Topology and differential calculus

Lecture notes for the course FMA_3F002_EP, Fall 2025
Bachelor of Science, École polytechnique

Kleber Carrapatoso
kleber.carrapatoso@polytechnique.edu

Contents

Chapter 1

Topological and metric spaces

1.1 Topological spaces

Definition 1.1. A *topology* on a set X is a family $\mathscr{O} \subseteq \mathscr{P}(X)$ of subsets of X verifying:

(i) The empty set $\varnothing$ and X belong to $\mathscr{O}$;

(ii) Every finite intersection of elements of $\mathscr{O}$ belongs to $\mathscr{O}$;

(iii) Every arbitrary union of elements of $\mathscr{O}$ belongs to $\mathscr{O}$.

We say that $(X, \mathscr{O})$ is a *topological space*.

Remark 1.2. One can replace (ii) by: "every intersection of two elements of $\mathscr{O}$ belongs to $\mathscr{O}$".

Elements of X are called *points*. Elements of $\mathscr{O}$ are called *open sets* in X (with respect to the topology $\mathscr{O}$). We often denote by X the couple $(X, \mathscr{O})$.

Definition 1.3. A subset C of a topological space X is *closed* if its complement $X \setminus C$ is open.

It easily follows from Definition 1.1 that in a topological space X there holds:

- $\varnothing$ and X are closed;
- Any arbitrary intersection of closed sets is closed;
- Any finite union of closed sets is closed.

Definition 1.4. Let $\mathscr{O}_1$ and $\mathscr{O}_2$ be two topologies on a set X. If $\mathscr{O}_1 \subseteq \mathscr{O}_2$ one says that $\mathscr{O}_1$ is *coarser* than $\mathscr{O}_2$, or that $\mathscr{O}_2$ is *finer* than $\mathscr{O}_1$. If moreover $\mathscr{O}_1 \neq \mathscr{O}_2$, then $\mathscr{O}_1$ is said to be *strictly coarser* than $\mathscr{O}_2$, or $\mathscr{O}_2$ is said to be *strictly finer* than $\mathscr{O}_1$.

Example 1.5. (1) The collection of subsets $\{\varnothing, X\}$ is a topology on any set X, called the *trivial topology*.

(2) The powerset $\mathscr{P}(X)$ of X is a topology on any set X, called the *discrete topology*. A topological space endowed with the discrete topology is called a *discrete space*.

(3) The *standard topology* on $\mathbf{R}$ is given by the collection of arbitrary unions of open intervals. We shall always consider this topology on $\mathbf{R}$ (unless otherwise specified).

(4) Let $(X, \mathcal{O})$ be a topological space and $Y \subseteq X$ a subset. The collection of subsets $\{U \cap Y : U \in \mathcal{O}\}$ is a topology on Y, called the *subspace topology*.

(5) Let X be a set, Y a topological space and $f : X \to Y$ a map. The collection of subsets $\{f^{-1}(U) : U \text{ open in } Y\}$ is a topology on X, called the *preimage topology*.

(6) The intersection of an arbitrary family of topologies on a set is a topology on this set.

1.2 Metric spaces

Definition 1.6. A *metric* (or *distance*) on a set X is a map $d : X \times X \to \mathbf{R}^+$ such that for all $x, y, z \in X$ one has:

(i) (Separation) $d(x, y) = 0$ if and only if $x = y$;

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

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

The pair (X, d) is called a *metric space*.

Remark 1.7. If (i) is replaced by the weaker assumption "$d(x, x) = 0$", then d is called a *pseudo-metric*.

Remark 1.8. If d satisfies $d(x, z) \leqslant \sup(d(x, y), d(y, z))$ for any $x, y, z \in X$, which is stronger than the triangle inequality, then we say that d is an *ultrametric* and (X, d) an *ultrametric space*.

Remark 1.9. For any $x, y, z \in X$ we also have $|d(x, z) - d(y, z)| \leqslant d(x, y)$.

Let (X, d) be a metric space. For any $x \in X$ and $r > 0$, we define the open ball centred at x of radius r by

$$B(x, r) := \{y \in X \mid d(x, y) < r\},$$

the closed ball centred at x of radius r by

$$\overline{B}(x, r) := \{y \in X \mid d(x, y) \leqslant r\},$$

and the sphere centred at x of radius r by

$$S(x, r) := \{y \in X \mid d(x, y) = r\}.$$

Definition 1.10. A subset Y of a metric space (X, d) is said to be *bounded* if Y is contained in some open ball, that is, if there are $a \in X$ and $r > 0$ such that $Y \subseteq B(a, r)$.

The following result gives an equivalent way of characterizing bounded subsets.

Proposition 1.11. *Let (X, d) be a metric space. A non-empty subset $Y \subseteq X$ is bounded if and only if* $\sup_{x,y \in Y} d(x, y) < \infty$.

Proof. Assume Y is bounded, then $Y \subseteq B(a, r)$ for some $a \in X$ and $r > 0$. Therefore the set $\{d(x, y) : x, y \in Y\}$ is bounded, hence it admits a supremum.

Conversely, assume that $\sup_{x,y \in Y} d(x, y) < \infty$. We choose some $a \in Y$ and denote $r = \sup_{x,y \in Y} d(x, y) + 1 > 0$, thus $Y \subseteq B(a, r)$. □

1.2.1 Topology associated to a metric

Definition 1.12. Let (X, d) be a metric space. We define a topology $\mathscr{O}_d$ on X by

$$\mathscr{O}_d := \{U \subseteq X \mid \forall x \in U, \exists r > 0, B(x, r) \subseteq U\},$$

called the *topology associated to* or *induced by* the metric d.

Let us verify that $\mathscr{O}_d$ is indeed a topology on X. It is clear that $\varnothing$ and X belong to $\mathscr{O}_d$. Let $U_1, \ldots, U_n \in \mathscr{O}_d$ and $x \in \bigcap_{1 \leqslant i \leqslant n} U_i$, then for each $1 \leqslant i \leqslant n$ we have $x \in U_i \in \mathscr{O}_d$, hence there is $r_i > 0$ such that $B(x, r_i) \subseteq U_i$. Consider $r = \inf_{1 \leqslant i \leqslant n} r_i$, we have $r > 0$ since there are finitely many r_i, then $B(x, r) \subseteq U_i$ for each $1 \leqslant i \leqslant n$, which implies that $B(x, r) \subseteq \bigcap_{1 \leqslant i \leqslant n} U_i$, hence $\bigcap_{1 \leqslant i \leqslant n} U_i \in \mathscr{O}_d$. Now consider a family $\{U_i\}_{i \in I}$ of elements of $\mathscr{O}_d$. Let $x \in \bigcup_{i \in I} U_i$, then there is $i_0 \in I$ such that $x \in U_{i_0} \in \mathscr{O}_d$, hence there is $r > 0$ such that $B(x, r) \subseteq U_{i_0} \subseteq \bigcup_{i \in I} U_i$, which implies $\bigcup_{i \in I} U_i \in \mathscr{O}_d$.

In this way, we shall always consider a metric space as a topological space endowed with the topology associated to the metric.

It is easy to check that open balls (resp. closed balls) are open sets (resp. closed sets) for the topology associated to the metric.

Definition 1.13. A topological space $(X, \mathscr{O})$ is *metrizable* if there is a metric d on X such that the topology associated to d coincides with $\mathscr{O}$.

Definition 1.14. We say that two metrics d_1 and d_2 on a set X are:

(i) *Topologically equivalent* if they induce the same topology.

(ii) *Equivalent* if there exist constants $c_1, c_2 > 0$ such that for any $x, y \in X$ one has

$$c_1 d_1(x, y) \leqslant d_2(x, y) \leqslant c_2 d_1(x, y).$$

It is easy to verify that if two metrics on a set are equivalent, then they are also topologically equivalent.

Example 1.15. (1) On any non-empty set X, we can define the distance $d(x, y) = 1$ if $x \neq y$ and $d(x, y) = 0$ if $x = y$, which is called the *discrete metric*. We can easily check that the topology associated to the discrete metric is the discrete topology (Example 1.5–(2)).

(2) The set of real numbers $\mathbf{R}$ (resp. complex numbers $\mathbf{C}$) is endowed with the metric $d(x, y) = |x - y|$ for any x, y in $\mathbf{R}$ (resp. in $\mathbf{C}$) where $|\cdot|$ denotes the absolute value (resp. the modulus), which is called the *standard metric*. It is easy to check that the topology on $\mathbf{R}$ associated to its standard metric coincides with its standard topology (Example 1.5–(3)).

(3) If (X, d) is as metric space and Y is a subset of X, we define the map $d_Y : Y \times Y \longrightarrow \mathbf{R}^+$ as the restriction of d to $Y \times Y$. Then d_Y is a metric on Y and hence (Y, d_Y) is a metric space, called a *metric subspace* of (X, d).

(4) Any normed vector space $(X, \|\cdot\|)$ is a metric space with the distance $d(x, y) := \|x - y\|$, for any $x, y \in X$.

(5) Let (X, d_X) and (Y, d_Y) be metric spaces. On the product set $X \times Y$ we can define the following distances: for any $(x, y), (x', y') \in X \times Y$

$$d_p\big((x, y), (x', y')\big) := \big[d_X(x, x')^p + d_Y(y, y')^p\big]^{1/p} \quad \text{for} \quad p \geqslant 1,$$

and

$$d_\infty\big((x, y), (x', y')\big) := \max\big(d_X(x, x'), d_Y(y, y')\big).$$

1.3 Neighborhoods

Let X be a topological space, $A \subseteq X$ a subset and $x \in X$ a point.

Definition 1.16. A *neighborhood* of x (resp. A) is a subset $V \subseteq X$ that contains an open set containing x (resp. A).

In other words, denoting by $\mathscr{O}$ the topology of X, a subset V of X is a neighborhood of x if there is $U \in \mathscr{O}$ such that $x \in U \subseteq V$.

We denote by $\mathscr{V}(x)$ the family of all neighborhoods of a point x, which is called the *neighborhood system* or *neighborhood filter* for x. Similarly, we denote by $\mathscr{V}(A)$ the family of all neighborhoods of A, which is called the *neighborhood system* or *neighborhood filter* for A.

One easily verifies the following properties:

Proposition 1.17. *A subset of a topological space X is open if and only if it is a neighborhood of each of its points.*

Proof. Let U be a subset of X. If U is open then it is clear by definition that U is a neighborhood of each of its points. Conversely, assume that U is a neighborhood of each of its points. For each $x \in U$ there is an open set U_x in X such that $x \in U_x \subseteq U$. This implies that $U = \bigcup_{x \in U} U_x$, and $\bigcup_{x \in U} U_x$ is open as the union of open sets. □

Proposition 1.18. *Let X be a topological space. For any $x \in X$, the neighborhood system $\mathscr{V}(x)$ satisfies:*

(i) *Any subset of X containing an element of $\mathscr{V}(x)$ belongs to $\mathscr{V}(x)$;*

(ii) *Any finite intersection of elements of $\mathscr{V}(x)$ belongs to $\mathscr{V}(x)$;*

(iii) *x belongs to any element of $\mathscr{V}(x)$;*

(iv) *For all $V \in \mathscr{V}(x)$, there exists $W \in \mathscr{V}(x)$ such that $V \in \mathscr{V}(y)$ for all $y \in W$.*

Proof. Consider $U \subseteq X$ such that there is $V \in \mathscr{V}(x)$ that satisfies $V \subseteq U$. Then there exists an open set $W \subseteq X$ such that $x \in W \subseteq V$, hence $x \in W \subseteq U$ and U is a neighborhood of x, which proves (i). Now let $V_1, \ldots, V_n \in \mathscr{V}(x)$, then there exist open sets $U_1, \ldots, U_n$ such that $x \in U_j \subseteq V_j$ for any $j \in \{1, \ldots, n\}$. The set $\bigcap_{1 \leqslant j \leqslant n} U_j$ is open, included in $\bigcap_{1 \leqslant j \leqslant n} V_j$, and contains x, thus $\bigcap_{1 \leqslant j \leqslant n} V_j$ is a neighborhood of x, and this gives (ii). Property (iii) is trivial. Finally, let $V \in \mathscr{V}(x)$. Then there exists an open set U such that $x \in U \subseteq V$, and U is also a neighborhood of x. We hence obtain that, for any $y \in U$, we have $y \in U \subseteq V$, thus $V \in \mathscr{V}(y)$ which proves (iv). □

We also observe that, thanks to the above properties on the neighborhood system, it is possible to define a topology on a set by X by prescribing the neighborhood system $\mathscr{V}(x)$ for all points $x \in X$, provided it satisfies properties (i) to (iv) of Proposition 1.18.

Proposition 1.19. *Let X be a set and $\mathscr{V} : X \to \mathscr{P}(X)$ a map that sends a point $x \in X$ to a collection $\mathscr{V}(x)$ of subsets of X satisfying properties* (i)–(ii)–(iii)–(iv) *of Proposition 1.18. Then there exists a unique topology on X such that, for any $x \in X$, $\mathscr{V}(x)$ is the neighborhood system of x associated to this topology.*

Proof. The uniqueness follows from the fact that a set is open if and only if it is a neighborhood of each of its points (Proposition 1.17). Indeed, assume that there exists a topology verifying the properties of the statement, then it must be equal to

$$\mathscr{O} = \{U \subseteq X \mid \forall x \in U, U \in \mathscr{V}(x)\}.$$

We easily verify that $\mathscr{O}$ is a topology on X: $\varnothing, X \in \mathscr{O}$ is trivial; the fact that finite intersections of elements of $\mathscr{O}$ belong to $\mathscr{O}$ follows from property (i) of Proposition 1.18; and the fact that arbitrary unions of elements of $\mathscr{O}$ belong to $\mathscr{O}$ follows from property (ii) of Proposition 1.18.

It remains to verify that, for any $x \in X$, $\mathscr{V}(x)$ corresponds to the collection of neighborhoods of x for the topology $\mathscr{O}$. Thanks to property (i) of Proposition 1.18, every neighborhood of x belongs to $\mathscr{V}(x)$. Conversely, let $V \in \mathscr{V}(x)$ and U be the set of points $y \in X$ such that $V \in \mathscr{V}(y)$. We shall prove that $x \in U \subseteq V$ and $U \in \mathscr{O}$ (i.e. V is a neighborhood of x) which completes the proof. We have $x \in U$ since $V \in \mathscr{V}(x)$. Let $y \in U$, then $V \in \mathscr{V}(y)$. On the one hand, from property (iii) of Proposition 1.18 we deduce $y \in V$ and hence $U \subseteq V$. On the other hand, from property (iv) of Proposition 1.18 there is $W \in \mathscr{V}(y)$ such that for any $z \in W$ we have $V \in \mathscr{V}(z)$, which implies $z \in U$ and hence $W \subseteq U$, which finally yields $U \in \mathscr{V}(y)$ thanks to (i), that is $U \in \mathscr{O}$. $\square$

1.3.1 Neighborhood basis

Let X be a topological space, $A \subseteq X$ a subset and $x \in X$ a point.

Definition 1.20. A *neighborhood basis* for x (resp. A) is a family $\mathscr{S}$ of neighborhoods of x (resp. A) such that every neighborhood of x (resp. A) contains an element of $\mathscr{S}$.

In other words, a collection of subsets $\mathscr{S} \subseteq \mathscr{V}(x)$ is a neighborhood basis for x if for every $V \in \mathscr{V}(x)$ there exists $W \in \mathscr{S}$ such that $W \subseteq V$.

We also remark that a neighborhood basis $\mathscr{S}(x)$ for x characterizes the neighborhood system $\mathscr{V}(x)$ of x since

$$\mathscr{V}(x) = \{V \subseteq X \mid \exists B \in \mathscr{S}(x), B \subseteq V\}.$$

In view of this and Proposition 1.19, one can also define a topology on a set X by prescribing the sets $\mathscr{S}(x)$ of neighborhood basis for all points of X in the following way.

Proposition 1.21. *Let X be a set and $\mathcal{S} : X \to \mathcal{P}(X) \setminus \{\varnothing\}$ a map that sends a point $x \in X$ to a collection $\mathcal{S}(x)$ of subsets of X satisfying*

(i) *x belongs to all elements of $\mathcal{S}(x)$;*

(ii) *For any $U, V \in \mathcal{S}(x)$, there is $W \in \mathcal{S}(x)$ such that $W \subseteq U \cap V$;*

(iii) *For any $V \in \mathcal{S}(x)$, there is $W \in \mathcal{S}(x)$ such that, for any $y \in W$, there is $U \in \mathcal{S}(y)$ such that $U \subseteq V$.*

Then there exists a unique topology on X such that, for any $x \in X$, $\mathcal{V}(x) := \{V \subseteq X \mid \exists B \in \mathcal{S}(x), B \subseteq V\}$ is the neighborhood system of x associated to this topology.

Example 1.22. (1) In a topological space the neighborhood system $\mathcal{V}(x)$ is clearly a neighborhood basis for every point x. Moreover, the collection of all open neighborhoods of x is a neighborhood basis for x.

(2) In a topological space with the trivial topology, the only neighborhood of any point is the whole set itself.

(3) In a discrete space (Example 1.5–(2)), the set $\{x\}$ is a neighborhood basis for any point x.

(4) Consider $\mathbf{R}$ endowed with its standard topology (Example 1.5–(3)). For any point $x \in \mathbf{R}$, the neighborhood system $\mathcal{V}(x)$ of x is the collection of all subsets containing an open interval that contains x. Moreover, each one of the following families is a neighborhood basis for a point $x \in \mathbf{R}$:

$$\Big\{(x-r, x+r) \ : \ r > 0\Big\}, \quad \Big\{[x-r, x+r) \ : \ r > 0\Big\},$$

$$\Big\{(x-\tfrac{1}{n}, x+\tfrac{1}{n}) \ : \ n \in \mathbf{N}^*\Big\}, \quad \Big\{[x-\tfrac{1}{n}, x+\tfrac{1}{n}) \ : \ n \in \mathbf{N}^*\Big\}.$$

(5) In a metric space (X, d), a subset $V \subseteq X$ is a neighborhood of a point $x \in X$ if there is $r > 0$ such that $B(x, r) \subseteq V$. Furthermore, the family $\{B(x, r) : r > 0\}$ is a neighborhood basis for any point $x \in X$.

1.4 Basis

Let $(X, \mathcal{O})$ be a topological space.

Definition 1.23. A family $\mathcal{B} \subseteq \mathcal{O}$ of open sets in X is a *basis* for the topology $\mathcal{O}$ if every open set in X can be written as a union of elements of $\mathcal{B}$.

Equivalently, we can caractherize a basis for a topology in the following way.

Proposition 1.24. *A subset $\mathcal{B} \subseteq \mathcal{O}$ is a basis for the topology $\mathcal{O}$ if and only if for any open set $U \in \mathcal{O}$ and any $x \in U$, there exists $B \in \mathcal{B}$ such that $x \in B \subseteq U$.*

Proof. Suppose that $\mathcal{B}$ is a basis for $\mathcal{O}$, namely that every open set $U \in \mathcal{O}$ is a union of elements of $\mathcal{B}$. Let $U \in \mathcal{O}$ and $x \in U$. Then there is $\mathcal{B}_1 \subseteq \mathcal{B}$ such that $U = \bigcup_{B \in \mathcal{B}_1} B$, therefore there is $B \in \mathcal{B}_1$ such that $x \in B \subseteq U$.

Conversely, assume $\mathscr{B} \subseteq \mathscr{O}$ satisfies the property: For any $U \in \mathscr{O}$ and any $x \in U$ there is $B \in \mathscr{B}$ such that $x \in B \subseteq U$. Let $U \in \mathscr{O}$, then for each $x \in U$ there is $B_x \in \mathscr{B}$ such that $x \in B_x \subseteq U$. This implies that $U = \bigcup_{x \in U} B_x$, that is, U is a union of elements of $\mathscr{B}$ and hence $\mathscr{B}$ is a basis for $\mathscr{O}$. □

1.4.1 Sub-basis and topology generated by a family of subsets

Recall that if $\{\mathscr{O}_i\}_{i \in I}$ is a family of topologies on a set X, then their intersection $\mathscr{O} := \bigcap_{i \in I} \mathscr{O}_i$ still is a topology on X (Example 1.5–(6)). Indeed, it is clear that $\varnothing, X \in \mathscr{O}$; if $\{U_\alpha\}_{1 \leqslant \alpha \leqslant n}$ is a finite family of elements of $\mathscr{O}$, then $U_\alpha \in \mathscr{O}_i$ for any $1 \leqslant \alpha \leqslant n$ and any $i \in I$, hence $\bigcap_{1 \leqslant \alpha \leqslant n} U_\alpha \in \mathscr{O}_i$ for any $i \in I$, thus $\bigcap_{1 \leqslant \alpha \leqslant n} U_\alpha \in \mathscr{O}$; and if $\{U_\alpha\}_{\alpha \in \mathscr{A}}$ is an arbitrary family of elements of $\mathscr{O}$, then $U_\alpha \in \mathscr{O}_i$ for any $\alpha \in \mathscr{A}$ and any $i \in I$, hence $\bigcup_{\alpha \in \mathscr{A}} U_\alpha \in \mathscr{O}_i$ for any $i \in I$, thus $\bigcup_{\alpha \in \mathscr{A}} U_\alpha \in \mathscr{O}$.

Moreover, given a family $\mathscr{S}$ of subsets of X there exists at least one topology on X containing $\mathscr{S}$, namely the discrete topology $\mathscr{P}(X)$.

Definition 1.25. Let $\mathscr{S}$ be a family of subsets of a set X. The *topology generated by $\mathscr{S}$* is the intersection of all topologies on X containing $\mathscr{S}$, which also corresponds to the coarsest topology on X containing $\mathscr{S}$.

One says that $\mathscr{S}$ is a *sub-basis* for the topology $\mathscr{O}_{\mathscr{S}}$ generated by it. It is easy to verify that $\mathscr{O}_{\mathscr{S}}$ corresponds to the collection of arbitrary unions of finite intersections of elements of $\mathscr{S}$. In other words, the family of finite intersections of elements of $\mathscr{S}$ is a basis for $\mathscr{O}_{\mathscr{S}}$.

The following result gives a condition under which a collection of subsets of X (that is a sub-basis) is actually a basis for the topology it generates.

Proposition 1.26. *Let $\mathscr{B}$ be a family of subsets of a set X satisfying*

(i) $\bigcup \mathscr{B} := \bigcup_{B \in \mathscr{B}} B = X$;

(ii) *For any $U, V \in \mathscr{B}$ and any $x \in U \cap V$, there is $B \in \mathscr{B}$ such that $x \in B \subseteq U \cap V$.*

Then the family $\mathscr{O}$ of arbitrary unions of elements of $\mathscr{B}$ coincides with the topology generated by $\mathscr{B}$.

Proof. One easily observes that if $\mathscr{O}$ is a topology then it coincides with the topology generated by $\mathscr{B}$. Therefore, we only need to show that $\mathscr{O}$ is a topology.

It is clear that $\varnothing \in \mathscr{O}$ and $X \in \mathscr{O}$. Let $\mathscr{U}$ be a family of elements of $\mathscr{O}$, then each member of $\mathscr{O}$ is a union of elements of $\mathscr{B}$, therefore the union $\bigcup \mathscr{U}$ is also a union of elements of $\mathscr{B}$. Let $U, V \in \mathscr{O}$, then there is $\mathscr{B}_1, \mathscr{B}_2 \subseteq \mathscr{B}$ such that $U = \bigcup \mathscr{B}_1$ and $V = \bigcup \mathscr{B}_2$. Therefore

$$U \cap V = \left(\bigcup_{A \in \mathscr{B}_1} A \right) \cap \left(\bigcup_{B \in \mathscr{B}_2} B \right) = \bigcup_{A \in \mathscr{B}_1,\, B \in \mathscr{B}_2} A \cap B,$$

and the proof will be complete if we show that any intersection of two elements of $\mathscr{B}$ is a union of elements of $\mathscr{B}$. Consider then $A, B \in \mathscr{B}$. For each $x \in A \cap B$, there is $W_x \in \mathscr{B}$ such that $x \in W_x \subseteq A \cap B$. Thus $A \cap B = \bigcup_{x \in A \cap B} W_x$, which completes the proof. □

Example 1.27. (1) The collection of all singletons is a basis for the discrete topology (Example 1.5–(2)).

(2) The collection of all open intervals in $\mathbf{R}$ is a basis for its standard topology (Example 1.5–(3)).

(3) In a metric space, the collection of all open balls is a basis for the topology associated to the metric.

1.5 Interior and closure

Let X be a topological space, A a subset of X, and $x \in X$ a point.

Definition 1.28. One says that x is an *interior point* of A if A is a neighborhood of x. The set of all interior points of A is called the *interior of* A, and is denoted by $\mathring{A}$ or $\operatorname{int}(A)$.

It is easy to verify that $\mathring{A}$ is the union of all open sets contained in A, therefore also the largest open set contained in A. Indeed let B be the union of all open sets contained in A. Since an arbitrary union of open sets is open, B is also the largest open set contained in A. We now check that $\mathring{A} = B$. Let $x \in \mathring{A}$, then there is an open set U such that $x \in U \subseteq A$, but $U \subseteq B$ thus $x \in B$. Conversely let $x \in B$, then there is an open set U such that $x \in U \subseteq A$, hence $x \in \mathring{A}$.

Definition 1.29. One says that x is an *adherent point* of A if every neighborhood of x intersects A. The set of all adherent points of A is called the *closure of* A, and is denoted by $\overline{A}$ or $\operatorname{cl}(A)$.

It is easy to check that $\overline{A}$ is closed. Indeed let $x \in X \setminus \overline{A}$, then there is an open neighborhood V of x that does not intersect A. If $y \in V$ then $y \notin \overline{A}$ since $V \cap A = \varnothing$, which implies $V \subseteq X \setminus \overline{A}$. Hence $X \setminus \overline{A}$ is open and thus $\overline{A}$ is closed.

Moreover $\overline{A}$ is the intersection of all closed sets containing A, hence also the smallest closed set containing A. Indeed let C be the intersection of all closed sets containing A. Since an arbitrary intersection of closed sets is closed, C is also the smallest closed set containing A. We now check that $\overline{A} = C$. Let $x \in \overline{A}$ and D be a closed set containing A: If $x \in X \setminus D$ then $(X \setminus D) \cap A \neq \varnothing$ because $X \setminus D$ is open, but $A \cap (X \setminus D) = \varnothing$ since $A \subseteq D$; therefore $x \in D$. Since this holds for any closed set D containing A, we deduce $\overline{A} \subseteq C$. The reverse inclusion $C \subseteq \overline{A}$ is trivial since $\overline{A}$ is closed.

Definition 1.30. One says that x is a *limit point* of A if every neighborhood of x intersects $A \setminus \{x\}$.

Definition 1.31. One says that x is a *boundary point* of A if every neighborhood of x intersects both A and its complement $X \setminus A$. The set of all boundary points of A is called the *boundary of* A, and is denoted by ∂A.

It readily follows by the definition that $\partial A = \overline{A} \cap \overline{X \setminus A}$. Moreover we also have $\partial A = \overline{A} \setminus \mathring{A}$. Indeed for $x \in X$ and a neighborhood U of x, we observe that $U \cap (X \setminus A) \neq \emptyset$ is not empty if and only if U is not cointaned in A.

Definition 1.32. One says that x is an *isolated point* of A if $x \in A$ and there exists a neighborhood of x which does not contain any other points of A.

Equivalently, x is an isolated point of A if and only if x is not a limit point of A. Another equivalent formulation is the following: the singleton $\{x\}$ is open in the topological space A endowed with the subspace topology (Example 1.5–(4)).

Definition 1.33. One says that A is dense in X if $\overline{A} = X$.

Equivalently, a subset $A \subseteq X$ is dense if and only if every non-empty open set in X intersects A. Indeed, assume that A is dense and let U be a non-empty open set in X. Let $x \in U$, then, since U is a neighborhood of x in X and $x \in X = \overline{A}$, one obtains that $U \cap A$ is non-empty. Conversely, assume that every non-empty open set in X intersects A. Let $x \in X$ and V be a neighborhood of x in X. Since $\mathring{V} \cap A$ is non-empty, one deduces that $V \cap A$ is non-empty, thus $x \in \overline{A}$.

One can easily verify the following properties, which are left as an exercise.

Proposition 1.34. *Let A and B be subsets of a topological space X, then:*

(i) *A is open if and only if $A = \mathring{A}$.*

(ii) *A is closed if and only if $A = \overline{A}$.*

(iii) *If $A \subseteq B$ then $\mathring{A} \subseteq \mathring{B}$ and $\overline{A} \subseteq \overline{B}$.*

(iv) *$\mathring{A} \cap \mathring{B} = \overbrace{A \cap B}^{\circ}$ and $\mathring{A} \cup \mathring{B} \subseteq \overbrace{A \cup B}^{\circ}$.*

(v) *$\overline{A} \cup \overline{B} = \overline{A \cup B}$ and $\overline{A \cap B} \subseteq \overline{A} \cap \overline{B}$.*

(vi) *$X \setminus \mathring{A} = \overline{X \setminus A}$.*

(vii) *$X \setminus \overline{A} = \overbrace{X \setminus A}^{\circ}$.*

1.6 Separation and countability

1.6.1 Axioms of separation

We list below some of the axioms of separation.

Definition 1.35. A topological space X is said to be:

(i) T_0 (or *Kolmogorov*[1]) if for any two distinct points there exist a neighborhood of one of the points that does not contain the other.

[1] Andrey Kolmogorov (1903–1987) was a Soviet mathematician who made important contributions to topology, foundations of probability theory, logic, and turbulence in fluid mechanics..

(ii) T_1 (or *Fréchet*[2]) if for any two distinct points $x, y \in X$ there exist a neighborhood U of x not containing y, and a neighborhood V of y not containing x.

(iii) *Hausdorff*[3] (or T_2) if for any two distinct points $x, y \in X$ there exist a neighborhood U of x and a neighborhood V of y such that $U \cap V = \varnothing$.

(iv) *Regular* (or T_3) if it is T_0 and if for any point $x \in X$ and any closed set $F \subseteq X$ such that $x \notin F$, there exist a neighborhood U of x and a neighborhood V of F such that $U \cap V = \varnothing$.

(v) *Normal* (or T_4) if it is Hausdorff and if for any disjoint closed sets $E, F \subseteq X$, there exist a neighborhood U of E and a neighborhood V of F such that $U \cap V = \varnothing$.

