an univariate function. To find $\frac{\partial f}{\partial x_i}$ of a given function f, simply use the usual differentiation rules as if f were solely a function of x_i and treat all other variables as if they were constants.

Example 6.2 *Consider $f(x,y) = xy$. To compute $\frac{\partial f}{\partial x}$, we simply treat y as a constant and apply the usual differentiation rules to x. Hence $\frac{\partial f}{\partial x}(x,y) = y$.*

Example 6.3 *Consider $f(x,y) = xy^2 + y$. To compute $\frac{\partial f}{\partial y}$, we simply treat x as a constant and apply the usual differentiation rules to xy. Hence $\frac{\partial f}{\partial y}(x,y) = 2xy + 1$.*

Total differentiation. Sometimes we may be interested in how a function changes as several of its variables simultaneously change. We usually denote the total derivative of a function as df and it is defined as:

$$df = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} dx_i.$$

Intuitively, the total change of the function is the result of how each change of variable dx_i affects the function as captured by the partial derivative $\frac{\partial f}{\partial x_i}$.

Example 6.4 *Consider $f(x,y) = xy$. Its total derivative writes $df = ydx + xdy$.*

6.2. Higher-order derivatives

As we have seen before, differentiating the derivative of a univariate function yields a second-order derivative. Repeatedly applying this process allowed us to compute higher-order derivatives.

We can also compute higher-order derivatives in the case of multivariate function but the main difference is that it is now not obvious what we mean by 'repeatedly differentiating'. Indeed, as they are many variables, there are several ways in which we can differentiate an already differentiated function.

We can obviously decide to compute the *direct* second-order partial derivative $\partial f^2/\partial x_i^2$ which consists in partially differentiating f with respect to x_i two times.

Example 6.5 *Consider $f(x,y) = x^2y + 4x$. The first-order partial derivative with respect to x is $\frac{\partial f}{\partial x} = 2xy + 4$ and the second-order partial derivative with respect to x is $\frac{\partial f^2}{\partial x^2} = 2y$.*

But we can also first differentiate a function with respect to x_i and then differentiate a second time with respect to another variable x_j. We call this derivative the mixed (or cross) partial derivative and denote by $\frac{\partial f^2}{\partial x_i \partial x_j}$.

Example 6.6 *Consider $f(x,y) = x^2y + 4x$. The first-order partial derivative with respect to x is $\frac{\partial f}{\partial x} = 2xy + 4$ and the mixed partial derivative with respect to x and y is $\frac{\partial f^2}{\partial x \partial y} = 2x$.*

If a function has n variables, then, it will have n^2 second order partial derivatives. It is common to arrange these n^2 partial derivatives into an $n \times n$ matrix whose (i, j)th entry is the $\left(\partial^2 f/\partial x_j \partial x_i\right)(x^*)$. This matrix is called **Hessian matrix**:

$$D^2 f_{\mathbf{x}} = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_2 \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n} & \frac{\partial^2 f}{\partial x_2 \partial x_n} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

7. Optimization: Multivariate functions and constrained optimization

7.1. Unconstrained optimization

Univariate case. We have already seen how to find and characterize extreme points of univariate functions. We briefly recall the rules we established in that case.

Type	First-order condition	Second-order condition
Local maximum	$f'(x^*) = 0$	$f''(x^*) > 0$
Local minimum	$f'(x^*) = 0$	$f''(x^*) < 0$
Saddle point	$f'(x^*) = 0$	$f''(x^*) = 0$
Global maximum	$f'(x^*) = 0$	$f''(x) < 0$, for all x
Global minimum	$f'(x^*) = 0$	$f''(x^*) > 0$, for all x

Multivariate case. When considering function of several variables, $f : \mathbb{R}^n \to \mathbb{R}$, finding extreme points is relatively similar to the univariate case. Let us first introduce the following definition.

