Math PhD Brush-Up Lecture Notes

Alessandro Ruggieri*

July 23, 2019

Abstract

These lecture notes are compendium for a PhD brush-up course in Real Analysis and Optimization. The whole text is divided into six sections, covering basic topological analysis —metric and normed spaces, sequences and convergence, notion of open, closed sets, boundedness and compactness—, the notion of continuity, convexity and differentiability for real functions, solution of optimization problems and basic integration.

*Contact: Departament de Economia i Historia Economica, Facultat D′Economia, Edifici B, Universitat Autònoma de Barcelona, 08193, Bellaterra, Spain, aruggierimail@gmail.com. I acknowledge financial support from the La Caixa-Severo Ochoa Program for Centers of Excellence in R&D of the Spanish Ministry of Economy and Competitiveness. The usual disclaimers apply.

Contents

1 Topology

1.1. Metric Spaces

Let X be a non-empty space. A function $d : X \times X \to \mathbb{R}_+$ is called a distance function or metric defined on X if $\forall x, y, z \in X$ if the following conditions are met:

- **Positivity**: the distance between two points in any space is positive, i.e. $d(x, y) \geq 0 \, \forall x, y \in X$.
- **Non-degeneration** of the space: if the distance between two objects is 0, the objects are the same. This means that $d(x, y) = 0 \iff x = y$.
- **Symmetry**: the distance between x and y is the same as the distance between y and x, i.e. $d(x, y) = d(y, x)$
- **Triangle inequality**: two sides of a triangle are always larger than the remaining side, such that $d(x, y) \leq d(x, z) + d(z, y)$.

If d is a distance function on X, the couple (d, X) defines a metric space. Thus, a metric space is a couple made of a set X and a distance function d that describes how elements in X are spatially related. A trivial way to define a metric on X is locating all points at the same distance. The discrete metric defines points as either equal or different, such that:

$$\begin{aligned} d(x, y) &= 0 \text{ if } x = y \\ d(x, y) &= 1 \text{ if } x \neq y \end{aligned} \tag{1.1}$$

Proposition 1. *The discrete metric defines* (d, X) *as a metric space.*

Proof. The first three properties are satisfied by construction of the metric. To check property 4, select three generic points, say $x, y, z \in X$. Notice that:

- If $x = y$, then $d(x, y) = 0 \leq d(x, z) + d(z, y) \quad \forall z \in X$.
- if $x \neq y$ then it must be either

  a. $x \neq y \Longrightarrow d(x, y) = 1 \leq 1 + d(z, y)$,

  b. $y \neq z \Longrightarrow d(x, y) = 1 \leq d(x, z) + 1$.

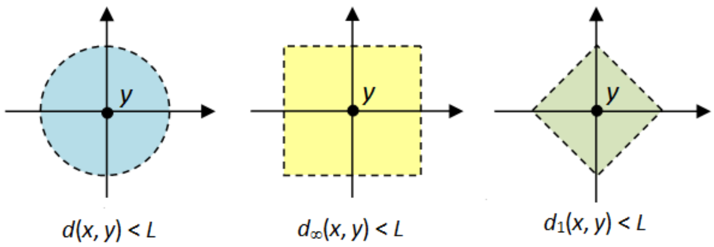
□

Consider the case of a space $X = \mathbb{R}^n$, for a generic $n \in \mathbb{N}$. It can be proved that the following functions satisfy the four properties that are required for the definition of a metric (distance function):

- **Taxicab distance**: $d_1 = \sum_{i=1}^{n} |x_i - y_i|$.
- **Euclidean distance**: $d_2 = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$.
- **Maximum distance**: $d_\infty = \max_{i=1...n} |x_i - y_i|$.

Notice that the way we define a metric space — i.e. which metric function we use — varies the properties of the metric space itself. Fig. 1.1 shows a graphical representation of points in $\mathbb{R}^2$ located at a distance smaller than $L \in \mathbb{R}$ from the origin, under the metrics d_1, d_2 and d_∞.

Figure 1.1. Metric Spaces



Proposition 2. *The maximum distance metric defines $(d_\infty, \mathbb{R}^n)$ as a metric space.*

Proof. Properties 1 and 3 are satisfied by the properties of the absolute value. Property 2 is satisfied by construction of the metric. To check property 4, consider three generic points $x, y, z \in \mathbb{R}^n$. Then:

$$\begin{gathered} d_\infty(x, y) = \max\{|x_i - y_i| : i = 1, ..., n\} = \\ \max\{|x_i - z_i + z_i - y_i| : i = 1, ..., n\} \leq \\ \max\{|x_i - z_i| + |z_i - y_i| : i = 1, ..., n\} \leq \\ \max\{|x_i - z_i| : i = 1, ..., n\} + \max\{|z_i - y_i| : i = 1, ..., n\} = \\ d_\infty(x, z) + d_\infty(z, y) \end{gathered} \tag{1.2}$$

□

A well-defined distance between pairs of elements in a given space induces what is called topological space, as it is informative of how elements in the space spatially relate to each other. Notice that different metrics might induce different notions of open sets and therefore different topologies on the same space.

Definition 1. *A norm over $\mathbb{R}^n$ is a function $\| \cdot \|: \mathbb{R}^n \to \mathbb{R}_+$ such that these properties are met:*

- ***Zero vector**, such that: $\| x \| = \mathbf{0} \iff x = \mathbf{0}$.*
- ***Triangle inequality**, such that: $\forall x, y \in V \subset \mathbb{R} : \| x + y \| \leq \| x \| + \| y \|$.*

- ***Absolute homogeneity****, such that:* $\forall \alpha \in \mathbb{R}, \forall x \in \mathbb{R} : \parallel \alpha x \parallel = |\alpha| \parallel x \parallel$.

By absolute homogeneity, notice that $\parallel \mathbf{0} \parallel = 0$, and $\forall x \in \mathbb{R}^n, \parallel -x \parallel = \parallel x \parallel$, so that by the triangle inequality, we have $\parallel x \parallel \geq 0$ (positivity). Examples of norm functions defined in $\mathbb{R}^n$ are:

- **Absolute value norm**: $\parallel x \parallel = |x|$.
- **Taxicab norm**: $\parallel x \parallel_1 = \sum_{i=1}^{n} |x_1|$.
- **Euclidean norm**: $\parallel x \parallel_2 = \sqrt{x_1^2 + ... + x_n^2}$.
- **Maximum norm**: $\parallel x \parallel_\infty = \max(|x_1|, ..., |x_n|)$.

Notice that a norm function is quite similar to a distance function. In fact, a norm is a function that assigns a strictly positive length or size to each vector in a vector space, whereas a distance function is a function that explains how two different points spatially relate. A norm function $\parallel \cdot \parallel$ defined on the space $\mathbb{R}^n$ generates a **normed space**. Given any normed space, a metric function $d : \mathbb{R}^n \to \mathbb{R}_+$ can always be induced by a norm as follows:

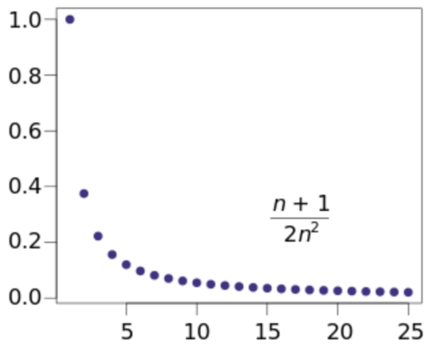
$$d(x, y) = \parallel x - y \parallel = \parallel x + (-y) \parallel \tag{1.3}$$

There are two caveats: first, the discrete metric does not correspond to any norm; second, a metric function needs not to be a norm function.

1.2. Sequences and Convergence

Definition 2. *A sequence in a set X is a function $x : \mathbb{N} \to X$. This is usually denoted by $(x_n)_{n\in\mathbb{N}}$. Consider now the space spanned by all the sequences of scalars, i.e. $(s_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$; this is a space whose elements are sequences of real numbers.*

Figure 1.2. Example of a sequence



Definition 3. *A sequence of real numbers $(x_n)_{n\in\mathbb{N}}$ is said to be bounded if $\exists M \in \mathbb{R}$ for which every term x_n satisfies $|x_n| \leq M$.*

The most important spaces of sequences are the l^P spaces, consisting of all the p-power summable sequences.

Definition 4. *For $0 < p < \infty$, denote l^p as the subspace of $\mathbb{R}^{\mathbb{N}}$ composed by all the p-bounded sequences in $\mathbb{R}^{\mathbb{N}}$, i.e.:*

$$l^P = \{(x_n)_{n \in N} \in \mathbb{R}^{\mathbb{N}} : \sum_{i=1}^{\infty} |x_i|^p < \infty\} \tag{1.4}$$

In this space, the p-distance defined by:

$$d_p((x_n),(y_n)) : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}_+ \text{ such that } d(x,y) = \left(\sum_{i=1}^{\infty} |x_i - y_i|^p \right)^{\frac{1}{p}} \tag{1.5}$$

is a well defined metric function.

Definition 5. *Given the metric space $(d, \mathbb{R})$, a sequence $(x_n)_{n \in \mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ is said to converge to a limit point $\bar{x} \in \mathbb{R}$ under the metric $d(x,y)$ if:*

$$\forall \epsilon > 0, \exists \bar{n} \in \mathbb{N} : n > \bar{n} \Longrightarrow d(x_n, \bar{x}) < \epsilon \tag{1.6}$$

Examples:

- The sequence $(x_n)_{n \in \mathbb{N}} = \frac{1}{n}$ converges to 0.
- The sequence $(x_n)_{n \in \mathbb{N}} = (-1)^n$ does not converge.
- A convergent sequence is bounded under any p-metric. However, a bounded sequence does not need to converge.

Proposition 3. *A sequence $(x_n)_{n \in \mathbb{N}}$ can converge to at most one limit.*

Proof. By contradiction, suppose that $(x_n)_{n \in \mathbb{N}}$ converges both to a and b, with $a \neq b$. By triangle inequality, $\forall n \in \mathbb{N}$, we have $d(a,b) \leq d(a,x_n) + d(b,x_n)$. By definition of convergence, $\forall \epsilon > 0, \exists M_a, M_b \in \mathbb{R}$ such that $d(a,x_n) < \frac{\epsilon}{2}, \forall n \geq M_a$, and $d(b,x_n) < \frac{\epsilon}{2}, \forall n > M_b$. Therefore, $\forall n \geq \max\{M_a, M_b\}$, it must be that $d(a,b) \leq d(a,x_n) + d(b,x_n) < \epsilon$. Taking the limit for ϵ going to zero, $d(a,b) \to 0$, implying $a = b$. □

Figure 1.3. Convergence

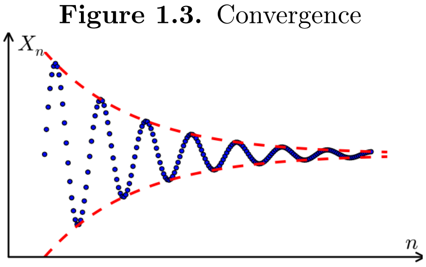
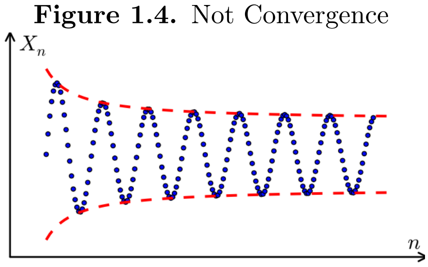


Figure 1.4. Not Convergence



Proposition 1. *The sequence $(x_n)_{n \in \mathbb{N}} = (x_n^1, x_n^2, ..., x_n^k)$ converges to $x = (x^1, x^2, ..., x^k)$ in $(d_p, \mathbb{R}^k)$ iff each series x_n^i converges to x_i in $(d, \mathbb{R})$.*

1.3. Open and Closed Sets

Definition 6. *Let (d, X) be a metric space. Then, $\forall x \in X$ and $\epsilon \in \mathbb{R}_{++}$, we define an ϵ-neighborhood of $x \in X$ as:*

$$N_{\epsilon,X}(x) = \{y \in X : d(x, y) < \epsilon\} \tag{1.7}$$

Notice that the ϵ-neighborhood of a point $x \in X$ is always not-empty, as it always contains at least x. Moreover, its definition not only depends on x and ϵ but also —and especially— on the metric d and the space X.