Proposition 1.36. *A topological space X is T_0 if and only if for any two distinct points one of them is not an adherent point of the singleton formed by the other.*

Proof. If $x, y \in X$ are distinct points, observe that $x \notin \overline{\{y\}}$ if and only if there is a neighborhood V of x in X such that $y \notin V$. □

Proposition 1.37. *A topological space X is T_1 if and only if every singleton is closed.*

Proof. Assume that X is T_1. Let $x \in X$. For any $y \in X \setminus \{x\}$ there is an open neighborhood V_y of y not containing x, hence $V_y \subseteq X \setminus \{x\}$, that is $X \setminus \{x\}$ is a neighborhood of y. Hence $X \setminus \{x\}$ is a neighborhood of each of its points, which means that $X \setminus \{x\}$ is open, thus $\{x\}$ is closed.

Conversely, assume that every singleton is closed. Let $x, y \in X$ be distinct points, then $\{x\}$ is a neighborhood of x not containing y, and $\{y\}$ is a neighborhood of y not containing x. □

Proposition 1.38. *A Hausdorff topological space X is regular if and only if for any $x \in X$ and any neighborhood U of x, there exists a neighborhood V of x such that $x \in V \subseteq \overline{V} \subseteq U$.*

Proof. Assume that X is regular. Let $x \in X$ and U be a neighborhood of x. The set $A = \overline{X \setminus U}$ is closed and $x \notin A$. By hypothesis there are open sets V, W such that $x \in V$, $A \subseteq W$ and $V \cap W = \varnothing$. We hence obtain that $\overline{V} \cap A = \varnothing$ and thus $x \in V \subseteq \overline{V} \subseteq U$ as desired.

Conversely, assume that for any $x \in X$ and any neighborhood U of x, there exists a neighborhood V of x such that $x \in V \subseteq \overline{V} \subseteq U$. Let $x \in X$ and $F \subseteq X$ be a closed set not containing x. Consider the open set $U = X \setminus F$. By hypothesis there is a neighborhood V of x such that $\overline{V} \subseteq U$. Therefore V and $X \setminus \overline{V}$ are disjoint open sets containing x and F respectively, as desired. □

[2] René Maurice Fréchet (1878–1973) was a French mathematician who made major contributions to topology, probability and calculus. In particular, he introduced the definition of metric spaces. Several concepts are named after him, for instance *Fréchet derivative* and *Fréchet filter*.

[3] Felix Hausdorff (1843–1896) was a German mathematician who made significant contributions to set theory, topology, measure theory, and functional analysis. Several concepts were named after him, for instance *Hausdorff maximal principle*, *Hausdorff distance*, *Hausdorff measure*. Under the pseudonym Paul Mongré, he also wrote poems and philosophical essays.

Proposition 1.39. *A Hausdorff topological space X is normal if and only if for any closed set F in X and any neighborhood U of F, there exists a neighborhood V of F such that $F \subseteq V \subseteq \overline{V} \subseteq U$.*

Proof. Similar as the proof of Proposition 1.38. □

Proposition 1.40. *Every metric space is normal.*

Proof. Let (X, d) be a metric space. Recall that for any $x \in X$ and any non-empty subset A of X, $d_A(x) := \inf_{a\in A} d(x, a) = 0$ if and only if $x \in \overline{A}$.

Let A, B be non-empty disjoint closed sets of X. For any $a \in A$ we define $\varepsilon_a := d_B(a)/3 > 0$, then we set $U = \bigcup_{a\in A} B(a, \varepsilon_a)$. For any $b \in A$ we define $\varepsilon_b := d_A(b)/3 > 0$, then we set $V = \bigcup_{b\in B} B(b, \varepsilon_b)$.

The sets U and V are clearly open and contain A and B respectively. Let us check that they are disjoint. By way of contradiction suppose that there is $x \in U \cap V$. Then there is $a \in A$ and $b \in B$ such that $x \in B(a, \varepsilon_a) \cap B(b, \varepsilon_b)$. This implies that $d(a, b) \leqslant d(a, x) + d(x, b) \leqslant \varepsilon_a + \varepsilon_b$. Without loss of generality we can suppose that $\varepsilon_a \geqslant \varepsilon_b$, then we obtain $d(a, b) \leqslant 2\varepsilon_a$, but by construction $d(a, b) \geqslant 3\varepsilon_a$, hence a contradiction. □

1.6.2 Axioms of countability

We list below some of the axioms of countability.

Definition 1.41. We say that a topological space X is:

(i) *First-countable* if every point has a countable neighborhood basis.

(ii) *Separable* if there exists a countable dense subset of X.

(iii) *Second-countable* if there is a countable basis for the topology of X.

One can also remark that in a first-countable space every point admits a neighborhood basis composed by a decreasing sequence of open sets. Indeed let x be a point in a first-countable space, then there exists a neighborhood basis $\{W_i\}_{i\in\mathbf{N}}$ of x. For any $n \in \mathbf{N}$ let $V_n = \bigcap_{0\leqslant i\leqslant n} \mathring{W}_i$, therefore the family $\{V_n\}_{n\in\mathbf{N}}$ is a decreasing sequence of open sets and it is a neighborhood basis for x.

Proposition 1.42. *Every second-countable space is first-countable and separable.*

Proof. Let X be a second-countable space, then there is a countable basis $\mathscr{B} = \{B_n\}_{n\in\mathbf{N}}$ for the topology of X.

We first show that X is separable. For any $n \in \mathbf{N}$ we pick some $x_n \in B_n$, and then define the countable subset $A = \{x_n : n \in \mathbf{N}\}$ of X. We now verify that $\overline{A} = X$. Let U be a non-empty open set in X, then there is some non-empty set $J \subseteq \mathbf{N}$ such that $U = \bigcup_{n\in J} B_n$. Therefore $U \cap A = \{x_n : n \in J\}$ is non-empty, thus $\overline{A} = X$.

We now show that X is first-countable. For any $x \in X$, define $\mathscr{S}(x) = \{B_n \in \mathscr{B} \mid x \in B_n\}$ which is a countable family of neighborhoods of x. Let V be a neighborhood of x, then there is a open set U in X such that $x \in U \subseteq V$. Since $\mathscr{B}$ is a basis, U is a union of elements of $\mathscr{B}$, therefore there is $n \in \mathbf{N}$ such that $x \in B_n \subseteq U \subseteq V$, hence $\mathscr{S}(x)$ is a neighborhood basis for x. □

Proposition 1.43. *Every metric space is first-countable.*

Proof. Let (X, d) be a metric space. For any $x \in X$ define $\mathcal{S}(x) = \{B(x, \frac{1}{n})\}_{n\in\mathbf{N}^*}$, which is a countable family of neighborhoods of x. Let V be a neighborhood of x, then there exists a open set U in X such that $x \in U \subseteq V$. Since U is open, there is $n \in \mathbf{N}^*$ large enough such that $B(x, \frac{1}{n}) \subseteq U \subseteq V$, thus $\mathcal{S}(x)$ is a neighborhood basis for x. □

Proposition 1.44. *For metric spaces second-countability and separability are equivalent.*

Proof. Let (X, d) be a metric space. Thanks to Proposition 1.42 we only need to show that separable implies second-countable. Assume X is separable, then there is a countable subset $A \subseteq X$ such that $\overline{A} = X$. Define $\mathscr{B} = \{B(y, \frac{1}{n})\}_{y\in A, n\in\mathbf{N}^*}$, which is a countable family of open subsets of X.

Let us show that $\mathscr{B}$ is a basis for the topology of X. Let U be a open set in X and $x \in U$. Since U is open, there is $\varepsilon > 0$ such that $B(x, \varepsilon) \subseteq U$, hence there exists $n \in \mathbf{N}^*$ large enough such that $B(x, \frac{2}{n}) \subseteq U$. Since $\overline{A} = X$ and $B(x, \frac{1}{n})$ is a neighborhood of x, there is a point $y \in X$ such that $y \in A \cap B(x, \frac{1}{n})$. Therefore $B(y, \frac{1}{n}) \subseteq B(x, \frac{2}{n}) \subseteq U$. □

Chapter 2

Continuity

2.1 Continuous maps

Definition 2.1. A map $f : X \to Y$ between two topological spaces is *continuous at a point* $a \in X$ if for any neighborhood V of $f(a)$ in Y, there is a neighborhood U of a in X such that $f(U) \subseteq V$.

We say that f is *continuous on* X (or simply *continuous*) if f is continuous at every point of X.

Observing that $f(U) \subseteq V$ is equivalent to $U \subseteq f^{-1}(V)$, we can thus reformulate the definition of continuity of f at the point a by: for any neighborhood V of $f(a)$ in Y, the set $f^{-1}(V)$ is a neighborhood of a in X.

Thanks to the definition of a neighborhood basis, one can observe that if $\mathscr{U}$ is a neighborhood basis for a in X and $\mathscr{V}$ a neighborhood basis for $f(a)$ in Y, then f is continuous at a if and only if for any $V \in \mathscr{V}$ there exists $U \in \mathscr{U}$ such that $f(U) \subseteq V$.

If (Y, d_Y) is a metric space, then $f : X \to Y$ is continuous at $a \in X$ if and only if for any $\varepsilon > 0$, there is a neighborhood U of a such that, for any $x \in U$, one has $d_Y(f(x), f(a)) < \varepsilon$.

Moreover, if both (X, d_X) and (Y, d_Y) are metric spaces, then $f : X \to Y$ is continuous at $a \in X$ if and only if for all $\varepsilon > 0$ there is $\delta > 0$ such that, for any $x \in X$, if $d_X(x, a) < \delta$ then $d_Y(f(x), f(a)) < \varepsilon$.

It is easy to obtain that the composition of continuous maps is continuous.

Proposition 2.2. *Let X, Y and Z be topological spaces. Suppose that $f : X \to Y$ is continuous at $a \in X$ and that $g : Y \to Z$ is continuous at $f(a) \in Y$. Then $g \circ f : X \to Z$ is continuous at $a \in X$.*

Proof. Let V be a neighborhood of $(g \circ f)(a) = g(f(a))$ in Z. Since g is continuous at $f(a)$, one has that $g^{-1}(V)$ is a neighborhood of $f(a)$ in Y. Since f is continuous at a, $f^{-1}(g^{-1}(V)) = (g \circ f)^{-1}(V)$ is a neighborhood of a in X. □

We prove now equivalent formulations of continuity.

Theorem 2.3. *Let $f : X \to Y$ be a map between two topological spaces. The following are equivalent:*

(i) *f is continuous on X;*

(ii) *For any open set U in Y, $f^{-1}(U)$ is an open set in X;*

(iii) *For any closed set C in Y, $f^{-1}(C)$ is a closed set in X.*

If $\mathscr{B}$ is a basis for the topology of Y and $\mathscr{S}$ is a sub-basis for the topology of Y, then one can easily verify that (ii) is equivalent to each one of the following properties:

(ii') For any $U \in \mathscr{B}$, $f^{-1}(U)$ is an open set in X ;

(ii'') For any $U \in \mathscr{S}$, $f^{-1}(U)$ is an open set in X.

Indeed the implications (ii) $\Rightarrow$ (ii') $\Rightarrow$ (ii'') are trivial. The converse follows from the fact that an open set in X is a union of elements of $\mathscr{B}$, and also a union of finite intersections of elements of $\mathscr{S}$, together with the property that the preimage under f commutes with unions and intersections.

Proof of Theorem 2.3. (i) $\Leftrightarrow$ (ii) Assume that f is continuous on X. Let V be an open set in Y, we shall prove that $f^{-1}(V)$ is a neighborhood of each of its points, which implies that $f^{-1}(V)$ is open in X. Let $x \in f^{-1}(V)$, then $f(x) \in V$ and thus V is a neighborhood of $f(x)$. Since f is continuous at x, there is a neighborhood U of x in X such that $f(U) \subseteq V$, that is $U \subseteq f^{-1}(V)$, and hence $f^{-1}(V)$ is a neighborhood of x. Conversely, assume that for any open set U in Y, $f^{-1}(U)$ is open in X. Let $x \in X$ and V be a neighborhood of $f(x)$ in Y. There is an open set $U \subseteq Y$ such that $f(x) \in U \subseteq V$, which implies that $x \in f^{-1}(U) \subseteq f^{-1}(V)$ and $f^{-1}(U)$ is open in X, hence $f^{-1}(V)$ is a neighborhood of x in X.

(ii) $\Leftrightarrow$ (iii) It follows from the relation $X \setminus f^{-1}(B) = f^{-1}(Y \setminus B)$ for any subset B of Y. □

Example 2.4. (1) Every map from a discrete space into an arbitrary topological space is continuous.

(2) Every map from an arbitrary topological space into a space with the trivial topology is continuous.

(3) A constant map between topological spaces is continuous.

(4) The identity map from a topological space into itself is continuous.

2.2 Open and closed maps, homeomorphisms

Definition 2.5. Let X and Y be topological spaces. A map $f : X \to Y$ is said to be *open* (resp. *closed*) if the image under f of every open (resp. closed) set in X is open (resp. closed) in Y.

Definition 2.6. A bijective map $f : X \to Y$ between topological spaces is a *homeomorphism* if both f and its inverse map f^{-1} are continuous. Two topological spaces X and Y are said to be *homeomorphic* if there exists a homeomorphism between them.

In other words, a bijection $f : X \to Y$ is an homeomorphism if and only if the image under f of every open set in X is a open set in Y and the preimage of every open set in Y is a open set in X, that is, if and only if f is continuous and open.

In the framework of metric spaces one defines the following:

Definition 2.7. A map $f : X \to Y$ between metric spaces (X, d_X) and (Y, d_Y) is called an *isometry* if for any $x, y \in X$ one has $d_Y(f(x), f(y)) = d_X(x, y)$. An *isometric isomorphism* is a bijective isometry. Two metric spaces X and Y are said to be *isometric* if there exists an isometric isomorphism between them.

2.3 Construction of topologies

2.3.1 Initial topology

Definition 2.8. Let X be a set, $\{Y_i\}_{i\in I}$ a collection of topological spaces and $f_i : X \to Y_i$ a map, for any $i \in I$. The *initial topology on X defined by* $\{f_i\}_{i\in I}$ is the coarsest topology on X for which every map f_i is continuous. This topology is generated by $\{f_i^{-1}(U_i) : i \in I, U_i \text{ open in } Y_i\}$.

The properties below easily follow from the definition:

Proposition 2.9. *Let X be endowed with the initial topology defined by the family of maps* $\{f_i : X \to Y_i\}_{i\in I}$ *as above.*

(i) *If $\mathscr{B}_i$ is a basis for the topology on Y_i for all $i \in I$, then the family $\mathscr{B}$ of finite intersection of elements of $\{f_i^{-1}(B_i) : i \in I, B_i \in \mathscr{B}_i\}$ is a basis for the initial topology on X. In particular, if I is countable and Y_i is second-countable for all $i \in I$,then X is second-countable.*

(ii) *If $x \in X$ and $\mathcal{S}_i$ is a neighborhood basis for $f_i(x)$ on Y_i, then the family $\mathcal{S}$ of finite intersection of elements of $\{f_i^{-1}(V_i) : i \in I, V_i \in \mathcal{S}_i\}$ is a neighborhood basis for x on X. In particular, if I is countable and Y_i is first-countable for all $i \in I$, then X is first-countable.*

(iii) *Let Z be a topological space and $g : Z \to X$ be a map. Then g is continuous if and only if every $f_i \circ g$ is continuous.*

Proof. Let us denote by $\mathscr{O}$ the initial topology on X and by $\mathscr{O}_i$ the topology of Y_i, for all $i \in I$.

(i) Let $U \in \mathscr{O}$ and $x \in U$. By definition of $\mathscr{O}$, U is a union of finite intersections of elements of $\{f_i^{-1}(U_i) : i \in I, U_i \in \mathscr{O}_i\}$, hence there is a finite set $J \subseteq I$ and, for each $i \in J$, an open set $U_i \in \mathscr{O}_i$ such that

$$x \in \bigcap_{i\in J} f_i^{-1}(U_i) \subseteq U.$$

For any $i \in J$ one has $x \in f_i^{-1}(U_i)$, thus $f_i(x) \in U_i$. Since $\mathscr{B}_i$ is a basis for the topology $\mathscr{O}_i$, there is $B_i \in \mathscr{B}_i$ such that $f_i(x) \in B_i \subseteq U_i$, hence $x \in f_i^{-1}(B_i) \subseteq f^{-1}(U_i)$.

Therefore,

$$x \in \bigcap_{i \in J} f_i^{-1}(B_i) \subseteq \bigcap_{i \in J} f_i^{-1}(U_i) \subseteq U,$$

and the set in the left-hand side belongs to $\mathscr{B}$.

(ii) Let $x \in X$ and V be a neighborhood of x. Then there is $U \in \mathscr{O}$ such that $x \in U \subseteq V$. By definition, there is a finite set $J \subseteq I$ and, for each $i \in J$, an open set $U_i \in \mathscr{O}_i$ such that

$$x \in \bigcap_{i \in J} f_i^{-1}(U_i) \subseteq V.$$

For any $i \in J$ one has $x \in f_i^{-1}(U_i)$, thus $f_i(x) \in U_i$. Since $\mathscr{S}_i$ is a neighborhood basis for $f_i(x)$ in Y_i and U_i is a neighborhood of $f_i(x)$, there is some $V_i \in \mathscr{S}_i$ such that $f_i(x) \in V_i \subseteq U_i$, hence $x \in f_i^{-1}(V_i) \subseteq f^{-1}(U_i)$. Therefore,

$$x \in \bigcap_{i \in J} f_i^{-1}(V_i) \subseteq \bigcap_{i \in J} f_i^{-1}(U_i) \subseteq V,$$

and the set in the left-hand side belongs to $\mathscr{S}$.

(iii) Suppose g is continuous. Since f_i is continuous for each $i \in I$, by composition each $f_i \circ g$ is continuous.

Conversely, suppose that $f_i \circ g$ is continuous for each $i \in I$. Let $z \in Z$ and V be a neighborhood of $g(z)$ in X. Hence there is a finite set $J \subseteq I$ and for each $i \in J$ an open set $U_i \in \mathscr{O}_i$ such that

$$g(z) \in \bigcap_{i \in J} f_i^{-1}(U_i) \subseteq V.$$

Therefore

$$z \in \bigcap_{i \in J} g^{-1}(f_i^{-1}(U_i)) = \bigcap_{i \in J} (f_i \circ g)^{-1}(U_i) \subseteq g^{-1}(V),$$

and the set in the left-hand side is open in X since it is a finite intersection of open sets, hence $g^{-1}(V)$ is a neighborhood of z in Z, which completes the proof. □

Two particular examples of initial topologies are the *subspace topology* and the *product topology* that we describe below.

2.3.1.1 Subspace topology

Definition 2.10. Let $(X, \mathscr{O})$ be a topological space and $Y \subseteq X$. We define the *subspace topology* $\mathscr{O}_Y$ on Y as the trace on Y of the topology of X, namely

$$\mathscr{O}_Y := \{U \cap Y \mid U \in \mathscr{O}\}.$$

The topological space $(Y, \mathscr{O}_Y)$ is called a (topological) subspace of $(X, \mathscr{O})$.

Let us verify that $\mathscr{O}_Y$ is indeed a topology on Y. Since $\varnothing, X \in \mathscr{O}$, we get $\varnothing = \varnothing \cap Y$, $Y = X \cap Y \in \mathscr{O}_Y$. Let $\{V_i\}_{1 \leqslant i \leqslant n}$ be a finite family of elements of $\mathscr{O}_Y$, then there exist $U_i \in \mathscr{O}$ such that $V_i = U_i \cap Y$ for any $i \in \{1, \ldots, n\}$. We have $U := \bigcap_{1 \leqslant i \leqslant n} U_i \in \mathscr{O}$

and $V := \bigcap_{1\leqslant i\leqslant n} V_i = \left(\bigcap_{1\leqslant i\leqslant n} U_i\right) \cap Y$, whence $V \in \mathscr{O}_Y$. Let $\{V_i\}_{i\in I}$ be an arbitrary family of elements of $\mathscr{O}_Y$, then there exist $U_i \in \mathscr{O}$ such that $V_i = U_i \cap Y$ for any $i \in I$. We have $U := \bigcup_{i\in I} U_i \in \mathscr{O}$ and $V := \bigcup_{i\in I} V_i = (\bigcup_{i\in I} U_i) \cap Y$, whence $V \in \mathscr{O}_Y$.

In this way, given a topological space X and a subset $Y \subseteq X$, we shall always consider Y endowed with the subspace topology, unless otherwise specified.

The following properties are easy to verify.

Proposition 2.11. *Let $(X, \mathscr{O})$ be a topological space and $(Y, \mathscr{O}_Y)$ a subspace of X.*

(i) *The topology $\mathscr{O}_Y$ is the coarsest topology on Y for which the inclusion map $\iota : Y \to X$, $x \mapsto x$ is continuous. In other words, the subspace topology is the initial topology on Y defined by the inclusion map.*

(ii) *For any $x \in Y$, $V \subseteq Y$ is a neighborhood of x in Y if and only if there is a neighborhood V' of x in X such that $V = V' \cap Y$. In particular, if X is Hausdorff then so is Y.*

(iii) *If $\mathscr{B}$ is a basis for the topology on X, then $\mathscr{B}_Y := \{U \cap Y \mid U \in \mathscr{B}\}$ is a basis for the subspace topology on Y. In particular, if X is second-countable then so is Y.*

(iv) *For any $x \in Y$, if $\mathcal{S}$ is a neighborhood basis for x in X, then $\mathcal{S}_Y := \{U \cap Y \mid U \in \mathcal{S}\}$ is a neighborhood basis for x in Y. In particular, if X is first-countable then so is Y.*

Proof. (i) It is clear that $\iota : (Y, \mathscr{O}_Y) \to (X, \mathscr{O})$ is continuous, indeed if $U \in \mathscr{O}$, then $\iota^{-1}(U) = U \cap Y \in \mathscr{O}_Y$. Now let $\mathscr{T}$ be a topology on Y such that $\iota : (Y, \mathscr{T}) \to (X, \mathscr{O})$ is continuous, then for any $U \in \mathscr{O}$ we have $U \cap Y \in \mathscr{T}$, hence $\mathscr{T} \supseteq \mathscr{O}_Y$.

(ii) Let V be a neighborhood of x in Y, then there exists $U \in \mathscr{O}_Y$ such that $x \in U \subseteq V$. Hence there is $U' \in \mathscr{O}$ such that $U = U' \cap Y$ and U' is a neighborhood of x in X. Thus the set $V' = V \cup U'$ is a neighborhood of x in X and satisfies $V = V' \cap Y$.

Conversely, let V' be a neighborhood of x in X, then there is $U' \in \mathscr{O}$ such that $x \in U' \subseteq V'$. Since $x \in Y$, one has $x \in U' \cap Y \subseteq V' \cap Y$ and $U' \cap Y \in \mathscr{O}_Y$, thus $V' \cap Y$ is a neighborhood of x in Y.

(iii) Let $U \in \mathscr{O}_Y$ and $x \in U$. We want to show that there exists $W \in \mathscr{B}_Y$ such that $x \in W \subseteq U$. There is $U' \in \mathscr{O}$ such that $U = U' \cap Y$. Since $\mathscr{B}$ is a basis for $\mathscr{O}$, there is $W' \in \mathscr{B}$ such that $x \in W' \subseteq U'$. Hence $W' \cap Y \in \mathscr{B}_Y$ and $x \in W' \cap Y \subseteq U' \cap Y$.

(iv) Let V be a neighborhood of x in Y. We want to show that there exists $W \in \mathcal{S}_Y$ such that $x \in W \subseteq V$. There is a neighborhood V' of x in X such that $V = V' \cap Y$. Since $\mathcal{S}$ is a neighborhood basis for x in X, there is $W' \in \mathcal{S}$ such that $x \in W' \subseteq V'$. Therefore $W' \cap Y \in \mathcal{S}_Y$ and $x \in W' \cap Y \subseteq V' \cap Y = V$. □

2.3.1.2 Product topology

Definition 2.12. Consider a collection $\{X_i\}_{i\in I}$ of topological spaces. On the product set $X = \prod_{i\in I} X_i$, we define the *product topology* as the coarsest topology on X for which the canonical projection $\mathrm{pr}_j : X \to X_j$, $(x_i)_{i\in I} \mapsto x_j$ is continuous for all $j \in I$.

In other words, the product topology is the initial topology on X defined by the canonical projections $\{\mathrm{pr}_i\}_{i\in I}$.

The properties below follow from the properties of the initial topology (see Proposition 2.9).

(i) A basis for this topology is given by the collection of sets of the form

$$\{(x_i)_{i\in I} \in X \mid \forall j \in J, x_j \in U_j\} = \bigcap_{j\in J} \mathrm{pr}_j^{-1}(U_j)$$

where $J \subseteq I$ is finite and U_j is open in X_j for any $j \in J$.

(ii) If $\mathscr{B}_i$ is a basis for the topology of X_i for all $i \in I$, then the family of sets of the form

$$\{(x_i)_{i\in I} \in X \mid \forall j \in J, x_j \in U_j\} = \bigcap_{j\in J} \mathrm{pr}_j^{-1}(U_j),$$

where $J \subseteq I$ is finite and $U_j \in \mathscr{B}_j$ for any $j \in J$, is a basis for the product topology.

(iii) Let $x = (x_i)_{i\in I} \in X$ and $\mathscr{V}_i$ be a neighborhood basis for x_i in X_i for any $i \in I$, then the family of sets of the form

$$\{(y_i)_{i\in I} \in X \mid \forall j \in J, y_j \in V_j\} = \bigcap_{j\in J} \mathrm{pr}_j^{-1}(V_j),$$

where $J \subseteq I$ is finite and $V_j \in \mathscr{V}_j$ for any $j \in J$, is a neighborhood basis for x in X for the product topology.

The next result shows that any product of Hausdorff spaces is also Hausdorff.

Proposition 2.13. *Let $\{X_i\}_{i\in I}$ be a family of Hausdorff topological spaces. Then the product space $X = \prod_{i\in I} X_i$ is Hausdorff.*

Proof. Let $x = (x_i)_{i\in I} \neq (y_i)_{i\in I} = y$ belong to X. There is $k \in I$ such that $x_k \neq y_k$ thus, since X_k is Hausdorff, there are disjoint open sets U_k and V_k in X_k such that $x_k \in U_k$ and $y_k \in V_k$. Therefore the sets $\mathrm{pr}_k^{-1}(U_k)$ and $\mathrm{pr}_k^{-1}(V_k)$ are disjoint open sets in X containing x and y, respectively. □

2.3.2 Final topology

Definition 2.14. Let X be a set, $\{Y_i\}_{i\in I}$ a collection of topological spaces and $f_i : Y_i \to X$ a map, for any $i \in I$. The *final topology on X defined by $\{f_i\}_{i\in I}$* is the finest topology on X for which every map f_i is continuous. This topology is given by the collection of all subsets $U \subseteq X$ such that $f_i^{-1}(U)$ is open in Y_i for any $i \in I$.

We easily verify the following property:

Proposition 2.15. *Let X be endowed with the final topology associated to the family of maps $\{f_i : Y_i \to X\}_{i\in I}$ as above. Let Z be a topological space and $g : X \to Z$ a map. Then g is continuous if and only if $g \circ f_i$ is continuous for every $i \in I$.*

Proof. Assume that g is continuous. Since f_i is continuous for every $i \in I$, we deduce that $g \circ f_i$ is continuous by composition.

Conversely, assume that $g \circ f_i$ is continuous for every $i \in I$. Let V be open in Z, then $(g \circ f_i)^{-1}(V) = f_i^{-1}(g^{-1}(V))$ is open in Y_i for every $i \in I$, hence by definition $g^{-1}(V)$ is open for the final topology. This implies that g is continuous. □

A particular example of final topology is the *quotient topology* described below.

2.3.2.1 Quotient topology

Let X be a topological space and $\mathscr{R}$ be an equivalence relation on X. Denote by $X/\mathscr{R}$ the quotient set and consider the canonical projection $\pi : X \to Y$, $x \mapsto [x]$, that sends every point to its equivalence class.

Definition 2.16. We define the *quotient topology* on $X/\mathscr{R}$ as the finest topology on $X/\mathscr{R}$ for which π is continuous.

In other words, the quotient topology is the final topology on $X/\mathscr{R}$ defined by the map π. Therefore, this topology is given by the collection of subsets $U \subseteq X/\mathscr{R}$ such that $\pi^{-1}(U)$ is open in X.

2.4 Limits and cluster points of a sequence

Let X be a topological space. A sequence $(x_n)_{n\in\mathbf{N}}$ in X corresponds to a map $n \mapsto x_n$ from $\mathbf{N}$ into X.

One says that a sequence $(y_n)_{n\in\mathbf{N}}$ in X is a *subsequence* of $(x_n)_{n\in\mathbf{N}}$ if there is a strictly increasing function $\varphi : \mathbf{N} \to \mathbf{N}$ such that

$$y_n = x_{\varphi(n)} \quad \text{for all } n \in \mathbf{N}.$$

The function φ is called an *extraction map* and it verifies $\varphi(n) \geqslant n$ for all $n \in \mathbf{N}$.

Let $(x_n)_{n\in\mathbf{N}}$ be a sequence in X and $x \in X$ a point.

Definition 2.17. One says that $(x_n)_{n\in\mathbf{N}}$ *converges to x when n goes to* $+\infty$ if for any neighborhood V of x in X, there is $N_0 \in \mathbf{N}$ such that for any $n \geqslant N_0$, one has $x_n \in V$.

When $(x_n)_{n\in\mathbf{N}}$ converges to x, we denote it by

$$\lim_{n\to+\infty} x_n = x \quad \text{or} \quad x_n \xrightarrow[n\to+\infty]{} x.$$

If $\mathscr{V}$ is a neighborhood basis for x in X, then $(x_n)_{n\in\mathbf{N}}$ converges to x if and only if for any $V \in \mathscr{V}$ there is $N_0 \in \mathbf{N}$ such that $x_n \in V$ for all integer $n \geqslant N_0$. In particular, if (X, d) is a metric space, then $(x_n)_{n\in\mathbf{N}}$ converges to x if and only if the real sequence $(d(x_n, x))_{n\in\mathbf{N}}$ converges to 0.

Definition 2.18. One says that x *is a cluster point of the sequence* $(x_n)_{n\in\mathbf{N}}$ if for any neighborhood V of x in X and any $N \in \mathbf{N}$, there exists an integer $n \geqslant N$ such that $x_n \in V$.

It easily follows from the definition that the set of all cluster points of a sequence $(x_n)_{n\in\mathbf{N}}$ is given by

$$\bigcap_{n\in\mathbf{N}} \overline{\{x_p : p \geqslant n\}}.$$

We now prove some basic properties concerning the sequential caracterization of closed sets and the closure of a set.