Definition 7.1 *The* ***gradient vector*** *of a function $f : \mathbb{R}^n \to \mathbb{R}$ is the vector whose components are the partial derivatives of f:*

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{bmatrix}.$$

The first-order condition in the multivariate case simply corresponds to the condition $\nabla f(x) = 0$.

Example 7.1 *Consider $f(x,y) = -(x-2)^2 - y^2 + 4$. The gradient of the function writes as $\nabla f(x) = [-2(x-2), -2y]$. Hence the unique candidate to an extreme point is $x^* = (2,0)$ as $\nabla f(x) = 0$.*

Second-order conditions are, however, a bit more complicate to use than in the univariate case.

Definition 7.2 *The **Hessian Matrix** of f is the $n \times n$ matrix whose (i, j)-th entry is the partial crossed derivative $\left(\partial^2 f / \partial x_j \partial x_i\right)(\mathbf{x}^*)$. That is,*

$$D^2 f_{\mathbf{x}} = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_2 \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n} & \frac{\partial^2 f}{\partial x_2 \partial x_n} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}.$$

Second-order conditions can be summarized as:

- If the Hessian is positive definite at x^*, then f attains a local minimum at x^*;
- If the Hessian is negative definite at x^*, then f attains a local maximum at x^*;
- If the Hessian has both positive and negative eigenvalues then x^* is a saddle point for f (and in fact this is true even if x^* is degenerate).

In those cases not listed above, the test is inconclusive.

7.2. Constrained optimization

Optimizing a function may be subject to some constraint on the admissible value of the solutions. This is typically the case in economics when a consumer must choose the best possible consumption bundle while being constrained by their budget. We briefly consider equality and inequality constraints in the case of a two-dimensional objective function (results naturally extend to the case of n-variable function).

The abstract formulation of a constrained maximization problem we are considering is the following:

$$\max_{x_1, x_2} f(x_1, x_2)$$
$$\text{s.t. } (x_1, x_2) \in C,$$

where C is the *constraint set*, that is, the set of admissible values for $x = (x_1, x_2)$.

Example 7.2 *Consider $f(x, y) = xy$ and $C = \{x, y \mid x + y = 1\}$.*

Example 7.3 *Consider $f(x, y) = x + y$ and $C = \{x, y \mid x + y \leqslant 5\}$.*

It is usually not obvious how to find the solution without a proper method. We first introduce the *substitution method* which is relatively simple but only works with equality constraints that are *well-behaved*. We then introduce Lagrange and Kuhn-Tucker multipliers to solve more general problems.

Substitution method. Consider the case in which the constraint is an equality constraint of the type $h(x_1,x_2) = c$ where h is function and c is a constant in $\mathbb{R}$. We do not enter into too many technical details but the substitution method essentially works when there is a way to invert h to solve for x_1 or x_2, that is, find a function φ such that $x_2 = \varphi(x_1,c)$ that satisfies $h(x_1,\varphi(x_1,c)) = c$ for all x_1. Then, one substitute this equation into the objective function and solves

$$\max_{x_1} f(x_1,\varphi(x_1,c)).$$

We can then use the rules we know for finding the solution of an unconstrained problem. That is, we can simply find the first-order condition with respect to x_1 which writes as follows:

$$\frac{\partial f}{\partial x_1} f(x_1,\varphi(x_1,c)) + \frac{\partial \varphi}{\partial x_1}(x_1,c)\frac{\partial f}{\partial x_2}(x_1,\varphi(x_1,c)) = 0.$$

Example 7.4 *Consider maximizing $f(x_1,x_2) = x_1x_2$ subject to $h(x_1,x_2) = 1$ where $h(x_1,x_2) = x_1 + x_2$. We can simply solve the equality constraint for x_2, that is, $x_2 = 1 - x_1$ (here it means that $\varphi(x_1,c1) = c - x_1$). Plugging this equation into the objective function we therefore have to maximize $f(x_1,1-x_1) = x_1(1-x_1)$. We immediately get that $\partial f/\partial x_1 = 1-2x_1$ and $\partial f^2/\partial x_1^2 = -2$. Finally, the first-order condition yields that $x^* = 0.5$ which is a global maximum as the second-order derivative is always negative.*

Lagrange method (equality constraints). Let f and h be differentiable functions of two variables, where f is the objective function and $h(x_1,x_2) = c$ is an equality constraint.