Definition 7. *A set $S \in X$ is said to be open in (d, X) if $\forall x \in S, \exists \epsilon > 0$, such that $N_{\epsilon,X}(x) \subset S$. By negation, a set $S \in X$ is not open if $\exists x \in S$ for which $\nexists \epsilon > 0$ such that $N_{\epsilon,X}(x) \subset S$.*

Consider a point in the reals, $x \in X = \mathbb{R}$. By definition, an ϵ-neighborhood of x is composed by all those points $y \in \mathbb{R}$ that are less than ϵ from x, meaning:

$$N_{\epsilon,X}(x) = \{y \in \mathbb{R} : y \in (x - \epsilon, x + \epsilon)\} \tag{1.8}$$

Proposition 4. *Real intervals of the form (a, b) are open sets in the metric space $(d_2, \mathbb{R})$.*

Proof. Take $x = \frac{a+b}{2}$ and $\delta = \frac{b-a}{2}$. Then $(a, b) = (x - \delta, x + \delta)$. Pick $z \in (x - \delta, x + \delta)$. By construction, $d_2(z, x) = |z - x| < \delta$. Define $\epsilon = \delta - d_2(z, x) > 0$. Notice that an interval $(z - \epsilon, z + \epsilon)$ is fully included into $(x - \delta, x + \delta)$. To see this, let $y \in (z - \epsilon, z + \epsilon)$. By the triangular inequality, we have that:

$$\begin{aligned} d_2(x, y) = |x - y| \leq |x - z| + |z - y| < |x - z| + \epsilon = \\ = |x - z| + \delta - d_2(z, x) = |x - z| + \delta - |z, x| = \delta. \end{aligned} \tag{1.9}$$

It follows that $y \in (x - \delta, x + \delta), \forall y \in (z - \epsilon, z + \epsilon)$, which implies that we $\forall z \in (x - \delta, x + \delta)$ we can always find a $\epsilon > 0$ s.t. $N_{\epsilon,X}(z) \subset X$. Thus $(x - \delta, x + \delta)$ is open. □

Proposition 5. *The union $\cup_{i \in I} O_i$ of an arbitrary family of open intervals is open.*

Proof. Pick a point $x \in \cup_{i \in I} O_i$. By definition of the union operator, it must be that $\exists j \in I$ such that $x \in O_j$. Since O_j is open $\forall j$, it must be that $\exists \epsilon > 0$ such that $(x - \epsilon, x + \epsilon) \subset O_j \in U_{i \in I} O_i$. This proves that $\cup_{i \in I} O_i$ is open. □

Proposition 6. *The intersection $\cap_{i \in \{1,2,...,n\}} O_i$ of a finite family of open intervals is open.*

Proof. Pick a point $x \in \cap_{i \in \{1,2,...,n\}} O_i$. By definition of intersection operator, it must be that $x \in O_i \, \forall i \in \{1, 2, ..., n\}$. Since O_i is open $\forall i$, it must be that $\exists \epsilon_i > 0$ such that $(x - \epsilon_i, x + \epsilon_i) \subset O_i \, \forall i$. Define $\epsilon = \min \epsilon_i : i \in \{1, 2, ..., n\}$. By construction, it must be that $(x - \epsilon, x + \epsilon) \subset O_i \, \forall i$, which implies $(x - \epsilon, x + \epsilon) \subset \cap_{i \in \{1,2,...,n\}} O_i$. This proves that $\cap_{i \in \{1,2,...,n\}} O_i$ is open. □

Definition 8. *A set $S \in X$ is said to be closed in (d, X) if $X \backslash S$ is open. By negation, a set is not closed if its complement is not open.*

In simple words, a set is said to be open if all its limiting points are included in the set. Thus, a set is not open if there exists at least one x in the set such that there is no $\epsilon > 0$ for which a neighborhood of x is included in the set.

Notice that the notion of openness and closeness are not mutually exclusive: any set might be open, closed, both, or neither. In any topological space (d, X), the entire set X is open by definition, as it is the empty set. On the other hand, the complement of the entire set X is the empty set; since X has an open complement, X must be closed too. It follows that, in any topology, the entire space is simultaneously open and closed, either called clopen. Thus, the real line $\mathbb{R}$, as well as the empty set $\emptyset$ are both clopen. Notice that the key is to define open and closed sets in opposition to their complements.

Figure 1.5. Open Interval & Disk

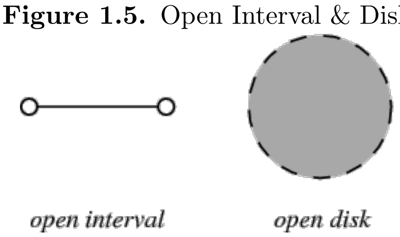
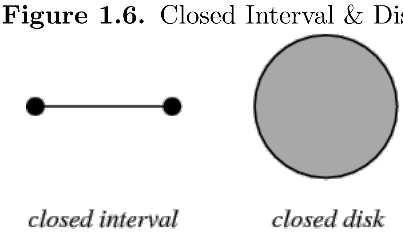


Figure 1.6. Closed Interval & Disk



Proposition 7. *The set $S = [a, b] \subset$ is closed in $(d_2, \mathbb{R})$.*

Proof. Consider the complement of S, $S^c = (-\infty, a) \cap (b, \infty)$. The set S^c is the union of two open intervals, so must be. Since S is the complement of an open set, by definition must be closed. □

Proposition 8. *The set $\{x\}$ is closed in $(d_2, \mathbb{R})$.*

Proof. Consider the complement of $\{x\}, \mathbb{R}\backslash x$. Pick a point $y \in \mathbb{R}\backslash x$. Take $\epsilon < |y - x|$. Then, $x \notin (y - \epsilon, y + \epsilon)$, which implies $(y - \epsilon, y + \epsilon) \in \mathbb{R}\backslash x$. □

Proposition 9. *For any metric space endowed with a discrete metric (d, X), any subset is both open and closed in X (this is not necessarily true in the case of other metric spaces).*

Proof. Let $S \subset X$ and $x \in S$. Set $\epsilon \in (0, 1)$. Then, $N_{\epsilon,X}(x) = x \in S$. By construction of the discrete metric, the space is made of points at a distance no less than 1 from x. Therefore, any neighborhood of x with radius less than one can only be made of a single point, x. Thus, any subset is open in X and hence, any subset (complementary of an open set in X) is also closed in X. □

Definition 9. *The interior of $S \subseteq X, int(S)$ is the largest open set contained in S. Equivalently, it can be defined as the union of all the open subsets of S.*

The notion of interior allows to characterize that of open set: a set S is open iff $S = int(S)$.

Definition 10. *The closure of* $S \subseteq X, cl(S)$ *is the smallest closed set that contains* S*. Equivalently, it can be defined as the intersections of all closed subsets that contain* S*.*

Definition 11. *The boundary of* $S \subseteq X$ *denoted by* $bd(S)$ *is the set* $cl(S) \backslash int(S)$*. Analogously, it can be defined as* $\{x \in X : N_{\epsilon_1,x}(X) \cap S \neq \emptyset \wedge N_{\epsilon_2,x}(X) \cap S^c \neq \emptyset\}$ *for* $\epsilon_1, \epsilon_2 > 0$*.*

The notion of closure allows us to characterize that of closed set: a set S is closed IFF it coincides with its closure. This implies that any closed set must contain its own boundary. Notice that the notion of closed set is anyway defined in terms of open sets, a concept that makes sense for topology induced by metric spaces. Finally notice that the boundary of S must be the set of points x such that, for any $\epsilon > 0$, ϵ-neighborhoods intersect both S and $R \backslash S$.

Definition 12 (**Sequential Characterization of Closed Sets**)**.** *A set* S *is closed iff any convergent sequence formed by elements of* S *has a limit in* S*. By negation, a set* S *is not closed if there exists at least one convergent sequence formed by elements of* S *that has limit outside* S*.*

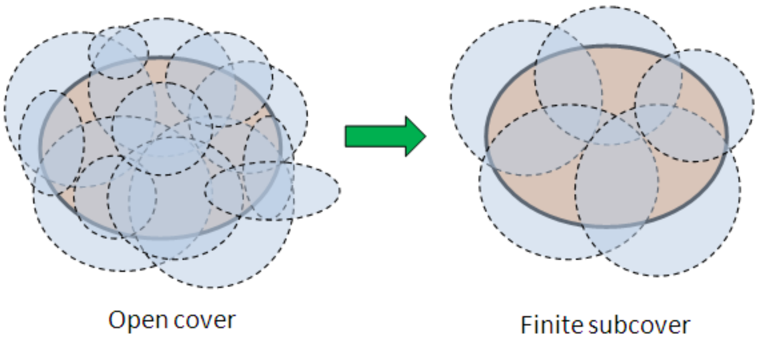
1.4. Boundedness and Compactness

Let (d, X) be a metric space and $S \subseteq X$. A class O of subsets of X is said to cover S if $S \subseteq U_{i \in I} O_i$. If the sets $O_i \in O$ are open, then we say that O is an open cover.

Definition 13. *A metric space* (d, X) *is said to be compact if every open cover has a finite subset that also covers* X*. A subset* S *of* X *is said to be compact in* X *if every open cover of* S *has a finite subset that also covers* S*. By negation, a space is not-compact if there exists at least one open cover of* S *with no finite subset that covers* S*.*

In other words, for a set to be compact it needs to be both bounded and closed. Sets that are infinite or that are not closed cannot have a finite set of open covers. This is so because an infinite set cannot have a finite cover; and because there exist infinite open covers from which a finite sub-cover cannot be selected. Thus, a finite number of open sets that cover S make S compact. A not compact set can be defined analogously. Notice that a continuous function in a compact set will always have a minimum and a maximum.

Figure 1.7. Compact Set



Proposition 10. *The real space, $\mathbb{R}$, endowed with the Euclidean norm, is not a compact metric space.*

Proof. To show it, we need to find an open cover of $\mathbb{R}$ with no finite sub-cover. Consider the collection of sets $O = \{O_i = (-i, i) : i = 1, 2, ...\}$. This is an open cover of $\mathbb{R}$. However, any finite sub-cover $O_n = \{O_i = (-i, i) : i = 1, 2, ...n\}$ won't be able to cover $\mathbb{R}$, but only the elements in $(-t, t)$, where t is the largest interval in the finite collection. Hence, $\mathbb{R}$ is not compact. □

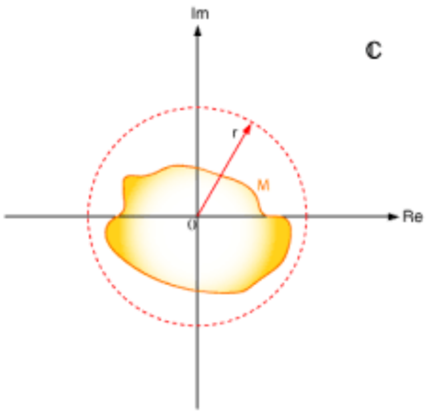
Proposition 11. *The set $X = (0, 1) \subset \mathbb{R}$ is not compact.*

Proof. Consider the open cover defined by $O = \{O_i = (\frac{1}{i}, 1) : i = 1, 2, ...\}$. By the same token as before, $X = (0, 1)$ is not compact. □

Definition 14. *Let (d, X) be a metric space. A subset $S \subseteq X$ is said to be bounded in X if $\exists\epsilon > 0$ and $x \in S$ such that $S \subseteq N_{\epsilon,X}(x)$.*

Thus, a set is bounded if we can draw a ball that completely includes it.

Figure 1.8. Bounded Set



What sort of sets are bounded? For instance, finite sets are bounded, as we can consider the maximum distance between two objects in the set and fix an ϵ equal to that.

Example: A set $S \subseteq \mathbb{R}$ is said to be bounded from above if $\exists k \in \mathbb{R}$ such that $k > s \, \forall s \in S$. Analogously, S is said to be bounded from below if $\exists k \in \mathbb{R}$ such that $k < s \, \forall s \in S$. A set S is bounded if it has both upper and lower bounds. Therefore, a set of real numbers is bounded if it is contained in a finite interval.

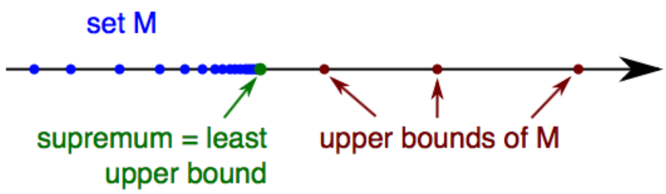
Proposition 2 (**Completeness Axiom**). *Suppose that S is a non-empty set of real numbers and is bounded above. Then, there is a number $M \in \mathbb{R}$, called supremum —sup, or least upper bound—, such that:*

- *M is an upper bound of S.*
- *Given any $\epsilon > 0$, there exists $s \in S$ such that $s > M - \epsilon$.*

Analogously, suppose S is bounded below. Then, there is a number $m \in \mathbb{R}$ called infimum —inf, or greatest lower bound—, such that:

- *m is a lower bound of S.*
- *Given any $\epsilon > 0$, there exists $s \in S$ such that $s < m + \epsilon$.*

Thus, a set is bounded above if it has an upper bound and it is bounded below if it has a lower bound. The greatest lower bound as, its name indicates, the smallest of the upper bounds. A lower bound is defined analogously. Notice that in open sets, maxima and minima are not well-defined. But supremums and infimums always exist. In closed sets, supremums and infimums are always maxima and minima.



Proposition 3 (**Archimedean Property**). *Given any real number $x, \exists n \in \mathbb{N}$ such that $n > x$. It follows that the set of natural numbers if not bounded above. Notice that the Completeness Axiom implies the Archimedean Property. To see this, assume $\mathbb{N}$ were bounded above. By the Completeness Axiom, $\mathbb{N}$ must have a supremum, say $\alpha = \sup \mathbb{N}$. If $\alpha \in \mathbb{N}$, then also $\alpha + 1 > \alpha \in \mathbb{N}$, contradicting the fact that α is a supremum of S. Suppose $\alpha \notin \mathbb{N}$. Then, $\alpha + 1 \notin \mathbb{N}$. Pick the element $x \in \mathbb{N}$ such that $d(x, \alpha) < 1$ is minimized. Then $x + 1 > \alpha$, contradicting again the notion of supremum.*

Definition 15 (**Sup. vs. Max.**). *The maximum (minimum) of a set S is its largest (smallest) element if such an element exists. If a set has a maximum (minimum), that is a supremum (infimum) for that set; the converse, however, might not be true. By construction, in any non-empty bounded set, both the least upper bound and the greatest lower bound are well-defined and finite, though they might not belong to the set itself. However, if exist, maximum (and minimum) always belong to the set.*

Proposition 12. *Let $S \subseteq \mathbb{R}$ denote a set that is bounded below. Let L denote a set of its lower bounds. Then, L is bounded above and* $\sup L = \inf S$.