Proposition 2.19. *Let A be a subset of X. If there is a sequence of elements of A converging to $x \in X$, then $x \in \overline{A}$. If X is first-countable then the converse also holds.*

Proof. Let $(x_n)_{n\in\mathbf{N}}$ be a sequence of elements of A that converges to $x \in X$. Then for any neighborhood V of x, there is $N \in \mathbf{N}$ such that $x_N \in V$ and $x_N \in A$, hence $V \cap A \neq \emptyset$. This proves that $x \in \overline{A}$.

Assume now that X is first-countable and let $x \in \overline{A}$. Let $\{V_n\}_{n\in\mathbf{N}}$ be a countable neighborhood basis for x composed by a decreasing sequence of opens sets. For any $n \in \mathbf{N}$ there exists $x_n \in V_n \cap A \neq \emptyset$. We now verify that the sequence $(x_n)_{n\in\mathbf{N}}$ converges to x. Let V be a neighborhood of x in X, then there is $N_0 \in \mathbf{N}$ such that $x \in V_{N_0} \subseteq V$. Therefore for any $n \geqslant N_0$ one has $x_n \in V_n \subseteq V_{N_0} \subseteq V$. □

As a consequence, closed sets in a first-countable space X can be characterized by sequences: A subset $A \subseteq X$ is closed if and only if for every sequence of elements of A that converges in X, the limit belongs to A.

Proposition 2.20. *Let $(x_n)_{n\in\mathbf{N}}$ be a sequence in X. If there exists some subsequence of $(x_n)_{n\in\mathbf{N}}$ converging to $x \in X$, then x is a cluster point of $(x_n)_{n\in\mathbf{N}}$. If X is first-countable, then the converse also holds.*

Proof. Let $(x_{\varphi(n)})_{n\in\mathbf{N}}$ be a subsequence of $(x_n)_{n\in\mathbf{N}}$ that converges to x. Let V be a neighborhood of x and $N_0 \in \mathbf{N}$. Since $(x_{\varphi(n)})_{n\in\mathbf{N}}$ converges to x, there is $N_1 \in \mathbf{N}$ such that for all integer $n \geqslant N_1$ one has $x_{\varphi(n)} \in V$. Choosing $n = \max(N_0, N_1) \in \mathbf{N}$, one has $\varphi(n) \geqslant N_0$ and $x_{\varphi(n)} \in V$. This proves that x is a cluster point of $(x_n)_{n\in\mathbf{N}}$.

Assume now that X is first-countable and let x be a cluster point of the sequence $(x_n)_{n\in\mathbf{N}}$. Let $\{V_n\}_{n\in\mathbf{N}}$ be a countable neighborhood basis for x composed by a decreasing sequence of opens sets. Since V_0 is an open neighborhood of x, there is $n_0 \in \mathbf{N}$ such that $x_{n_0} \in V_0$. By induction one can therefore construct a sequence $(x_{n_k})_{k\in\mathbf{N}}$ such that $n_k > n_{k-1}$ and $x_{n_k} \in V_k$ for all $k \in \mathbf{N}^*$, thus $(x_{n_k})_{k\in\mathbf{N}}$ is a subsequence of $(x_n)_{n\in\mathbf{N}}$.

Let us now verify that $(x_{n_k})_{k\in\mathbf{N}}$ converges to x. Let V be a neighborhood of x, then there is $N_0 \in \mathbf{N}$ such that $x \in V_{N_0} \subseteq V$. Hence for any $k \geqslant N_0$ one has $n_k \geqslant k$ and thus $x_{n_k} \in V_{n_k} \subseteq V_{N_0} \subseteq V$. □

As a consequence, in a first-countable space X the closure of a subset can be characterized using sequences: If A is a subset of X, then the closure $\overline{A}$ is the set of all limits of all convergent sequences of points in A.

Chapter 3

Completeness

In this chapter we shall always consider metric spaces.

3.1 Cauchy sequences

Definition 3.1. A sequence $(x_n)_{n\in\mathbf{N}}$ in a metric space (X, d) is a *Cauchy*[1] *sequence* if for all $\varepsilon > 0$ there is an integer $N_0 \in \mathbf{N}$ such that, for all integers $n, m \geqslant N_0$, one has $d(x_n, x_m) < \varepsilon$.

One easily verifies that every convergent sequence is a Cauchy sequence, and that every Cauchy sequence is bounded.

Proposition 3.2. *Let (X, d) be a metric space, then the following properties hold:*

(i) *Every convergent sequence is a Cauchy sequence.*

(ii) *Every Cauchy sequence is bounded.*

Proof. (i) Let $(x_n)_{n\in\mathbf{N}}$ be a sequence in X that converges to $x \in X$. Let $\varepsilon > 0$, then there is $N_0 \in \mathbf{N}$ such that $d(x_n, x) < \varepsilon/2$ for all $n \geqslant N_0$. Thanks to the triangle inequality, for all $n, m \geqslant N_0$ one obtains

$$d(x_n, x_m) \leqslant d(x_n, x) + d(x_m, x) < \varepsilon.$$

(ii) Let $(x_n)_{n\in\mathbf{N}}$ be a Cauchy sequence in X. We want to show that there is some $a \in X$ and $M > 0$ such that $d(x_n, a) \leqslant M$ for all $n \in \mathbf{N}$. There is $N_0 \in \mathbf{N}$ such that $d(x_n, x_{N_0}) \leqslant 1$ for all $n \geqslant N_0$. Denote

$$M' := \max\left\{d(x_0, x_{N_0}), \ldots, d(x_{N_0-1}, x_{N_0})\right\}$$

then for all $n \in \mathbf{N}$ we get $d(x_n, x_{N_0}) \leqslant \max\{M', 1\}$. □

We now prove another property concerning the convergence of Cauchy sequences.

[1] Augustin-Louis Cauchy (1789–1857) was a French mathematician who worked in all fields of mathematics of his period. He was a prolific worker and, in particular, made significant contributions to real analysis, complex analysis, algebra, geometry, mechanics and optics. He was a student and later professor at École polytechnique.

Proposition 3.3. *If a Cauchy sequence in a metric space (X, d) has a cluster point, then it converges to it.*

Proof. Let $(x_n)_{n\in\mathbf{N}}$ be a Cauchy sequence in X that has a cluster point $x \in X$. From Proposition 2.20, there is some subsequence $(x_{\varphi(n)})_{n\in\mathbf{N}}$ that converges to x. Let $\varepsilon > 0$. On the one hand, since $(x_{\varphi(n)})_{n\in\mathbf{N}}$ converges to x, there is $N_0 \in \mathbf{N}$ such that $d(x_{\varphi(n)}, x) < \varepsilon/2$ for any integer $n \geqslant N_0$. On the other hand, since $(x_n)_{n\in\mathbf{N}}$ is a Cauchy sequence, there is $N_1 \in \mathbf{N}$ such that $d(x_n, x_m) < \varepsilon/2$ for any integers $n, m \geqslant N_1$. Therefore for any $n \geqslant \max(N_0, N_1)$ it follows

$$d(x_n, x) \leqslant d(x_n, x_{\varphi(n)}) + d(x_{\varphi(n)}, x) < \varepsilon.$$

□

3.2 Complete metric spaces

Definition 3.4. A metric space (X, d) is said to be *complete* if every Cauchy sequence converges.

We say that a subset Y of a metric space X is complete if the metric subspace Y (Example 1.15–(3)) is complete.

Proposition 3.5. *A closed subset of a complete metric space is complete.*

Proof. Let (X, d) be a complete metric space and $Y \subseteq X$ be closed. Let $(x_n)_{n\in\mathbf{N}}$ be a Cauchy sequence in Y. Then $(x_n)_{n\in\mathbf{N}}$ is a Cauchy sequence in X and hence there is $y \in X$ such that $(x_n)_{n\in\mathbf{N}}$ converges to y in X. Since Y is closed, one concludes that $y \in Y$ by Proposition 2.19. □

Proposition 3.6. *A complete subspace of a metric space is closed.*

Proof. Let (X, d) be a metric space and $Y \subseteq X$ be a complete subspace. Let $y \in \overline{Y}$, then there is a sequence $(x_n)_{n\in\mathbf{N}}$ in Y that converges to y in X by Proposition 2.19. Therefore $(x_n)_{n\in\mathbf{N}}$ is a Cauchy sequence in Y that is complete, hence it converges to some $\bar{y} \in Y$. By uniqueness of the limit, one obtains $y = \bar{y} \in Y$. □

Definition 3.7. A topological space $(X, \mathcal{O})$ is *completely metrizable* if there is a metric d on X that induces the topology $\mathcal{O}$ and for which (X, d) is complete.

Example 3.8. (1) The set of real numbers $\mathbf{R}$ endowed with its standard metric (Example 1.15–(2)) is complete. However the set of rational numbers $\mathbf{Q}$ endowed with the restriction of this metric is not complete.

(2) If $(X_1, d_1), \dots, (X_n, d_n)$ are complete metric spaces, then the product space $X = \prod_{i=1}^n X_i$ is complete for any of the following metrics

$$d_p((x_1, \dots, x_n), (y_1, \dots, y_n)) = \left\{ \sum_{i=1}^n d_i(x_i, y_i)^p \right\}^{1/p}, \quad 1 \leqslant p < +\infty,$$
$$d_\infty((x_1, \dots, x_n), (y_1, \dots, y_n)) = \max_{1\leqslant i\leqslant n} d_i(x_i, y_i).$$

3.3 Fixed-point theorem

Let (X, d_X) and (Y, d_Y) be metric spaces.

Definition 3.9. A map $f : X \longrightarrow Y$ is *uniformly continuous* if for all $\varepsilon > 0$ there is $\delta > 0$ such that for all $x, x' \in X$, if $d_X(x, x') < \delta$ then $d_Y(f(x), f(x')) < \varepsilon$.

Definition 3.10. A map $f : X \longrightarrow Y$ is called *k-Lipschitz*[2], with $k \geqslant 0$, if for all $x, x' \in X$ there holds

$$d_Y(f(x), f(x')) \leqslant k d_X(x, x').$$

A map $f : X \longrightarrow Y$ is *Lipschitz* if it is k-Lipschitz for some $k \geqslant 0$, and it is a *contraction* if it is k-Lipschitz with $0 \leqslant k < 1$.

It is clear that "Lipschitz" implies "uniformly continuous", which implies "continuous".

Theorem 3.11 (Banach[3]-Picard[4] fixed-point theorem)**.** *Let (X, d) be a complete metric space and consider a contraction map $f : X \longrightarrow X$. Then f has a unique fixed point, that is, there is a unique $\bar{x} \in X$ such that $f(\bar{x}) = \bar{x}$.*

Proof. Since f is a contraction map, there is $0 \leqslant k < 1$ such that $d(f(x), f(y)) \leqslant k d(x, y)$ for any $x, y \in X$.

We first prove uniqueness. Assume $\bar{x}, \bar{y} \in X$ are fixed points of f, then $d(\bar{x}, \bar{y}) = d(f(\bar{x}), f(\bar{y})) \leqslant k d(\bar{x}, \bar{y})$, which implies $d(\bar{x}, \bar{y}) = 0$ and hence $\bar{x} = \bar{y}$.

We shall now construct a convergent sequence in X whose limit is a fixed point of f. Let $x_0 \in X$ and define $x_n = f(x_{n-1})$ for any $n \in \mathbf{N}^*$. Let us verify that $(x_n)_{n \in \mathbf{N}}$ is a Cauchy sequence. For any $n \in \mathbf{N}^*$ one has

$$d(x_{n+1}, x_n) = d(f(x_n), f(x_{n-1})) \leqslant k d(x_n, x_{n-1}),$$

hence one gets by induction

$$d(x_{n+1}, x_n) \leqslant k^n d(x_1, x_0),$$

therefore for any integers $n, p \in \mathbf{N}$ one obtains

$$\begin{aligned} d(x_{n+p}, x_n) &\leqslant d(x_{n+p}, x_{n+p-1}) + d(x_{n+p-1}, x_{n+p-2}) + \cdots + d(x_{n+1}, x_n) \\ &\leqslant \left(k^{n+p-1} + k^{n+p-2} + \cdots + k^n\right) d(x_1, x_0) \\ &\leqslant \frac{k^n}{1-k} d(x_1, x_0). \end{aligned}$$

[2] Rudolf Lipschitz (1832–1903) was a German mathematician who worked in the fields of analysis, differential geometry, algebra, and mechanics.

[3] Stefan Banach (1892–1945) was a Polish mathematician who made important contributions to functional analysis. Several concepts bear his name, in particular *Banach space*, *Banach-Steinhaus theorem*, *Banach-Alaoglu theorem*, *Hahn-Banach theorem*, *Banach-Tarski paradox*.

[4] Emile Picard (1856–1941) was a French mathematician who made contributions to analysis.

Fron this inequality we deduce that $(x_n)_{n\in\mathbf{N}}$ is a Cauchy sequence in X. Indeed, let $\varepsilon > 0$, then we choose $N_0 \in \mathbf{N}$ large enough such that $\frac{k^{N_0}}{1-k} d(x_1, x_0) < \varepsilon$, then for all integers $n, m \geqslant N_0$ one has

$$d(x_m, x_n) \leqslant \frac{k^{\min(n,m)}}{1-k} d(x_1, x_0) \leqslant \frac{k^{N_0}}{1-k} d(x_1, x_0) < \varepsilon.$$

Since X is complete one deduces that there is $\bar{x} \in X$ such that $(x_n)_{n\in\mathbf{N}}$ converges to $\bar{x}$. Since f is continuous at $\bar{x}$ and X is Hausdorff (so that limits are unique), one can take the limit $n \longrightarrow +\infty$ in the relation $x_{n+1} = f(x_n)$, which yields $\bar{x} = f(\bar{x})$. □

3.4 Baire's category theorem

Definition 3.12. A topological space is said to be a *Baire*[5] *space* if the intersection of any countable family of dense open sets is dense.

Equivalently, by passing to complement sets, a topological space is a Baire space if the union of any countable family of closed sets with empty interior has empty interior. In other words, in a Baire space, if the union of a countable collection of closed sets has non-empty interior, then at least one of these closed sets has non-empty interior.

Theorem 3.13 (Baire's category theorem, first version). *Every complete metric space is a Baire space.*

Proof. Let X be a complete metric space and $\{U_n\}_{n\in\mathbf{N}}$ be a countable family of dense open sets in X. Denoting $U := \bigcap_{n\in\mathbf{N}} U_n$ the intersection, we want to show that U is dense in X, that is, every non-empty open set W in X intersects U.

Let W be a non-empty open set in X. Since U_0 is dense in X, the set U_0 intersects W and $U_0 \cap W$ is open as a finite intersection of open sets; therefore there exist a point x_0 and a radius $0 < \varepsilon_0 < 1$ such that $\overline{B}(x_0, \varepsilon_0) \subseteq U_0 \cap W$. Since U_1 is dense in X, the set U_1 intersects $B(x_0, \varepsilon_0)$ and $B(x_0, \varepsilon_0) \cap U_1$ is open; therefore there exist a point x_1 and a radius $0 < \varepsilon_0 < \frac{1}{2}$ such that $\overline{B}(x_1, \varepsilon_1) \subseteq B(x_0, \varepsilon_0) \cap U_1$. One can repeat this construction recursively so that, for any integer $n \in \mathbf{N}$, there are a point x_n and a radius $0 < \varepsilon_n < \frac{1}{n+1}$ such that $\overline{B}(x_n, \varepsilon_n) \subseteq B(x_{n-1}, \varepsilon_{n-1}) \cap U_n$. The sequence $(x_n)_{n\in\mathbf{N}}$ is a Cauchy sequence in X which is complete, hence it converges to some $y \in X$. For any $n \in \mathbf{N}$, the sequence $(x_k)_{k\geqslant n}$ is in $\overline{B}(x_n, \varepsilon_n)$ which is closed, and it also is a subsequence of $(x_n)_{n\in\mathbf{N}}$, therefore $y \in \overline{B}(x_n, \varepsilon_n) \subseteq W \cap U_n$. □

[5] René-Louis Baire (1874–1932) was a French mathematician who made important contributions to analysis.

Chapter 4

Compactness

4.1 Compact spaces

Let X be a topological space and A a subset of X.

Definition 4.1. A family $\{U_i\}_{i\in I}$ of subsets of X is said to be a *cover of* X if its union is the whole space X, that is $X = \bigcup_{i\in I} U_i$. If moreover every U_i is open in X, then we say that $\{U_i\}_{i\in I}$ is an *open cover of* X. Finally, a *sub-cover of* $\{U_i\}_{i\in I}$ is any sub-collection $\{U_j\}_{j\in J}$, with $J \subseteq I$, that still is a cover of X.

Similarly, if A is a subset of X and $\{U_i\}_{i\in I}$ a family of subsets of X, we say that $\{U_i\}_{i\in I}$ is a *cover of* A if its union contains A, that is $A \subseteq \bigcup_{i\in I} U_i$.

Definition 4.2. A collection $\{F_i\}_{i\in I}$ of subsets of X is said to have the *finite intersection property* if the intersection over any finite sub-collection of $\{F_i\}_{i\in I}$ is non-empty.

More precisely, $\{F_i\}_{i\in I}$ has the finite intersection property if for every finite subset $J \subseteq I$ the intersection $\bigcap_{j\in J} F_j$ is non-empty.

Definition 4.3. A topological space X is *quasi-compact* if every open cover of X has a finite sub-cover. A topological space X is *compact* if it is Hausdorff and quasi-compact.

By taking complement sets, a topological space X is quasi-compact if and only if every family of closed sets in X with the finite intersection property has non-empty intersection.

A subset $A \subseteq X$ is said to be *compact* if the topological subspace A is compact (endowed with the subspace topology). This is equivalent to the fact that the topological subspace A is Hausdorff and every cover of A by open sets in X has a finite sub-cover. By passing to the complement sets, it is still equivalent to the fact that the topological subspace A is Hausdorff and that for every family of closed sets in X such that the intersection of any finite sub-family intersects A, the intersection of the whole family intersects A.

Let us now prove some basic properties of compact spaces.

Proposition 4.4. *Let X be a Hausdorff topological space.*

(i) *If A is a compact subset of X, then A is closed in X.*

(ii) *A finite union of compact subsets of X is compact.*

Proof. (i) We shall prove that $X \setminus A$ is open, so let $x \in X \setminus A$. Since X is Hausdorff, for any $y \in A$ there exist an open neighborhood U_y of x in X and an open neighborhood V_y of y in X such that $U_y \cap V_y = \varnothing$. Hence $A \subseteq \bigcup_{y\in A} V_y$ and there exist $\{y_1, \dots, y_N\}$ such that $A \subseteq \bigcup_{1\leqslant i\leqslant N} V_{y_i}$. Then the set $U = \bigcap_{1\leqslant i\leqslant N} U_{y_i}$ is open in X and $U \cap A = \varnothing$, hence $x \in U \subseteq X \setminus A$.

(ii) Let $A_1, \dots, A_n$ be compact in X. The union $A := \bigcup_{1\leqslant k\leqslant n} A_k$ is Hausdorff, since it is a subset of a Hausdorff space. Let $\{U_i\}_{i\in I}$ be a cover of A by open sets in X. For any $k \in \{1, \dots, n\}$ we have $A_k \subseteq \bigcup_{i\in I} U_i$, hence there exists a finite $J_k \subseteq I$ such that $A_k \subseteq \bigcup_{i\in J_k} U_i$. Then $J = \bigcup_{1\leqslant k\leqslant n} J_k$ is a finite set contained in I and $A \subseteq \bigcup_{i\in J} U_i$, hence A is compact. □

Proposition 4.5. *A closed subset of a compact topological space is compact.*

Proof. Let A be a closed set of a compact topological space X. Recall that we already know that A is Hausdorff. Let $\{U_i\}_{i\in I}$ be a cover of A by open sets in X. Let $k \notin I$ and define $U_k = X \setminus A$ which is open since A is closed. Therefore $\{U_i\}_{i\in I\cup\{k\}}$ is a open cover of X, and since X is compact, there exists a finite set $J \subseteq I \cup \{k\}$ such that $\bigcup_{i\in J} U_i$ is a finite cover of X. This implies that $\bigcup_{i\in J\setminus\{k\}} U_i$ is a finite cover of A. □

Proposition 4.6 (Cantor's[1] intersection theorem)**.** *Let X be a Hausdorff topological space. The intersection of a decreasing sequence of non-empty compact sets in X is non-empty.*

Proof. Let $\{K_n\}_{n\in\mathbf{N}}$ be a decreasing sequence of non-empty compact sets in X. By way of contradiction, suppose that $\bigcap_{n\in\mathbf{N}} K_n = \varnothing$. Consider, for any $n \in \mathbf{N}$, the open set $U_n = X \setminus K_n$ in X. The collection $\{U_n\}_{n\in\mathbf{N}}$ is a open cover of K_0, which is compact, therefore there is a finite set $J \subseteq \mathbf{N}$ such that $\{U_n\}_{n\in J}$ is a open cover of K_0. Denoting $p = \max J$, it follows that $K_0 \subseteq U_p$ which in turn implies $K_p \subseteq U_p$, hence a contradiction. □

Theorem 4.7. *Let $f : X \to Y$ be a continuous maps from a compact topological space X into a Hausdorff topological space Y. Then $f(X)$ is compact in Y.*

Proof. One has that $f(X)$ is Hausdorff since it is a subspace of Y which is Hausdorff. Let $\{U_i\}_{i\in I}$ be an open cover of $f(X)$ by open sets in Y. Therefore, since $X = f^{-1}(f(X))$ and f is continuous, one obtains that $\{f^{-1}(U_i)\}_{i\in I}$ is an open cover of X. Since X is compact, there is a finite $J \subseteq I$ such that $\{f^{-1}(U_i)\}_{i\in J}$ is a finite cover of X, which implies that $\{U_i\}_{i\in J}$ is a finite sub-cover of $f(X)$. □

Proposition 4.8. *In a compact topological space, every sequence has at least one cluster point.*

[1] Georg Cantor (1845–1918) was a German mathematician who created set theory. In particular he defined infinite sets and well-ordered sets.

Proof. Let $(x_n)_{n\in\mathbf{N}}$ be a sequence in a compact topological space X. Recall that the set of cluster points of $(x_n)_{n\in\mathbf{N}}$ is given by $F = \bigcap_{n\in\mathbf{N}} \overline{\{x_p : p \geqslant n\}}$. Denote $F_n := \overline{\{x_p : p \geqslant n\}}$ and consider the family of closed sets $\{F_n\}_{n\in\mathbf{N}}$. The intersection of any finite family of $\{F_n\}_{n\in\mathbf{N}}$ is non-empty, indeed if $J \subseteq \mathbf{N}$ is finite then $\bigcap_{n\in J} F_n$ contains $x_{\max J}$. Therefore, since X is compact, we deduce that $F = \bigcap_{n\in\mathbf{N}} F_n$ is non-empty. □

4.2 Compact metric spaces

A topological space is said to be *sequentially compact* if every sequence has a convergent subsequence. A metric space X is said to be *totally bounded* if for any positive $\varepsilon > 0$ there exists a finite family of open balls of radius ε that covers X.

Theorem 4.9. *In a metric space (X, d) the following are equivalent:*

(i) *X is compact;*

(ii) *X sequentially compact;*

(iii) *X is complete and totally bounded.*

Proof. (i) $\Rightarrow$ (ii) It follows from Proposition 4.8 and Proposition 2.20.

(ii) $\Rightarrow$ (iii) Assume that X is sequentially compact. We want to show that X is complete and totally bounded. Let $(x_n)_{n\in\mathbf{N}}$ be a Cauchy sequence in X, then it admits a subsequence that converges to some $x \in X$, and one deduces that x is a cluster point of $(x_n)_{n\in\mathbf{N}}$ from Proposition 2.20. Therefore, by Proposition 3.3, one concludes that $(x_n)_{n\in\mathbf{N}}$ converges to x.

Let $\varepsilon > 0$. We argue by contradiction and assume that X does not admit any finite cover by a finite family of open balls of radius ε. We fix $x_0 \in X$, then by hypothesis we can chose some $x_1 \in X \setminus B(x_0, \varepsilon)$. By induction, we can then construct a sequence $(x_n)_{n\in\mathbf{N}}$ of elements of X such that $x_n \notin \bigcup_{i=1}^{n-1} B(x_i, \varepsilon)$ for all $n \geqslant 1$. This sequence satisfies $d(x_i, x_j) \geqslant \varepsilon$ for any integers $i \neq j$ and hence does not admit any convergent subsequence, which is a contradiction.

(ii) $\Rightarrow$ (i) Assume that X is sequentially compact. We want to show that X is compact. We already know that X is Hausdorff, so we need to prove that X is quasi-compact.

Let $\{U_i\}_{i\in I}$ be an open cover of X. We now show that there is $\delta > 0$ such that, for any $x \in X$, there is $i \in I$ such that $B(x, \delta) \subseteq U_i$. We argue by contradiction and assume that it is not the case, hence for any $n \in \mathbf{N}^*$ there is $x_n \in X$ such that $B(x_n, \frac{1}{n}) \nsubseteq U_i$ for any $i \in I$. The sequence $(x_n)_{n\geqslant 1}$ has a subsequence that converges to some $x \in X$, which implies that x is a cluster point of $(x_n)_{n\in\mathbf{N}^*}$. Let $i_0 \in I$ such that $x \in U_{i_0}$, thus for n large enough one obtains $B(x_n, \frac{1}{n}) \subseteq U_{i_0}$, which is a contradiction.

By repeating the above argument used to show that (ii) implies (iii), we show that X can be recovered by a finite number of ball of radius δ.

(iii) $\Rightarrow$ (ii) Assume now that X is complete and totally bounded. We wan to prove that X is sequentially compact. Let $\mathbf{x} = (x_n)_{n\in\mathbf{N}}$ be a sequence in X.

For $j = 1$ there exists a finite family of open ball of radius 1 covering X, hence there is a subsequence $\mathbf{x}^1 = (x_{\varphi_1(n)})_{n\in\mathbf{N}}$ of $\mathbf{x}$ that is included in some open ball of radius 1. By induction, for any integer $j \geqslant 2$, one can then find a subsequence $\mathbf{x}^j = (x_{\varphi_1\circ\cdots\circ\varphi_j(n)})_{n\in\mathbf{N}}$ of $\mathbf{x}^{j-1}$ (hence also a subsequence of $\mathbf{x}$) that is included in some open ball of radius $\frac{1}{j}$. By a diagonal extraction argument, one defines the sequence $(y_n)_{n\in\mathbf{N}^*}$ with $y_n = x_{\varphi_1\circ\cdots\circ\varphi_n(n)}$, which is a subsequence of $(x_n)_{n\in\mathbf{N}}$ and satisfies $d(y_n, y_m) \leqslant \frac{2}{n}$ for any positive integers $n \leqslant m$. Therefore $(y_n)_{n\in\mathbf{N}^*}$ is a Cauchy sequence, thus it converges since X is complete. □

4.3 Product of compact spaces

Lemma 4.10 (Alexander's sub-basis theorem)**.** *Let X be a topological space and $\mathscr{S}$ be a sub-basis for the topology of X. If every open cover of X by elements of $\mathscr{S}$ has a finite sub-cover, then X is quasi-compact.*

Proof. Assume that every open cover of X by elements of $\mathscr{S}$ has a finite sub-cover. By way of contradiction, suppose that there is a open cover of X with no finite sub-cover.

Consider the collection $\mathfrak{P}$ of open covers of X with no finite sub-cover, which is non-empty, partially ordered by inclusion. Let $\mathfrak{U}$ be a chain (totally ordered subset) in $\mathfrak{P}$. We now define $\mathscr{U} = \bigcup \mathfrak{U}$ which is an upper bound of $\mathfrak{U}$, and we shall verify that $\mathscr{U}$ has no finite sub-cover. Consider any finite sub-collection $U_1, \dots, U_n$ of $\mathscr{U}$, then for any $1 \leqslant i \leqslant n$ there is $\mathscr{U}_i \in \mathfrak{U}$ such that $U_i \in \mathscr{U}_i$. Since $\mathfrak{U}$ is totally ordered, there is some $k \in \{1, \dots, n\}$ such that $\mathscr{U}_k$ contains all the sets of $\{U_1, \dots, U_n\}$, therefore this finite sub-collection cannot cover X. By Zorn's Lemma (Theorem A.23), one deduces that $\mathfrak{P}$ has a maximal element $\mathscr{M}$. Since $\mathscr{M}$ is a maximal element, for any open $U \notin \mathscr{M}$ the cover $\mathscr{M} \cup \{U\}$ has a finite sub-cover, necessarily of the form $\mathscr{M}' \cup \{U\}$ for some finite subset $\mathscr{M}' \subseteq \mathscr{M}$.

We claim that if U, V are open sets in X such that $U, V \notin \mathscr{M}$, then $U \cap V \notin \mathscr{M}$. Indeed, there is finite subsets $\mathscr{M}' \subseteq \mathscr{M}$ and $\mathscr{M}'' \subseteq \mathscr{M}$ such that $\mathscr{M}' \cup \{U\}$ and $\mathscr{M}'' \cup \{V\}$ both cover X. Therefore $\{U \cap V\} \cup \mathscr{M}' \cup \mathscr{M}''$ is a finite cover of X, hence $U \cap V \notin \mathscr{M}$ since $\mathscr{M}$ has no finite sub-cover.

We also claim that if U, V are open sets in X such that $U \notin \mathscr{M}$ and $U \subseteq V$, then $V \notin \mathscr{M}$. Indeed, if $\mathscr{M}' \cup \{U\}$ is a finite cover, with $\mathscr{M}' \subseteq \mathscr{M}$ finite, then also is $\mathscr{M}' \cup \{V\}$, thus $V \notin \mathscr{M}$.

We now show that $\mathscr{M} \cap \mathscr{S}$ covers X. Let $x \in X$, since $\mathscr{M}$ is a cover, there is some open set $U \in \mathscr{M}$ such that $x \in U$. Since $\mathscr{S}$ is a sub-basis, there are $S_1, \dots, S_n \in \mathscr{S}$ such that $x \in S_1 \cap \cdots \cap S_n \subseteq U$. Thanks to the two claims above, there is some i such that $S_i \in \mathscr{M}$, whence $x \in S_i \subseteq \mathscr{M} \cap \mathscr{S}$.

Since $\mathscr{M} \cap \mathscr{S}$ is a cover of X, it follows that $\mathscr{M} \cap \mathscr{S}$ has a finite sub-cover, which contradicts the fact that $\mathscr{M}$ has no finite sub-cover. Therefore the original collection $\mathfrak{P}$ is empty and thus every open cover of X has a finite sub-cover. □

Theorem 4.11 (Tikhonov's[2] theorem)**.** *Any product of compact topological spaces is*

[2] Andrey Tikhonov (1906–1993) was Soviet mathematician who made important contributions to topology, functional analysis, and mathematical physics.

compact.