Definition 7.3 *The* ***Lagrangian*** *of an optimization problem is a function denoted by $\mathscr{L}$ defined as follows:*

$$\mathscr{L}(x_1,x_2,\lambda) = f(x_1,x_2) + \lambda(h(x_1,x_2) - c),$$

where $\lambda \in \mathbb{R}$ is called the ***Lagrangian multiplier.***

Under some conditions, solving the **unconstrained** problem $\max_{x_1,x_2,\lambda} \mathscr{L}(x_1^*,x_2^*,\lambda^*)$ is equivalent to solving the **constrained** problem $\max_{x_1,x_2} f(x_1,x_2)$ s.t. $h(x_1,x_2) = c$. Hence, we can apply the method of first-order conditions, the following equations are necessary conditions for an extreme point (omitting the arguments for convenience):

$$\frac{\partial \mathscr{L}}{\partial x_1} = \frac{\partial \mathscr{L}}{\partial x_2} = \frac{\partial \mathscr{L}}{\partial \lambda} = 0,$$

or, equivalently, we have to solve the following system of equations (still omitting the arguments for clarity):

$$\frac{\partial f}{\partial x_1} + \lambda \frac{\partial h}{\partial x_1} = 0$$
$$\frac{\partial f}{\partial x_2} + \lambda \frac{\partial h}{\partial x_2} = 0$$
$$h(x_1, x_2) - c = 0.$$

Notice that the last first-order condition of the Lagrangian problem is simply the equality constraint itself.

Example 7.5 *Consider the same problem as earlier. We want to maximize $f(x_1, x_2) = x_1 x_2$ subject to $h(x_1, x_2) = 1$ where $h(x_1, x_2) = x_1 + x_2$. We can write the Lagrangian function as $\mathscr{L}(x_1, x_2, \lambda) = x_1 x_2 + \lambda(x_1 + x_2 - 1) = 0$. The first-order conditions of the Lagrangian write as follows:*

$$\frac{\partial \mathscr{L}}{\partial x_1} = x_2 + \lambda = 0$$
$$\frac{\partial \mathscr{L}}{\partial x_2} = x_1 + \lambda = 0$$
$$\frac{\partial \mathscr{L}}{\partial \lambda} = x_1 + x_2 - 1 = 0.$$

It is immediate that we have that $x_1 = x_2 = -\lambda$ from the first two equations. Hence, the third equation can be rewritten as $2x_1 - 1 = 0$ which yields that $x_1 = x_2 = 1/2$ and $\lambda = -1/2$. The point $(1/2, 1/2, \lambda)$ is a candidate extreme point of the Lagrangian. Notice that $x_1 = x_2 = 1/2$ coincides with our previous result using the substitution method.

Example 7.6 *Consider the problem of maximizing $f(x_1, x_2) = xy$ subject to $-x^2 - y + 2 = 0$. The Lagrangian writes as follows:*

$$\mathscr{L} = xy + \lambda(-x^2 - y + 2).$$

The first-order conditions of the Lagrangian are:

$$\frac{\partial \mathscr{L}}{\partial x} = y + \lambda(-2x) = 0$$
$$\frac{\partial \mathscr{L}}{\partial y} = x + \lambda(-1) = 0$$
$$\frac{\partial \mathscr{L}}{\partial \lambda} = -x^2 - y + 2 = 0.$$

From the second equation we immediately have that $\lambda = x$. Plugging this into the first equation yields that $y - 2x^2 = 0 \Leftrightarrow y = 2x^2$. Hence, the third equation rewrites as $-x^2 - 2x^2 = -2$ or equivalently that $3x^2 = 2$. This last equation yields that $x = \sqrt{2/3}$ or $x = -\sqrt{2/3}$, and therefore $y = 2\sqrt{2/3}^2 = 4/3$.