Proof. As S is bounded from below, then L is non-empty and is defined as $L := \{y \in S : y \leq x \, \forall x \in S\}$, which implies that every $x \in S$ is an upper bound for L. Then, L is bounded above. Since L is bounded above, it has an upper bound in S, say $\alpha = \sup L$. Pick a real number n. If $n < \alpha$, then n is not an upper bound for L, implying $n \notin S$. It follows that $\alpha \leq x, \forall x \in S$. If $n > \alpha$, then $n \notin L$. Then, α is a lower bound of S, while n is clearly not. Therefore, $\alpha = \inf S$. □

Notice that every compact metric space is both close and bounded. However, closed and bounded sets are not necessarily compact for some metric spaces.

Proposition 13 (Theorem Heine—Borel). *Consider the metric space $(d, \mathbb{R}^n)$, where $n \in \mathbb{N}$ and d is a non-discrete metric function. Then, any subset $S \subset \mathbb{R}^n$ is compact iff it is closed and bounded.*

Thus, in the reals, a set is compact iff it is closed and bounded, and otherwise. Notice also that most closed sets are bounded, but there are some closed sets, such as the emptyset, which is closed but not bounded.

Proof. To prove this proposition, let's prove all its implications.

- **Every compact metric space is bounded**. Let S be a compact space. Pick a point $x \in S$ and consider the following open cover of $S, \{N_{m,S}(x) : m \in \mathbb{N}\}$. Since S is compact, there exists a finite sub-cover of S, say $\{N_{m_1,S}, ..., N_{m_T,S}\}$. Fix $\epsilon = \max\{m_1, ..., m_T\}$ and create an ϵ-neighborhood around x: this is enough to cover S.

- **Every compact metric space is closed**. Let S be a compact subset of a metric space X. If $S = X$, then S is closed, as it is the reciprocal of the empty space, which by definition is open. Suppose $S \neq X$. We need to show that $X \backslash S$ is open. Pick $y \in X \backslash S$. Then it must be that $\forall x \in S, \exists \epsilon_x > 0$ such that $N_{\epsilon_x,X}(x) \cap N_{\epsilon_x,X}(y) = \emptyset$. Notice that the collection of open sets $\{N_{\epsilon_x,X}(x) : x \in S\}$ covers S and, by compactness, there exists a finite sub-cover that also covers S, meaning $S \subset \cup_i^n N_{\epsilon_{x_i},X}(x_i)$. Fix $\epsilon = \min\{\epsilon_{x_i} : i = 1, ..., n\} > 0$. Then, it must that $N_{\epsilon,X}(y) \cap N_{\epsilon_{x_i},X}(x) = \emptyset \, \forall i = 1, ..., n$, which implies that $N_{\epsilon,X}(y) \subset X \backslash S$. Therefore, $X \backslash S$ is open, implying S is closed.

- **A closed subset of a compact set is compact**. Let S be a closed subset of a compact metric space X. Let O be an open cover of S. Since S is closed, $X \backslash S$ is open and $O \cup X \backslash S$ is an open cover of X. As X is compact, there exists a finite sub-cover of X. Since a finite sub-cover of X is also a finite sub-cover of any subset of X, $S \subset X$ must be compact.

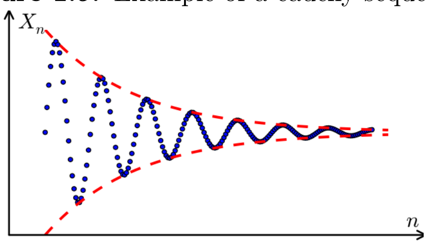
- **In any Euclidean space endowed with a non-discrete metric function, any closed and bounded space is compact**. Take a closed and bounded set $S \in \mathbb{R}^n$. By boundedness, $\exists \epsilon > 0, x \in S$ such that $S \subseteq N_{\epsilon,\mathbb{R}^n}(x)$. Notice that S can be enclosed within the n-box $[-k, k]^n$ for k large enough. By the property above, it is enough to show that $[-k, k]^n$ is compact. Using bisection, it can be proved that any n-dimensional box of the form $[a, b]^n \subset \mathbb{R}$, with $-\infty < a < b < \infty$ is compact. This would be enough to shows that S itself is compact.

□

1.5. Completeness

A sequence $(x_n)_{n\in\mathbb{N}}$ is called cauchy if $\forall \epsilon > 0, \exists M \in \mathbb{R}$ such that $d(x_k, x_l) < \epsilon$, $\forall k, l \geq M$. In other words, every convergent sequence is cauchy, and every cauchy sequence is convergent in complete metric spaces. Thus, in all complete metric spaces, cauchy and convergence can be taken as synonyms.

Figure 1.9. Example of a cauchy sequence



Proposition 14. *Every convergent sequence is cauchy.*

Proof. Let $(x_n)_{n\in\mathbb{N}}$ be a convergent sequence. Suppose the limit is $\bar{x}$. Let $\epsilon > 0$. Then, $\exists M \in \mathbb{N}$ such that $d(x_n, \bar{x}) < \frac{\epsilon}{2}\, \forall n \geq M$. Hence, $\forall k, l \geq M$, it must be that:

$$d(x_k, x_l) \leq d(x_k, \bar{x}) + d(x_l, \bar{x}) < \epsilon \tag{1.10}$$

□

Definition 16. *A set S is said to be complete if every cauchy sequence of S converges to a point in S.*

Example: $\mathbb{R}$ is a complete space.

Proposition 15. *A subset $S \subset \mathbb{R}$ is complete iff it is closed.*

Proof. Suppose S is complete. Consider a sequence of real scalars $(x_n)_{n\in\mathbb{N}}$, defined in S that converges to $\bar{x}$. Since $(x_n)_{n\in\mathbb{N}}$ is convergent, it is also cauchy; and since S is complete, it must converge to $\bar{x} \in S$. Therefore, S is closed. Suppose S is closed. Consider a cauchy sequence of real scalars in S. First notice that it is convergent, since $\mathbb{R}$ is complete. Then, since S is closed, the limit point must lie in S. □

2 Continuity

2.1. Continuity

Definition 17. *A map $f : X \to \mathbb{R}$ is said to be continuous at $x \in X$ if $\forall \epsilon > 0, \exists \delta > 0$ such that:*

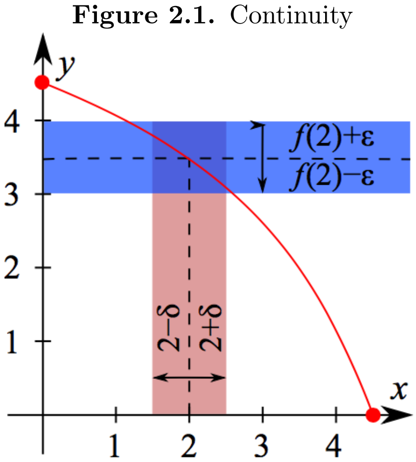
$$\forall y \in X \text{ such that } d(x, y) < \delta \Rightarrow d(f(x), f(y)) < \epsilon \tag{2.1}$$

If f is not continuous at x, we say that it is discontinuous at x. By negation, a function is not continuous at x if $\forall \epsilon > 0, \nexists \delta > 0$ such that the above condition is satisfied. If f is continuous at every point in a set X, we say that f is continuous on X. Notice that the definition of continuity given above is equivalent to:

$$f(N_{\delta,X}(x)) \subseteq (N_{\epsilon,X} f(x)) \tag{2.2}$$

Continuity is a critical concept in topology. Let's think of continuity in the following way. Define an incredibly small neighborhood in the image of a function, there where the image exists. Then, pick the point in the range of the function that generates that image. Then, define an arbitrarily small neighborhood around that point in the range. Finally, check what the image of the function is for the extreme values of the neighborhood in the range and check if they are still inside the previously defined neighborhood of the image. If it is inside, the function is continuous; if it is not, the function is not continuous at that point. Notice that this does not prevent the image of the function to vary depending on the slope and neighborhood of the image.

Figure 2.1. Continuity



Example: The map $f : \mathbb{R}_{++} \to \mathbb{R}_{++}$ defined by $f(x) = \frac{1}{x}$ is continuous. To see this, fix any $x > 0$. $\forall y > 0$, it must be that:

$$|f(x) - f(y)| = |\frac{1}{x} - \frac{1}{y}| = \frac{|x-y|}{xy} \tag{2.3}$$

Fix δ small and let $|x - y| < \delta$. By definition of continuity, we need to prove that given δ, there exists a small $\epsilon > 0$ such that $|f(x), f(y)| < \epsilon$. Notice that $\frac{|x-y|}{xy} < \frac{|x-y|}{x(x-\delta)} < \frac{\delta}{x(x-\delta)}$. For a given $\epsilon > 0$, fix $\delta < \frac{\epsilon x^2}{1+\epsilon x}$. Then it must be that $\frac{\delta}{x(x-\delta)} < \epsilon$. Since x is arbitrary, f is continuous.

Proposition 4. *Let $X \subseteq \mathbb{R}$ and let $f : X \to \mathbb{R}$ and $g : f(X) \to \mathbb{R}$ be continuous functions defined, respectively, on X and $f(X)$. Then, the composition function $h = g(f)$ is a continuous function on X.*

Notice that continuity is a local property. Consider a point x in the domain of a function f, and its image $f(x)$ in the co-domain. For any point x, the image of points nearby x under f are close to $f(x)$. That is, given $\epsilon > 0$, we can find $\delta > 0$ such that all points in the δ-neighborhood of x are mapped into the ϵ-neighborhood of $f(x)$. From the definition of continuity, different x might have different δ. To achieve global continuity, a unique δ must exist regardless of what x is.

Figure 2.2. Continuous Function

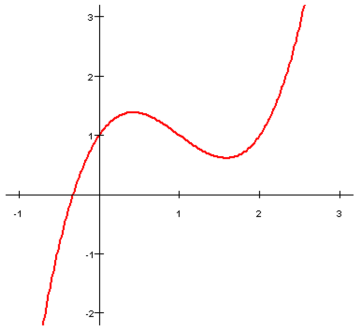
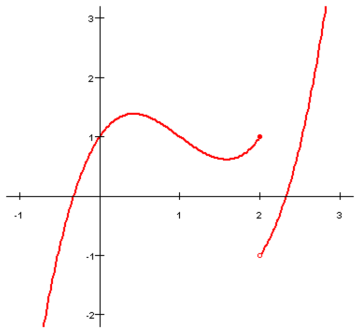


Figure 2.3. Discontinuous Function



Definition 18. *A map $f : X \to \mathbb{R}$ is said to be uniformly continuous if $\forall \epsilon > 0, \exists \delta > 0$, such that:*

$$\forall y, x \in X \text{ such that } d(x, y) < \delta \Longrightarrow d(f(x), f(y)) < \epsilon \tag{2.4}$$

Obviously, a function is not uniformly continuous if it is not continuous $\forall x \in X$. Moreover, a continuous function is not uniformly continuous if there exist at least two different points x in the domain of f for which only different δ-neighborhood can span the same ϵ-neighborhood through the function f. A function is continuous if it is continuous at all the points in the domain. In the same way, a function is uniformly continuous if the neighborhood of the domain is not a function of the value of the domain itself. For a function to be uniformly continuous, it also has to be fully continuous as well.

Proposition 16. *If $f : X \to Y$ is continuous and X is compact, then it is uniformly continuous.*

Proof. Since f is continuous in X, then $\forall\epsilon > 0, \exists\delta_x$ such that $f(N_{\delta,X}(x)) \subset N_{\epsilon,X}(f(x))$. Remember that d_x might vary across x. Notice that $\{N_{\delta_x,X}(x)\}_{x\in X}$ is an open cover of X, and so is $\{N_{\frac{\delta_x}{2},X}(x)\}_{x\in X}$. Since X is compact, X has a finite sub-cover $\{N_{\frac{\delta_{x_i}}{2},X}(x_i)\}_{i=1}^n$. Now, pick two elements $x, y \in X$. For some i, it must be that x lies within $N_{\frac{\delta_{x_i}}{2},X}(x_i)$, therefore within $N_{\delta_{x_i},X}(x_i)$ as well. To have uniform continuity, we need y to lie in $N_{\delta_{x_i},X}(x_i)$ also. Notice that:

$$d(x_i, y) \leq d(x_i, x) + d(x, y) < \frac{d_{x_i}}{2} + \min_i \delta_i \leq \frac{d_{x_i}}{2} + \frac{d_{x_i}}{2} = d_{x_i} \tag{2.5}$$

Since x, y are generic points, this holds for any pair without loss of generality. □

Notice that it is possible to define continuity according to the properties of sequences. In fact, if the domain is a converging sequence to $\bar{x}$ and the function is continuous, the image of the function is converging to the image of $\bar{x}$.

Proposition 5 (**Sequential Characterization**). *A function $f : X \subset \mathbb{R} \to Y \subset \mathbb{R}$ is continuous iff for any sequence $(x_n)_{n\in\mathbb{N}} \subset X^{\mathbb{N}}$ converging to $\bar{x} \in \mathbb{R}$, the sequence $(f(x_n))_{n\in\mathbb{N}}$ converges to $f(\bar{x}) \in \mathbb{R}$.*

The notion of continuity can be characterized further as follows.

Proposition 17. *Let $f : X \subset \mathbb{R} \to Y \subset \mathbb{R}$. The following are equivalent:*

- *f is continuous.*
- *$\forall O \subseteq Y$ open, the set called inverse image, $f^{-1}(O) \subseteq X$, is open.*
- *$\forall S \subseteq Y$ closed, the set $f^{-1}(S) \subseteq X$ is closed.*

Proof. Let's prove that the first statement implies the second.