Proof. Let $\{X_i\}_{i\in I}$ be a family of compact topological spaces and consider the product space $X = \prod_{i\in I} X_i$, equipped with the product topology, and denote by $\mathrm{pr}_j : X \to X_j$, $(x_i)_{i\in I} \mapsto x_j$, the canonical projections. Since every X_i is Hausdorff, we already know that X is also Hausdorff (Proposition 2.13). Recall that the collection $\mathcal{S}$ of subsets of the form $\mathrm{pr}_i^{-1}(U_i)$, where $i \in I$ and U_i is open in X_i, is a sub-basis for the product topology on X.

We argue by contradiction and assume that X is not compact. Thanks to Lemma 4.10, there is a cover $\mathcal{U}$ of X by elements of $\mathcal{S}$ with no finite sub-cover. For each $i \in I$, define the collection $\mathcal{S}_i$ of open sets $U \subseteq X_i$ such that $\mathrm{pr}_i^{-1}(U) \in \mathcal{S}$.

We now show that $\mathcal{S}_i$ does not cover X_i. Indeed, since X_i is compact, if $\mathcal{S}_i$ covers X_i then there is a finite collection $U_1, \dots, U_n \in \mathcal{S}_i$ that covers X_i. If follows that

$$\mathrm{pr}_i^{-1}(U_1) \cup \cdots \cup \mathrm{pr}_i^{-1}(U_n) = \mathrm{pr}_i^{-1}(X_i) = X,$$

which contradicts the fact that $\mathcal{S}$ has no finite sub-cover.

Then let $x_i \in X_i$ such that $x_i \notin \bigcup \mathcal{S}_i$. Define $x = (x_i)_{i\in I}$, then, since $\mathcal{S}$ is a cover, there is $i \in I$ and an open set U in X_i such that $x \in \mathrm{pr}_i^{-1}(U)$, but this contradicts the fact that $x_i \notin \bigcup \mathcal{S}_i$. □

4.4 Local compactness

Definition 4.12. A topological space X is *locally compact* if it is Hausdorff and if every point admits a compact neighbourhood.

Proposition 4.13. *In a locally compact space X, every point has a neighbourhood basis composed by compact sets.*

Proof. Let $x \in X$ and V be a neighborhood of x. We want to find a compact neighborhood of x contained in V.

By definition, there exists a compact $K \subseteq X$ that is a neighborhood of x in X. Since $V \cap K$ is a neighborhood of x, there is an open set U such that $x \in U \subseteq V \cap K \subseteq K$. Remark that the set $K \setminus U$ is compact, since it is closed and contained in the compact K. Since X is Hausdorff, there are disjoint open sets W_1 and W_2 such that $x \in W_1$ and $K \setminus U \subseteq W_2$. The set $F := \overline{U \cap W_1}$ contains x, is compact by construction (since it is a closed set included in the compact K) and is contained in U, hence in $V \cap K \subseteq V$. □

Example 4.14. (1) $\mathbf{R}$ is locally compact.

(2) A discrete space is locally compact.

(3) A finite product of locally compact spaces is locally compact.

Theorem 4.15 (Baire's category theorem, second version)**.** *Every locally compact topological space is a Baire space.*

Proof. Let X be a locally compact topological space and $\{U_n\}_{n\in\mathbf{N}}$ be a countable family of dense open sets in X. Denote $U := \bigcap_{n\in\mathbf{N}} U_n$ the intersection.

Let W be a non-empty open set in X. Since U_0 is dense in X, W intersects U_0, which implies that there is $x_0 \in U_0 \cap W$. Since X is locally compact, there exists a compact set K_0 that is a neighborhood of x_0, hence with non-empty interior, such that $x_0 \in K_0 \subseteq W \cap U_0$.

Since U_1 is dense in X, one has $U_1 \cap \mathring{K}_0 \neq \varnothing$, which implies that there is $x_1 \in U_1 \cap \mathring{K}_0$. As before, there exists a compact set K_1 with non-empty interior such that $K_1 \subseteq U_1 \cap \mathring{K}_0$. Repeating this construction recursively, for any integer $n \in \mathbf{N}$ there is a compact set K_n with non-empty interior such that $K_n \subseteq U_n \cap \mathring{K}_{n-1}$. Therefore $\bigcap_{n\in\mathbf{N}} K_n \subseteq W \cap \bigcap_{n\in\mathbf{N}} U_n$, and the intersection $\bigcap_{n\in\mathbf{N}} K_n$ is non-empty by Proposition 4.6. □

Chapter 5

Connectedness

5.1 Connected spaces

Definition 5.1. A topological space X is said to be *connected* if there is no subset of X, other than the empty set $\varnothing$ and X itself, that is both open and closed.

We say that a subset A of a topological space X is connected if the topological subspace A, equipped with the subspace topology, is connected.

Definition 5.2. A topological space X is *locally connected* if every point has a neighborhood basis formed by connected sets.

Theorem 5.3. *Let X be a topological space. The following are equivalent:*

(i) *X is connected;*

(ii) *X cannot be written as the union of two non-empty disjoint open sets;*

(iii) *X cannot be written as the union of two non-empty disjoint closed sets;*

(iv) *Any continuous map from X into $\{0, 1\}$ equipped with the discrete topology is constant;*

(v) *Any continuous map from X into a discrete topological space is constant.*

Proof. Assume that Y is discrete topological space and that there is a continuous function $f : X \to Y$ that is not constant. Since f is not constant, there is $y \in Y$ such that $f^{-1}(\{y\})$ is a non-empty proper subset of X. But $f^{-1}(\{y\})$ is both open and closed in X because f is continuous, hence X is not connected. This proves that (i) implies (v).

The implication (v) $\Rightarrow$ (iv) is trivial.

Now assume that there are two non-empty disjoint closed sets A, B such that $X = A \cup B$. Define the map $f : X \to \{0, 1\}$ by $f(x) = 0$ if $x \in A$, and $f(x) = 1$ if $x \in B$. Then f is continuous and non-constant. This proves that (iv) implies (iii).

Assume there are two non-empty disjoint open sets A, B such that $X = A \cup B$. Hence $A = X \setminus B$ is the complement of B, which implies that X also is the union of two non-empty disjoint closed sets. This proves that (iii) implies (ii).

Assume that X is not connected, then there is a non-empty proper subset $A \subseteq X$ that is both open and closed. Therefore we can write $X = A \cup (X \setminus A)$ as the union of two non-empty disjoint open sets. This proves that (ii) implies (i). □

Let us now prove some basic properties of connected spaces.

Proposition 5.4. *Consider subsets $A \subseteq B \subseteq \overline{A}$ of a topological space X and assume that A is connected. Then B is also connected.*

Proof. By way of contradiction assume that B is not connected. Then there are two non-empty disjoint open sets U and V in B such that $B = U \cup V$. Since $B \subseteq \overline{A}$, we deduce that $U \cap A$ and $V \cap A$ are also non-empty. This implies that $A = (U \cap A) \cup (V \cap A)$ is the union of two non-empty disjoint open sets in A, which is a contradiction since A is connected. □

Proposition 5.5. *Let $\{A_i\}_{i \in I}$ be a collection of connected subsets of a topological space X such that $A_i \cap A_j \neq \emptyset$ for any $i, j \in I$. Then $A = \bigcup_{i \in I} A_i$ is connected.*

Proof. Consider a continuous map f from A to the discrete space $\{0, 1\}$. For any $i \in I$, the restriction $f_{|A_i}$ of f to A_i is continuous hence constant. For any $i, j \in I$, since $A_i \cap A_j \neq \emptyset$, it follows that f is constant on $A_i \cup A_j$. Therefore f is constant on A and hence A is connected. □

From the above result, given a topological space X and a point $x \in X$, we define the set C_x, called the *connected component* of x, as the union of all connected subsets containing x, which is also the largest connected subset containing x. Remark that, given $y \in X$, $y \in \mathrm{C}_x$ if and only if $\mathrm{C}_y = \mathrm{C}_x$, hence the connected components form a partition of X.

A topological space is said to be *totally disconnected* if all connected components are singletons.

Example 5.6. (1) Consider the set of real numbers $\mathbf{R}$ endowed with its standard topology (Example 1.5–(3)). Any interval of $\mathbf{R}$ is connected; the set of real numbers deprived of the origin $\mathbf{R} \setminus \{0\}$ is not connected; and the set of rational numbers $\mathbf{Q}$ is not connected.

(2) If $f : X \to Y$ is continuous and X is connected, then $f(X)$ is connected.

(3) If X and Y are connected, then $X \times Y$ is connected.

5.2 Path-connected spaces

Definition 5.7. A topological space X is *path-connected* if for any points $x, y \in X$ there is a path joining x to y, that is, a continuous map $\gamma : [0, 1] \to X$ such that $\gamma(0) = x$ and $\gamma(1) = y$.

A subset $A \subseteq X$ is said to be path-connected if every points $x, y \in A$ can de joined by a path taking values in A, that is, a continuous map $\gamma : [0, 1] \to A$ with $\gamma(0) = x$ and $\gamma(1) = y$.

Definition 5.8. A topological space X is *locally path-connected* if every point has a neighborhood basis formed by path-connected sets.

As above, given a topological space X and a point $x \in X$, we define the *path-connected component* of x as the largest path-connected subset containing x.

Theorem 5.9. *Every path-connected space is connected.*

Proof. Let X be a path-connected topological space. By way of contradiction, assume that X is not connected. Then there are non-empty open disjoint sets $A, B \subseteq X$ such that $X = A \cup B$. Let $a \in A$ and $b \in B$. There exists a path γ joining a and b. Therefore $[0, 1] = \gamma^{-1}(A) \cup \gamma^{-1}(B)$ is the union of non-empty disjoint open sets, which is a contradiction since $[0, 1]$ is connected (see Example **??**). □

Proposition 5.10. *If X is connected and locally path-connected, then X is path-connected.*

Proof. Let $a \in X$ and denote by A the set of all points that can be joined to a by a path. We will show that A is both open and closed, which concludes the proof since X is connected.

For any $x \in X$ there is a open neighborhood B_x of x such that every point $y \in B_x$ can be joined to x by a path. Therefore, for any $x \in A$ and $y \in B_x$ there is a path joining a to y, this implies $y \in A$ and thus A is open.

Let $z \in \overline{A}$. Let $y \in A \cap B_z$, then there is a path joining y to z and a path joining a to y, hence the existence of a path joining a to z. Therefore $z \in A$ and hence A is closed. □

Chapter 6

Function spaces

If X and Y are sets, we denote by $\mathscr{F}(X,Y) = Y^X$ the set of maps from X to Y. If X and Y are topological space, we then denote by $\mathscr{C}(X,Y)$ the set of continuous maps from X to Y.

6.1 Pointwise and uniform convergences

Let X be a set and Y a topological space. Consider a sequence $(f_n)_{n\in\mathbf{N}}$ of maps from X into Y and let f be a map from X to Y.

One says that the the sequence $(f_n)_{n\in\mathbf{N}}$ *converges pointwisely to f on X* if for any $x \in X$ the sequence $(f_n(x))_{n\in\mathbf{N}}$ converges to $f(x)$ in Y. The map f is called the *pointwise limit* of $(f_n)_{n\in\mathbf{N}}$. In particular, if (Y,d) is a metric space, then $(f_n)_{n\in\mathbf{N}}$ converges pointwisely to f on X if

$$\forall x \in X,\ \forall \varepsilon > 0,\ \exists N \in \mathbf{N},\ \forall n \geqslant N,\ d(f_n(x), f(x)) < \varepsilon.$$

When (Y,d) is a metric space, then one says that $(f_n)_{n\in\mathbf{N}}$ *converges uniformly to f on X* if

$$\forall \varepsilon > 0,\ \exists N \in \mathbf{N},\ \forall n \geqslant N,\ \forall x \in X,\ d(f_n(x), f(x)) < \varepsilon.$$

The map f is called the *uniform limit* of $(f_n)_{n\in\mathbf{N}}$.

It is clear that the uniform convergence implies pointwise convergence.

6.1.1 Topology of pointwise convergence

Consider a set X and a topological space Y. The topology associated to the pointwise convergence described is called the *topology of pointwise convergence* and it is nothing but the product topology on the set Y^X. Indeed let us verify that $(f_n)_{n\in\mathbf{N}}$ converges to f in Y^X endowed with the product topology if and only if $(f_n)_{n\in\mathbf{N}}$ converges pointwisely to f on X.

Assume that $(f_n)_{n\in\mathbf{N}}$ converges to f for the product topology. Let $x \in X$. Let V be a neighbourhood of $f(x)$ in Y, then there is an open set U in Y such that $f(x) \in U \subseteq V$. Consider the set $\mathrm{pr}_x^{-1}(U) = \{g \in Y^X \mid g(x) \in U\}$, which is open in Y^X and contains f, then $\mathrm{pr}_x^{-1}(U)$ is a neighbourhood of f. Therefore, there is

$n_0 \in \mathbf{N}$ such that, for any integer $n \geqslant n_0$ one has $f_n \in \mathrm{pr}_x^{-1}(U)$. Hence for any $n \geqslant n_0$ one has $f_n(x) \in U \subseteq V$.

Conversely, now assume that $(f_n)_{n\in\mathbf{N}}$ converges pointwisely to f. Let V be a neighbourhood of f in Y^X. Then there is a finite number of points $x_1, \ldots, x_k \in X$ and a finite number of open sets $U_1, \ldots, U_k$ in Y such that $f \in \mathrm{pr}_{x_1}^{-1}(U_1) \cap \cdots \cap \mathrm{pr}_{x_k}^{-1}(U_k) \subseteq V$. Therefore, for any $i = 1, \ldots, k$, one has $f(x_i) \in U_i$ and hence there is $n_i \in N$ such that $f_n(x_i) \in U_i$ for any integer $n \geqslant n_i$. Hence, taking $N = \max_{1\leqslant i\leqslant k} n_i$, for any integer $n \geqslant N$ one has $f_n \in \mathrm{pr}_{x_1}^{-1}(U_1) \cap \cdots \cap \mathrm{pr}_{x_k}^{-1}(U_k) \subseteq V$.

6.1.2 Topology of uniform convergence

Consider a set X and a metric space (Y, d). For any $f, g \in \mathscr{F}(X, Y)$ define the map

$$d_\infty(f, g) := \min\left(\sup_{x\in X} d(f(x), g(x)), 1\right).$$

One can easily check that d_∞ is a metric on $\mathscr{F}(X, Y)$, and that a sequence of maps $(f_n)_{n\in\mathbf{N}}$ converges uniformly to f on X if and only if $(f_n)_{n\in\mathbf{N}}$ converges to f with respect the metric d_∞. The metric d_∞ is called the *metric of uniform convergence* or the *uniform metric*, and its associated topology is called the *topology of uniform convergence*.

Proposition 6.1. *Let X be a set and (Y, d) a complete metric space. Then $(\mathscr{F}(X, Y), d_\infty)$ is complete.*

Proof. Let $(f_n)_{n\in\mathbf{N}}$ be a Cauchy sequence in $(\mathscr{F}(X, Y), d_\infty)$. Then for any $x \in X$, $(f_n(x))_{n\in\mathbf{N}}$ is a Cauchy sequence in (Y, d), since $d(f_n(x), f(x)) \leqslant \sup_{y\in X}(f_n(y), f(y))$ if $d_\infty(f_n, f) < 1$. Therefore, for any $x \in X$, $(f_n(x))_{n\in\mathbf{N}}$ converges to some point $f(x)$ in Y. We have hence constructed a map $f : X \to Y$, $x \mapsto f(x)$, and it remains to show that $(f_n)_{n\in\mathbf{N}}$ converges to f with respect to the metric d_∞.

Let $\varepsilon > 0$, then there is $n_0 \in \mathbf{N}$ such that one has, for any for any integers $n, m \geqslant n_0$ and any $x \in X$, $d(f_n(x), f_m(x)) < \varepsilon$. Hence, with $x \in X$ fixed and letting $m \to +\infty$, we obtain $d(f_n(x), f(x)) \leqslant \varepsilon$ for all $n \geqslant n_0$. Thus $(f_n)_{n\in\mathbf{N}}$ converges uniformly to f on X. □

6.2 Space of continuous functions on a compact set

Let X be a compact space and (Y, d) a metric space, and consider the space $\mathscr{C}(X, Y)$ of continuous maps from X to Y. We define, for any $f, g \in \mathscr{C}(X, Y)$, the map

$$d_\infty(f, g) = \sup_{x\in X} d(f(x), g(x)).$$

As before, one can easily check that d_∞ is a metric on $\mathscr{C}(X, Y)$. Furthermore, a sequence of maps $(f_n)_{n\in\mathbf{N}}$ in $\mathscr{C}(X, Y)$ converges uniformly on X to a function $f : X \to Y$ if and only if $(f_n)_{n\in\mathbf{N}}$ converges to f with respect the metric d_∞. The metric d_∞ is called the *metric of uniform convergence* or the *uniform metric*, and its associated topology is called the *topology of uniform convergence*.

Theorem 6.2. *Let X be compact and (Y, d) a complete metric space. Then the space $(\mathscr{C}(X, Y), d_\infty)$ is complete.*

Proof. Let $(f_n)_{n\in\mathbf{N}}$ be a Cauchy sequence in $(\mathscr{C}(X, Y), d_\infty)$. Then for any $x \in X$, $(f_n(x))_{n\in\mathbf{N}}$ is a Cauchy sequence in (Y, d), since $d(f_n(x), f_m(x)) \leqslant d_\infty(f_n, f_m)$. Therefore, for any $x \in X$, $(f_n(x))_{n\in\mathbf{N}}$ converges to some $f(x)$ in Y. We have hence constructed a map $f : X \to Y$, $x \mapsto f(x)$.

It remains to show that $(f_n)_{n\in\mathbf{N}}$ converges to f with respect to the metric d_∞, and that $f \in \mathscr{C}(X, Y)$. Let $\varepsilon > 0$, then there is $n_0 \in \mathbf{N}$ such that one has, for any integers $n, m \geqslant n_0$ and any $x \in X$, $d(f_n(x), f_m(x)) < \varepsilon$. Hence, letting $m \to +\infty$, we obtain $d(f_n(x), f(x)) \leqslant \varepsilon$ for all $n \geqslant n_0$ and $x \in X$. Thus $(f_n)_{n\in\mathbf{N}}$ converges uniformly to f on X.

We now prove that f is continuous on X. Let $a \in X$ and $\varepsilon > 0$. There is $n_0 \in \mathbf{N}$ such that $\sup_{x\in X} d(f_{n_0}(x), f(x)) < \varepsilon/3$. Since f_{n_0} is continuous at a, there is a neighbourhood V of a in X such that for any $x \in V$ one has $d(f_{n_0}(x), f_{n_0}(a)) < \varepsilon/3$. Therefore for any $x \in V$ one has

$$d(f(x), f(a)) \leqslant d(f(x), f_{n_0}(x)) + d(f_{n_0}(x), f_{n_0}(a)) + d(f_{n_0}(a), f(a)) < \varepsilon,$$

thus f is continuous at a. We conclude that f is continuous on X since a is arbitrary. □

6.3 Equicontinuity

Let X be a topological space and (Y, d) a metric space.

Definition 6.3. Let $\mathscr{H}$ be a subset of $\mathscr{F}(X, Y)$ and $x \in X$ a point. One says that $\mathscr{H}$ is *equicontinuous at* x if for any $\varepsilon > 0$, there is a neighborhood U of x in X such that for all $y \in U$ and all $f \in \mathscr{H}$ one has

$$d(f(y), f(x)) < \varepsilon.$$

One says that $\mathscr{H}$ is *equicontinuous* on X if $\mathscr{H}$ is equicontinuous at every point of X.

Example 6.4. (1) A finite family $\mathscr{H} = \{f_1, \ldots, f_p\}$ of functions that are continuous at a point $x \in X$ is equicontinuous at x.

(2) If X is a metric space and $k > 0$, the set of all k-Lipschitz functions from X to Y is equicontinuous on X.

Proposition 6.5. *Let $(f_n)_{n\in\mathbf{N}}$ be a sequence of maps from X to Y. Suppose that $\{f_n : n \in \mathbf{N}\}$ is equicontinuous at $a \in X$ and that $(f_n)_{n\in\mathbf{N}}$ converges pointwisely on X to some map $f : X \to Y$. Then f is continuous at a.*

Proof. Let $\varepsilon > 0$. Since $\{f_n : n \in \mathbf{N}\}$ is equicontinuous at a, there is a neighbourhood U of a such that for all $x \in U$ and any $n \in \mathbf{N}$ one has $d(f_n(x), f_n(a)) < \varepsilon/3$. Since $(f_n)_{n\in\mathbf{N}}$ converges pointwisely to f, for any $x \in U$ there is $N_0 \in \mathbf{N}$ such that for any

$n \geqslant N_0$ one has $d(f_n(x), f(x)) < \varepsilon/3$. Hence, for any $x \in U$ and any $n \geqslant N_0$, one obtains

$$d(f(x), f(a)) \leqslant d(f(x), f_n(x)) + d(f_n(x), f_n(a)) + d(f_n(a), f(a)) < \varepsilon,$$

thus f is continuous at a. □

Proposition 6.6. *Let $(f_n)_{n\in\mathbf{N}}$ be a sequence of maps from X to Y and suppose that A is dense in X and that Y is complete. Assume that $\{f_n : n \in \mathbf{N}\}$ is equicontinuous on X and that $(f_n)_{n\in\mathbf{N}}$ converges pointwisely on A. Then $(f_n)_{n\in\mathbf{N}}$ converges pointwisely on X.*

Proof. We show that, for any $x \in X$, the sequence $(f_n(x))_{n\in\mathbf{N}}$ is a Cauchy sequence in Y. Let $x \in X$ and $\varepsilon > 0$. Since $\{f_n : n \in \mathbf{N}\}$ is equicontinuous at x, there is a open neighbourhood U of x such that for all $y \in U$ and any $n \in \mathbf{N}$ one has $d(f_n(x), f_n(y)) < \varepsilon/3$. By density of A, there exists a point $z \in A \cap U$ and in particular one has $d(f_n(x), f_n(z)) < \varepsilon/3$. Thanks to the pointwise convergence, the sequence $(f_n(z))_{n\in\mathbf{N}}$ is a Cauchy sequence in Y, therefore there is $N_0 \in \mathbf{N}$ such that for any $n, m \geqslant N_0$ one has $d(f_n(z), f_m(z)) < \varepsilon/3$. One finally gets, for any $n, m \geqslant N_0$, that

$$d(f_n(x), f_m(x)) \leqslant d(f_n(x), f_n(z)) + d(f_n(z), f_m(z)) + d(f_m(z), f_m(x)) < \varepsilon,$$

hence $(f_n(x))_{n\in\mathbf{N}}$ is a Cauchy sequence in Y. □

Proposition 6.7. *Let $(f_n)_{n\in\mathbf{N}}$ be a sequence of maps from X to Y and suppose that X is compact. Assume that $\{f_n : n \in \mathbf{N}\}$ is equicontinuous on X and that $(f_n)_{n\in\mathbf{N}}$ converges pointwisely on X to some map $f : X \to Y$. Then $(f_n)_{n\in\mathbf{N}}$ converges uniformly on X to f.*

Proof. We already know that f is continuous on X by Proposition 6.5.

Let $\varepsilon > 0$. By equicontinuity of $\{f_n : n \in \mathbf{N}\}$, for any $x \in X$ there is a open neighbourhood U_x of x such that for all $y \in U_x$ and any $n \in \mathbf{N}$ one has $d(f_n(x), f_n(y)) < \varepsilon/3$. Moreover, since f is continuous, for any $x \in X$ there is a open neighbourhood V_x of x such that for all $y \in U_x$ one has $d(f(x), f(y)) < \varepsilon/3$.

Considering the open neighbourhood $W_x := U_x \cap V_x$ of x, we write $X = \bigcup_{x\in X} W_x$. Since X is compact, there is a finite set $\{x_1, \dots, x_p\} \subseteq X$ such that $X = W_{x_1} \cup \dots \cup W_{x_p}$. Thanks to the pointwise convergence, there is $N_0 \in \mathbf{N}$ such that for all $i = 1, \dots, p$ and any $n \geqslant N_0$ one has $d(f_n(x_i), f(x_i)) < \varepsilon/3$.

Now let $x \in X$. Then there is $i \in \{1, \dots, p\}$ such that $x \in W_{x_i}$, therefore for any $n \geqslant N_0$ one has

$$d(f_n(x), f(x)) \leqslant d(f_n(x), f_n(x_i)) + d(f_n(x_i), f(x_i)) + d(f(x_i), f(x)) < \varepsilon,$$

which concludes the proof. □

6.4 Arzelà-Ascoli theorem

Let X be a compact space and (Y, d) a complete metric space. Consider the space $\mathscr{C}(X, Y)$ of continuous maps from X to Y endowed with the uniform metric

$$d_\infty(f, g) = \sup_{x\in X} d(f(x), g(x)).$$

The following result characterizes the compact subsets of $\mathscr{C}(X,Y)$.

Theorem 6.8 (Arzelà[1]-Ascoli[2])**.** *Let $\mathscr{H}$ be a subset of $\mathscr{C}(X,Y)$. Then the closure of $\mathscr{H}$ is compact in $\mathscr{C}(X,Y)$ if and only if $\mathscr{H}$ satisfies:*

(i) *The closure of the set $\mathscr{H}(x) := \{f(x) : f \in \mathscr{H}\}$ is compact in Y for any $x \in X$;*

(ii) *$\mathscr{H}$ is equicontinuous on X.*

Remark 6.9. If $Y = \mathbf{R}$ or $\mathbf{C}$ then property (i) just means that $\mathscr{H}(x)$ is bounded in $\mathbf{R}$ or $\mathbf{C}$ for any $x \in X$.

Proof of Theorem 6.8. We start with the easier part of the result and assume that $\mathscr{H}$ has compact closure in $\mathscr{C}(X,Y)$. For any $x \in X$, the evaluation map $f \mapsto f(x)$ is continuous from $\mathscr{C}(X,Y)$ into Y since $d(f(x),g(x)) \leqslant d_\infty(f,g)$, thus the set $\mathscr{H}(x)$ is compact as the image of a compact under a continuous function. This gives (i).

Let $x \in X$ and $\varepsilon > 0$. Since $\mathscr{H}$ is compact it is totally bounded, hus there is a finite collection $\{f_1,\dots,f_p\} \subseteq \mathscr{H}$ such that, for any $f \in \mathscr{H}$ there is a label $i \in \{1,\dots,p\}$ such that $d_\infty(f,f_i) < \varepsilon/3$. Since the finite family $\{f_1,\dots,f_p\}$ is equicontinuous at x, there is U a neighbourhood of x such that for all $y \in U$ and any $i \in \{1,\dots,p\}$ one has $d(f_i(y),f_i(x)) < \varepsilon/3$. Therefore, for any $y \in U$ and any $f \in \mathscr{H}$ one obtains, with the choice of label $i \in \{1,\dots,p\}$ described above,

$$d(f(y),f(x)) \leqslant d(f(y),f_i(y)) + d(f_i(y),f_i(x)) + d(f_i(x),f(x)) < \varepsilon.$$

This proves (ii).

Conversely, assume that $\mathscr{H}$ satisfies (i) and (ii). Since $\mathscr{C}(X,Y)$ is complete, the subset $\overline{\mathscr{H}} \subseteq \mathscr{C}(X,Y)$ is complete thanks to Proposition 3.5. Therefore, thanks to Theorem 4.9, it remains to prove that $\overline{\mathscr{H}}$ is totally bounded, and for this it suffices to prove that $\mathscr{H}$ is totally bounded.

Let $\varepsilon > 0$. For any $x \in X$, there is a open neighbourhood U_x of x in X such that, for any $y \in U_x$ and any $f \in \mathscr{H}$, one has $d(f(y),f(x)) < \varepsilon/4$. Since $X = \bigcup_{x\in X} U_x$ and X is compact, there is a finite number of points $x_1,\dots,x_p \in X$ such that $X = U_{x_1} \cup \dots \cup U_{x_p}$. Since Y is Hausdorff, the set $K := \overline{\mathscr{H}(x_1)} \cup \dots \cup \overline{\mathscr{H}(x_p)}$ is compact, thus there is a finite number of points $y_1,\dots,y_q \in Y$ such that $K \subseteq B(y_1,\varepsilon/4) \cup \dots \cup B(y_q,\varepsilon/4)$.

Define the set Φ of all maps from $\{1,\dots,p\}$ to $\{1,\dots,q\}$, which is finite, and for any $\phi \in \Phi$ define

$$\mathscr{H}_\phi = \left\{ f \in \mathscr{H} \mid \forall i \in \{1,\dots,p\}, f(x_i) \in B(y_{\phi(i)},\tfrac{\varepsilon}{4}) \right\}.$$

By construction, the family $\{\mathscr{H}_\phi\}_{\phi\in\Phi}$ is a finite cover of $\mathscr{H}$. Moreover, for any $\phi \in \Phi$, if $f,g \in \mathscr{H}_\phi$ then for any $x \in X$ we consider $i \in \{1,\dots,p\}$ such that $x \in U_{x_i}$,

[1] Cesare Arzelà (1847–1912) was a Italian mathematician who made contributions to functional analysis.

[2] Giulio Ascoli (1843–1896) was a Italian mathematician who contributed to functional analysis. In particular, he introduced the notion of equicontinuity.

and thus we get

$$\begin{aligned} d(f(x),g(x)) &\leqslant d(f(x),f(x_i)) + d(f(x_i),y_{\phi(i)}) + d(y_{\phi(i)},g(x_i)) + d(g(x_i),g(x)) \\ &< \frac{\varepsilon}{4}+\frac{\varepsilon}{4}+\frac{\varepsilon}{4}+\frac{\varepsilon}{4} = \varepsilon, \end{aligned}$$

therefore $\mathscr{H}_\phi$ is contained in a ball of radius ε for the uniform metric d_∞, which concludes the proof. □

6.5 Stone-Weierstrass theorem

Let X be a compact space and consider the space $\mathscr{C}(X,\mathbf{R})$ of continuous real-valued maps defined on X endowed with the uniform metric.