Which of the two is our solution? A closer look at our previous result shows that $y > 0$ *for both admissible values of* x*. Therefore, it is immediate to see that* $f(\sqrt{2/3}, y) = \sqrt{2/3}y > f(-\sqrt{2/3}, y) = -\sqrt{2/3}y$ *for all* $y > 0$*. Hence, our solution can only be with* $x^* = \sqrt{2/3}$ *and* $y^* = 4/3$*.*

Karush-Kuhn-Tucker conditions (inequality constraints). Let f and g be two differentiable functions of two variables. As before f is the objective function but we now consider inequality constraints of the form $g(x_1, x_2) \geqslant c$. The Karush-Kuhn-Tucker conditions (hereafter, KKT) extend the Lagrangian method to these type of constraint. As before, we rely on the Lagrangian function:

$$\mathscr{L}(x_1, x_2, \mu) = f(x_1, x_2) + \mu(g(x_1, x_2) - c),$$

where $\mu \in \mathbb{R}_+$ is the **KKT multiplier.**

There are two main differences with the Lagrange method: (i) the KKT multiplier has to be positive or null and (ii) we replace the first-order condition with respect with the multiplier by what is called a *complementary slackness condition.*

Formally, necessary conditions for (x_1^*, x_2^*) to solve the constrained problem are the following in the unconstrained KKT problem:

$$\text{First-order conditions: } \frac{\partial \mathscr{L}}{\partial x_1} = \frac{\partial \mathscr{L}}{\partial x_2} = 0,$$
$$\text{Complementary slackness: } \mu(g(x_1, x_2) - c) = 0.$$

The complementary slackness condition should be understood as follows: if the inequality constraint strictly holds at a critical point (i.e. $g(x) < c$) then the KKT multiplier μ has to be zero. If otherwise the inequality constraint holds with equality (i.e. $g(x) = c$) then the KKT multiplier μ can take any nonnegative value.

Example 7.7 *Consider maximizing* $f(x, y) = -x^2 - y^2 + 3$ *subject to* $-2x - y \geqslant 1$*. The Lagrangian writes as follows:*

$$\mathscr{L} = -x^2 - y^2 + 3 + \mu(-2x - y - 1).$$

First-order conditions are:

$$\frac{\partial \mathscr{L}}{\partial x} = -2x + \mu(-2) = 0$$
$$\frac{\partial \mathscr{L}}{\partial y} = -2y + \mu(-1) = 0,$$

and the complementary slackness condition is $\mu(-2x - y - 1) = 0$*.*

The first-order conditions immediately give that $x = -\mu$ *and* $y = -\mu/2$*. It is clear that if we had* $\mu = 0$ *then we would also have* $x = y = 0$ *and* $g(x_1, x_2) = 0 < 1$*. It is therefore impossible* *that* $\mu = 0$*, meaning that we must have* $\mu > 0$*. By complementary slackness, we must then have that* $-2x - y - 1 = 0$*. Using previous results this rewrites as* $-2x - x/2 - 1 = 0 \Leftrightarrow x = -2/5$ *and* $y = -1/5$.

Be careful though, as in the previous methods these conditions are only necessary for optimality. In principle, one must also check second-order conditions or make sure that the initial problem is itself concave or convex. These issues are beyond the scope of this course as you will mostly encountered well-behaved problems. The interested reader can find more on these topics in the textbooks of reference.

References

Carter, M., 2001. *Foundations of mathematical economics.* MIT press.

Simon, Carl P., and Lawrence Blume. *Mathematics for economists.* Vol. 7. New York: Norton, 1994