- **If $f : X \to Y$ is continuous, then $\forall O \subseteq Y$ open, $f^{-1}(O) \subseteq X$ is open**. Consider any open subset $O \subseteq Y$ and any $x \in f^{-1}(O)$. Then, $f(x) \in O$, and since O is open, $\exists\epsilon > 0$ such that $N_{\epsilon,Y}(f(x)) \subseteq O$. By continuity of f, $\exists\delta > 0$ such that $f(N_{\delta,X}(x)) \subseteq N_{\epsilon,Y}(f(x) \subseteq O$, which implies that $N_{\delta,X}(x) \subseteq f^{-1}(O)$, which implies that $f^{-1}(O)$ is open.
- **Therefore, if $\forall O \subseteq Y$ open, the set $f^{-1}(O)$ is open, then $\forall S \subseteq Y$ closed $f^{-1}(S) \subseteq X$ is closed**. If S is a closed subset of Y, $Y\backslash S$ must be open in Y. Then, $f^{-1}(Y\backslash S)$ is an open subset of X. By continuity of f, $f^{-1}(Y\backslash S) = X\backslash f^{-1}(S)$. Therefore, $f^{-1}(S)$ is the complement of an open set, which proves the statement[1].

[1] Analogously, it can be proved that the third statement implies the second one. Finally, the last statement implies the first

- **If $\forall S \subseteq Y$ closed, the set $f^{-1}(S) \subseteq X$ is closed, then f is continuous**. Suppose by contradiction that f were not continuous, everything else equal. Pick a point x in X. If f were not continuous, then for a given $\epsilon > 0$, we would have that $\forall \delta > 0$, such that:

$$\begin{aligned} f(x-\delta, x+\delta) \nsubseteq (f(x)-\epsilon, f(x)+\epsilon) \Longrightarrow N_{\delta,X}(x) = \\ = (x-\delta, x+\delta) \nsubseteq f^{-1}(f(x)-\epsilon, f(x)+\epsilon) \end{aligned} \tag{2.6}$$

Notice however that $(f(x)-\epsilon, f(x)+\epsilon)$ is an open in Y. By assumption, $f^{-1}(f(x)-\epsilon, f(x)+\epsilon)$ must be open in X. This implies that $\forall x \in f^{-1}(f(x)-\epsilon, f(x)+\epsilon), \exists \delta > 0$, such that a δ-neighborhood of x, $(x-\delta, x+\delta)$, belongs to $f^{-1}(f(x)-\epsilon, f(x)+\epsilon)$, which contradicts the above statement. Therefore, it must be that f is continuous. □

Following the characterization of continuity given above, it can be proved that the set $S = \{x \in X : f(x) \leq \alpha\}$ is closed for any continuous function $f : X \to \mathbb{R}$ and $\forall \alpha \in \mathbb{R}$. To see this, consider that S can be written as follows: $S = \{x \in X : f(x) \leq \alpha\} = f^{-1}((-\infty, \alpha])$. The same argument can be used to show that the set $s = \{x \in X : f(x) < \alpha\}$ is open.

Proposition 6 (Intermediate Value Theorem). *If $f : \mathbb{R} \to \mathbb{R}$ is continuous on the closed interval $[a, b]$ and $k \in \mathbb{R}$ is a scalar between $f(a)$ and $f(b)$, then $\exists c \in [a, b]$ such that $f(c) = k$.*

Proposition 18. *Given any $-\infty < a \leq b < \infty$, for any continuous map $f : [a, b] \to [a, b], \exists c \in [a, b]$ such that $f(c) = c$. We call c a fixed point of f.*

Proof. Consider any continuous map f and define a new map $g : [a, b] \to \mathbb{R}$ as follows: $g(t) = f(t) - t$. Since g is the composition of two continuous maps, so it is. Without loss of generality, suppose $g(a) = f(a) - a > 0$ and $g(b) = f(b) - b < 0$. By the intermediate value theorem $\exists z \in [a, b]$ such that $g(z) = 0$. This implies $\exists z \in [a, b]$ such that $f(z) - z = 0$. Therefore, z is a fixed point. □

2.2. The Weierstrass' Theorem

Proposition 19 (Weierstrass' Theorem). *If X is a compact metric space —both closed and bounded— and $f : X \to \mathbb{R}$ is a continuous function, then there exists a maximum and minimum of f in X.*

Proof. Notice that the set spanned by the images of the whole domain X under f, $f(X)$, is closed and bounded by continuity of f. Since $f(X)$ is bounded, both $\sup\{f(x) : x \in S\}$ and $\inf\{f(x) : x \in S\}$ exist. As $\sup\{f(x) : x \in S\}$ and $\inf\{f(x) : x \in S\}$ belongs to $\mathrm{cl}(f(X))$ and $f(X)$ is closed, then they belong to $f(X)$ too. □

Notice that this theorem applies only under these two conditions are met. If the range of the function is not compact —the set is either infinite, like the real line; or it is open—, we cannot know if there is a maximum or a minimum in a different location of the range. If the range is not continuous, there could be vertical asymptotes that would challenge the existence of a maximum or a minimum.

Example: Consider a consumer whose preferences can be represented by a continuous utility function $u : X \to \mathbb{R}$ where $X \subset \mathbb{R}^n$ is the set spanned by the following budget relation: $X = \{x \in \mathbb{R}^n_+ : p_1x_1 + p_2x_2 + ... + p_nx_n \leq m\}$, where $p \in \mathbb{R}^n$ and $m \in \mathbb{R}_+$. Therefore, the consumer maximization problem is:

$$\begin{gathered} \max_x u(x) \text{ such that} \\ p_1x_1 + p_2x_2 + ... + p_nx_n \leq m \\ x \in \mathbb{R}^n_+ \end{gathered} \tag{2.7}$$

is well-defined and has a maximum.

2.3. Hemi-continuity

Functions are single-valued. Given a function $f : X \to Y$, any $x \in X$ is mapped to one and only one point $y \in Y$, the point $y = f(x)$. A correspondence is however a multi-valued function.

Definition 19. *A correspondence is a map* $\Gamma : X \rightrightarrows 2^Y \backslash \emptyset$

This implies that $\forall x \in X, \Gamma(x)$ represents a non-empty replace of Y. The set X is the domain of Γ, Y is the co-domain while $\Gamma(X)$ is the range of Γ, where $\Gamma(S) = \cup\{\Gamma(x) : x \in S\}$. Functions are single-valued correspondences ($|\Gamma(x)| = 1$).

Example: The budget set of a consumer with income $m \in \mathbb{R}_{++}$ and prices $p \in \mathbb{R}^n_{++}$, with $n \in \mathbb{N}$ is given by:

$$\boldsymbol{B}(p, m) = \{x \in \mathbb{R}^n_{++} : \sum_{i=1}^{n} p_ix_i \leq m\} \tag{2.8}$$

If we consider p and m variables in the consumer problem, then $\boldsymbol{B}$ is a correspondence that maps $\mathbb{R}^{n+1}_{++}$ (the space spanned by p and m) into $\mathbb{R}^n_+$ (the space spanned by x).

Unfortunately, the notion of continuity that applies to functions does not apply to correspondences. However, there are two weaker notions of continuity for correspondences, called upper hemi-continuity and lower hemi-continuity, each of which is a partial analogue of continuity of a single-valued function.

Definition 20. *For any two metric spaces X and Y, a correspondence $\Gamma : X \rightrightarrows Y$ is said to be:*

- ***Upper hemi-continuous (UHC)*** *at $x \, \forall X$ if, $\forall O \subset Y$ open, with $\Gamma(x) \subseteq O, \exists \delta > 0$, such that $\Gamma(N_{\delta,X}(x)) \subseteq O$.*
- ***Lower hemi-continuous (LHC)*** *at $x \in X$ if, $\forall O \subset Y$ open, with $\Gamma(x) \cap O \neq \emptyset, \exists \delta > 0$, such that $\Gamma(x') \cap O \neq \emptyset \, \forall x' \in N_{\delta,X}(x)$.*

- ***Continuous*** *at $x \in X$ if it is both UHC and LHC at $x \in X$.*

Γ is called UHC(LHC) if it is UHC(LHC) at each $x \in X$. The property of upper (lower) hemi-continuity tells that a small perturbation of x does not cause the image set to suddenly explode (implode). A further characterization of upper and lower continuity can be made in terms of upper and lower inverse image (as for continuous single-valued functions).

Proposition 7. *Let X and Y be two metric spaces and $\Gamma : X \rightrightarrows Y$ a correspondence.*

- *Define the upper inverse image of O under Γ by $\Gamma^{-1}(O) = \{x \in X : \Gamma(x) \subseteq O\}$. Then, Γ is UHC $\iff \Gamma^{-1}(O)$ is open for every open subset O of Y.*

- *Define the lower inverse image of O under Γ by $\Gamma^{-1}(O) = \{x \in X : \Gamma(x) \cap O \neq \emptyset\}$. Then, Γ is LHC $\iff \Gamma^{-1}(O)$ is open for every open subset O of Y.*

The notion of correspondence is tightly linked to the notion of graph. The graph of a correspondence $\Gamma : X \rightrightarrows Y$, denoted by $Gr(\Gamma)$ is defined by the following set:

$$\mathrm{Gr}(\Gamma) := \{(x, y) \in X \times Y | y \in \Gamma(x)\} \tag{2.9}$$

It follows that a correspondence Γ has a closed (open) graph if $\mathrm{Gr}(\Gamma)$ is a closed (open) subset of $X \times Y$.

Proposition 8. *A correspondence $\Gamma : X \rightrightarrows Y$ with closed values (i.e. $Gr(\Gamma)$ closed $\forall x \in X$), closed domain and compact range, is upper hemi-continuous iff it has a closed graph. If Γ has an open graph, then it is lower hemi-continuous.*

2.4. The Maximum Theorem

Proposition 9 (**Maximum Theorem**). *Let Θ and X be two metric spaces, and $\Gamma : \Theta \rightrightarrows X$ a compact-valued correspondence and let $f : X \times \Theta \to \mathbb{R}$ a continuous function. Define:*

$$\sigma(\theta) := \arg\max\{f(x, \theta) : x \in \Theta(x)\} \, \forall \theta \in \Theta \tag{2.10}$$

and

$$f^*(\theta) := \max\{f(x, \theta) : x \in \Theta(x)\} \, \forall \theta \in \Theta \tag{2.11}$$

Assume that Γ is continuous (both UHC and LHC) at some point $\theta \in \Theta$. Then:

- *$\sigma(\theta) : \Theta \rightrightarrows X$ is non-empty, compact-valued and upper hemi-continuous at θ.*

- *$f^* : \Theta \times X \to \mathbb{R}$ is continuous at $\theta \in \Theta$.*

Example: Consider the consumer problem presented before:

$$\max_x u(x) \text{ such that } x \in \boldsymbol{B}(p, m) = \{x \in \mathbb{R}^n_{++} : \sum_{i=1}^n p_i x_i \leq m\} \tag{2.12}$$

Parameters p and m determine the constraint set and this the optimal choice and the utility level achieved. Now, consider the demand correspondence:

$$\boldsymbol{d}(p, m) := \arg\max\{u(x) : x \in \boldsymbol{B}(p, m)\} \tag{2.13}$$

By the Weierstrass theorem, the demand correspondence is well-defined. Since the objective function is assumed to be continuous and the set spanned by the budget correspondence to be closed and bounded, a solution to this maximization problem exists. Assume $\boldsymbol{B}(p, m)$ is a continuous correspondence. Then, by the Maximum Theorem, $\boldsymbol{d}(p, m)$ is compact-valued and upper hemi-continuous. Now, consider the indirect utility function:

$$u^*(p, m) = \max\{u(x) : x \in \boldsymbol{B}(p, m)\} \tag{2.14}$$

The Maximum Theorem guarantees that u^* is a continuous function.