Theorem 6.10 (Stone[3]-Weierstrass[4])**.** *Let $\mathscr{A} \subseteq \mathscr{C}(X,\mathbf{R})$ satisfy*

(i) *$\mathscr{A}$ is a subalgebra: that is, $\mathscr{A}$ is closed under addition, multiplication, and scalar multiplication;*

(ii) *$\mathscr{A}$ contains the constant maps;*

(iii) *$\mathscr{A}$ separates points: that is, for any distinct $x,y \in X$, there is $f \in \mathscr{A}$ such that $f(x) \neq f(y)$.*

Then $\mathscr{A}$ is dense in $\mathscr{C}(X,\mathbf{R})$.

The proof of this theorem will use the following:

Lemma 6.11. *There exists a sequence of real polynomials $(P_n)_{n\in\mathbf{N}}$ in t that converges uniformly to the function $t \mapsto |t|$ on $[-1,1]$.*

Proof. Define $P_0(t) = 0$ for any $t \in [-1,1]$ and, for any $n \in \mathbf{N}$, $P_{n+1}(t) = P_n(t) + \frac{1}{2}(t^2 - P_n(t)^2)$ for any $t \in [-1,1]$. By induction we easily obtain that $0 \leqslant P_n(t) \leqslant P_{n+1}(t) \leqslant |t|$ for any $t \in [-1,1]$. Therefore, for any $t \in [-1,1]$, the real sequence $(P_n(t))_{n\in\mathbf{N}}$ is increasing and bounded above, hence it converges to some limit that we denote $P(t)$. By passing to the limit $n \to \infty$, we obtain that $0 \leqslant P(t) \leqslant |t|$ and $P(t) = P(t) - \frac{1}{2}(t^2 - P(t)^2)$, thus $P(t) = |t|$.

This implies that the sequence $(P_n)_{n\in\mathbf{N}}$ converges pointwisely to $t \mapsto |t|$ defined on $[-1,1]$. By Dini's theorem (see Exercise sheet) we deduce that $(P_n)_{n\in\mathbf{N}}$ converges uniformly to $t \mapsto |t|$ defined on $[-1,1]$. □

Proof of Theorem 6.10. First of all, observe that $\overline{\mathscr{A}}$ is also a subalgebra of $\mathscr{C}(X,\mathbf{R})$. We split the proof into three steps.

Step 1. We shall prove that if $f,g \in \mathscr{A}$ then $|f| \in \overline{\mathscr{A}}$, $\min(f,g) \in \overline{\mathscr{A}}$, and $\max(f,g) \in \overline{\mathscr{A}}$.

[3] Marshall H. Stone (1903–1989) was an American mathematician who made contributions to real analysis, functional analysis, and topology.

[4] Karl Weierstrass (1815–1897) was a German mathematician who made significant contributions to analysis. He is considered as one of the father's of modern analysis. He is also known for the *Weierstrass function* and the *Bolzano–Weierstrass theorem*.

Let $f \in \mathscr{A}$. If $f = 0$ then it is clear that $f \in \overline{\mathscr{A}}$, so we assume that $f \neq 0$. The sequence $P_n(f/\|f\|_\infty)$ converges uniformly to $|f|/\|f\|_\infty$ thanks to Lemma 6.11. Since $P_n(f/\|f\|_\infty) \in \mathscr{A}$ for any $n \in \mathbf{N}$, it follows that $|f|/\|f\|_\infty \in \overline{\mathscr{A}}$ and hence $|f| \in \overline{\mathscr{A}}$.

Let $f, g \in \mathscr{A}$. Since

$$\min(f,g) = \frac{1}{2}(f + g - |f-g|) \text{ and } \min(f,g) = \frac{1}{2}(f+g+|f-g|),$$

we deduce that $\min(f,g) \in \overline{\mathscr{A}}$ and $\max(f,g) \in \overline{\mathscr{A}}$.

Step 2. We shall prove now that for any distinct $x, y \in X$ and any $\alpha, \beta \in \mathbf{R}$, there is $h \in \mathscr{A}$ such that $h(x) = \alpha$ and $h(y) = \beta$.

Since $\mathscr{A}$ separates points, there is $g \in \mathscr{A}$ such that $g(x) \neq g(y)$. We define

$$h(z) = \frac{\beta - \alpha}{g(y) - g(x)} g(z) + \frac{\alpha g(y) - \beta g(x)}{g(y) - g(x)}, \quad \forall z \in X,$$

so that $h \in \mathscr{A}$, $h(x) = \alpha$ and $h(y) = \beta$.

Step 3. Let $f \in \mathscr{C}(X, \mathbf{R})$ and $\varepsilon > 0$. Let $x \in X$. For any $y \in X \setminus \{x\}$ there is a function $f_y \in \mathscr{A}$ such that $f_y(x) = f(x)$ and $f_y(y) = f(y)$. Denote

$$U_y = \{z \in X \mid f_y(z) < f(z) + \varepsilon\}$$

which is a non-empty open subset of X containing x. Since $X = \bigcup_{y \in X \setminus \{x\}} U_y$ and X is compact, there are $y_1, \ldots, y_n \in X$ such that $X = U_{y_1} \cup \cdots \cup U_{y_n}$. Define the function $g_x = \min(f_{y_1}, \ldots, f_{y_n})$, so that $g_x \in \overline{\mathscr{A}}$ and $g_x(x) = f(x)$. Moreover, for any $z \in X$ there is $i \in \{1, \ldots, n\}$ such that $z \in U_{y_i}$, hence $g_x(z) \leqslant f_{y_i}(z) < f(z) + \varepsilon$.

We now consider the set

$$V_x = \{z \in X \mid f(z) - \varepsilon < g_x(z)\}$$

which is non-empty and open in X. Since $X = \bigcup_{x \in X} V_x$ and X is compact, there are $x_1, \ldots, x_m \in X$ such that $X = V_{x_1} \cup \cdots \cup V_{x_m}$. Consider the function $g = \max(g_{x_1}, \ldots, g_{x_m})$ so that $g \in \overline{\mathscr{A}}$. On the one hand we have $g < f + \varepsilon$ since $g_{x_i} < f + \varepsilon$ for any $i \in \{1, \ldots, m\}$. On the other hand, for any $z \in X$ there is $i \in \{1, \ldots, m\}$ such that $z \in V_{x_i}$, hence $g_x(z) \geqslant g_{x_i}(z) > f(z) - \varepsilon$.

Therefore for any $f \in \mathscr{C}(X, \mathbf{R})$ and $\varepsilon > 0$ there exists $g \in \overline{\mathscr{A}}$ such that $d_\infty(f, g) < \varepsilon$, which means that $\overline{\mathscr{A}}$ is dense in $\mathscr{C}(X, \mathbf{R})$, hence $\mathscr{A}$ is also dense in $\mathscr{C}(X, \mathbf{R})$. □

It is worth noting that the above result regards the space of continuous real-valued maps and one can prove that it is false in the case of complex-valued maps. One can however prove a complex version of the Stone-Weierstrass theorem by adding a further hypothesis.

Consider the space $\mathscr{C}(X, \mathbf{C})$ of complex-valued continuous maps defined on X endowed with the metric of uniform convergence.

If $z \in \mathbf{C}$, we denote $\bar{z}$ its complex conjugate. If $f \in \mathscr{C}(X, \mathbf{C})$, we denote by $\bar{f} \in \mathscr{C}(X, \mathbf{C})$ its complex conjugate function defined by $\bar{f} : z \mapsto \overline{f(z)}$. Moreover, any $f \in \mathscr{C}(X, \mathbf{C})$ can be decomposed into $f = \Re(f) + \mathrm{i}\Im(f)$ where $\Re(f)$ and $\Im(f)$ are real-valued functions.

Corollary 6.12 (Stone-Weierstrass, complex version)**.** *Let $\mathscr{B}$ be a subset of $\mathscr{C}(X, \mathbf{C})$ that satisfies*

(i) *$\mathscr{B}$ is a subalgebra: that is $\mathscr{B}$ is closed under addition, multiplication, and scalar multiplication;*

(ii) *$\mathscr{B}$ contains the constant maps;*

(iii) *$\mathscr{B}$ separates points: that is, for any distinct $x, y \in X$, there is $f \in \mathscr{B}$ such that $f(x) \neq f(y)$;*

(iv) *$\mathscr{B}$ is closed under complex conjugation: that is, if $f \in \mathscr{B}$ then $\bar{f} \in \mathscr{B}$.*

Then $\mathscr{B}$ is dense in $\mathscr{C}(X, \mathbf{C})$.

Proof. Consider the set $\mathscr{B}_{\mathbf{R}} := \{f \in \mathscr{B} \mid f(X) \subseteq \mathbf{R}\}$. Therefore $\mathscr{B}_{\mathbf{R}} \subseteq \mathscr{C}(X, \mathbf{R})$ is a subalgebra and contains the constant maps. Furthermore $\mathscr{B}_{\mathbf{R}}$ separates points, indeed if $x, y \in X$ are distinct there exists a map $f \in \mathscr{B}$ such that $f(x) \neq f(y)$, hence the maps $\Re(f) = \frac{1}{2}(f + \bar{f})$ and $\Im(f) = \frac{1}{2i}(f - \bar{f})$ belong to $\mathscr{B}_{\mathbf{R}}$ and, since $f = \Re(f) + \mathrm{i}\Im(f)$, at least one of them separates the points x and y. Therefore $\mathscr{B}_{\mathbf{R}}$ is dense in $\mathscr{C}(X, \mathbf{R})$ by the Stone-Weierstrass theorem (Theorem 6.10), and hence we can conclude since $\mathscr{B} = \mathscr{B}_{\mathbf{R}} + \mathrm{i}\mathscr{B}_{\mathbf{R}}$. ☐

Chapter 7

Banach spaces and continuous linear maps

Through this chapter we shall always consider **K**-vector spaces where **K** is the field of real numbers **R** or complex numbers **C**. Moreover, by a linear map we shall always mean a **K**-linear map.

7.1 Normed vector spaces and Banach spaces

Definition 7.1. A *norm* on a vector space X is a map $\|\cdot\| : X \to \mathbf{R}^+$ such that, for all $x, y \in X$ and $\lambda \in \mathbf{K}$, one has:

(i) *separation*: $\|x\| = 0$ if and only if $x = 0$;

(ii) *homogeneity*: $\|\lambda x\| = |\lambda|\|x\|$;

(iii) *triangle inequality*: $\|x + y\| \leqslant \|x\| + \|y\|$.

The pair $(X, \|\cdot\|)$ is called a *normed vector space*.

Remark 7.2. We deduce that $\big|\|x\| - \|y\|\big| \leqslant \|x - y\|$ for any $x, y \in X$.

Remark 7.3. A *semi-norm* on X is a map $\|\cdot\| : X \to \mathbf{R}^+$ satisfying (ii) and (iii).

If $(X, \|\cdot\|)$ is a normed space then we can define a metric on X by setting $d(x, y) = \|x - y\|$ for $x, y \in X$, and therefore we can endow X with the topology associated to the metric d. If Y is a vector subspace of X, then Y is also a normed space endowed with the restriction to Y of the norm on X.

In particular, for any $x \in X$ and $r > 0$, we define the open ball centred at x of radius r by

$$B(x, r) := \{y \in X \mid \|x - y\| < r\}$$

as well as the closed ball centred at x of radius r by

$$\overline{B}(x, r) := \{y \in X \mid \|x - y\| \leqslant r\}.$$

Definition 7.4. Let $(X, \|\cdot\|)$ be a normed vector space and A a subset of X. One says that:

(i) A is *bounded* if $\sup_{x\in A} \|x\| < \infty$;

(ii) A is *convex* if for any $x, y \in A$ and any $t \in [0,1]$ one has $tx + (1-t)y \in A$.

One easily remarks that A is bounded if and only if A is contained in some open ball.

Definition 7.5. Two norms $\|\cdot\|_1$ and $\|\cdot\|_2$ on a vector space X are said to be *equivalent* if there are positive constants $\alpha, \beta > 0$ such that

$$\alpha\|x\|_1 \leqslant \|x\|_2 \leqslant \beta\|x\|_1 \quad \text{for all } x \in X.$$

Definition 7.6. A *Banach space* is a normed space that is complete with respect to the metric associated to its norm.

Example 7.7. (1) On $\mathbf{K}^N$ we can define the norms, for $x = (x_1, \dots, x_N) \in \mathbf{K}^N$,

$$\|x\|_p = \left(\sum_{k=1}^{N} |x_k|^p\right)^{1/p} \quad \text{for} \quad 1 \leqslant p < \infty, \quad \|x\|_\infty = \sup_{1\leqslant k\leqslant N} |x_k|.$$

Endowed with any of these norms, $\mathbf{K}^N$ is a Banach space.

(2) For any $p \geqslant 1$, the vector space $\ell^p(\mathbf{N};\mathbf{K})$ of sequences $(x_n)_{n\in\mathbf{N}}$ in $\mathbf{K}$ such that the series with general term $|x_n|^p$ is convergent can be endowed with the norm

$$\|(x_n)_{n\in\mathbf{N}}\|_p = \left(\sum_{n\in\mathbf{N}} |x_n|^p\right)^{\frac{1}{p}}.$$

The space $(\ell^p(\mathbf{N};\mathbf{K}), \|\cdot\|_p)$ is a Banach space.

(3) The vector space $\ell^\infty(\mathbf{N};\mathbf{K})$ of bounded sequences in $\mathbf{K}$ can be endowed with the norm

$$\|(x_n)_{n\in\mathbf{N}}\|_\infty = \sup_{n\in\mathbf{N}} |x_n|.$$

The space $(\ell^\infty(\mathbf{N};\mathbf{K}), \|\cdot\|_\infty)$ is a Banach space.

(4) On the vector space $\mathscr{C}([0,1];\mathbf{K})$ of continuous functions from $[0,1]$ into $\mathbf{K}$, we can define the following norms:

$$\|f\|_p := \left(\int_0^1 |f(x)|^p \, \mathrm{d}x\right)^{\frac{1}{p}} \quad \text{for} \quad 1 \leqslant p < \infty, \quad \|f\|_\infty := \sup_{x\in[0,1]} |f(x)|.$$

The space $(\mathscr{C}([0,1];\mathbf{K}), \|\cdot\|_\infty)$ is a Banach space, however $(\mathscr{C}([0,1];\mathbf{K}), \|\cdot\|_1)$ is not.

(5) If X is a non-empty compact space, we can consider the space $\mathscr{C}(X;\mathbf{K})$ of continuous functions from X to $\mathbf{K}$, endowed with the norm

$$\|f\|_\infty := \sup_{x\in X} |f(x)|.$$

The space $(\mathscr{C}(X;\mathbf{K}), \|\cdot\|_\infty)$ is a Banach space.

7.1.1 Basic properties

We gather below some basic properties of normed vector spaces and Banach spaces.

Proposition 7.8. *Let $(X, \|\cdot\|)$ be a normed space.*

(i) *The map $x \mapsto \|x\|$ from X to $\mathbf{R}^+$ is continuous.*

(ii) *The map $(\lambda, x) \mapsto \lambda x$ from $\mathbf{K} \times X$ to X is continuous.*

(iii) *The map $(x, y) \mapsto x + y$ from $X \times X$ to X is continuous.*

Proof. (i) Let $x_0 \in X$, we deduce that the map $x \mapsto \|x\|$ is continuous at x_0 thanks to the inequality

$$|\|x\| - \|x_0\|| \leqslant \|x - x_0\|.$$

(ii) Let $(\lambda_0, x_0) \in \mathbf{K} \times X$, we deduce that the map $(\lambda, x) \mapsto \lambda x$ is continuous at (λ_0, x_0) since one has

$$\|\lambda x - \lambda_0 x_0\| \leqslant |\lambda - \lambda_0|\|x\| + |\lambda_0|\|x - x_0\|.$$

(iii) Let $(x_0, y_0) \in X \times X$, we deduce that the map $(x, y) \mapsto x + y$ is continuous at (x_0, y_0) since one has

$$\|(x + y) - (x_0 + y_0)\| \leqslant \|x - x_0\| + \|y - y_0\|.$$

□

Proposition 7.9. *Let $(X, \|\cdot\|)$ be a normed vector space.*

(i) *Every open (resp. closed) ball is convex.*

(ii) *The closure of any open ball is equal to the corresponding closed ball.*

Proof. (i) Let $x \in X$ and $r > 0$. We will prove that the open ball $B(x, r) = \{y \in X \mid \|x - y\| < r\}$ is convex. Let $y, z \in B(x, r)$ and $t \in [0, 1]$, then $ty + (1 - t)z \in X$ and

$$\begin{aligned} \|x - (ty + (1 - t)z)\| &= \|t(x - y) + (1 - t)(x - z)\| \\ &\leqslant \|t(x - y)\| + \|(1 - t)(x - z)\| \\ &\leqslant t\|x - y\| + (1 - t)\|x - z\| < tr + (1 - t)r = r, \end{aligned}$$

hence $ty + (1 - t)z \in B(x, r)$. The proof for the closed ball is similar, hence we omit it.

(ii) Let $x \in X$ and $r > 0$. Consider the open ball $B(x, r) = \{y \in X \mid \|x - y\| < r\}$ and the corresponding closed ball $\overline{B}(x, r) = \{y \in X \mid \|x - y\| \leqslant r\}$. Since $B(x, r)$ is included in $\overline{B}(x, r)$ which is a closed set, one has that $\overline{B(x, r)} \subseteq \overline{B}(x, r)$.

We now prove that $\overline{B}(x, r) \subseteq \overline{B(x, r)}$. Let $y \in \overline{B}(x, r)$. Consider the sequence $(z_n)_{n \geqslant 1}$ defined by $z_n = x/n + y(1 - 1/n)$. We can easily check that $z_n \in B(x, r)$ and that $\|z_n - y\| \to 0$, hence $y \in \overline{B(x, r)}$. □

Proposition 7.10. *A closed vector subspace of a Banach space is a Banach space.*

Proof. It is a direct consequence of Proposition 3.5. □

If $(X, \|\cdot\|)$ is a normed vector space, we say that a X-valued series $\sum x_n$ is *convergent* if the sequence of partial sums has a finite limit, which is called the sum of the series and is denoted by $\sum_{n=0}^{\infty} x_n = \lim_{p\to\infty} \sum_{n=0}^{p} x_n$. We say that the series $\sum x_n$ is *absolutely convergent* if the real-valued series $\sum \|x_n\|$ is convergent in $\mathbf{R}$.

Proposition 7.11. *A normed vector space $(X, \|\cdot\|)$ is a Banach space if and only if every absolutely convergent series is convergent*

Proof. Assume that $(X, \|\cdot\|)$ is a Banach space. Let $\sum x_n$ be an absolutely convergent series, then the sequence of partial sums of the series $\sum \|x_n\|$ is a Cauchy sequence in $\mathbf{R}$. It is sufficient to show that the sequence of partial sums $s_p := \sum_{n=0}^{p} x_n$ is a Cauchy sequence in X. Let $\varepsilon > 0$, then there is $N_0 \in \mathbf{N}$ such that for any integers $q > p \geqslant N_0$ one has

$$\sum_{n=p+1}^{q} \|x_n\| < \varepsilon,$$

which implies that

$$\|s_q - s_p\| = \left\| \sum_{n=p+1}^{q} x_n \right\| \leqslant \sum_{n=p+1}^{q} \|x_n\| < \varepsilon.$$

Therefore $(s_p)_{p\in\mathbf{N}}$ is a Cauchy sequence in X and thus the series converges.

Conversely, assume that every absolutely convergent series in X is convergent. Let $(x_n)_{n\in\mathbf{N}}$ be a Cauchy sequence in X, then there is $n_0 \in \mathbf{N}$ such that $\|x_n - x_m\| < 1$ for any $n, m \geqslant n_0$. Moreover there is $n_1 \in \mathbf{N}$ such that $\|x_n - x_m\| < 2^{-1}$ for any $n, m \geqslant n_1$. By induction, one obtains that for any $k \in \mathbf{N}^*$ there is an integer $n_k > n_{k-1}$ such that $\|x_n - x_m\| < 2^{-k}$ for any $n, m \geqslant n_k$. Thus $(x_{n_k})_{k\in\mathbf{N}}$ is a subsequence of $(x_n)_{n\in\mathbf{N}}$. Define $y_0 = x_{n_0}$ and $y_k = x_{n_k} - x_{n_{k-1}}$ for all $k \in \mathbf{N}^*$, then

$$\sum_{k=0}^{\infty} \|y_k\| < \sum_{k=0}^{\infty} 2^{-k} = 2$$

and by hypothesis $\sum y_k$ converges to some limit $y \in X$. Since $x_{n_k} = \sum_{i=0}^{k} y_i$ for any $k \geqslant 0$, the sequence $(x_{n_k})_{k\in\mathbf{N}}$ converges to y. It follows that the Cauchy sequence $(x_n)_{n\in\mathbf{N}}$ has a cluster point, hence it converges by Proposition 3.3. □

7.1.2 Finite-dimensional normed vector spaces

In this section we shall focus on finite-dimensional vector spaces and prove some properties of them.

Theorem 7.12. *Let X be a finite-dimensional vector space over $\mathbf{K}$. Then all norms on X are equivalent and X is complete for any of them.*

Proof. Let $(e_1, \dots, e_n)$ be a basis of X, and for $x \in X$ we write $x = x_1 e_1 + \cdots + x_n e_n$. Since $\mathbf{K}$ is complete, it suffices to show that every norm on X is equivalent to the norm $\|\cdot\|_\infty$ defined by $\|x\|_\infty = \sup(|x_1|, \dots, |x_n|)$.

Let $\|\cdot\|$ be a norm on X and $x \in X$, then we have

$$\|x_1 e_1 + \cdots + x_n e_n\| \leqslant (\|e_1\| + \cdots + \|e_n\|) \sup(|x_1|, \dots, |x_n|),$$

which implies that $\|x\| \leqslant c_1 \|x\|_\infty$, where $c_1 = \|e_1\| + \cdots + \|e_n\| > 0$. This gives the first inequality.

Moreover, last inequality implies that the map $\|\cdot\| : (X, \|\cdot\|_\infty) \to \mathbf{R}$ is continuous. Since $S := \{x \in X \mid \|x\|_\infty = 1\}$ is compact, the image of S under $\|\cdot\|$ is a compact subset of $\mathbf{R}$ and therefore it possess a minimum, that is, there is some $x_0 \in S$ such that $\|x\| \geqslant \|x_0\| > 0$ for any $x \in S$. By homogeneity of the norm and setting $c_2 = \|x_0\|$, we finally get that $c_2 \|x\|_\infty \leqslant \|x\|$ for all $x \in X$, which completes the proof. □

Proposition 7.13. *If X is a finite-dimensional vector space, then the compact sets in X are the closed and bounded sets.*

Proof. We will show that the compact sets in $(\mathbf{R}^N, \|\cdot\|_\infty)$ are the closed and bounded sets, and the case of a finite-dimensional vector space follows from it since all norms are equivalent.

We already know that if $A \subseteq \mathbf{R}^N$ is compact then it is closed and bounded. Conversely, let $A \subseteq \mathbf{R}^N$ be bounded and closed, then there is $M > 0$ such that $A \subseteq [-M, M]^N$, and the set $[-M, M]^N$ is compact as the product of compact sets. Therefore A is a closed subset of a compact set, thus A is compact. □

Theorem 7.14 (Riesz[1] theorem)**.** *The closed unit ball $\overline{B}(0,1)$ of a normed vector space X is compact if and only if X is finite-dimensional.*

Proof. If X is finite-dimensional, we already know that the closed unit ball $\overline{B}(0,1)$ is compact since it is bounded and closed.

Conversely, assume that $\overline{B}(0,1)$ is compact. Since $\{B(x, 1/2)\}_{x \in \overline{B}(0,1)}$ is a cover of $\overline{B}(0,1)$ by open sets in X, there is a finite subset $\{e_1, \dots, e_N\} \subseteq \overline{B}(0,1)$ such that $\overline{B}(0,1) \subseteq B(e_1, \frac{1}{2}) \cup \cdots \cup B(e_N, \frac{1}{2})$. Define $Y = \operatorname{span}(e_1, \dots, e_N)$, we will show that $Y = X$ which will complete the proof. Since Y is closed, it suffices to show that Y is dense in X. Let $x \in X$, then there are $y \in Y$ and $a \in \mathbf{Z}$ such that $\|x - y\| \leqslant 2^{-a}$. Then we have $2^a(x - y) \in B(0,1)$ and hence, by definition of Y, there is $j \in \{1, \dots, N\}$ such that $\|2^a(x - y) - e_j\| \leqslant 1/2$. Therefore one has that $y_1 := y + 2^{-a} e_j \in Y$ and $\|x - y_1\| \leqslant 2^{-a-1}$. By induction we can construct a sequence $(y_n)_{n \in \mathbf{N}}$ of elements of Y such that $\|x - y_n\| \leqslant 2^{-a-n}$, thus x belongs to the closure of Y. □

[1] Frigyes Riesz (1880–1956) was a Hungarian mathematician who made significant contributions to functional analysis. He is also known for the *Riesz representation theorem*, *Riesz rearrangement inequality*, *Radon–Riesz property*. His younger brother, Marcel Riesz, also contributed to functional analysis; toghether they proved the *F. and M. Riesz theorem*.

7.2 Continuous linear maps

Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be normed spaces. We recall that a map $T : X \to Y$ is linear if for any two vectors $x, y \in X$ and any scalar $\lambda \in \mathbf{K}$ the following conditions are satisfied

$$T(x+y) = T(x) + T(y) \quad \text{and} \quad T(\lambda x) = \lambda T(x).$$

In that case one usually denotes $T(x)$ simply by Tx.

Theorem 7.15. *Let $T : X \to Y$ be a linear map. The following are equivalent:*

(i) *T is continuous ;*

(ii) *T is continuous at 0 ;*

(iii) *T is bounded: that is, there is $C > 0$ such that for any $x \in X$ one has $\|Tx\|_Y \leqslant C\|x\|_X$.*

Proof. (i) $\Rightarrow$ (ii) Trivial.

(ii) $\Rightarrow$ (iii) Assume that T is continuous at 0, then there is $\delta > 0$ such that for all $x \in X$, if $\|x\|_X \leqslant \delta$ then $\|u(x)\|_Y \leqslant 1$. Now for any $x \neq 0$ we write $x = \frac{\|x\|_X}{\delta} \frac{\delta}{\|x\|_X} x$, hence

$$\|Tx\|_Y = \frac{\|x\|_X}{\delta} \left\| T(\tfrac{\delta}{\|x\|_X} x) \right\|_Y \leqslant \frac{\|x\|_X}{\delta}.$$

(iii) $\Rightarrow$ (i) Assume that T is bounded. Then, for any $x, y \in X$, one has

$$\|Tx - Ty\|_Y = \|T(x-y)\|_Y \leqslant C\|x-y\|_X,$$

which implies that T is Lipschitz, hence continuous. □

7.2.1 Space of continuous linear maps

Given two normed vector spaces $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$, we denote by $\mathscr{L}(X,Y)$ the vector space of all continuous linear maps from X to Y. For any $T \in \mathscr{L}(X,Y)$ we define

$$\|T\|_{\mathscr{L}(X,Y)} := \sup_{x \in X \setminus \{0\}} \frac{\|Tx\|_Y}{\|x\|_X}.$$

Proposition 7.16. *The map $\|\cdot\|_{\mathscr{L}(X,Y)}$ is a norm on $\mathscr{L}(X,Y)$. Moreover*

$$\|T\|_{\mathscr{L}(X,Y)} = \sup_{\|x\|_X \leqslant 1} \|Tx\|_Y = \sup_{\|x\|_X = 1} \|Tx\|_Y,$$

and also

$$\|Tx\|_Y \leqslant \|T\|_{\mathscr{L}(X,Y)} \|x\|_X.$$

Proof. It is clear that $\|0\|_{\mathscr{L}(X,Y)} = 0$. Let $f \in \mathscr{L}(X,Y)$, if $\|f\|_{\mathscr{L}(X,Y)} = 0$ then $\|f(x)\|_Y = 0$ for all $x \in X$, which implies $f(x) = 0$ for all $x \in X$ since $\|\cdot\|_Y$ is a norm, thus $f \equiv 0$ is the null function. Let $f \in \mathscr{L}(X,Y)$ and $\lambda \in \mathbf{K}$. For any $x \in X \setminus \{0\}$ we have

$$\|\lambda f(x)\|_Y = |\lambda| \|f(x)\|_Y,$$

then

$$\|\lambda f\|_{\mathscr{L}(X,Y)} = \sup_{x\in X\setminus\{0\}} \frac{\|\lambda f(x)\|_Y}{\|x\|_X} = \sup_{x\in X\setminus\{0\}} \frac{|\lambda|\|f(x)\|_Y}{\|x\|_X}$$
$$= |\lambda| \sup_{x\in X\setminus\{0\}} \frac{\|f(x)\|_Y}{\|x\|_X} = |\lambda|\|f\|_{\mathscr{L}(X,Y)}.$$

Let $f, g \in \mathscr{L}(X,Y)$ and $\lambda \in \mathbf{K}$. For any $x \in X \setminus \{0\}$ we have

$$\frac{\|f(x)+g(x)\|_Y}{\|x\|_X} \leqslant \frac{\|f(x)\|_Y}{\|x\|_X} + \frac{\|g(x)\|_Y}{\|x\|_X} \leqslant \|f\|_{\mathscr{L}(X,Y)} + \|g\|_{\mathscr{L}(X,Y)},$$

then

$$\|f+g\|_{\mathscr{L}(X,Y)} = \sup_{x\in X\setminus\{0\}} \frac{\|f(x)+g(x)\|_Y}{\|x\|_X} \leqslant \|f\|_{\mathscr{L}(X,Y)} + \|g\|_{\mathscr{L}(X,Y)}.$$

Denote

$$A := \sup_{x\in X\setminus\{0\}} \frac{\|f(x)\|_Y}{\|x\|_X}, \quad B := \sup_{\|x\|_X=1} \|f(x)\|_Y, \quad C := \sup_{\|x\|_X\leqslant 1} \|f(x)\|_Y.$$

For any $x \in X \setminus \{0\}$ we have

$$\frac{\|f(x)\|_Y}{\|x\|_X} = \left\| f\left(\frac{x}{\|x\|_X}\right)\right\|_Y \leqslant B.$$

thus $A \leqslant B$. It is clear that $B \leqslant C$. For $x \in X$ with $\|x\|_X \leqslant 1$, we have

$$\|f(x)\|_Y \leqslant \frac{\|f(x)\|_Y}{\|x\|_X},$$

thus $C \leqslant A$.

Finally, the last inequality is a direct consequence of the definition. □

Proposition 7.17. *If $(Y, \|\cdot\|_Y)$ is a Banach space, then $(\mathscr{L}(X,Y), \|\cdot\|_{\mathscr{L}(X,Y)})$ is a Banach space.*

Proof. Let $(T_n)_{n\in\mathbf{N}}$ be a Cauchy sequence in $\mathscr{L}(X,Y)$. For any $\varepsilon > 0$ there is $N_0 \in \mathbf{N}$ such that for any integers $n, m \geqslant N_0$ we have $\|T_n - T_m\|_{\mathscr{L}(X,Y)} < \varepsilon$. Thus, for any $x \in X$, we have

$$\|T_n x - T_m x\|_Y \leqslant \|T_n - T_m\|_{\mathscr{L}(X,Y)} \|x\|_X < \varepsilon \|x\|_X. \tag{7.1}$$

It follows that, for any $x \in X$, the sequence $(T_n x)_{n\in\mathbf{N}}$ is a Cauchy sequence in Y, which is complete, hence there is $y_x \in Y$ such that $(Tx_n)_{n\in\mathbf{N}}$ converges to y_x in Y. Define the map $T : x \mapsto y_x$ from X to Y, which clearly is linear. It remains to show that T is continuous and that $(T_n)_{n\in\mathbf{N}}$ converges to T in $\mathscr{L}(X,Y)$.

From (7.1) we obtain that, for any $x \in X$,

$$\|Tx\|_Y \leqslant \|Tx - T_{N_0}x\|_Y + \|T_{N_0}x\|_Y \leqslant \left(\varepsilon + \|T_{N_0}\|_{\mathscr{L}(X,Y)}\right) \|x\|_X,$$

which implies that T is continuous. Finally, taking the limit $m \to \infty$ in (7.1) we obtain, for any $x \in X$ and $n \geqslant N_0$,

$$\|Tx - T_n x\|_Y = \lim_{m\to\infty} \|T_n x - T_m x\|_Y \leqslant \varepsilon \|x\|_X,$$

which implies that $\|T - T_n\|_{\mathscr{L}(X,Y)} \leqslant \varepsilon$, thus $(T_n)_{n\in\mathbf{N}}$ converges to T in $\mathscr{L}(X,Y)$. □

We now state a result concernig the extension of continuous linear maps.

Theorem 7.18 (Bounded linear transformation theorem)**.** *Let X be a normed space, Y a Banach space, and D a dense vector subspace of X. Then every continuous linear map $T : D \to Y$ can be uniquely extended to a continuous linear map $\widehat{T} : X \to Y$, that is, $\widehat{T} \in \mathscr{L}(X,Y)$ and $\widehat{T}_{|D} = T$.*

Proof. Let $x \in X$, then there is a sequence $(x_n)_{n\in\mathbf{N}}$ in D converging to x in X. For all $n, m \in \mathbf{N}$ we have

$$\|Tx_n - Tx_m\|_Y \leqslant \|T\|_{\mathscr{L}(D,Y)} \|x_n - x_m\|_X,$$

thus $(Tx_n)_{n\in\mathbf{N}}$ is a Cauchy sequence in Y which is complete, hence it converges to some limit.

We define the map $\widehat{T} : X \to Y$, $x \mapsto \lim_{n\to\infty} Tx_n$, where $(x_n)_{n\in\mathbf{N}}$ is a sequence in D converging to x. We first check that $\widehat{T}$ is well-defined, that is, it does not depend on the choice of the sequence converging to x. Indeed if $(x'_n)_{n\in\mathbf{N}}$ is another sequence in D converging to x, then

$$\|Tx_n - Tx'_n\|_Y \leqslant \|T\|_{\mathscr{L}(D,Y)} \|x_n - x'_n\|_X \xrightarrow[n\to\infty]{} 0,$$

thus $\lim_{n\to\infty} Tx_n = \lim_{n\to\infty} Tx'_n$.

The map $\widehat{T}$ is clearly linear. For all $x \in D$, we consider the constant sequence $(x)_{n\in\mathbf{N}}$ so that $\widehat{T}x = \lim_{n\to\infty} Tx = Tx$, that is, $\widehat{T}$ is an extension of T.

We now prove that $\widehat{T}$ is continuous. Let $x \in X$ and $(x_n)_{n\in\mathbf{N}}$ a sequence in D converging to x, then

$$\|\widehat{T}x\|_Y = \|\lim_{n\to\infty} Tx_n\|_Y = \lim_{n\to\infty} \|Tx_n\|_Y \leqslant \lim_{n\to\infty} \|T\|_{\mathscr{L}(D,Y)} \|x_n\|_X \leqslant \|T\|_{\mathscr{L}(D,Y)} \|x\|_X,$$

which implies that $\widehat{T}$ is continuous and $\|\widehat{T}\|_{\mathscr{L}(X,Y)} \leqslant \|T\|_{\mathscr{L}(D,Y)}$. Moreover, we have

$$\|\widehat{T}\|_{\mathscr{L}(X,Y)} = \sup_{x\in X,\ \|x\|_X\leqslant 1} \|\widehat{T}x\|_Y \geqslant \sup_{x\in D,\ \|x\|_X\leqslant 1} \|\widehat{T}x\|_Y = \sup_{x\in D,\ \|x\|_X\leqslant 1} \|Tx\|_Y = \|T\|_{\mathscr{L}(D,Y)},$$

thus $\|\widehat{T}\|_{\mathscr{L}(X,Y)} = \|T\|_{\mathscr{L}(D,Y)}$.

We finally prove that $\widehat{T}$ is unique. Let $\widetilde{T} \in \mathcal{L}(X,Y)$ be an extension of T. Let $x \in X$ and consider a sequence $(x_n)_{n\in\mathbf{N}}$ in D converging to x. Thus

$$\widehat{T}x = \lim_{n\to\infty} Tx_n = \lim_{n\to\infty} \widetilde{T}x_n = \widetilde{T}\left(\lim_{n\to\infty} x_n\right) = \widetilde{T}x$$

using the continuity of $\widetilde{T}$, hence $\widehat{T} = \widetilde{T}$. □

7.2.2 Dual space

Let $(X, \|\cdot\|_X)$ be a normed vector space.

Definition 7.19. We define the *(topological) dual space of* X as the vector space $X' = \mathscr{L}(X, \mathbf{K})$ of all continuous linear functionals on X.

We can endow the dual space X' with the norm

$$\|f\|_{X'} = \|f\|_{\mathscr{L}(X,\mathbf{K})} = \sup_{x \in X,\, \|x\|_X \leqslant 1} |f(x)|.$$

We observe that, since $\mathbf{K}$ is complete, it follows from Proposition 7.17 that $(X', \|\cdot\|_{X'})$ is a Banach space.

Remark 7.20. We will prove later in Section 7.4, a few properties of the dual space X' as a consequence of the Hahn-Banach theorems. In particular there holds (see Corollary 7.34 and Corollary 7.43):

- For any $x \in X$ one has $\|x\|_X = \sup_{\|f\|_{X'} \leqslant 1} |f(x)|$.
- The dual space X' separates points in X: for all distinct $x, y \in X$ there is $f \in X'$ such that $f(x) \neq f(y)$.
- For any vector subspace $Y \subseteq X$ such that $\overline{Y} \neq X$, there exists $f \in X' \setminus \{0\}$ such that $f(x) = 0$ for any $x \in Y$.

7.2.3 Isomorphisms

Definition 7.21. Let X and Y be normed vector spaces. A map $f : X \to Y$ is:

(i) An *isometry* if it preserves the distance, that is, if $\|f(x) - f(y)\|_Y = \|x - y\|_X$ for all $x, y \in X$.

(ii) An *isomorphism* if it is bijective and if both f and its inverse map f^{-1} are linear and continuous. Two normed vector spaces are said to be *isomorphic* if there is a isomorphism between them.

(iii) An *isometric isomorphism* if it is an isomorphism that is also an isometry. Two normed vector spaces are said to be *isometrically isomorphic* if there is a isometric isomorphism between them.

Remark 7.22. If $f : X \to Y$ is linear and bijective, then its inverse map is also linear.

We shall denote by $\mathrm{Isom}(X, Y)$ the set of all isomorphisms from X into Y.

Proposition 7.23. *Every normed vector space* Y *is isometrically isomorphic to* $\mathscr{L}(\mathbf{R}, Y)$.

Proof. For any $x \in Y$, we can associate the linear map $\lambda \mapsto \lambda x$ from $\mathbf{R}$ to Y, which clearly is continuous since $\|\lambda x\|_Y = |\lambda| \|x\|_Y$. We hence consider the map $\phi : x \mapsto \{\lambda \mapsto \lambda x\}$ from Y to $\mathscr{L}(\mathbf{R}, Y)$, which is linear and continuous since

$$\|\phi(x)\|_{\mathscr{L}(\mathbf{R},Y)} = \sup_{\lambda \in \mathbf{R} \setminus \{0\}} \frac{\|\lambda x\|_Y}{|\lambda|} = \|x\|_Y.$$

In the other way, define the map $\psi : f \mapsto f(1)$ from $\mathscr{L}(\mathbf{R}, Y)$ to Y. It is clear that ψ is linear and continuous, and moreover we have $\phi \circ \psi = \mathrm{id}_{\mathscr{L}(\mathbf{R},Y)}$ and $\psi \circ \phi = \mathrm{id}_Y$, thus ϕ is bijective with inverse ψ. We deduce that ϕ is a isometric isomorphism from Y into $\mathscr{L}(\mathbf{R}, Y)$. □

We now prove the following result concerning the set of isomorphisms and the inversion map.

Proposition 7.24. *Let X and Y be Banach spaces.*

(i) *If $T \in \mathscr{L}(X, X)$ with $\|T\|_{\mathscr{L}(X,X)} < 1$, then $\mathrm{id}_X - T \in \mathrm{Isom}(X, X)$ with inverse $\sum_{n\in\mathbf{N}} T^n$.*

(ii) *$\mathrm{Isom}(X, Y)$ is open in $\mathscr{L}(X, Y)$.*

(iii) *The inversion map $T \mapsto T^{-1}$ from $\mathrm{Isom}(X, Y)$ into $\mathrm{Isom}(Y, X)$ is continuous.*

Proof. (i) The series $\sum T^k$ in the Banach space $\mathscr{L}(X, X)$ is absolutely convergent since $\|T^k\|_{\mathscr{L}(X,X)} \leqslant \|T\|^k_{\mathscr{L}(X,X)}$ for all $k \in \mathbf{N}$, hence it converges to $S = \sum_{n\in\mathbf{N}} T^n \in \mathscr{L}(X, X)$ by Proposition 7.11.

For any $n \in \mathbf{N}$ we denote $S_n = \sum_{k=0}^n T^k$. We easily check that $S_n T = T S_n = S_{n+1} - \mathrm{id}_X$. Therefore taking the limit $n \to \infty$ implies $ST = TS = S - \mathrm{id}_X$, and we deduce that S is the inverse of $\mathrm{id}_X - T$. Moreover one has

$$\|(\mathrm{id}_X - T)^{-1}\|_{\mathscr{L}(X,X)} \leqslant (1 - \|T\|_{\mathscr{L}(X,X)})^{-1}.$$

(ii) Let $T_0 \in \mathrm{Isom}(X, Y)$. Let $r := \|T_0^{-1}\|^{-1}_{\mathscr{L}(Y,X)} > 0$, we will show that $\{T \in \mathscr{L}(X, Y) \mid \|T - T_0\|_{\mathscr{L}(X,Y)} < r\}$ is contained in $\mathrm{Isom}(X, Y)$, which implies that $\mathrm{Isom}(X, Y)$ is open.

Let $T \in \mathscr{L}(X, Y)$ be such that $\|T - T_0\|_{\mathscr{L}(X,Y)} < r$. Define

$$S = -T_0^{-1}(T - T_0) = \mathrm{id}_X - T_0^{-1}T,$$

then

$$\|S\|_{\mathscr{L}(X,X)} \leqslant \|T_0^{-1}\|_{\mathscr{L}(Y,X)} \|T - T_0\|_{\mathscr{L}(X,Y)} < 1. \tag{7.2}$$

Therefore $\mathrm{id}_X - S = T_0^{-1}T$ is an isomorphism from X to X by question 1, which implies that T is invertible and $T^{-1} = (\mathrm{id}_X - S)^{-1}T_0^{-1}$ is continuous, hence $T \in \mathrm{Isom}(X, Y)$. Moreover one has

$$\begin{aligned}\|T^{-1}\|_{\mathscr{L}(Y,X)} &\leqslant \|(\mathrm{id}_X - S)^{-1}\|_{\mathscr{L}(X,X)} \|T_0^{-1}\|_{\mathscr{L}(Y,X)} \\ &\leqslant \frac{\|T_0^{-1}\|_{\mathscr{L}(Y,X)}}{1 - \|T_0^{-1}(T - T_0)\|_{\mathscr{L}(X,X)}}.\end{aligned}$$

(iii) Let $T_0 \in \mathrm{Isom}(X, Y)$. For $T \in \mathscr{L}(X, Y)$ such that $\|T - T_0\|_{\mathscr{L}(X,Y)} < r$, the expression of T^{-1} above yields

$$\begin{aligned}\|T^{-1} - T_0^{-1}\|_{\mathscr{L}(Y,X)} &\leqslant \|(\mathrm{id}_X - S)^{-1} - \mathrm{id}_X\|_{\mathscr{L}(X,X)} \|T_0^{-1}\|_{\mathscr{L}(Y,X)} \\ &= \left\| \sum_{n=1}^{\infty} S^n \right\|_{\mathscr{L}(X,X)} \|T_0^{-1}\|_{\mathscr{L}(Y,X)} \\ &\leqslant \frac{\|S\|_{\mathscr{L}(X,X)} \|T_0^{-1}\|_{\mathscr{L}(Y,X)}}{1 - \|S\|_{\mathscr{L}(X,X)}},\end{aligned}$$

which goes to 0 as T tends to T_0 thanks to (7.2). This proves that $T \mapsto T^{-1}$ is continuous at T_0. □

7.2.4 Continuous bilinear maps

Let $(X_1, \|\cdot\|_{X_1})$, $(X_2, \|\cdot\|_{X_2})$ and $(Y, \|\cdot\|_Y)$ be normed vector spaces. The topological space $X_1 \times X_2$ is a normed vector space, and the product topology is associated to the norm

$$\|(x_1, x_2)\|_{X_1 \times X_2} = \sup(\|x_1\|_{X_1}, \|x_2\|_{X_2})$$

or

$$\|(x_1, x_2)\|_{X_1 \times X_2} = \|x_1\|_{X_1} + \|x_2\|_{X_2},$$

or still any other equivalent norm.

Definition 7.25. A map $T : X_1 \times X_2 \to Y$ is *bilinear* if

- for any $x \in X_1$, the map $y \mapsto T(x, y)$ from X_2 to Y is linear;
- for any $y \in X_2$, the map $x \mapsto T(x, y)$ from X_1 to Y is linear.

Theorem 7.26. *Let $T : X_1 \times X_2 \to Y$ be a bilinear map. The following are equivalent:*

(i) *T is continuous;*

(ii) *T is continuous at $(0, 0)$;*

(iii) *T is bounded: there is $C > 0$ such that, for any $(x_1, x_2) \in X_1 \times X_2$, one has*

$$\|T(x_1, x_2)\|_Y \leqslant C\|x_1\|_{X_1}\|x_2\|_{X_2}.$$

Proof. The proof is similar to the proof of Theorem 7.15, and hence is left as an exercise. □

We denote by $\mathscr{L}_2(X_1, X_2; Y)$ the vector space of all bilinear continuous maps from $X_1 \times X_2$ to Y. The space $\mathscr{L}_2(X_1, X_2; Y)$ can be endowed with the norm

$$\|T\|_{\mathscr{L}_2(X_1,X_2;Y)} := \sup_{x_1 \in X_1 \setminus \{0\},\, x_2 \in X_2 \setminus \{0\}} \frac{\|T(x_1, x_2)\|_Y}{\|x_1\|_{X_1}\|x_2\|_{X_2}}$$

and thus $(\mathscr{L}_2(X_1, X_2; Y), \|\cdot\|_{\mathscr{L}_2(X_1,X_2;Y)})$ is a normed vector space. As for the case of continuous linear maps, one can easily obtain that

$$\|T\|_{\mathscr{L}_2(X_1,X_2;Y)} = \sup_{\|x_1\|_{X_1} \leqslant 1, \|x_2\|_{X_2} \leqslant 1} \|T(x_1, x_2)\|_Y = \sup_{\|x_1\|_{X_1} = 1, \|x_2\|_{X_2} = 1} \|T(x_1, x_2)\|_Y.$$

Proposition 7.27. *The spaces $\mathscr{L}_2(X_1, X_2; Y)$ and $\mathscr{L}(X_1, \mathscr{L}(X_2, Y))$ are isomorphic. More precisely, the map*

$$\begin{array}{rcl} \mathscr{L}_2(X_1, X_2; Y) & \to & \mathscr{L}(X_1, \mathscr{L}(X_2, Y)) \\ T & \mapsto & \{x \mapsto (y \mapsto T(x, y))\} \end{array}$$

is an isometric isomorphism between normed spaces, and its inverse map is given by

$$\begin{array}{rcl} \mathscr{L}(X_1, \mathscr{L}(X_2, Y)) & \to & \mathscr{L}_2(X_1, X_2; Y) \\ S & \mapsto & \{(x, y) \mapsto S(x)(y)\}. \end{array}$$

Proof. For any $x \in X_1$ and $y \in X_2$ one has, denoting $T_x : y \mapsto T(x, y)$,

$$\|T_x(y)\|_Y = \|T(x, y)\|_Y \leqslant \|T\|_{\mathscr{L}_2(X_1,X_2;Y)} \|x\|_{X_1} \|y\|_{X_2},$$

thus the linear map $T_x : X_2 \to Y$ is continuous with

$$\|T_x\|_{\mathscr{L}(X_2,Y)} \leqslant \|T\|_{\mathscr{L}_2(X_1,X_2;Y)} \|x\|_{X_1}.$$

Therefore the map $x \mapsto T_x$ is continuous with norm less or equal than $\|T\|_{\mathscr{L}_2(X_1,X_2;Y)}$. Since one has

$$\sup_{x \in X_1 \setminus \{0\}} \frac{\|T_x\|_{\mathscr{L}(X_2,Y)}}{\|x\|_{X_1}} = \sup_{x \in X_1 \setminus \{0\}} \sup_{y \in X_2 \setminus \{0\}} \frac{\|T(x, y)\|_Y}{\|x\|_{X_1} \|y\|_{X_2}} = \|T\|_{\mathscr{L}_2(X_1,X_2;Y)},$$

it follows that the map $T \mapsto \{x \mapsto T_x\}$ is an isometry.

Finally, for $S \in \mathscr{L}(X_1, \mathscr{L}(X_2, Y))$ we define $T : (x, y) \mapsto S(x)(y)$ which is a bilinear map. Since

$$\|S(x)(y)\|_Y \leqslant \|S(x)\|_{\mathscr{L}(X_2,F)} \|y\|_{X_2} \leqslant \|v\|_{\mathscr{L}_2(X_1,X_2;Y)} \|x\|_{X_1} \|y\|_{X_2},$$

we deduce that T is continuous and $S(x) = T_x$, which completes the proof. □

7.3 Applications of Baire's theorem

As a consequence of Baire's theorem (Theorem 3.13) for complete metric spaces, we shall state and prove below three fundamental results in functional analysis: the Banach-Steinhaus theorem or uniform boundedness principle, the open mapping theorem, and the closed graph theorem.

7.3.1 Banach-Steinhaus theorem

Theorem 7.28 (Banach-Steinhaus[2] theorem or uniform boundedness principle)**.** *Let $(X, \|\cdot\|_X)$ be a Banach space and $(Y, \|\cdot\|_Y)$ a normed vector space. Let $\{T_\alpha\}_{\alpha \in \mathcal{A}} \subseteq \mathscr{L}(X, Y)$ be an arbitrary family of continuous linear maps from X into Y such that for any $x \in X$ one has*

$$\sup_{\alpha \in \mathcal{A}} \|T_\alpha x\|_Y < +\infty.$$

Then there holds

$$\sup_{\alpha \in \mathcal{A}} \|T_\alpha\|_{\mathscr{L}(X,Y)} < +\infty.$$

Proof. Define the set $X_n = \{x \in X \mid \forall \alpha \in \mathcal{A}, \|T_\alpha x\|_Y \leqslant n\}$ for any $n \in \mathbf{N}$. Then X_n is closed and, by hypothesis, one has $X = \bigcup_{n \in \mathbf{N}} X_n$. By Baire's theorem (Theorem 3.13), we deduce that there exists $N \in \mathbf{N}$ such that the interior of X_N is not empty, thus there are $x_0 \in X$ and $r > 0$ such that $\overline{B_X}(x_0, r) \subseteq X_N$.

[2] Hugo Steinhaus (1887–1972) was a Polish mathematician who made important contributions to functional analysis, geometry, probability, and logic.

This implies that for any $\alpha \in \mathcal{A}$ and any $x \in X$ such that $\|x\|_X \leqslant r$, one has $\|T_\alpha x\|_Y \leqslant \|T_\alpha(x + x_0)\|_Y + \|T_\alpha x_0\|_Y \leqslant 2N$. Finally, one obtains that for any $\alpha \in \mathcal{A}$ there holds

$$\|T_\alpha\|_{\mathscr{L}(X,Y)} = \sup_{\|x\|_X \leqslant 1} \|T_\alpha x\|_Y \leqslant \frac{2N}{r},$$

which concludes the proof. □

7.3.2 Open mapping theorem

Theorem 7.29 (Open mapping theorem)**.** *Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be two Banach spaces, and $T \in \mathscr{L}(X,Y)$ be a surjective continuous linear map from X into Y. Then T is an open map.*

Proof. Define $A_n = n\,\overline{T(B_X(0,1))}$ for any $n \in \mathbf{N}^*$, so that we have $Y = \bigcup_{n\in\mathbf{N}^*} A_n$ since T is surjective. Thanks to Baire's theorem (Theorem 3.13), we deduce that there is $N \in \mathbf{N}^*$ such that the interior of A_N is not empty. It follows that the interior of $\overline{T(B_X(0,1))}$ is not empty, hence there are $r > 0$ and $y_0 \in Y$ such that $B_Y(y_0, r) \subseteq \overline{T(B_X(0,1))}$. In particular $y_0 \in \overline{T(B_X(0,1))}$ which implies that $-y_0 \in \overline{T(B_X(0,1))}$, hence

$$\begin{aligned} B_Y(0,r) = -y_0 + B_Y(y_0, r) &\subseteq \overline{T(B_X(0,1))} + \overline{T(B_X(0,1))} \\ &\subseteq \overline{T(B_X(0,2))} = 2\overline{T(B_X(0,1))}, \end{aligned}$$

and thus $B_Y(0, r/2) \subseteq \overline{T(B_X(0,1))}$. Observe that this implies that for any $y \in Y$ such that $\|y\|_Y < \lambda \frac{r}{2}$, with $\lambda > 0$, and any $\varepsilon > 0$ there is $z \in X$ such that $\|z\|_X < \lambda$ and $\|y - Tz\|_Y < \varepsilon$.

Now let $y \in Y$ with $\|y\|_Y < r/4$, then taking $\varepsilon = r/4$ we find $z_1 \in X$ such that $\|z_1\|_X < 1/2$ and $\|y - Tz_1\|_Y < r/4$. By applying the same argument to $y - Tz_1$ that verifies $\|y - Tz_1\|_Y < r/4$ and taking $\varepsilon = r/8$, we find $z_2 \in X$ such that $\|z_2\|_X < 1/4$ and $\|y - Tz_1 - Tz_2\|_Y < r/8$. By induction we construct a sequence $(z_n)_{n\in\mathbf{N}^*}$ in X such that for any $n \in \mathbf{N}^*$ we have

$$\|z_n\|_X < \frac{1}{2^n} \quad \text{and} \quad \|y - T(z_1 + \cdots + z_n)\|_Y < \frac{r}{2^{n+1}}.$$

Therefore the sequence $(x_n)_{n\in\mathbf{N}^*}$, defined by $x_n = z_1 + \cdots + z_n$, is a Cauchy sequence in X, which is complete. Hence there is some $x \in X$ such that $x_n \to x$, and we have $\|x\|_X < 1$ and $y = Tx$ since T is continuous. We have hence proved that for any $y \in Y$ with $\|y\|_Y < r/4$ there is $x \in X$ with $\|x\|_X < 1$ such that $y = Tx$, which implies that

$$B_Y(0, r/4) \subseteq T(B_X(0,1)).$$

We now prove that the above estimate implies that T is an open map. Let V be open in X. Let $y \in T(V)$, then there is $x \in V$ such that $y = Tx$. Since V is open, there is $r > 0$ so that $B_X(x,r) \subseteq V$. Hence $x + B_X(0,r) = B_X(x,r) \subseteq V$ from which we obtain $y + T(B_X(0,r)) \subseteq u(V)$. By hypothesis we have $B_Y(0, rc) \subseteq T(B_X(0,r))$, then it follows $B_Y(y, rc) = y + B_Y(0, rc) \subseteq y + T(B_X(0,r)) \subseteq T(V)$. This proves that $T(V)$ is open in Y. □

We now state a direct consequence of the open mapping theorem.

Corollary 7.30 (Banach's isomorphism theorem)**.** *Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be two Banach spaces, and $T \in \mathscr{L}(X, Y)$ be a bijective continuous linear map from X into Y. Then T is an isomorphism.*

Proof. The map T is bijective, continuous and open by the open mapping theorem (Theorem 7.29), hence T is a homeomorphism. Therefore T^{-1} is linear and continuous, hence T is an isomorphism. □

7.3.3 Closed graph theorem

Theorem 7.31 (Closed graph theorem)**.** *Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be two Banach spaces and $T : X \longrightarrow Y$ be a linear map. Assume that the graph $G(T) = \{(x, y) \in X \times Y \mid y = Tx\}$ of T is closed in $X \times Y$. Then T is continuous.*

Recall that the converse is true, indeed we already know that the graph of a continuous map (not necessarily linear) is closed.

We remark that saying that the graph $G(T)$ of T is closed in $X \times Y$ is equivalent to: For every sequence $(x_n)_{n \in \mathbf{N}}$ in X that converges to $x \in X$ and is such that $(Tx_n)_{n \in \mathbf{N}}$ converges to $y \in Y$, then one has $y = Tx$.

Proof of Theorem 7.31. Recall that the product space $X \times Y$, endowed with the norm $\|(x, y)\|_{X \times Y} := \|x\|_X + \|y\|_Y$, is a Banach space. Since T is linear, it follows that its graph $G(T)$ is a closed vector subspace of $X \times Y$, thus also a Banach space.

The projection $\pi : (x, y) \longmapsto x$ is a continuous linear bijective map from $G(T)$ onto X, hence by Corollary 7.30 the map $\pi^{-1} : x \longmapsto (x, Tx)$ from X onto $G(T)$ is also continuous. The projection $\theta : (x, y) \longmapsto y$ from $X \times Y$ into Y is also a continuous linear map, therefore one deduces that $T = \theta \circ \pi^{-1}$ is continuous as the composition of continuous maps. □

7.4 Hahn-Banach theorems

7.4.1 Analytic version

We shall prove the analytic form of the Hahn[3]-Banach theorem, which concerns the extension of linear functionals.

Theorem 7.32 (Analytic form of the Hahn-Banach theorem: first version)**.** *Let X be a $\mathbf{R}$-vector space and $p : X \longrightarrow \mathbf{R}$ be a map verifying, for any $x, y \in X$ and $\lambda \in \mathbf{R}_+$,*

$$p(\lambda x) = \lambda p(x) \quad \textit{and} \quad p(x + y) \leqslant p(x) + p(y).$$

Consider a vector subspace V of X and a linear functional $\varphi : V \longrightarrow \mathbf{R}$ such that

$$\varphi(x) \leqslant p(x), \quad \forall x \in V.$$

[3] Hans Hahn (1879–1934) was an Austrian mathematician who contributed to set theory, topology, functional analysis, and calculus of variation. He is also known for the *Hahn decomposition theorem*, *Hahn–Mazurkiewicz theorem*, and *Vitali–Hahn–Saks theorem*.

Then there is a linear functional $f : X \to \mathbf{R}$ such that

(i) *f extends φ, i.e. $f(x) = \varphi(x)$ for any $x \in V$;*

(ii) *$f(x) \leqslant p(x)$ for any $x \in X$.*

Proof. Define $\mathscr{P}$ as the set of all linear functionals $h : D(h) \to \mathbf{R}$ where $D(h)$ is a vector subspace of X containing V, h extends φ, and $h(x) \leqslant p(x)$ for any $x \in D(h)$, with the partial order

$$h_1 \preceq h_2 \quad \Longleftrightarrow \quad D(h_1) \subseteq D(h_2) \ \text{ and } \ h_1(x) = h_2(x), \quad \forall x \in D(h_1).$$

It is clear that $\mathscr{P}$ is non-empty since it contains φ. We now verify that every chain (totally ordered subset) in $\mathscr{P}$ has an upper bound. Let $\mathscr{Q} \subseteq \mathscr{P}$ be a chain, then $\bigcup_{h \in \mathscr{Q}} D(h)$ is a vector subspace of X since $\mathscr{Q}$ is totally ordered. Define $q : D(q) \to \mathbf{R}$ by $q(x) = h(x)$ if $x \in D(h)$, where $D(q) = \bigcup_{h \in \mathscr{Q}} D(h)$. Then $q \in \mathscr{P}$ and is an upper bound of $\mathscr{Q}$.

Therefore by Zorn's Lemma (Theorem A.23) the set $\mathscr{P}$ has a maximal element f. We shall check that $D(f) = X$, which completes the proof.

By way of contradiction, assume that $D(f) \neq X$, thus there is $x_0 \in X \setminus D(f)$. For any $x, y \in D(f)$ one has

$$f(x) + f(y) = f(x + y) \leqslant p(x + y) \leqslant p(x + x_0) + p(y - x_0)$$

thus

$$f(y) - p(y - x_0) \leqslant \alpha := \sup_{y \in D(f)} \{f(y) - p(y - x_0)\} \leqslant p(x + x_0) - f(x). \tag{7.3}$$

Define the map $g(y + tx_0) = f(y) + \alpha t$ for any $t \in \mathbf{R}$ and $y \in D(f)$, then g is a linear functional that extends f, whence that extends φ, over $D(g) = D(f) + \mathbf{R}x_0$ and $D(f) \subsetneq D(g)$, thus $f \preceq g$ and $f \neq g$. We obtain that $g \in \mathscr{P}$ by using (7.3), which contradicts the fact that Y is a maximal element of $\mathscr{P}$. □

We give now another version of the Hahn-Banach theorem considering normed vector spaces.