3 Convexity

3.1. Convex Sets

Consider the real space $\mathbb{R}^n$. For any two points $x, y \in \mathbb{R}^n$, the set $\{\lambda x + (1-\lambda)y : \lambda \in \mathbb{R}\}$ is called line through x and y. Similarly, the set $\{\lambda x + (1-\lambda)y : \lambda \in [0,1]\}$ is called segment between x and y. Moreover, $\forall m \in \mathbb{N}$ and $\forall x^1, ..., x^m \in \mathbb{R}^n$, their convex combination are points of the following form:

$$\sum_{i=1}^{m} \lambda_i x^i \tag{3.1}$$

where $\lambda_i > 0 \, \forall i$ and $\sum_{i=1}^{m} \lambda_i = 1$. Thus, the segment between x and y is the set of all convex combinations of x and y. A subset $S \subset \mathbb{R}^n$ is said to be convex if $\forall x, y \in S$ and $\forall \lambda \in [0,1]$ we have that

$$\lambda x + (1-\lambda)y \in S \tag{3.2}$$

Figure 3.1. Convex Set

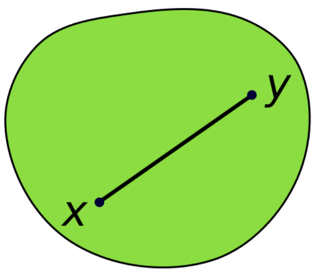
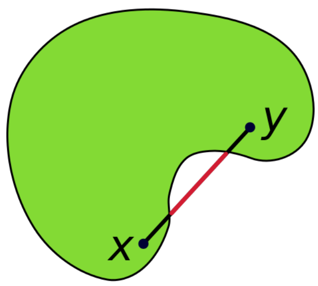


Figure 3.2. Not Convex Set



Proposition 20. *A set S is convex IFF every combination of points of S lies in S.*

Proof. Suppose every combination of points in S lie in S. If so, this would hold for each pair of points in S, and in particular for each points on the segment between pairs of points in S, implying convexity. Suppose now S is convex. This guarantees that segments between $m = 2$ points in S lie in S. By induction, suppose proved that this holds for a generic m. Consider the case of $m + 1$ points. Pick $z = (z_1, z_2, ..., z_n)$ where

$z_j = \sum_{i=1}^{m+1} \lambda_i x_j^i$ and $\lambda_i > 0$. Then:

$$\begin{gathered} z_j = \lambda_1 z_j^1 + \lambda_2 x_j^2 + ...\lambda_{m+1} x_j^{m+1} = \\ \left(\sum_{i=1}^{m} \lambda_i\right)\left(\frac{\lambda_1}{\sum_{i=1}^{m} \lambda_i} z_j^1 + \frac{\lambda_2}{\sum_{i=1}^{m} \lambda_i} x_j^2 + ... \frac{\lambda_{m-1}}{\sum_{i=1}^{m} \lambda_i} x_j^m\right) + \lambda_{m+1} x_j^{m+1} = \\ \left(\sum_{i=1}^{m} \lambda_i\right) z_j^{'} + \lambda_{m+1} x_j^{m+1} \end{gathered} \tag{3.3}$$

which is a segment between $z^{'}$ and x_j^{m+1}, both belonging to S. By convexity, the segment is in S too. This proves the statements. □

Definition 21. *Given a set $S \in \mathbb{R}^n$, the smallest convex set $\mathbb{R}^n$ that contains S is called convex hull of S and is denoted by $co(S)$.*

A convex hull of S is the smallest convex set that includes S in it. Analogously, a convex hull of S is the set of all convex combinations of points in the set S, meaning:

$$co(S) = \{\theta_1 x_1 + ... + \theta_k x_k : x_i \in S, \theta_i \geq 0, \sum_{i=1}^{k} \theta_i = 1\} \tag{3.4}$$

Notice that any closed convex set can be written as the convex hull of itself. Moreover, a number of properties of convex sets are useful to mention. For instance:

- The intersection of any collection of convex sets is a convex set.
- Let S and T be convex subsets in $X \subseteq \mathbb{R}^n$ and α and β be real numbers. Then, the set $Z = \alpha S + \beta T = \{z \in X : \alpha x + \beta y,\, x \in S, y \in T\}$ is convex.

3.2. Convex Functions

Definition 22. *Let T be a non-empty convex subset of $\mathbb{R}^n$. A real map $f : T \to \mathbb{R}$ is said to be concave if $\forall x, y \in T$ and $\lambda \in [0, 1]$, then:*

$$f(\lambda x + (1-\lambda)y) \geq \lambda f(x) + (1-\lambda) f(y) \tag{3.5}$$

We say that the function is convex if $-f$ is concave.

Example: A function $f : x \to \mathbb{R}$ is concave iff:

- $f(\lambda(z - x) + x) \geq \lambda(f(z) - f(x)) + f(x)$, $\forall x, z \in X$, $\forall \lambda \in (0, 1)$.
- $f(\lambda \Delta x + x) \geq \lambda(f(\Delta x + x) + f(x))$, $\forall x, (x + \Delta x) \in X$, $\forall \lambda \in (0, 1)$

Figure 3.3. Convex Function

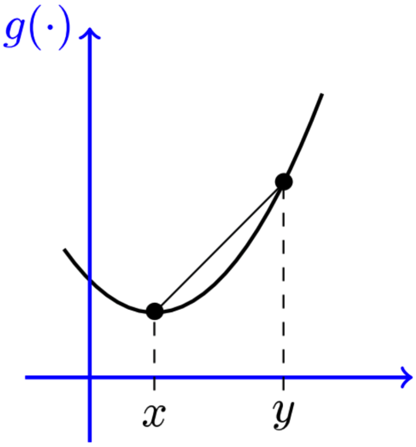
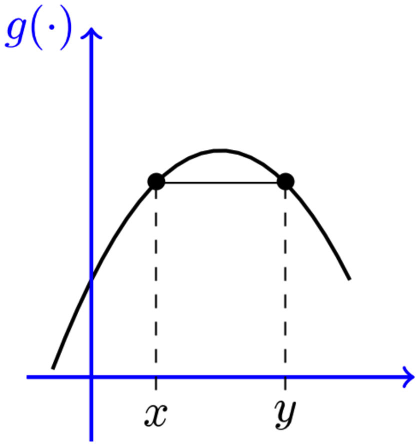


Figure 3.4. Concave Function



Proposition 21 (Jensen Theorem). *Suppose that $f : X \to \mathbb{R}$ is concave. If $x^1, ..., x^m \in X$ and $\alpha_1, ..., \alpha_m \in \mathbb{R}_+$ with $\sum_{i=1}^{m} \alpha_i = 1$, then:*

$$f\left(\sum_{i=1}^{m} \alpha_i x^i\right) \geq \sum_{i=1}^{m} \alpha_i f(x^i) \tag{3.6}$$

Proof. Let's prove this theorem by induction. Consider the case of $m = 2$. Then:

$$f(\alpha_1 x^1 + \alpha_2 x^2) \geq \alpha_1 f(x^1) + \alpha_2 f(x^2) \tag{3.7}$$

is true by definition of concavity. Suppose it is true also for a generic m, then we need to prove it for $m + 1$. Notice that $\alpha_i > 0$ for at least one i. Without loss of generality, set $\alpha_1 > 0$. Then, by definition of convexity:

$$\begin{gathered} f\left(\sum_{i=1}^{m+1} \alpha_i x_i\right) = \\ f(\alpha_1 x_1 + (1-\alpha_1) \sum_{i=2}^{m+1} \frac{\alpha_i}{(1-\alpha_1)} x_i) \leq \alpha_1 f(x_1) + (1-\alpha_1) f\left(\sum_{i=2}^{m+1} \frac{\alpha_i}{(1-\alpha_1)} x_i\right) \end{gathered} \tag{3.8}$$

Since $\sum_{i=2}^{m+1} \frac{\alpha_i}{(1-\alpha_1)} = 1$, induction applies and the statement of the theorem holds. □

Proposition 22. *Let $f : [a, b] \subseteq \mathbb{R} \to \mathbb{R}$ be a continuous and convex function. Let $g : \mathbb{R} \to \mathbb{R}$ be a continuous and increasing function, such that $\forall x \geq y \in \mathbb{R} \implies f(x) \geq f(y)$. Then, $g \cdot f$ is a convex function.*

Proof. Since f is convex, then:

$$\begin{gathered} f(\alpha x + (1-\alpha) y) \leq \alpha f(x) + (1-\alpha) f(y), \\ \forall x, y \in [a, b],\ \forall \alpha \in (0, 1) \end{gathered} \tag{3.9}$$

Since g is increasing, then:

$$(g \cdot f)(\alpha x + (1-\alpha)y) \leq g(\alpha f(x) + (1-\alpha)f(y)), \\ \forall x, y \in [a,b], \forall \alpha \in (0,1) \tag{3.10}$$

Since g is convex, then:

$$g(\alpha f(x) + (1-\alpha)f(y)) \leq g(\alpha f(x)) + g((1-\alpha)f(y)), \\ \forall x, y \in [a,b], \forall \alpha \in (0,1) \tag{3.11}$$

Since:

$$g(\alpha f(x)) + g((1-\alpha)f(y)) = \alpha(g \cdot f)(x) + (1-\alpha)(g \cdot f)(y), \\ \forall x, y \in [a,b], \forall \alpha \in (0,1) \tag{3.12}$$

Then:

$$(g \cdot f)(\alpha x + (1-\alpha)y) \leq \alpha(g \cdot f)(x) + (1-\alpha)(g \cdot f)(y), \\ \forall x, y \in [a,b], \forall \alpha \in (0,1) \tag{3.13}$$

Which proves convexity of $g \cdot f$. □

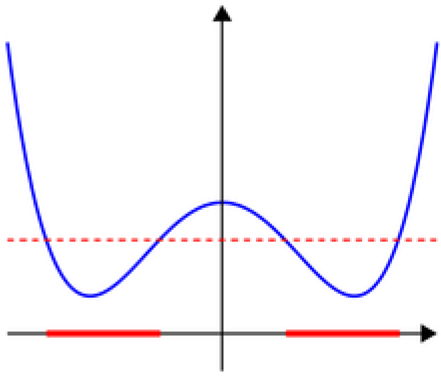
Definition 23. *Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$. Assume U is convex. Then:*

- ***Epigraph** of f is defined as: $epif := \{(x,y) \in \mathbb{R}^{n+1} : x \in U, y \geq f(x)\}$.*
- ***Hypograph** of f is defined as: $hypf := \{(x,y) \in \mathbb{R}^{n+1} : x \in U, y \leq f(x)\}$.*

Notice that any function $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ is convex iff its epigraph is a convex set. Conversely, f is concave iff its hypograph is a convex set. Moreover, we define the:

- ***Upper contour set** of f at a point $\alpha \in \mathbb{R}$ as $U_f(\alpha) = \{x \in U : f(x) \geq \alpha\}$.*
- ***Lower contour set** of f at a point $\alpha \in \mathbb{R}$ as $L_f(\alpha) = \{x \in U : f(x) \leq \alpha\}$.*

Figure 3.5. Lower Contour Sets



We can use the epigraph and hypograph to figure out if a function is convex. Basically:

- If the epigraph is convex, the function is convex.
- If the hypograph is convex, the function is concave.
- If the neither the hypograph or the epigraph are convex, the function is neither convex or concave.

Actually, the lower and the upper contour set can also be used with the same purpose:

- If the upper contour set is convex, the function is concave.
- If the lower contour set is convex, the function is convex.
- If neither the lower or the upper contour set is convex, the function is neither convex or concave.

Definition 24. *Let T be a non-empty convex subset of $\mathbb{R}^n$.*

- *A real map $f : T \to \mathbb{R}$ is said to be quasi-concave if $\forall x, y \in T$ and $\lambda \in [0, 1]$, then:*

$$f(\lambda x + (1 - \lambda)y) \geq \min\{f(x), f(y)\} \tag{3.14}$$

We say that a function is quasi-convex if

$$f(\lambda x + (1 - \lambda)y) \leq \min\{f(x), f(y)\} \tag{3.15}$$

- *If the inequality is strict $\forall x \neq y$ then f is strictly quasi-concave (convex).*

Notice that a function f is said to be quasi-concave if $U_f(\alpha)$ is a convex set $\forall \alpha \in \mathbb{R}$. By the same token, a function f is said to be quasi-convex if $L_f(\alpha)$ is a convex set $\forall \alpha \in \mathbb{R}$.

Proposition 23. *Let $f : \mathbb{R} \to \mathbb{R}$ be an increasing function. The f is both quasi-concave and quasi-convex.*

Proof. To show this, consider $x, y \in \mathbb{R}, \lambda \in (0, 1)$. Assume,without loss of generality, that $x > y$. Then:

$$x > \lambda x + (1 - \lambda)y > y \tag{3.16}$$

Since f is increasing, then:

$$f(x) \geq f(\lambda x + (1 - \lambda)y) \geq f(y) \tag{3.17}$$

Since $f(x) = \max\{f(x), f(y)\}$, then, from the first inequality, $f(y) \leq \max\{f(x), f(y)\}$, which proves quasi-convexity. Since $f(y) = \min\{f(x), f(y)\}$, then, from the second inequality, $f(x) \geq \min\{f(x), f(y)\}$, which proves quasi-concavity. $\square$

Proposition 10. *Consider the Maximum Theorem stated above. In addition to the listed conditions, assume:*

- *f is quasi-concave in its argument $x\,\forall\theta \in \Theta$.*
- *Θ is convex-valued.*

Then, $\sigma(\theta)$ is convex-valued as well. Moreover, if f is strictly quasi-concave in its argument $x\,\forall\theta \in \Theta$, then $\sigma(\theta)$ is single-valued, and, thus, it is a continuous function rather than a correspondence.

4 Differentiability

4.1. Derivatives

Consider the case of a function f defined on the reals.

Definition 25. *The derivative of a function $f : \mathbb{R} \to \mathbb{R}$ at a point $x \in \mathbb{R}$ is usually defined as a real number that describes the instantaneous change of the value of f as x changes:*

$$f^{'}(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \tag{4.1}$$

Definition 26. *A point x_0 is said to be a local maximum or maximizer if there exists $\delta > 0$ such that $f(x_0) \geq f(x) \forall x \in N_{\delta,\mathbb{R}}(x_o)$, that is, $\forall x \in (x_0 - \delta, x_0 + \delta)$. A local minimum or minimizer is defined in an analogous way.*

Proposition 24. *Let O be an open subset of $\mathbb{R}$, and let $f : O \to \mathbb{R}$ a function with a derivative at x, and let x_0 be a local maximum (minimum). Then, $f^{'}(x) = 0$.*

Proof. Let x_0 be a local maximum. Then, $f(x_0 + h) - f(x_0) \leq 0 \forall h > 0$ small enough, $(|h| < \delta)$. Notice that we have:

$$\frac{f(x_0 + h) - f(x_0)}{h} \leq 0 \forall h \in (0, \delta) \tag{4.2}$$

Taking limits as h goes to 0 from the right, it must be that:

$$\lim_{h \to 0^+} \frac{f(x_0 + h) - f(x_0)}{h} \leq 0 \tag{4.3}$$

Moreover, we have:

$$\frac{f(x_0 + h) - f(x_0)}{h} \geq 0 \;\; \forall h \in (-\delta, 0) \tag{4.4}$$

Taking limits as h goes to 0 from the left, it must be that:

$$\lim_{h \to 0^-} \frac{f(x_0 + h) - f(x_0)}{h} \geq 0 \tag{4.5}$$

Since the function has a derivative at x then it must be that:

$$\lim_{h \to 0^+} \frac{f(x_0 + h) - f(x_0)}{h} = \lim_{h \to 0^-} \frac{f(x_0 + h) - f(x_0)}{h} = f^{'}(x) = 0 \tag{4.6}$$

□

Proposition 25 (Roll Theorem). *Let $f : [a, b] \to \mathbb{R}$ be continuous and have a derivative on (a, b). If $f(a) = f(b) = 0$, then $\exists x \in (a, b)$ such that $f^{'}(x) = 0$.*

Proof. If f is continuous on the compact set $I = [a, b]$, then, by the Weierstrass Theorem, we know that the function has a well-defined maximum x_M and a minimum x_m in I. If $f(x_M) = f(x_m) = 0$, then the function must be constant and $f^{'}(x) = 0 \forall x \in I$. If this is not the case, then either $f(x_m) < 0$ or $f(x_M) > 0$. In the first case, it must be that x_m lie in the interior of $I, x_m \in (a, b)$, which implies $f^{'}(x_m) = 0$. The second case is analogous. ☐

Proposition 26 (Mean Value Theorem). *Let $f : [a, b] \to \mathbb{R}$ be continuous and have a derivative on (a, b). Then, $\exists x \in (a, b)$ such that $f^{'}(x) = \frac{f(b)-f(a)}{b-a}$.*

Proof. Define the function $g : [a, b] \to \mathbb{R}$ as:

$$g(x) = f(x) - f(a) - \frac{f(b) - f(a)}{b - a}(x - a) \tag{4.7}$$

It follows that:

- $g(a) = g(b) = 0$.
- g is continuous.
- g has a derivative on (a, b).

Then, $\exists x \in (a, b)$ such that $g^{'}(x) = 0$. That is, $g^{'}(x) = f^{'}(x) - \frac{f(b)-f(a)}{b-a} = 0$ and thus, $f^{'}(x) = \frac{f(b)-f(a)}{b-a}$. ☐

Consider now the case of a function defined on $\mathbb{R}^n$.

Definition 27. *The directional derivative of $f : \mathbb{R}^n \to \mathbb{R}$ in the direction of u at point x is defined by:*

$$Df(x, u) = \lim_{\alpha \to 0} \frac{f(x + \alpha u) - f(x)}{\alpha} \tag{4.8}$$

Where $\alpha \in \mathbb{R}$ and $\parallel u \parallel = 1$, whenever the limit exists.

Notice that partial derivatives are specific cases of directional derivatives or, in different words, they are a directional derivative in a single direction (dimension).

Definition 28. *The partial derivative of f with respect to the $i - th$ argument, x_i, is defined by:*

$$f_i(x) = D_{x_i} f(x) = Df(x, u_i) \tag{4.9}$$

With $u_i = (0, ..., 0, 1, 0, ..., 0)$

If all the partial derivatives of f exist and are continuous in some neighborhood of x_0, then all the directional derivatives exist and can be written as a linear combination of the partial derivatives:

$$Df(x_0, u) = \sum_{i=1}^{n} f_i(x_0) u_i \tag{4.10}$$

4.2. Differentiability

On the one hand, the derivative of a function $f : \mathbb{R} \to \mathbb{R}$ at a point $x \in \mathbb{R}$ is defined as a real number that describes the instantaneous change of the value of f as x changes. On the other hand, the notion of derivative is tightly linked to the line that best approximates f near x.

Definition 29. *Let f be a function from a vector space $V \subseteq \mathbb{R}^n$ to another vector space $W \subseteq \mathbb{R}^n$. We say that f is a linear map if for any two vectors $x, y \in V$ and any scalar $\alpha \in \mathbb{R}$, we have:*

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

Example: Let O be an open set $\mathbb{R}, x \in \mathbb{R}$ and $f : O \to \mathbb{R}$. The derivative of f at x exists $\iff$ there exists a linear map L on a neighborhood of x that approximates f near x. To see this, notice that:

$$\begin{aligned} f^{'}(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} = \\ 0 = \lim_{h \to 0} \frac{f(x+h) - f(x) - f^{'}(x)h}{h} \\ 0 = \lim_{t \to x} \frac{f(t) - f(x) - f^{'}(x)(t-x)}{|t-x|} \end{aligned} \tag{4.11}$$

Therefore, we can only consider the map L such that $L(t) = f(x) - f^{'}(x)(t - x)$. A function f is said to be differentiable at x if there exists such a linear map that approximates $f(x)$ near x. For functions on the reals, this is equivalent to the existence of a derivative: the linear map will be the one intersecting $f(x)$ with the slope $f^{'}(x)$. For functions defined on different spaces, the definition of differentiability is more general.

Definition 30. *A function $f : O \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is said to be differentiable at $x \in O$ if there exists a matrix A_x, called Jacobian, such that:*

$$\lim_{\|h\| \to 0} \frac{\| f(x+h) - f(x) - A_x h \|}{\| h \|} = 0 \tag{4.12}$$

Notice the following:

- *The derivative $Df : O \to \mathbb{R}^{m \times n}$ assigns a matrix A_x to each x.*
- *The differential $df_x : \mathbb{R}^n \to \mathbb{R}^m$ is a mapping that assigns to each x the linear operator $df_x(h) = A_x h$.*

In different words, the Jacobian is a matrix of partial derivatives. In addition, it is relevant to see that, in a single dimension space, differentiability implies the existence of a derivative; while in multidimensional spaces, differentiability implies the existence of a Jacobian. Notice that $f : O \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is differential at x iff every component $f^1, .., f^m$ is differentiable at x. Moreover, if f is differentiable, the partial derivatives exist at x, and the derivatives of f at x is the matrix of first partial derivatives of the component functions evaluated at x: $Df(x) = [Df^1(x), ..., Df^m(x)]'$. Finally, notice that if all functions in the Jacobian are continuous at x, $f(x)$ is differentiable. Thus differentiability at x implies continuity at x.

Proposition 11. *Consider the function $f : O \subseteq \mathbb{R}^n \to \mathbb{R}^m$. If f is differentiable at $x \in O$, then f is continuous at x. Suppose f has well-defined partial derivatives and those are continuous. Then, f is differentiable.*

Definition 31. *A function $f : O \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is said to be continuously differentiable and Df is a continuous function. Therefore, a function is continuous differentiable iff the partial derivatives of the component function of f exist and are continuous.*

4.3. Inverse Function Theorem

Proposition 12. *Let $f : O \subseteq \mathbb{R}^n \to \mathbb{R}^n$ be a continuously differentiable function and let $x^0 \in O$. If the determinant of the Jaccobian of f is not equal to zero at x^0, i.e. $|Df(x^0)| \neq 0$, then there exists an open neighborhood of x^0, such that:*

- *f is one-to-one function in U and f^{-1} is well-defined.*
- *$V = f(U)$ is an open set containing $f(x^0)$.*
- *f^{-1} is continuously differentiable with $D(f^{-1}(x^0)) = (Df(x^0))^{-1}$.*

Example: Let $f : O \times \Sigma \subseteq \mathbb{R}^{n+p} \to \mathbb{R}^m$, with $O \times \Sigma$ open set. Consider the model $F(x, \alpha) = \mathbf{0}$. The solution is a correspondence mapping the parameter space into the choice space, $S : \Sigma \subseteq \mathbb{R}^p \rightrightarrows O \subseteq \mathbb{R}^n$; i.e., assigning to each parameter vector α, the set of elements $S(\alpha)$ that solves the following equation:

$$S(\alpha) := \{x \in O : F(x, \alpha) = \mathbf{0}\} \tag{4.13}$$

How is the solution for a given parameter α? Suppose $n = m$. Then $f_\alpha = F(x, \alpha)$ can be thought as a function mapping $\mathbb{R}^n$ into itself. If, for a certain point (x^0, α^0), we have:

- $F(x^0, \alpha^0) = \mathbf{0}$.
- F is continuously differentiable.
- The Jacobian of $f_{\alpha^0}, |D_x F(x, \alpha^0)|$ is not zero at $x = x^0$.

Then, f_{α^0} is a one-to-one map in some neighborhood of x^0 and therefore x^0 is a locally unique solution of the system $F(x, \alpha^0) = \mathbf{0}$.

4.4. Implicit Function Theorem

How does the solution change with a change of parameters? For each α we might find different solution sets (maybe empty). By restricting the analysis to sufficiently small neighborhood of α^0, as F is continuously differentiable, f_α will be sufficiently close to the original one, thus making the solution close to x^0. Moreover, since f_α is locally invertible, the solution will be unique. Suppose that the solution correspondence is a differentiable function, i.e., $x^* = x(\alpha) = x(\alpha_1, ..., \alpha_p)$. How would the solution x^* change for small changes of α_k? Consider the solution function: $F[x(\alpha), \alpha] = \mathbf{0}$, or equivalently:

$$
\begin{gathered}
F^1[x_1(\alpha), ..., x_n(\alpha), \alpha] = 0 \\
. \\
. \\
. \\
F^n[x_1(\alpha), ..., x_n(\alpha), \alpha] = 0
\end{gathered} \tag{4.14}
$$

By differentiating with respect to α_k, we have:

$$
\begin{gathered}
F^1_{x_1}\frac{\partial x_1(\alpha)}{\partial \alpha_k} + ... + F^1_{x_n}\frac{\partial x_n(\alpha)}{\partial \alpha_k} + F^1_{\alpha_k} = 0 \\
. \\
. \\
. \\
F^n_{x_1}\frac{\partial x_1(\alpha)}{\partial \alpha_k} + ... + F^n_{x_n}\frac{\partial x_n(\alpha)}{\partial \alpha_k} + F^n_{\alpha_k} = 0
\end{gathered} \tag{4.15}
$$

This can be written as:

$$
J\left[\frac{x(\alpha)}{\alpha_k}\right]' = -\left[F^1_{\alpha_k}, ..., F^n_{\alpha_k}\right] \tag{4.16}
$$

If the Jacobian is invertible ($|J| \neq 0$), then the system can be solved for the th partial derivatives of solution.

Proposition 13. *Let $F : O \times \Sigma \subseteq \mathbb{R}^{n+p} \to \mathbb{R}^n$ be a continuously differentiable function defined on an open set $O \times \Sigma$. Consider the system of equations $F(x, \alpha) = \mathbf{0}$, and assume that it has solution x^0 for a parameter α^0. If the determinant of the Jacobian is not zero at (x^0, α^0), then:*

- $\exists U \subseteq \mathbb{R}^{n+p}, U_\alpha \subseteq \mathbb{R}^p$, *both open, with* $(x^0, \alpha^0) \in U, \alpha^0 \in U_\alpha$ *such that* $\forall \alpha \in U_\alpha, \exists! x_\alpha$ *with* $(x_\alpha, \alpha) \in U, F(x_\alpha, \alpha) = \mathbf{0}$.
- *The solution function $x : U_\alpha \to \mathbb{R}^n$ defined by $x(\alpha) = x_\alpha$ is continuously differentiable, with derivative given by $-(D_x F(x_\alpha, \alpha))^{-1} D_\alpha F(x_\alpha, \alpha)$.*
- *If f is of k-order continuously differentiable, so is the function x.*

5 Static Optimization

Consider the following non-linear optimization program:

$$\begin{aligned} &\max_x f(x,\theta) \text{ such that} \\ &x \in \Gamma(\theta) \subseteq A \subseteq \mathbb{R}^n \\ &\theta \in \Theta \subset \mathbb{R}^p \\ &f : A \times \Theta \to \mathbb{R} \end{aligned} \tag{5.1}$$

In what follows, let's analyse three different cases:

- **Convex constraint set**: $\Gamma(\theta)$ is a convex subset of $\mathbb{R}^n$.
- **Lagrange problem**: $\Gamma(\theta) = \{x \in A : g(x,\theta) = \mathbf{0}\}$.
- **Kuhn-Tucker problem**: $\Gamma(\theta) = \{x \in A : g(x,\theta) \leq \mathbf{0}\}$

5.1. Convex Constraint Set

Let S be a convex set in $\mathbb{R}^n$, $f : S \subseteq \mathbb{R}^n \to \mathbb{R}$ and consider the problem:

$$\max_x \{f(x) : x \in S\} \tag{5.2}$$

Suppose S is an open set. Then, a necessary condition for the optimum x^* is:

$$Df(x^*) = \mathbf{0} \tag{5.3}$$

However, in the more general case, this condition is neither necessary nor sufficient. For instance, suppose S were a closed set. If a maximum happens to be on the boundary of S, some partial derivatives may not be equal to zero. As the value of the function has to decrease for different values, directional derivatives in feasible directions need to be non-positive.