Theorem 7.33 (Analytic form of the Hahn-Banach theorem: second version)**.** *Let $(X, \|\cdot\|)$ be a normed vector space over $\mathbf{K} = \mathbf{R}$ or $\mathbf{C}$. Let V be a vector subspace of X and $\varphi : V \to \mathbf{K}$ be a continuous linear functional. Then there exists a continuous linear functional $f \in X' = \mathscr{L}(X, \mathbf{K})$ such that*

(i) *f extends φ, i.e. $f(x) = \varphi(x)$ for any $x \in V$;*

(ii) *$\|f\|_{X'} = \|\varphi\|_{V'}$, more precisely*

$$\sup_{x \in X,\, \|x\| \leqslant 1} |f(x)| = \sup_{x \in V,\, \|x\| \leqslant 1} |\varphi(x)|.$$

Proof. When $K = \mathbf{R}$ the result is a consequence of Theorem 7.32. Indeed taking $p(x) = \|\varphi\|_{V'}\|x\|$ for any $x \in X$, we deduce the existence of a linear functional $f : X \to \mathbf{R}$ that extends φ and such that $f(x) \leqslant \|\varphi\|_{V'}\|x\|$ for any $x \in X$.

Exchanging x in $-x$ we hence obtain that $|f(x)| \leqslant \|\varphi\|_{V'}\|x\|$ for any $x \in X$. Therefore f is continuous and we have

$$\|f\|_{X'} \leqslant \|\varphi\|_{V'},$$

which implies $\|f\|_{X'} = \|\varphi\|_{V'}$.

When $\mathbf{K} = \mathbf{C}$, we shall use the result in the real case by considering the $\mathbf{R}$-vector space $X_{\mathbf{R}}$ associated to X, that is, $X_{\mathbf{R}}$ is the same set as X endowed with the vector space structure of X where scalars are restricted to $\mathbf{R}$. Let $V_{\mathbf{R}}$ be the $\mathbf{R}$-vector space associated to V. Consider u the real part of φ defined by $u(x) = \operatorname{Re}\varphi(x)$ for any $x \in V$, so that u is a $\mathbf{R}$-linear functional on $V_{\mathbf{R}}$ and $\varphi(x) = u(x) - \mathrm{i}u(\mathrm{i}x)$ for any $x \in V$. By hypothesis u is a continuous $\mathbf{R}$-linear functional on $V_{\mathbf{R}}$ with norm equal to $\|\varphi\|_{V'}$. Thanks to the result in the real case, there is a continuous $\mathbf{R}$-linear functional U on $X_{\mathbf{R}}$ that extends u and $\|U\|_{(X_{\mathbf{R}})'} = \|\varphi\|_{V'}$. For any $x \in X$ we define

$$f(x) = U(x) - \mathrm{i}U(\mathrm{i}x),$$

so that f is a continuous $\mathbf{C}$-linear functional on X that extends φ and

$$\|f\|_{X'} = \|U\|_{(X_{\mathbf{R}})'} = \|\varphi\|_{V'}.$$

□

Let us state some consequences of the Hahn-Banach theorem.

Corollary 7.34. *Let $(X, \|\cdot\|_X)$ be a normed vector space over $\mathbf{K} = \mathbf{R}$ or $\mathbf{C}$. The following properties hold:*

(i) *For any $x_0 \in X$ there exists $f_0 \in X'$ such that $\|f_0\|_{X'} = \|x_0\|_X$ and $f_0(x_0) = \|x_0\|_X^2$.*

(ii) *For any $x \in X$ one has*

$$\|x\|_X = \sup_{f \in X',\, \|f\|_{X'} \leqslant 1} |f(x)|.$$

(iii) *The topological dual space X' separates points in X, that is, for any distinct $x, y \in X$ there is $f \in X'$ such that $f(x) \neq f(y)$.*

Proof. (i) Apply Theorem 7.33 with $V = \mathbf{K}x_0$ and $\varphi(\lambda x_0) = \lambda\|x_0\|_X^2$ for any $\lambda \in \mathbf{K}$, so that $\|\varphi\|_{V'} = \|x_0\|_X$.

(ii) Assume $x \neq 0$, otherwise is trivial. Since $f(x) \leqslant \|f\|_{X'}\|x\|$ for any $f \in X'$, one already has

$$\sup_{f \in X',\, \|f\|_{X'} \leqslant 1} |f(x)| \leqslant \|x\|.$$

Moreover, from (i) we know that there is some $f_0 \in X'$ such that $\|f_0\|_{X'} = \|x\|_X$ and $f_0(x) = \|x\|_X^2$. Define $f_1 = f_0/\|x\|_X$, then $\|f_1\|_{X'} = 1$ and $f_1(x) = \|x\|_X$, which concludes the proof.

(iii) Direct consequence of (i) by taking $x_0 = x - y$. □

7.4.2 Geometric version

We shall now prove the geometric form of the Hahn-Banach theorem, regarding the separation of convex sets. Observe that in this part we shall consider only normed vector spaces over $\mathbf{R}$.

Let $(X, \|\cdot\|)$ be a normed vector space over $\mathbf{R}$.

Definition 7.35. An *affine hyperplane* of X is a subset $H \subseteq X$ of the form

$$H = \{x \in X \mid f(x) = \alpha\}$$

for some linear functional $f : X \to \mathbf{R}$ that does not vanish identically and some constant $\alpha \in \mathbf{R}$.

Proposition 7.36. *The hyperplane $H = \{x \in X \mid f(x) = \alpha\}$ is closed if and only if f is continuous.*

Proof. The "if" part is trivial. Conversely, assume that H is closed. The complement set $X \setminus H$ is open and non-empty since f is not the null function. Thus consider $y \in X \setminus H$ and let $\delta > 0$ be such that $B(y, \delta) \subseteq X \setminus H$. Without loss of generality, we can suppose that $f(y) < \alpha$.

We will prove that for any $x \in B(y, \delta)$ one has $f(x) < \alpha$, and this implies that f is continuous by writing, for any $x \in B(0, 1)$,

$$f(x) = \frac{1}{\delta}\left(f(y + \delta x) - f(y)\right) \leqslant \frac{1}{\delta}(\alpha - f(y))$$

which gives $\|f\|_{X'} < \infty$.

By way of contradiction, we thus suppose that $f(z) \geqslant \alpha$ for some $z \in B(y, \delta)$. On the one hand, the line segment $[y, z] := \{(1-t)y + tz \mid 0 \leqslant t \leqslant 1\}$ is contained in $B(y, \delta)$ by convexity, hence $f((1-t)y + tz) \neq \alpha$ for all $0 \leqslant t \leqslant 1$. On the other hand, for $t = \frac{f(z)-\alpha}{f(z)-f(y)}$ we have $t \in [0, 1]$ and $f((1-t)y + tz) = \alpha$, thus a contradiction and the proof is complete. ☐

Given two subsets A and B of X, we say that the hyperplane $H = \{x \in X \mid f(x) = \alpha\}$ *separates* A and B if

$$A \subseteq H^- := \{x \in X \mid f(x) \leqslant \alpha\} \quad \text{and} \quad B \subseteq H^+ := \{x \in X \mid f(x) \geqslant \alpha\}.$$

Furthermore, we say that H *strictly separates* A and B if there is $\varepsilon > 0$ such that

$$A \subseteq \{x \in X \mid f(x) \leqslant \alpha - \varepsilon\} \quad \text{and} \quad B \subseteq \{x \in X \mid f(x) \geqslant \alpha + \varepsilon\}.$$

Theorem 7.37 (Geometric form of the Hahn-Banach theorem: first version)**.** *Let A and B be two non-empty convex subsets of a normed vector space X over* $\mathbf{R}$ *such that $A \cap B = \varnothing$. If A is open, then there exists a closed affine hyperplane separating A and B.*

Remark 7.38. If X is a normed vector space over $\mathbf{C}$, then the same result holds with a closed real affine hyperplane $H = \{x \in X \mid \operatorname{Re} f(x) = \alpha\}$ for some $f \in X' \setminus \{0\} = \mathscr{L}(X, \mathbf{C}) \setminus \{0\}$.

In the proof of the above theorem, we will need the following two lemmas.

Lemma 7.39. *Let C be a non-empty open convex subset of X containing* 0. *We define the* gauge *or the* Minkowski functional *of C as the map*

$$p_C(x) = \inf\left\{t > 0 \mid t^{-1}x \in C\right\}.$$

Then

(i) *for any $x \in X$ and $\lambda \in \mathbf{R}_+$ one has $p_C(\lambda x) = \lambda p_C(x)$;*

(ii) *there is $M > 0$ such that $0 \leqslant p_C(x) \leqslant M\|x\|$ for any $x \in X$;*

(iii) $C = \{x \in X \mid p_C(x) < 1\}$;

(iv) *for any $x, y \in X$ one has $p_C(x + y) \leqslant p_C(x) + p_C(y)$;*

Proof. (i) Trivial.

(ii) Take $\varepsilon > 0$ such that $B(0, \varepsilon) \subseteq C$, then $p_C(x) \leqslant \varepsilon^{-1}\|x\|$ for any $x \in X$.

(iii) Let $x \in C$ then there is $\varepsilon > 0$ small enough such that $x + \varepsilon x \in C$. Thus $p_C(x) \leqslant (1 + \varepsilon)^{-1} < 1$.

Conversely let $x \in X$ be such that $p_C(x) < 1$. Then there is $0 < t < 1$ such that $t^{-1}x \in C$, and hence $x = t(t^{-1}x) + (1 - t)0$ is a convex combination of elements of C, which implies that $x \in C$.

(iv) Let $x, y \in X$ and $\varepsilon > 0$. From (i) and (iii) we get

$$\frac{x}{p_C(x) + \varepsilon} \in C \quad \text{and} \quad \frac{y}{p_C(y) + \varepsilon} \in C.$$

Since C is convex we obtain that

$$\frac{x + y}{p_C(x) + p_C(y) + 2\varepsilon} \in C,$$

and using (i) and (iii) once more we deduce that $p(x + y) < p(x) + p(y) + 2\varepsilon$. We conclude the proof since $\varepsilon > 0$ is arbitrary. □

Lemma 7.40. *Let C be a non-empty open convex subset of X and let $y \in X \setminus C$. Then there exists $f \in X'$ such that $f(x) < f(y)$ for any $x \in C$, in particular the hyperplane $H = \{x \in X \mid f(x) = f(y)\}$ separates $\{y\}$ and C.*

Proof. By translations we can assume that C contains 0. Let p_C be the gauge of C, $V = \mathbf{R}y$ be the vector subspace generated by y and consider the linear functional $\varphi : V \to \mathbf{R}$, $\lambda y \mapsto \lambda$ for any $\lambda \in \mathbf{R}$. For any $t \in \mathbf{R}$ we have $\varphi(\lambda y) \leqslant p_C(\lambda y)$, indeed if $\lambda \leqslant 0$ we have $\varphi(\lambda y) = \lambda \leqslant 0 \leqslant p_C(\lambda y)$, otherwise if $\lambda > 0$ then $\varphi(\lambda y) = \lambda \leqslant \lambda p_C(y) = p_C(\lambda y)$ where we have used $p_C(y) \geqslant 1$ because $y \notin C$.

Therefore, from Hahn-Banach theorem (Theorem 7.32) we deduce the existence of a linear functional $f : X \to \mathbf{R}$ that extends φ and verifies

$$f(x) \leqslant p_C(x), \quad \forall x \in X.$$

In particular we obtain $f(y) = 1$ and Lemma 7.39 implies that f is continuous and that $f(x) < 1$ for any $x \in C$. □

We can now prove the first version of the geometric form of the Hahn-Banach theorem.

Proof of Theorem 7.37. Define the set $C = A - B := \{x - y : x \in A,\ y \in B\}$. We deduce that C is convex because A and B are convex; C is open since $C = \bigcup_{y \in B}(A - y)$ and $A - y$ is open for any $y \in B$ because it is a translation of A which is open; and C does not contain 0 since A and B are disjoint.

By Lemma 7.40 we there exists a continuous linear form $f \in X'$ such that $f(z) < 0$ for any $z \in C$, which implies $f(x) < f(y)$ for all $x \in A$ and $y \in B$. Take $\alpha \in \mathbf{R}$ such that $\sup_{x \in A} f(x) \leqslant \alpha \leqslant \inf_{y \in B} f(y)$, therefore the hyperplane $H = \{x \in X \mid f(x) = \alpha\}$ separates A and B. □

We prove now a second version of the geometric form of the Hahn-Banach theorem.

Theorem 7.41 (Geometric form of the Hahn-Banach theorem: second version)**.** *Let A and B be two non-empty convex subsets of a normed vector space X over $\mathbf{R}$ such that $A \cap B = \varnothing$. Assume that A is closed and B compact, then there exists a closed affine hyperplane strictly separating A and B.*

Remark 7.42. If X is a normed vector space over $\mathbf{C}$, then the same result holds with a closed real affine hyperplane $H = \{x \in X \mid \operatorname{Re} f(x) = \alpha\}$ for some $f \in X' \setminus \{0\} = \mathscr{L}(X, \mathbf{C}) \setminus \{0\}$.

Proof of Theorem 7.41. Define the set $C = A - B$ so that C is open, convex, and does not contain 0. There is $\delta > 0$ such that $B(0, \delta) \cap C$ is empty. Theorem 7.37 implies that there exists a closed hyperplane $H = \{x \in X \mid f(x) = \alpha\}$, with $f \in X' \setminus \{0\}$ and $\alpha \in \mathbf{R}$, that separates $B(0, \delta)$ and C, that is

$$B(0, \delta) \subseteq \{x \in X \mid f(x) \leqslant \alpha\} \quad \text{and} \quad C \subseteq \{x \in X \mid f(x) \geqslant \alpha\}.$$

It follows that for any $z \in B(0, 1)$, $x \in A$ and $y \in B$ we have $f(\delta z) \leqslant f(x - y)$ which implies $\delta \|f\|_{X'} \leqslant f(x) - f(y)$. Taking $\varepsilon = \frac{1}{2}\delta \|f\|_{X'} > 0$ yields $f(y) + \varepsilon \leqslant f(x) - \varepsilon$ for all $x \in A$ and $y \in B$. Choosing now $\alpha' \in \mathbf{R}$ such that $\sup_{y \in B} f(y) \leqslant \alpha \leqslant \inf_{x \in A} f(x)$, the hyperplane hyperplane $H' = \{x \in X \mid f(x) = \alpha'\}$ strictly separates A and B. □

We conclude this section with a consequence of the geometric form of the Hahn-Banach theorem that provides a simple way to determine whether a subspace is dense.

Corollary 7.43. *Let Y be a vector subspace of a normed vector space X over $\mathbf{R}$ such that $\overline{Y} \neq X$. Then there exists $f \in X' \setminus \{0\}$ such that $f(x) = 0$ for any $x \in Y$.*

Remark 7.44. For normed vector spaces over $\mathbf{C}$ the conclusion is the same.

Proof of Corollary 7.43. Let $y \in X \setminus \overline{Y}$. The Hahn-Banach theorem (Theorem 7.41) implies that there is a closed hyperplane $H = \{x \in X \mid f(x) = \alpha\}$, with $f \in X' \setminus \{0\}$ and $\alpha \in \mathbf{R}$, that strictly separates $\overline{Y}$ and $\{y\}$. Therefore one obtains

$$f(x) < \alpha < f(y), \quad \forall x \in Y.$$

Since $0 \in Y$ we have in particular that $\alpha > 0$. Now let $x \in Y$. Since Y is a vector subspace, for any $\lambda \in \mathbf{R}$ we get $\lambda f(x) = f(\lambda x) < \alpha$, which implies $f(x) = 0$. □

7.5 Weak topologies

Let $(X, \|\cdot\|_X)$ be a normed vector space and consider its dual space X' defined in Definition 7.19. For any $f \in X'$ and $x \in X$ we denote $\langle f, x\rangle = f(x)$.

We have already seen that we can endow X with the topology associated to the norm $\|\cdot\|_X$, that we call *strong topology* (or *norm topology*). We shall define now another topology on X.

Definition 7.45. The *weak topology* on X, denoted by $\sigma(X, X')$, is the initial topology on X associated to the family of maps $\{f\}_{f\in X'}$, that is, it is the coarsest topology on X for which all f remains continuous, where f ranges over X'.

Thanks to the material developed in Section 2.3.1, we know that a subbasis of this topology is given by the sets of the form

$$U(x, f, \varepsilon) = \left\{y \in X : |\langle f, x - y\rangle| < \varepsilon\right\}$$

for $x \in X$, $f \in X'$ and $\varepsilon > 0$. Moreover, for any point $x_0 \in X$, the family of subsets given by

$$V(f_1, \ldots, f_k, \varepsilon) = \left\{x \in X : |\langle f_i, x - x_0\rangle| < \varepsilon, \forall i = 0, \ldots, k\right\},$$

where $\varepsilon > 0$, $k \in \mathbf{N}$ and $f_1, \ldots, f_k \in X'$, is a neighborhood basis for x_0 with respect to the weak topology $\sigma(X, X')$ on X.

By definition, we observe that every map $f \in X'$ is continuous for the strong topology, therefore the weak topology on X is coarser than the strong topology on X.

Remark 7.46. In a finite-dimensional normed vector space the weak topology and the strong topology are the same. However, in a infinite-dimensional normed vector space the weak topology is strictly coarser than the strong topology. For instance, if $(X, \|\cdot\|_X)$ is infinite-dimensional then:

- The closure of the unit sphere $S = \{x \in X : \|x\|_X = 1\}$ in the weak topology is the closed unit ball $\overline{B}(0, 1) = \{x \in X : \|x\|_X \leqslant 1\}$.
- The open unit ball $B(0, 1) = \{x \in X : \|x\|_X < 1\}$ is not open in the weak topology.

We shall say that a subset $A \subseteq X$ is *weakly open* (resp. *weakly closed*) if it is open (resp. closed) in the weak topology $\sigma(X, X')$; and is *strongly open* (resp. *strongly closed*) if it is open (resp. closed) in the strong topology of X.

We shall now prove some basic properties of the weak topology.

Proposition 7.47. *The topological space $(X, \sigma(X, X'))$ is Hausdorff.*

Proof. Let $x, y \in X$ be distinct. Thanks to the second form of the geometric Hahn-Banach theorem (Theorem 7.41) applied to the sets $\{x\}$ and $\{y\}$, we deduce that there exists $f \in X'$ such that $\operatorname{Re} f(x) \neq \operatorname{Re} f(y)$, thus $f(x) \neq f(y)$. Since $\mathbf{K}$ is Hausdorff, there are disjoint open sets U and V in $\mathbf{K}$ containing $f(x)$ and $f(y)$ respectively. Therefore the sets $f^{-1}(U)$ and $f^{-1}(V)$ are disjoint open sets for the weak topology containing respectively x and y. □

We say that a sequence $(x_n)_{n\in\mathbf{N}}$ in X converges *strongly* to x in X if it converges with respect to the topology associated to the norm $\|\cdot\|_X$, which is denoted by $x_n \to x$; and converges *weakly* to x in X if it converges with respect to the weak topology $\sigma(X, X')$, in which case we denote $x_n \rightharpoonup x$.

Proposition 7.48. *Let $(x_n)_{n\in\mathbf{N}}$ be a sequence in X and $x \in X$.*

(i) *$x_n \rightharpoonup x$ weakly in X if and only if $\langle f, x_n\rangle \to \langle f, x\rangle$ for all $f \in X'$.*

(ii) *If $x_n \to x$ strongly in X, then $x_n \rightharpoonup x$ weakly in X.*

(iii) *If $x_n \rightharpoonup x$ weakly in X, then $(\|x_n\|_X)_{n\in\mathbf{N}}$ is bounded and $\|x\|_X \leqslant \liminf_{n\to\infty} \|x_n\|_X$.*

(iv) *If $x_n \rightharpoonup x$ weakly in X and $f_n \to f$ strongly in X', then $\langle f_n, x_n\rangle \to \langle f, x\rangle$.*

Proof. (i) Assume that $x_n \rightharpoonup x$ weakly in X. Since each $f \in X'$ is continuous, we deduce $\langle f, x_n\rangle \to \langle f, x\rangle$.

Conversely, assume that $\langle f, x_n\rangle \to \langle f, x\rangle$ for all $f \in X'$. Let U be a neighborhood of x for the weak topology $\sigma(X, X')$. By definition of the initial topology, there is a finite set $\{f_1, \dots, f_k\} \subseteq X'$ and $\varepsilon > 0$ such that $x \in \bigcap_{i=1,\dots,k}\{y \in X \mid |\langle f_i, x\rangle - \langle f_i, y\rangle| < \varepsilon\} \subseteq U$. For any $i = 1, \dots, k$ there is an integer $N_i \in \mathbf{N}$ such that $|\langle f_i, x_n\rangle - \langle f_i, x\rangle| < \varepsilon$ for all $n \geqslant N_i$. Therefore, defining $N = \max_{i=1,\dots,k} N_i \in \mathbf{N}$, we deduce that $x_n \in U$ for all $n \geqslant N$.

(ii) It is a direct consequence to the fact that weak topology on X is coarser than the strong topology on X.

(iii) For every $f \in X'$ the sequence $(\langle f, x_n\rangle)_{n\in\mathbf{N}}$ is bounded in $\mathbf{K}$, hence the Banach-Steinhaus theorem (Theorem 7.28) implies that $(\|x_n\|_X)_{n\in\mathbf{N}}$ is bounded. Moreover we have $|\langle f, x_n\rangle| \leqslant \|f\|_{X'}\|x_n\|_X$ for all $n \in \mathbf{N}$, thus taking the limit $n \to \infty$ yields $|\langle f, x\rangle| \leqslant \|f\|_{X'} \liminf_{n\to\infty} \|x_n\|_X$. We then conclude thanks to Corollary 7.34–(ii).

(iv) We write, for any $n \in \mathbf{N}$,

$$\begin{aligned} |\langle f_n, x_n\rangle - \langle f, x\rangle| &\leqslant |\langle f_n - f, x_n\rangle| + |\langle f, x_n - x\rangle| \\ &\leqslant \|f_n - f\|_{X'}\|x_n\|_X + |\langle f, x_n - x\rangle|, \end{aligned}$$

and conclude thanks to (i) and (iii). □

Theorem 7.49. *Let C be a convex set in X, then C is weakly closed if and only if it is strongly closed.*

Proof. We already know that if C is weakly closed then it is also strongly closed, since the weak topology is coarser than the strong topology.

Conversely, assume that C is strongly closed. We shall prove that its complement $X \setminus C$ is weakly open. Let $x_0 \in X \setminus C$, then by the Hahn-Banach theorem (Theorem 7.41) there is $f \in X'$ and $\alpha \in \mathbf{R}$ such that

$$\operatorname{Re} f(x_0) < \alpha < \operatorname{Re} f(x), \quad \forall x \in C.$$

The set $U = \{x \in X : \operatorname{Re} f(x) < \alpha\}$ is weakly open and verifies $x_0 \in U \subseteq X \setminus C$. □

Theorem 7.50. *Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be Banach spaces and $T : X \to Y$ a linear map. Then T is continuous from $(X, \|\cdot\|_X)$ into $(Y, \|\cdot\|_Y)$ if and only if T is continuous from $(X, \sigma(X, X'))$ into $(Y, \sigma(Y, Y'))$.*

Proof. Suppose that T is continuous from $(X, \|\cdot\|_X)$ into $(Y, \|\cdot\|_Y)$. Thanks to the definition of the weak topology as an initial topology (see Proposition 2.9), it is sufficient to check that for every $f \in X'$ the map $x \mapsto f(Tx)$ is continuous from $(X, \sigma(X, X'))$ into $\mathbf{K}$. Since the map $x \mapsto f(Tx)$ is continuous on $(X, \|\cdot\|_X)$ it is also continuous in $(X, \sigma(X, X'))$.

Conversely, assume that T is continuous from $(X, \sigma(X, X'))$ into $(Y, \sigma(Y, Y'))$. Then its graph $G(T) = \{(x, Tx) : x \in X\}$ is closed in $X \times Y$ endowed with the product topology of the weak topologies $\sigma(X, X') \times \sigma(Y, Y')$. But this topology clearly coincides with the weak topology $\sigma(X \times Y, (X \times Y)')$. This means that $G(T)$ is weakly closed and thus also strongly closed. By the closed graph theorem (Theorem 7.31) we deduce that T is continuous from $(X, \|\cdot\|_X)$ into $(Y, \|\cdot\|_Y)$. □

7.5.1 Weak topologies on a dual space

We are now interested on topologies of the dual space X'. We denote by $X'' = \mathscr{L}(X', \mathbf{K})$ the space of continuous linear functionals on X' (endowed with the topology associated to the norm $\|\cdot\|_{X'} = \|\cdot\|_{\mathscr{L}(X,\mathbf{K})}$), which is also called the bidual of X. For any $\xi \in X''$ and $f \in X'$ we denote $\langle \xi, f \rangle = \xi(f)$. We have already defined two topologies on X':

- The strong topology (or norm topology) associated to the norm $\|\cdot\|_{X'}$.
- The *weak topology* $\sigma(X', X'')$ defined in Definition 7.45 above, namely the initial topology associated to the family of maps $\{\xi\}_{\xi \in X''}$, that is, the coarsest topology on X' for which all ξ remains continuous, where ξ ranges over X''. A subbasis of this topology is given by the sets of the form

$$U(f, \xi, \varepsilon) = \left\{ g \in X' : |\langle \xi, f - g \rangle| < \varepsilon \right\}$$

for $f \in X'$, $\xi \in X''$ and $\varepsilon > 0$.

For every $x \in X$ we define the linear functional ev_x on X' by

$$\begin{aligned} \mathrm{ev}_x : X' &\to \mathbf{K} \\ f &\mapsto f(x), \end{aligned}$$

so that we have a canonical injection $X \subseteq X''$ via the map $x \mapsto \mathrm{ev}_x$. We can therefore consider a third topology on X':

Definition 7.51. The weak-$\star$ topology on X', denoted by $\sigma(X', X)$, is the initial topology on X' associated to the family of maps $\{\mathrm{ev}_x\}_{x\in X}$, that is, it is the coarsest topology on X' for which all ev_x remains continuous, where x ranges over X.

Thanks to Section 2.3.1, a subbasis of this topology is given by the sets of the form

$$\mathscr{U}(x, f, \varepsilon) = \left\{g \in X' : |\langle f - g, x\rangle| < \varepsilon\right\}$$

for $x \in X$, $f \in X'$ and $\varepsilon > 0$. Moreover, for any $f_0 \in X'$, the family of subsets given by

$$\mathscr{V}(x_1, \dots, x_k, \varepsilon) = \left\{f \in X' : |\langle f - f_0, x_i\rangle| < \varepsilon, \forall i = 0, \dots, k\right\},$$

where $\varepsilon > 0$, $k \in \mathbf{N}$ and $x_1, \dots, x_k \in X$, is a neighborhood basis for f_0 with respect to the weak-$\star$ topology $\sigma(X', X)$ on X'. Since $X \subseteq X''$, the weak-$\star$ topology $\sigma(X', X)$ on X' is coarser than the weak topology $\sigma(X', X'')$ on X'.

We shall now prove some basic properties of the weak-$\star$ topology.

Proposition 7.52. *The topological space $(X', \sigma(X', X))$ is Hausdorff.*

Proof. Let $f, g \in X'$ be distinct, thus there is $x \in X$ such that $f(x) \neq g(x)$. Since $\mathbf{K}$ is Hausdorff, there are disjoint open sets U and V in $\mathbf{K}$ containing $f(x)$ and $g(x)$ respectively. Hence the sets $\mathrm{ev}_x^{-1}(U)$ and $\mathrm{ev}_x^{-1}(V)$ are disjoint open sets for the weak-$\star$ topology containing respectively f and g. $\square$

We say that a sequence $(f_n)_{n\in\mathbf{N}}$ in X' converges *weakly-$\star$* to f in X' if it converges with respect to the weak-$\star$ topology $\sigma(X', X)$, in which case we denote $f_n \overset{\star}{\rightharpoonup} f$.

Proposition 7.53. *Let $(f_n)_{n\in\mathbf{N}}$ be a sequence in X' and $f \in X'$.*

(i) *$f_n \overset{\star}{\rightharpoonup} x$ weakly-$\star$ in X' if and only if $\langle f_n, x\rangle \longrightarrow \langle f, x\rangle$ for all $x \in X$.*

(ii) *If $f_n \longrightarrow f$ strongly in X', then $f_n \rightharpoonup f$ weakly in X'. Moreover, if $f_n \rightharpoonup f$ weakly in X' then $f_n \overset{\star}{\rightharpoonup} f$ weakly-$\star$ in X'.*

(iii) *If $f_n \overset{\star}{\rightharpoonup} f$ weakly-$\star$ in X', then $(\|f_n\|_X)_{n\in\mathbf{N}}$ is bounded and $\|f\|_{X'} \leqslant \liminf_{n\to\infty} \|f_n\|_{X'}$.*

(iv) *If $f_n \overset{\star}{\rightharpoonup} f$ weakly-$\star$ in X' and $x_n \longrightarrow x$ strongly in X, then $\langle f_n, x_n\rangle \longrightarrow \langle f, x\rangle$.*

Proof. The proof is exactly the same as the proof of Proposition 7.48. $\square$

7.6 Integration of vector-valued functions

Let $[a, b]$ be an interval of $\mathbf{R}$, and $(Y, \|\cdot\|_Y)$ be a Banach space. We denote by $\mathscr{B}([a, b]; Y)$ the Banach space of bounded Y-valued functions defined on $[a, b]$ endowed with the norm

$$\|f\|_\infty = \sup_{t\in[a,b]} \|f(t)\|_Y,$$

and by $\mathscr{C}([a, b]; Y)$ the subspace of continuous functions.

The aim of this section is to define the integral of (a class of) functions $f : [a,b] \to Y$. We recall that a *finite partition* of the interval $[a,b]$ is a collection $\sigma = \{t_0, \dots, t_{n+1}\}$, with $n \in \mathbf{N}$, such that $a = t_0 < t_1 < \cdots < t_n < t_{n+1} = b$.

We start by defining the integral of step functions.

Definition 7.54. One says that a map $f : [a,b] \to Y$ is a *step function* if there is a partition $\sigma = \{t_0, \dots, t_{n+1}\}$ of $[a,b]$ and $y_0, \dots, y_n \in Y$ such that, for all $k = 0, \dots, n$, one has

$$f(t) = y_k, \quad \text{for all } t \in (t_k, t_{k+1}).$$

One then defines the *integral of f on* $[a,b]$ by

$$\int_a^b f(t)\,\mathrm{d}t := \sum_{k=0}^{n} y_k(t_{k+1} - t_k). \tag{7.4}$$

One can check that this definition does not depend on the choice of partition, that is, if σ' is another finite partition of $[a,b]$ such that f is constant on the open intervals of σ', then the value of (7.4) associated to σ coincides with the value of (7.4) associated to σ'.