Definition 32. *Given a convex set $S \subset \mathbb{R}^n$ and $x \in \mathbb{R}$, we say that $h \in \mathbb{R}^n$ is a feasible direction from x if $\exists \delta > 0$ such that $(x + \alpha h) \in S$, $\forall \alpha \in (0, \delta)$.*

Proposition 27. *Suppose that f is continuously differentiable and let x^* be a solution of the optimization problem. Then, $Df(x^*)h \leq 0$ for every feasible direction $h \in \mathbb{R}^n$ from x^*.*

Proof. Let x^* be a solution of the problem. Let h be any feasible direction from x^*. Then, $\exists \delta > 0$ such that $(x^* + \alpha h) \in S$, $\forall \alpha \in (0, \delta)$. Since $f(x^* + \alpha h) \leq f(x^*)\, \forall h \in \mathbb{R}^n$, re-arranging and dividing by $\alpha > 0$, we get $\frac{f(x^*+\alpha h)-f(x^*)}{\alpha} \leq 0$. Taking the limit for $\alpha \to 0$, we have:

$$\lim_{\alpha \to 0} \frac{f(x^* + \alpha h) - f(x^*)}{\alpha} = Df(x^*, h) = Df(x^*)h \leq 0 \tag{5.4}$$

$\square$

Proposition 28. *Suppose that f is continuously differentiable and that x^* is a solution of a problem with an open constraint set. Then, $Df(x^*)h = 0$, i.e., all the partial derivatives are 0.*

Proof. In an open set, all directions are feasible. Then, for some direction $h, Df(x^*)h < 0$, while for other directions $-h, Df(x^*)h > 0$. This is a contradiction, implying $Df(x^*)h = 0$ $\square$

A solution to this problem can be characterized as follows.

Proposition 14. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable. Then:*

- *If f has a local maximum at x^*, then the Hessian matrix of partial derivatives $D^2f(x)$, evaluated at x^*, is negative semi-definite, i.e.: $h^{'}D^2f(x^*)h \leq 0 \, \forall h \in \mathbb{R}^n$*
- *If $x^* \in S$ open and convex, such that $Df(x^*) = \mathbf{0}$, then the Hessian matrix is negative definite, i.e.: $h^{'}D^2f(x^*)h < 0 \, \forall h \in \mathbb{R}^n$.*

And f achieves a strict local maximum at x^*.

We refer to a maximization problem as a convex maximization problem if:

- The constraint set is convex.
- The objective function is concave.

Similarly, a minimization problem is said to be convex if the objective function is convex. Let $X \subset \mathbb{R}^n$ be a convex set and let $f : X \to \mathbb{R}$ be concave. Then:

- Any local maximum of f is a global maximum of f.
- The set $\sigma(\theta) := \arg\max\{f(x) : x \in X\}$ is either empty or convex.

Similar results hold of convex minimization problems. Moreover, the second part of the result implies that there cannot be multiple isolated points as maximisers.

5.2. The Lagrange Problem

Consider the maximization of $f(x)$ subject to n building constraints $g_j(x) = 0$. Suppose we impose a penalizationλ_j for deviating from each constraint g_j, i.e.

$$L(x, \lambda) = f(x) + \sum_{j=1}^{n} \lambda_j g_j(x) \tag{5.5}$$

For a given λ_j we can consider the following first order conditions:

$$\begin{aligned}\frac{\partial L}{\partial x} &= \frac{\partial f}{\partial x} + \sum_{j=1}^{n} \lambda_j \frac{\partial g_j}{\partial x} = 0 \\ \frac{\partial L}{\partial \lambda_j} &= g_j(x) = 0 \, \forall j = 1, ..., n\end{aligned} \tag{5.6}$$

The last n equation guarantees we choose $\forall j = 1, ..., n$ the correct λ_j^* such that the constraint is satisfied. The $n+1$ equations determine stationary points of the Lagrangian function.

Proposition 15. *Let x^* be a solution of the Lagrangian program. Let f and $\{g_j\}_{j=1}^n$ be continuously differentiable functions and assume x^* is a regular points; that is, assume all the binding constraints evaluated at x^* are linearly independent, or, equivalently:*

$$D_g(x^*) = \begin{bmatrix} Dg_1(x^*) \\ . \\ . \\ . \\ Dg_n(x^*) \end{bmatrix}$$

have linearly independent rows. Then:

$$\exists! \{\lambda_j^*\}_{j=1}^n \in \mathbb{R}^c : Df(x^*) + \sum_{j=1}^{n} \lambda_j^* Dg_j(x^*) = 0 \tag{5.7}$$

Proposition 16. *Let f be concave and $\{g_i\}_{i=1}^n$ quasi-concave. If (x^*, λ^*) satisfy the first order conditions, with x^* feasible and $\lambda^* \geq 0$, then x^* is a solution to the Lagrange problem.*

5.3. The Kuhn-Tucker Problem

Consider the maximization of $f(x)$ subject to n constraints $g_j(x) \geq 0$, where both f and $\{g_j\}_{j=1}^n$ are continuously differentiable functions from $\mathbb{R}^n$ to $\mathbb{R}$. The constraints can be:

- **Binding**, or active, at a feasible point x^0 if they hold with equality, in which case we are in the Lagrangian situation.
- **Inactive** otherwise, in which case they won't have effects on the local properties of the solution.

To solve this problem, we proceed as if we wanted to maximise $L(x, \lambda) = f(x) + \sum_{j=1}^n \lambda_j g_j(x)$ with respect to x (no constraints) and with respect to $\{\lambda_j\}_{j=1}^n$ subject to $\lambda_j \geq 0 \, \forall j = 1, ..., n$.

Then, the first order conditions characterize the optimum solution s follows:

- $D_x L(x, \lambda) = Df(x) + \sum_{j=1}^n \lambda_j Dg_j(x) = 0$ for the endogenous variables.

- $D_{\lambda_j} L(x, \lambda) = g_j(x) \geq 0$ with $g_j(x) = 0$ if $\lambda_j > 0$ or $\lambda_j = 0$ if $g_j(x) > 0$.

Notice that the non-negativity program can be expressed as:

$$\begin{aligned} \lambda_j &\geq 0 \, \forall j = 1, ..., n \\ g_j(x) &\geq 0 \, \forall j = 1, ..., n \\ \lambda_j g_j(x) &= 0 \, \forall j = 1, ..., n \end{aligned} \tag{5.8}$$

Proposition 17. *Let x^* be a solution of the Kuhn-Tucker problem. Let f and g_j, $\forall j = 1, ..., n$ be continuously differentiable and suppose that the constraint qualification is satisfied, that is, suppose that the number of linearly independent constraints is equal to the number of binding constraints (this has the same interpretation as in the Lagrange problem). Then, $\exists \{\lambda_j^*\}_{j=1}^n \in X_{j=1}^n \mathbb{R}_+$ such that:*

$$\begin{aligned} Df(x^*) + \sum_{j=1}^n \lambda_j^* Dg_j(x^*) &= 0 \\ \lambda_j^* &\geq 0 \, \forall j = 1, ..., n \\ g_j(x^*) &\geq 0 \, \forall j = 1, ..., n \\ \lambda_j^* g_j(x^*) &= 0 \, \forall j = 1, ..., n \end{aligned} \tag{5.9}$$

Proposition 18. *Let x^* be a solution of the Kuhn-Tucker problem. Let f and g_j, $\forall j = 1, ..., n$ be continuously differentiable and concave functions. Suppose moreover that the slater condition holds, that is, suppose the constraint set has a non-empty interior. Then $\exists \{\lambda_j^*\}_{j=1}^n \in X_{j=1}^n \mathbb{R}_+$ such that:*

$$\begin{gathered} L(x^*, \lambda^*) = f(x^*) + \sum_{j=1}^n \lambda_j^* g_j(x^*) \geq f(x) + \sum_{j=1}^n \lambda_j^* g_j(x) = L(x, \lambda^*) \, \forall x \in X \\ \lambda^* j g_j(x^*) = 0 \, \forall j = 1, ..., n. \end{gathered} \tag{5.10}$$

5.4. The Envelope Theorem

Proposition 29 (Unconstrained Problem). *Let $\Theta \in \mathbb{R}$ and X be two compact metric spaces. Let $f : X \times \Theta \to \mathbb{R}$ be a continuously differentiable function. Define:*

$$\sigma(\theta) := \arg\max\{f(x, \theta) : x \in X\} \tag{5.11}$$

and

$$f^*(\theta) := \max\{f(x, \theta) : x \in X\} \tag{5.12}$$

Let $\theta \in int(\Theta)$ and $x^* \in \sigma(\theta)$. Then, $f_\theta^*(\theta)$ and $f_\theta(x^*, \theta)$ are well-defined and:

$$f_\theta^*(\theta) = f_\theta(x^*, \theta) \, \forall x \in \sigma(\theta) \tag{5.13}$$

Proof. $\forall x^* \in \sigma(\theta)$ it must be that: $\max_{t \in \Theta}(f(x^*, t) - \sigma(t)) = f(x^*, \theta) - \sigma(\theta) = 0$. As the objective function is differentiable, this proves the theorem. □

Notice that the only assumption required for the theorem to hold is differentiability of the value function. No further topological structure needs to be imposed on the choice set X.

Example: Consider the problem of a firm that has to choose the amount of input $x \in \mathbb{R}$, in order to maximise profits: $\max_x \pi(x, w)$; where w is the price for input, taken as given by firm. Denote $x^*(w)$ the unique and interior solution of this problem and assume $\pi(x, w)$ to be continuously differentiable in w. How does the firm's maximal profit $\pi^*(w)$ vary with price w? Notice that since $x^*(w) \in \text{int}X$, then the following first order condition must be true: $\pi_x(x^*(w), w) = 0$. Using the chain rule to take the derivative of the maximal profit $\pi(x^*(w), w)$ with respect to w, we have:

$$\pi^*_w(w) = \pi_x(x^*(w), w)x^*_w(w) + \pi_w(x^*(w), w) \tag{5.14}$$

The first term corresponds to the change in π caused by the change in the solution of the problem that occurs when w changes; the second term corresponds to the direct effect of a change in w the value of π. Combining the two conditions, we have:

$$\pi^*_w(w) = \pi_w(x^*(w), w) \tag{5.15}$$

This result says that the change in the maximal value of profits as a parameter changes is the change caused by the direct impact of the parameter on the function, holding the value of x fixed at its optimal value; the indirect effect, resulting from the change in the optimal value of x caused by a change in the parameter, is zero.

Proposition 19 (**The Lagrange Problem Theorem**). *Let $\Theta \in \mathbb{R}$ and $X \in \mathbb{R}^n$ be two compact metric spaces. Let $f : X \times \Theta \to \mathbb{R}$ be a continuously differentiable function and $\Gamma(\theta) \subseteq \mathbb{R}^n$ a constraint set defined by $\Gamma(\theta) = \{x \in X : g_j(x, \theta) = 0 \, \forall j = 1, ..., n\}$, with $g_j : X \to \mathbb{R}$ continuously differentiable. Define:*

$$\sigma(\theta) := \arg\max\{f(x, \theta) : x \in \Gamma(\theta)\} \tag{5.16}$$

And:

$$f^*(\theta) := \arg\max\{f(x, \theta) : x \in \Gamma(\theta)\} \tag{5.17}$$

Assume the constraint qualification property is met. Define the Lagrangian function as:

$$L(x, \lambda, \theta) = f(x, \theta) + \sum_{j=1}^{n} \lambda_j g_j(x, \theta) \tag{5.18}$$

Where λ_j are the Lagrangian multipliers. Let $\Theta \in int(\Theta)$ and $x^* \in \sigma(\theta)$. Then, it must be that:

$$f^*_\theta(\theta) = L_\theta(x^*, \lambda^*, \theta) \tag{5.19}$$

Where λ^* is the vector of Lagrange multipliers at the stationary solution.

5.5. Saddle Point Theory

Consider the following problem:

$$\begin{aligned} P: \max_{x} f(x) &\text{ such that} \\ g_i(x) = 0 \,\forall i &= 1, ..., s \\ g_i(x) \geq 0 \,\forall i &= s+1, ..., n \\ x &\in X \end{aligned} \tag{5.20}$$

Let $f, g_i : \mathbb{R}^n \to \bar{\mathbb{R}}$ be continuously differentiable. Denote the constraint region for P by Ω. The Lagrangian for this problem is:

$$L(x, \lambda) = f(x) + \sum_{i=1}^{n} \lambda_i f_i(x) \tag{5.21}$$

where it is always assumed that $\lambda_i \leq 0 \,\forall i$. Set $K = \mathbb{R}^s_+ \times \mathbb{R}^{m-s} \subset \mathbb{R}^m$. A pair $(x^*, \lambda^*) \in \mathbb{R}^n \times K$ is said to be saddle point for L if:

$$L(x, \lambda) \geq L(x^*, \lambda^*) \geq L(x, \lambda^*), \quad \forall (x, \lambda) \in \mathbb{R}^n \times K \tag{5.22}$$

Proposition 20 (**Saddle Point Theorem**). *Let $x^* \in \mathbb{R}^n$. If $\exists \lambda^* \in K$ such that (x^*, λ^*) is a saddle point for the Lagrangian L, then x^* solves P. Conversely, if x^* is a solution to P for which the slater condition holds, then $\exists \lambda^* \in K$ such that (x^*, λ^*) is a saddle point for P. It follows that if a saddle point for L exists on $\mathbb{R}^n \times K$, then:*

$$\inf_{\lambda \in K} \sup_{x \in \mathbb{R}^n} L(x, \lambda) = \sup_{x \in \mathbb{R}^n} \inf_{\lambda \in K} L(x, \lambda) \tag{5.23}$$

This result is called min-max theorem and provides the condition under which inferior and supremum can be exchanged. Moreover, notice that since:

$$\inf_{\lambda \in K} L(x, \lambda) = \begin{cases} f(x) & \text{if } x \in \Omega \\ -\infty & \text{if } x \notin \Omega \end{cases} \tag{5.24}$$

The solution to the problem P can be written as:

$$\sup_{x \in \Omega} f(x) = \sup_{x \in \mathbb{R}^n} \inf_{\lambda \in K} L(x, \lambda) \tag{5.25}$$

6 Integration

6.1. The Riemann and the Riemann-Stieltjes Integrals

Definition 33. *Consider a closed interval $[a, b] \subset \mathbb{R}$. A n-point partition of $[a, b]$ is a set $P := \{x_0, x_1, ..., x_n\}$, with $a = x_0 < ... < x_n = b$. Given two partitions P_1 and P_2, P_2 is said to be a refinement of P_1 if $P_1 \subseteq P_2$.*

Definition 34. *Consider a bounded function $f : [a, b] \to \mathbb{R}$ and a n-point partition P of $[a, b]$. Then, $\forall i = 1, ..., n$; we set:*

$$\begin{aligned} M_i &:= \sup\{f(x) : x_{i-1} \leq x \leq x_i\} \\ m_i &:= \inf\{f(x) : x_{i-1} \leq x \leq x_i\} \end{aligned} \tag{6.1}$$

Therefore:

- *The upper Reimann sum of f given P is defined as $U(f, P) = \sum_{i=1}^{n} M_i(x_i - x_{i-1})$.*
- *Analogously, the lower Reimann sum of of f given P is defined as $L(f, P) = \sum_{i=1}^{n} m_i(x_i - x_{i-1})$.*

A critical assumptuon to define the Riemann integrals is boundedness of f: without it, both M_i and m_i might not be finite. A generalisation of the Riemmann sum of f given P is:

$$\sum_{i=1}^{n} f(x_i^{'})(x_i - x_{i-1}) \tag{6.2}$$

Where $x^{'} \in [x_{i-1}, x_i]$, $\forall i = 1, .., n$. By construction, it must be that:

$$L(f, P) \leq \sum_{i=1}^{n} f(x_i^{'})(x_i - x_{i-1}) \leq U(f, P) \tag{6.3}$$

It can be proved that $U(f, P)$ is the supremum of all the general Riemann sum of f given P, while $L(f, P)$ is the infimum.

Definition 35. *Let $f : [a, b] \to \mathbb{R}$ be a bounded function and P a partition of $[a, b]$. Then:*

- *The upper Riemann integral of f is $\int_a^{\bar{b}} f(x)dx = \inf_P\{U(f, P)\}$.*
- *The lower Riemann integral of f is $\int_{\bar{a}}^{b} f(x)dx = \sup_P\{U(f, P)\}$.*

Where inf and sup are taken by considering an increasingly finer partition of $[a, b]$. Therefore, f is said to be Riemann integrable when the two integrals coincide.

Definition 36. *We say that a bounded function $f : [a,b] \to \mathbb{R}$ satisfies the Riemann integrability criterion provided that for every $\epsilon > 0, \exists P$ such that $U(f,P) - L(f,P) < \epsilon$.*

Proposition 30. *Let $f : [a,b] \subset \mathbb{R}$ be a continuous function. Then, f is Riemann integrable.*

Proof. Since f is continuous and $[a,b]$ is compact, then f is uniformly continuous. Fix $\epsilon > 0$. Since f is uniformly continuous, $\exists \delta > 0$ such that if $d(x,y) < \delta$, then $d(f(x), f(y)) < \frac{\epsilon}{b-a}, \forall x \in X$. Let $P = \{a = x_0, ..., x_n = b\}$ be any partition of $[a,b]$, with $\| P \| < \delta$. Pick $z, y \in [x_i - x_{i-1}]$. Then, it must be that $d(f(z) - f(y)) < \frac{\epsilon}{b-a}$. By boundedness, there must be two points $z_i, y_i \in [x_{i-1}, x_i]$ such that $M_i = f(z_i)$ and $m_i = f(y_i)$. Hence, $0 \leq M_i - m_i < \frac{\epsilon}{b-a}$, which implies:

$$0 \leq U(f,P) - L(f,P) = \sum_{i=1}^{n} (M_i - m_i)(x_i - x_{i-1}) < \sum_{i=1}^{n} \frac{\epsilon}{b-a}(x_i - x_{i-1}) = \epsilon \quad (6.4)$$

Which is the Riemann integrability condition. □

The Riemann-Stieltjes integral is a further generalisation of the Riemann integral. The novelty of this notion of integral lies on how each sub-interval in the partition P is measured: the length of each sub-interval $[x_{i-1}, x_i]$ is indeed determined by $\alpha(x_i) - \alpha(x_{i-1})$, where $\alpha : [a,b] \subset \mathbb{R} \to \mathbb{R}$, is a function assumed to be increasing in its argument, i.e $x \leq y \Longrightarrow \alpha(x) \leq \alpha(y)$. It follows that, given a bounded function $f : [a,b] \subset \mathbb{R} \to \mathbb{R}$, the Riemann-Stieltjes integral is denoted as:

$$\int_a^b f d\alpha \quad (6.5)$$

Lower and upper Riemann-Stieltjes sums have an analogous definition. Moreover, any bounded function f is said to be Riemann-Stieltjes integrable when the lower and the upper Riemann-Stieltjes sums coincide.

Proposition 21. *Let $f : [a,b] \subset \mathbb{R}$ be a continuous function. Let $\alpha : [a,b] \to \mathbb{R}$ be an increasing function. Then, f is Riemann-Stieltjes integrable with respect to α.*

Properties: Let $\alpha : [a,b] \to \mathbb{R}$ be an increasing function, let $f, g' : [a,b] \to \mathbb{R}$ be two bounded functions that are Riemann-Stieltjes integrable with respect to α and let $c \in \mathbb{R}$ be a constant. Then:

- cf is Riemann-Stieltjes integrable with respect to α and $\int_a^b cf d\alpha = c \int_a^b f d\alpha$.
- $f + g'$ is Riemann-Stieltjes integrable with respect to α and $\int_a^b (f + g') d\alpha = \int_a^b f d\alpha + \int_a^b g' d\alpha$.
- fg' is Riemann-Stieltjes integrable with respect to α and $\int_a^b (fg') d\alpha = [fg]_a^b - \int_a^b (f' g) d\alpha$.

6.2. The Fundamental Theorem of Calculus

Proposition 22 (**Integral Mean Value Theorem**). *Let $a : [a,b] \to \mathbb{R}$ be an increasing function and let $f : [a,b] \to \mathbb{R}$ be a continuous function. Then, there $\exists c \in [a,b]$ such that $\int_a^b f d\alpha(t) = f(c)(\alpha(b) - \alpha(a))$*

Proposition 23. *Let $a \leq c < d \leq b$, let $\alpha : [a,b] \to \mathbb{R}$ be an increasing function and let $f : [a,b] \to \mathbb{R}$ be a bounded function that is Riemann-Stieltjes integrable with respect to α on $[a,b]$. Then, f is Riemann-Stieltjes integrable with respect to α on $[c,d]$.*

Proposition 24 (**Fundamental Theorem of Calculus**). *Let $\alpha : [a,b] \to \mathbb{R}$ be an increasing function that is differentiable on (a,b) and let $f : [a,b] \to \mathbb{R}$ be continuous. If $F(x) = \int_a^x f(t)d\alpha(t)$ then F is differentiable on (a,b) and $F'(x) = f(x)\alpha^{'}(x)$.*

It follows that if $\alpha(t) = t$, then $F(t) = \int_a^b f(t)dt$ and $F'(t) = f(t) \, \forall t \in (a,b)$.

Proposition 25. *Let $\alpha : [a,b] \to \mathbb{R}$ be an increasing function defined by $\alpha(t) = t$ and let $f : [a,b] \to \mathbb{R}$ be continuous. If $F(x) : [a,b] \to \mathbb{R}$ is a continuous function that is differentiable on (a,b) and $F'(x) = f(x)$, then $\int_a^b f(t)dt = F(b) - F(a)$.*

7 Appendix

7.1. Notation

Let A, B be sets in a space X. Let m, n be nonnegative integers.

- $\mathbb{Z} := \{..., -3, -2, -1, 0, 1, 2, 3, ...\}$ denotes the set of integers.
- $\mathbb{N} := \{0, 1, 2, 3, 4, 5, ...\}$ denotes the set of natural numbers.
- $\mathbb{Z}_+ := \{1, 2, 3, 4, ...\}$ denotes the set of positive integers.
- $\mathbb{Q} := \{\frac{m}{n} : m, n \in \mathbb{Z}, n \neq 0\}$ denotes the set of rational numbers.
- $\mathbb{R}$ denotes the set of real numbers.
- $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty\} \cup \{\infty\}$ denotes the set of extended real numbers.
- $\emptyset$ denotes the empty set, or a set consisting of 0 elements.
- $\in$ means "is an element of". For example, $2 \in \mathbb{Z}$ is read as "2 is an element of $\mathbb{Z}$, this is, the set of integers".
- $\forall$ means "for all".
- $\exists$ means "there exists".
- $\nexists$ means "there does not exist".
- $\exists!$ means "there exists a unique".
- $A \backslash B := \{x \in A : x \notin B\}$. $A^c := X \backslash A$ denotes the complement of A.
- $\cup$ denotes the union of two sets.
- $\cap$ denotes the intersection of two sets.
- $\subset$ indicates that one set is a proper subset of another.
- $\subseteq$ indicates that one set is a subset of another; or that both sets are equivalent.

7.2. Set Theory

Let X, Y be sets, and let $f : X \to Y$ be a function.

- The function f is said to be injective (or one-to-one) iff $\forall x, x^{'} \in X, f(x) = f(x^{'}) \iff x = x^{'}$.

- The function f is said to be surjective (or onto) iff $\forall y \in Y, \exists x \in X$ such that $f(x) = y$.
- The function f is said to be bijective (or a one-to-one correspondence) iff $\forall y \in Y, \exists! x \in X$ such that $f(x) = y$. A function f is bijective if it is both injective and surjective.
- Two sets X, Y are said to have the same cardinality iff there exists a bijection from X onto Y.

7.3. Limit Laws

Let $(\alpha_n)_{n=0}^{\infty}$, $(b_n)_{n=0}^{\infty}$ be convergent sequences. Let x, y be real numbers such that $x = \lim_{n\to\infty} a_n$, $y = \lim_{n\to\infty} b_n$. Then:

- The sequence $(a_n + b_n)_{n=0}^{\infty}$ converges to $x + y$. That is, $\lim_{n\to\infty}(a_n + b_n) = (\lim_{n\to\infty} a_n) + (\lim_{n\to\infty} b_n)$.
- The sequence $(a_n b_n)_{n=0}^{\infty}$ converges to xy. That is, $\lim_{n\to\infty}(a_n b_n) = (\lim_{n\to\infty} a_n)(\lim_{n\to\infty} b_n)$
- For any real number c, the sequence $(ca_n)_{n=0}^{\infty}$ converges to cx. That is, $c \lim_{n\to\infty} a_n = \lim_{n\to\infty}(ca_n)$.
- Suppose $x \neq 0$ and that there exists m such that $a_n \neq 0$, $\forall n \geq m$. Then, $(a_n^{-1})_{n=0}^{\infty}$ converges to x^{-1}. That is, $\lim_{n\to\infty}(a_n^{-1}) = (\lim_{n\to\infty} a_n)^{-1}$.
- Suppose $x \neq 0$ and that there exists m such that an $(a_n) \neq 0$, $\forall n \geq m$. Then, $(\frac{b_n}{a_n})_{n=0}^{\infty}$ converges to $\frac{y}{x}$. That is, $\lim_{n\to\infty}(\frac{b_n}{a_n})_{n=0}^{\infty} = \frac{lim_{n\to\infty} b_n}{\lim_{n\to\infty} a_n}$.
- Suppose $a_n \geq b_n$, $\forall n \geq 0$. Then, $x \geq y$.

7.4. Properties of derivatives

Let X be a subset of $\mathbb{R}$, let x_0 be a limit point of X, and let $f : X \to \mathbb{R}$ be functions. Then:

- If f is constant, so that there exists $c \in \mathbb{R}$ such that $f(x) = c$ for all $x \in X$, then f is differentiable at x_0 and $f^{'}(x_0) = 0$.
- If f is the identity function, so that $f(x) = x$ for all $x \in X$, then f is differentiable at x_0 and $f^{'}(x_0) = 1$.
- If f, g are differentiable at x_0, then $f + g$ is differentiable at x_0, and $(f + g)^{'}(x_0) = f^{'}(x_0) + g^{'}(x_0)$ (Sum Rule).
- If f, g are differentiable at x_0, then fg is differentiable at x_0, and $(fg)^{'}(x_0) = f^{'}(x_0)g(x_0) + f(x_0)g^{'}(x_0)$ (Product Rule).

- If f is differentiable at x_0, and if $c \in \mathbb{R}$, then cf is differentiable at x_0, and $(cf)^{'}(x_0) = cf^{'}(x_0)$.
- If f, g are differentiable at x_0, then $f - g$ is differentiable at x_0 and $(f-g)^{'}(x_0) = f^{'}(x_0) - g^{'}(x_0)$.
- If g is differentiable at x_0, and if $g(x) \neq 0$ for all $x \in X$, then $\frac{1}{g}$ is differentiable at x_0 and $(\frac{1}{g})^{'}(x_0) = -\frac{g^{'}(x_0)}{(g(x_0))^2}$.
- If f, g are differentiable at x_0, and if $g(x) \neq 0$ for all $x \in X$, then $\frac{f}{g}$ is differentiable at x_0, and $(\frac{f}{g})^{'}(x_0) = \frac{g(x_0)f^{'}(x_0) - f(x_0)g^{'}(x_0)}{(g(x_0))^2}$ (Quotient Rule).