Denote by $S([a,b];Y) \subseteq \mathscr{B}([a,b];Y)$ the set of all step functions from $[a,b]$ to Y. The map

$$\begin{array}{rccc} \mathcal{I}: & S([a,b];Y) & \to & Y \\ & f & \mapsto & \displaystyle\int_a^b f(t)\,\mathrm{d}t \end{array}$$

is linear and continuous. Moreover we easily obtain from the definition the following properties:

Proposition 7.55. *For any step function $f \in S([a,b];Y)$ there holds:*

(i) *Let $a < c < b$ then*

$$\int_a^b f(t)\,\mathrm{d}t = \int_a^c f(t)\,\mathrm{d}t + \int_c^b f(t)\,\mathrm{d}t.$$

(ii) *We have*

$$\left\| \int_a^b f(t)\,\mathrm{d}t \right\|_Y \leqslant \int_a^b \|f(t)\|_Y \,\mathrm{d}t \leqslant (b-a)\|f\|_\infty.$$

We now define the integral of a larger class of functions by extending the continuous linear map $\mathcal{I}$.

Definition 7.56. One says that a map $f : [a,b] \to Y$ is a *regulated function* if it belongs to $\overline{S([a,b];Y)}$, the closure of $S([a,b];Y)$ in $\mathscr{B}([a,b];Y)$.

In other words, f is a regulated function if it is the uniform limit of a sequence of step functions. One can check that f is a regulated function if and only if

f admits left-side limits on $[a,b)$ and right-side limits on $(a,b]$. In particular $\mathscr{C}([a,b];Y) \subseteq \overline{S([a,b];Y)}$.

Since $\mathcal{I}$ is a continuous linear map from S into Y, by the bounded linear transformation theorem (Theorem 7.18) we can extend $\mathcal{I}$ uniquely to a linear continuous map from $\overline{S([a,b];Y)}$ to Y. In particular if $f \in \overline{S([a,b];X)}$ then its integral is defined by

$$\int_a^b f(t)\,\mathrm{d}t := \lim_{n\to\infty} \int_a^b f_n(t)\,\mathrm{d}t,$$

where $(f_n)_{n\in\mathbf{N}}$ is a sequence of step functions that converges uniformly to f, and we recall that, by the proof of Theorem 7.18, this limit is independent of the chosen sequence. Thanks to this procedure, properties (i) and (ii) above remain true for $f \in \overline{S([a,b];Y)}$.

Chapter 8

Differentiable maps

Through this chapter, let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be Banach spaces over $\mathbf{K}$. When there is no confusion, we omit the subscripts X and Y of the norms.

8.1 Fréchet-differentiable maps

Let $U \subseteq X$ be an open set.

Definition 8.1. A map $f : U \longrightarrow Y$ is *Fréchet-differentiable at a point* $a \in U$ if there exists a continuous linear map $L \in \mathscr{L}(X, Y)$ and a neighborhood V of 0 in X such that $a + V \subseteq U$ and

$$f(a+h) = f(a) + Lh + \|h\|\,\varepsilon(h), \quad \text{for all } h \in V,$$

where $\varepsilon : V \longrightarrow Y$ satisfies $\lim_{h\to 0} \varepsilon(h) = 0$.

Proposition 8.2. *The map L in Definition 8.1 is unique.*

Proof. Assume that there are two maps $L_1, L_2 \in \mathscr{L}(X, Y)$ and two neighborhoods V_1, V_2 of 0 in X such that $a + V_1 \subseteq U$, $a + V_2 \subseteq U$ and

$$\begin{aligned} f(a+h) &= f(a) + L_1 h + \|h\|\,\varepsilon_1(h), \quad \text{for all } h \in V_1, \\ f(a+h) &= f(a) + L_2 h + \|h\|\,\varepsilon_2(h), \quad \text{for all } h \in V_2, \end{aligned}$$

where $\varepsilon_1 : V_1 \longrightarrow Y$ and $\varepsilon_2 : V_2 \longrightarrow Y$ are maps satisfying $\lim_{h\to 0} \varepsilon_1(h) = 0$ and $\lim_{h\to 0} \varepsilon_2(h) = 0$. Then $V_1 \cap V_2$ is a neighborhood of 0 in X and one has

$$L_1 h - L_2 h = \|h\|(\varepsilon_2(h) - \varepsilon_1(h)), \quad \text{for all } h \in V_1 \cap V_2.$$

This implies that, for any $\eta > 0$, there is $\delta > 0$ such that for any $h \in V_1 \cap V_2$, if $\|h\| \leqslant \delta$ then one has

$$\|(L_1 - L_2)(h)\| \leqslant \eta \|h\|.$$

By homogeneity of the norm, this implies that $\|L_1 - L_2\| \leqslant \eta$, and we conclude that $L_1 = L_2$ since $\eta > 0$ is arbitrary. □

Definition 8.3. The map L in Definition 8.1 is called the *Fréchet derivative of f at the point a*, and it is denoted by $f'(a) = L$.

Remark 8.4. The differentiability of f and the value of $f'(a)$ only depend on the topologies of X and Y. In particular, they do not change if we replace the norms of X and Y by equivalent ones.

Definition 8.5. One says that a map $f : U \to Y$ is:

(i) *Fréchet-differentiable on* U if f is Fréchet-differentiable at every point of U. In that case we define the *Fréchet derivative* of f, denoted by f', as the map

$$\begin{array}{cccc} f' : & U & \to & \mathscr{L}(X,Y) \\ & a & \mapsto & f'(a) \end{array} .$$

(ii) *Continuously differentiable* or *of class* $\mathscr{C}^1$ on U if f is Fréchet-differentiable on U and the Fréchet derivative f' is continuous.

We denote by $\mathscr{C}^1(U,Y)$ the vector space of all continuously differentiable maps on U.

Hereafter, when there is no possible misunderstanding, *Fréchet differentiable* shall be referred to simply as *differentiable*, and *Fréchet derivative* to *derivative*.

8.1.1 Gateaux-differentiable maps

Let $U \subseteq X$ be an open set.

Definition 8.6. Let $f : U \to Y$ be a map, one says that:

(i) f admits a *directional derivative along* $h \in X$ *at a point* $a \in U$ if

$$D_h f(a) := \lim_{t\to 0} \frac{f(a+th) - f(a)}{t} \quad \text{exists in } Y.$$

(ii) f is *Gateaux*[1]*-differentiable at* $a \in U$ if f admits a directional derivative along any vector $h \in X$ at the point a, and if the map $D_{\mathrm{G}} f(a) : h \mapsto D_h f(a)$ from X to Y is linear and continuous. The map $D_{\mathrm{G}} f(a) \in \mathscr{L}(X,Y)$ is called the *Gateaux-derivative* of f at the point a.

(iii) f is *Gateaux-differentiable on* U if f is Gateaux-differentiable at every point of U.

It is easy to verify that if f is Fréchet-differentiable at a point a, then f is Gateaux-differentiable at a and

$$f'(a)(h) = D_h f(a), \quad \text{for all } h \in X,$$

but the converse is however not true. We also remark that Gateaux-differentiability does not imply continuity.

Example 8.7. (1) A constant map $f : U \to Y$ is differentiable on U with $f'(a) = 0$ for any $a \in U$.

[1] René Gateaux (1889–1914) was a French mathematician who contributed to functional calculus.

(2) If $f : U \to Y$ is the restriction to U of a continuous linear map $T \in \mathscr{L}(X, Y)$, then f is differentiable on U with $f'(a) = T$ for any $a \in U$.

(3) When $X = \mathbf{R}$, then $f'(a) \in \mathscr{L}(\mathbf{R}, Y)$ is identified with the vector $\lambda = f'(a)(1) \in Y$, through the canonical isomorphism $\mathscr{L}(\mathbf{R}, Y) \simeq Y$, where

$$\lambda := \lim_{x \to a} \frac{f(x) - f(a)}{x - a}$$

is the usual derivative for functions defined on $\mathbf{R}$.

8.2 Operations on differentiable maps

We prove some basic results regarding operations on differentiable maps. The first one shows that the derivative is linear.

Proposition 8.8. *Let $U \subseteq X$ be an open set.*

(i) *If $f : U \to Y$ is differentiable at $a \in U$ and $\lambda \in \mathbf{K}$, then the map $\lambda f : x \mapsto \lambda f(x)$ from U to Y is differentiable at a, with $(\lambda f)'(a) = \lambda f'(a)$.*

(ii) *If $f, g : U \to Y$ are differentiable at $a \in U$, then the map $f + g : x \mapsto f(x) + g(x)$ from U to Y is differentiable at a, with $(f + g)'(a) = f'(a) + g'(a)$.*

Proof. (i) There is V a neighborhood of 0 in X such that $a + V \subseteq U$ and

$$f(a + h) = f(a) + f'(a)(h) + \|h\|\, \varepsilon(h), \quad \text{for all } h \in V,$$

where $\varepsilon : V \to Y$ satisfies $\lim_{h \to 0} \varepsilon(h) = 0$. Then for any $h \in V$ there holds

$$(\lambda f)(a + h) = \lambda f(a + h) = (\lambda f)(a) + \lambda f'(a)(h) + \|h\| \lambda \varepsilon(h),$$

and one has $\lim_{h \to 0} \lambda \varepsilon(h) = 0$. The map $X \ni h \mapsto \lambda f'(a)(h) \in Y$ is clearly linear and continuous, thus the proof is complete.

(ii) There are V_1, V_2 neighborhoods of 0 in X such that $a + V_1 \subseteq U$, $a + V_2 \subseteq U$, and

$$\begin{aligned} f(a + h) &= f(a) + f'(a)(h) + \|h\|\, \varepsilon_1(h), \quad \text{for all } h \in V_1, \\ g(a + h) &= g(a) + g'(a)(h) + \|h\|\, \varepsilon_2(h), \quad \text{for all } h \in V_2, \end{aligned}$$

where $\varepsilon_1 : V_1 \to Y$ and $\varepsilon_2 : V_2 \to Y$ verify $\lim_{h \to 0} \varepsilon_1(h) = 0$ and $\lim_{h \to 0} \varepsilon_2(h) = 0$. For any $h \in V_1 \cap V_2$, which is a neighborhood of 0 on X, there holds

$$\begin{aligned} (f + g)(a + h) &= f(a + h) + g(a + h) \\ &= (f + g)(a) + f'(a)(h) + g'(a)(h) + \|h\|(\varepsilon_1(h) + \varepsilon_2(h)), \end{aligned}$$

and one has $\lim_{h \to 0}(\varepsilon_1(h) + \varepsilon_2(h)) = 0$. The map $X \ni h \mapsto f'(a)(h) + g'(a)(h) \in Y$ is clearly linear and continuous, thus the proof is complete. □

In the next result we show that the composition of differentiable maps is also differentiable.

Proposition 8.9. *Let X, Y, Z be Banach spaces, U an open subset of X, and V an open subset of Y. Let $f : U \to Y$ and $g : V \to Z$ be two maps. Assume that f is differentiable at $a \in U$ and g is differentiable at $b = f(a) \in V$. Then the composition map $g \circ f : U \to Z$ is differentiable at a and*

$$(g \circ f)'(a) = g'(f(a)) \circ f'(a).$$

Proof. By hypothesis there are $\delta, \delta' > 0$ such that

$$\begin{aligned} f(a+h) &= f(a) + f'(a)(h) + \|h\|\,\varepsilon(h), \quad \text{for all } h \in B_X(0,\delta),\\ g(b+k) &= g(b) + g'(b)(k) + \|k\|\eta(k), \quad \text{for all } k \in B_Y(0,\delta'), \end{aligned}$$

with $\varepsilon : B_X(0,\delta) \to Y$ and $\eta : B_Y(0,\delta') \to Z$ satisfying $\lim_{h\to 0} \varepsilon(h) = 0$ and $\lim_{k\to 0} \eta(k) = 0$. We set $k = f(a+h) - f(a) = f'(a)(h) + \|h\|\,\varepsilon(h)$ so that

$$\|k\| \leqslant (\|f'(a)\| + \|\varepsilon(h)\|)\|h\|.$$

Let $\delta > 0$ be small enough such that $\|h\| < \delta$ implies $\|k\| < \delta'$. Hence one obtains

$$\begin{aligned} g \circ f(a+h) - g \circ f(a) &= g(b+k) - g(b)\\ &= g'(b)(k) + \|k\|\eta(k)\\ &= g'(b) \circ f'(a)(h) + \|h\| g'(b)(\varepsilon(h)) + \big\|f'(a)(h) + \|h\|\,\varepsilon(h)\big\|\,\eta\left(f'(a)(h) + \|h\|\,\varepsilon(h)\right)\\ &=: g'(b) \circ f'(a)(h) + \|h\|\overline{\varepsilon}(h), \end{aligned}$$

and one observes that $\lim_{h\to 0} \overline{\varepsilon}(h) = 0$ since

$$\|\overline{\varepsilon}(h)\| \leqslant \|g'(b)\|\|\varepsilon(h)\| + (\|f'(a)\| + \|\varepsilon(h)\|)\|\eta(f'(a)(h) + \|h\|\,\varepsilon(h))\| \xrightarrow[h\to 0]{} 0,$$

where we have used the composition of limits in last term. □

We also obtain below a result concerning the differentiability of the inverse map.

Proposition 8.10. *Let U be an open set in X and V be an open set in Y. Let $f : U \to V$ be a homeomorphism and assume that f is differentiable at some point $a \in U$. Then the inverse map f^{-1} is differentiable at $b = f(a) \in V$ if and only if $f'(a) \in \mathrm{Isom}(X,Y)$ is a isomorphism from X into Y, and thus one has*

$$(f^{-1})'(b) = \left(f'(a)\right)^{-1}.$$

Proof. Assume that $g = f^{-1}$ is differentiable at b, then thanks to Proposition 8.9 one obtains

$$g'(b) \circ f'(a) = \mathrm{id}_X \quad \text{and} \quad f'(a) \circ g'(b) = \mathrm{id}_Y,$$

which implies that $f'(a)$ is a isomorphism from X into Y with inverse isomorphism given by $g'(b)$.

Conversely, assume that $f'(a) \in \mathrm{Isom}(X,Y)$, so in particular one has $c = \|(f'(a))^{-1}\| > 0$. We have

$$f(a+h) - f(a) - f'(a)(h) = \|h\|\,\varepsilon(h) \tag{8.1}$$

with $\lim_{h\to 0} \varepsilon(h) = 0$, thus

$$\|h\| = \|(f'(a))^{-1}(f'(a)(h))\| \leqslant \|(f'(a))^{-1}\| \left(\|f(a+h) - f(a)\| + \|h\|\varepsilon(h)\right).$$

Therefore, if $\|h\|$ is small enough so that $\|\varepsilon(h)\| < \frac{1}{2c}$, one has

$$\|h\| \leqslant 2c\|f(a+h) - f(a)\|.$$

Denote $b = f(a)$ and $k = f(a+h) - f(a)$, then applying $(f'(a))^{-1}$ to (8.1) we obtain

$$(f'(a))^{-1}(b+k) - (f'(a))^{-1}(b) - (f'(a))^{-1}(f'(a)(h)) = \|h\|(f'(a))^{-1}(\varepsilon(h)).$$

Remarking that $h = a + h - a = f^{-1}(b+k) - f^{-1}(b)$, for h small enough we get

$$\begin{aligned} &f^{-1}(b+k) - f^{-1}(b) - (f'(a))^{-1}(k) \\ &\quad = -\|f^{-1}(b+k) - f^{-1}(b)\| \, (f'(a))^{-1} \circ \varepsilon\left(f^{-1}(b+k) - f^{-1}(b)\right) \\ &\quad =: \|k\|\eta(k). \end{aligned}$$

It remains to check that $\lim_{k\to 0} \eta(k) = 0$ and the proof will be complete. Since $\|h\| \leqslant 2c\|k\|$ for h small enough, we obtain $\|\eta(k)\| \leqslant 2c^2\|\varepsilon(h)\|$ and we conclude by composition of limits since $k \longrightarrow 0$ implies $h \longrightarrow 0$. □

8.3 Maps with values in a product space

In this part we consider the case in which $Y = Y_1 \times \cdots \times Y_m$ is a finite product of Banach spaces. We endow Y with the with the norm

$$\|y\| = \max(\|y_1\|_{Y_1}, \ldots, \|y_m\|_{Y_m})$$

for any $y = (y_1, \ldots, y_m) \in Y$, and we recall that $(Y, \|\cdot\|)$ is a Banach space.

For any $1 \leqslant i \leqslant m$ we define the canonical projection

$$\begin{array}{rccl} p_i : & Y & \longrightarrow & Y_i \\ & (y_1, \ldots, y_m) & \longmapsto & y_i \end{array}$$

and the canonical injection

$$\begin{array}{rccl} \iota_i : & Y_i & \longrightarrow & Y \\ & y_i & \longmapsto & (0, \ldots, y_i, \ldots, 0) \end{array}$$

where there are 0's everywhere except at the i-th coordinate. One easily checks that p_i and ι_i are continuous linear maps and verify

$$p_i \circ \iota_i = \mathrm{id}_{Y_i} \quad \text{and} \quad \sum_{i=1}^{m} \iota_i \circ p_i = \mathrm{id}_Y.$$

We deduce the following result:

Proposition 8.11. *Let $U \subseteq X$ be an open set. A continuous map $f : U \to Y$ is differentiable at a point $a \in U$ if and only if the map $f_i = p_i \circ f : U \to Y_i$ is differentiable at a for any $1 \leqslant i \leqslant m$, and then we have*

$$f'(a) = \sum_{i=1}^{m} \iota_i \circ f_i'(a).$$

Proof. If f is differentiable at a, then each $f_i = p_i \circ f$ is differentiable at a as the composition of differentiable maps (see Proposition 8.9). Moreover we obtain that

$$f_i'(a) = p_i \circ f'(a) \in \mathscr{L}(X, Y_i).$$

Conversely, assume that every $f_i = p_i \circ f$ is differentiable at a. From the above relation we get

$$f = \sum_{i=1}^{m} \iota_i \circ p_i \circ f = \sum_{i=1}^{m} \iota_i \circ f_i,$$

thus f is differentiable at a as the sum of composition of differentiable maps (see Propositions 8.8 and 8.9) and moreover

$$f'(a) = \sum_{i=1}^{m} \iota_i \circ f_i'(a).$$

□

8.4 Partial derivatives

We consider now the case where $X = X_1 \times \cdots \times X_n$ is a finite product of Banach spaces. We endow X with the norm

$$\|x\| = \max(\|x_1\|_{X_1}, \ldots, \|x_n\|_{X_n})$$

for any $x = (x_1, \ldots, x_n) \in X$, for which X is a Banach space.

For each $a = (a_1, \ldots, a_n) \in X$ we define the injection, for any $i \in \{1, \ldots, n\}$,

$$\begin{array}{cccc} \lambda_i : & X_i & \to & X \\ & x_i & \mapsto & (a_1, \ldots, a_{i-1}, x_i, a_{i+1}, \ldots, a_n) \end{array}.$$

Let $f : U \to Y$ be a map, where U is an open subset of X. The composition map $f \circ \lambda_i$, called the i-th partial map at the point a, is then defined on $\lambda_i^{-1}(U) \subseteq X_i$, more precisely one has

$$\begin{array}{cccc} f \circ \lambda_i : & \lambda_i^{-1}(U) & \to & Y \\ & x_i & \mapsto & f(a_1, \ldots, a_{i-1}, x_i, a_{i+1}, \ldots, a_n) \end{array}.$$

Definition 8.12. Let $i \in \{1, \ldots, n\}$. One says that $f : U \to Y$ is *differentiable with respect to the i-th variable at the point $a \in U$* if the map $f \circ \lambda_i$ is differentiable at a_i. We then denote this derivative by

$$\partial_i f(a) \quad \text{or} \quad \partial_{x_i} f(a) \quad \text{or} \quad \frac{\partial f}{\partial x_i}(a).$$

As usual, one says that $f : U \to Y$ is *differentiable with respect to the i-th variable on U* if it is differentiable with respect to the i-th variable at every point of U, and we denote

$$\begin{array}{rccc} \partial_i f : & U & \to & \mathscr{L}(X,Y) \\ & a & \mapsto & \partial_i f(a). \end{array}$$

Proposition 8.13. *If $f : U \to Y$ is differentiable at $a \in U$, then f is differentiable with respect any variable at a and one has*

$$f'(a)(h_1, \ldots h_n) = \sum_{i=1}^{n} \partial_i f(a)(h_i), \quad \text{for all } (h_1, \ldots, h_n) \in X.$$

Remark 8.14. A map that is differentiable with respect to every variable at some point is not necessarily differentiable at this point.

Proof of Proposition 8.13. Let ι_i denote the canonical injection

$$\begin{array}{rccc} \iota_i : & X_i & \to & X \\ & x_i & \mapsto & (0, \ldots, x_i, \ldots, 0) \end{array}$$

where there are 0's everywhere except at the i-th coordinate, which clearly is a continuous linear map. Remark that

$$\lambda_i(x_i) = a + \iota_i(x_i - a_i) \quad \text{and} \quad \lambda_i(a_i) = a,$$

thus $\lambda_i'(x_i) = \iota_i$ for all $x_i \in X_i$. By composition of differentiable maps, we deduce that $f \circ \lambda_i$ is differentiable at the point $a_i \in X_i$ and moreover

$$(f \circ \lambda_i)'(a_i) = f'(a) \circ \iota_i,$$

in other words $\partial_i f(a) = f'(a) \circ \iota_i$. Denoting by $p_i : X \to X_i$, $(x_1, \ldots, x_n) \mapsto x_i$ the canonical projection and using the identity $\sum_{i=1}^{n} \iota_i \circ p_i = \mathrm{id}_X$, we deduce

$$f'(a) = \sum_{i=1}^{n} (f'(a) \circ \iota_i) \circ p_i$$

which completes the proof. □

Chapter 9

Mean-value theorem and applications

9.1 Mean-value theorems

Theorem 9.1. *Let Y be a Banach space and $a, b \in \mathbf{R}$ with $a < b$. Consider two maps $f : [a, b] \to Y$ and $g : [a, b] \to \mathbf{R}$ that are continuous on $[a, b]$ and differentiable on (a, b), and such that $\|f'(t)\| \leqslant g'(t)$ for any $t \in (a, b)$. Then*

$$\|f(b) - f(a)\| \leqslant g(b) - g(a).$$

Proof. We shall prove that for any $\varepsilon > 0$ and any $t \in [a, b]$ one has $\|f(t) - f(a)\| \leqslant g(t) - g(a) + \varepsilon(t - a + 1)$, which concludes the proof by taking $t = b$ and letting $\varepsilon \to 0$.

By way of contradiction, assume that the set

$$A = \{t \in [a, b] \mid \|f(t) - f(a)\| > g(t) - g(a) + \varepsilon(t - a + 1)\}$$

is non-empty. Then A is a non-empty open set (by continuity) bounded by below and such that $a \notin A$, thus A has an infimum $m = \inf A$ which satisfies $m \in (a, b]$.

By definition of differentiability, there is $\delta > 0$ such that $m + \delta \leqslant b$ and for any $t \in (m, m + \delta]$ one has

$$\|f'(m)\| \geqslant \frac{\|f(t) - f(m)\|}{t - m} - \frac{\varepsilon}{2} \quad \text{and} \quad g'(m) \leqslant \frac{g(t) - g(m)}{t - m} + \frac{\varepsilon}{2}.$$

Hence one obtains $\|f(t) - f(m)\| \leqslant g(t) - g(m) + \varepsilon(t - m)$. Since $m \notin A$, one deduces $\|f(m) - f(a)\| \leqslant g(m) - g(a) + \varepsilon(m - a + 1)$ and thus

$$\|f(t) - f(a)\| \leqslant \|f(t) - f(m)\| + \|f(m) - f(a)\| \leqslant g(t) - g(a) + \varepsilon(t - a + 1).$$

This is true for any $t \in (m, m + \delta]$, which contradicts the fact that m is the infimum of A. □

Theorem 9.2. *Let X, Y be Banach spaces, $U \subseteq X$ an open and convex set, and $f : U \to Y$ differentiable on U. Assume that there is $M > 0$ such that $\|f'(x)\| \leqslant M$ for any $x \in U$. Then for any $x, y \in U$ there holds*

$$\|f(y) - f(x)\| \leqslant M\|y - x\|.$$

Proof. Since U is convex, for any $x, y \in U$ and $t \in [0,1]$ one has $x + t(y-x) \in U$. Consider the map $\phi : t \mapsto f(x + t(y-x))$ from $[0,1]$ into Y, which is continuous on $[0,1]$ (by composition of continuous maps) and differentiable on $(0,1)$ (by composition of differentiable maps). Moreover, for any $t \in (0,1)$ one has

$$\phi'(t) = f'(x + t(y-x))(y-x),$$

hence $\|\phi'(t)\| \leqslant \|f'(x + t(y-x))\|\|y-x\| \leqslant M\|y-x\|$. We then conclude thanks to Theorem 9.1 applied to the maps ϕ and $g : [0,1] \ni t \mapsto tM\|y-x\| \in \mathbf{R}$. □

9.2 Relation between partial differentiability and differentiability

Let $X = X_1 \times \cdots \times X_n$ be a finite product of Banach spaces endowed with the norm $\|x\| = \max(\|x_1\|_{X_1}, \ldots, \|x_n\|_{X_n})$ for any $x = (x_1, \ldots, x_n) \in X$, for which X is a Banach space. Let Y be a Banach space and U be an open set in X.

Theorem 9.3. *A map $f : U \to Y$ is of class $\mathscr{C}^1$ on U if and only if f is differentiable with respect to every variable on U and, for any $i \in \{1, \ldots, n\}$, the map*

$$\begin{array}{rccc} \partial_i f : & U & \to & \mathscr{L}(X_i, Y) \\ & x & \mapsto & \partial_i f(x) \end{array}$$

is continuous.

Proof. Assume that f is of class $\mathscr{C}^1$ on U, then by Proposition 8.13 f is differentiable with respect to any variable on U. Moreover $\partial_i f$ is continuous by composition of continuous maps.

Conversely, assume that f is differentiable with respect to every variable on U and $\partial_i f$ is continuous for any $i \in \{1, \ldots, n\}$. Let $a = (a_1, \ldots, a_n) \in U$. We want to prove that for any $\varepsilon > 0$, there is $\delta > 0$ such that for any $h \in B_X(0, \delta)$ one has

$$\left\| f(a_1 + h_1, \ldots, a_n + h_n) - f(a_1, \ldots, a_n) - \sum_{i=1}^{n} \partial_i f(a)(h_i) \right\| < \varepsilon \|h\|.$$

We hence write

$$\begin{aligned} & f(a_1 + h_1, \ldots, a_n + h_n) - f(a_1, \ldots, a_n) - \sum_{i=1}^{n} \partial_i f(a)(h_i) \\ & \quad = f(a_1 + h_1, \ldots, a_n + h_n) - f(a_1, a_2 + h_2, \ldots, a_n + h_n) - \partial_1 f(a)(h_1) \\ & \qquad + f(a_1, a_2 + h_2, \ldots, a_n + h_n) - f(a_1, a_2, a_3 + h_3, \ldots, a_n) - \partial_2 f(a)(h_2) \\ & \qquad + \quad \cdots \quad + \\ & \qquad + f(a_1, \ldots, a_{n-1}, a_n + h_n) - f(a_1, \ldots, a_{n-1}, a_n) - \partial_n f(a)(h_n) \end{aligned}$$

so that it suffices to prove that for any $\varepsilon > 0$ there is $\delta > 0$ such that $\|h\| < \delta$ implies

$$\begin{aligned}\|f(a_1+h_1,\dots,a_n+h_n) - f(a_1,a_2+h_2,\dots,a_n+h_n) - \partial_1 f(a)(h_1)\| &< \frac{\varepsilon\|h_1\|}{n},\\ \|f(a_1,a_2+h_2,\dots,a_n+h_n) - f(a_1,a_2,a_3+h_3,\dots,a_n) - \partial_2 f(a)(h_2)\| &< \frac{\varepsilon\|h_2\|}{n},\\ &\vdots\\ \|f(a_1,\dots,a_{n-1},a_n+h_n) - f(a_1,\dots,a_{n-1},a_n) - \partial_n f(a)(h_n)\| &< \frac{\varepsilon\|h_n\|}{n}.\end{aligned} \tag{9.1}$$

Let then $\varepsilon > 0$. We will show that there is $\delta > 0$ such that $\|h\| < \delta$ implies the first inequality in (9.1). The other inequalities can be obtained in a similar fashion, hence at the end we can take $\delta > 0$ to be the smallest one so that all the above inequalities hold.

Define the map $g : t \mapsto f(a_1+t, a_2+h_2,\dots,a_n+h_n) - \partial_1 f(a)(t)$ for $t \in X_1$ small enough. Then g is differentiable and

$$g'(t) = \partial_1 f(a_1+t,a_2+h_2,\dots,a_n+h_n) - \partial_1 f(a_1,\dots,a_n).$$

Moreover $\partial_1 f$ is continuous at a, hence there is $\delta > 0$ such that for any $\|h\| < \delta$ one has

$$\|\partial_1 f(a+h) - \partial_1 f(a)\| < \frac{\varepsilon}{n}$$

and this implies that, for any $\|t\| \leqslant \|h_1\|$, there holds

$$\|g'(t)\| = \|\partial_1 f(a_1+t,a_2+h_2,\dots,a_n+h_n) - \partial_1 f(a_1,\dots,a_n)\| < \frac{\varepsilon}{n}.$$

By Theorem 9.2 one deduces that $\|g(h_1) - g(0)\| < \frac{\varepsilon}{n}\|h_1\|$ which is exactly the first inequality in (9.1).

We have then proved that f is differentiable at a, for any $a \in U$, and that

$$f'(a)(h_1,\dots,h_n) = \sum_{i=1}^n \partial_i f(a)(h_i),$$

and moreover $f'(a)$ is continuous as the sum of composition of continuous maps. □

9.3 Limits of differentiable maps

Let X and Y be Banach spaces.

Theorem 9.4. *Let $U \subseteq X$ be open and connected. Consider a sequence $f_n : U \to Y$ of differentiable maps on U and assume that:*

(a) *there is $x_0 \in U$ such that $(f_n(x_0))_{n\in\mathbf{N}}$ has a limit in Y;*

(b) *for any $a \in U$ there is $r > 0$ such that $(f_n')_{n\in\mathbf{N}}$ converges uniformly on $B_X(a,r)$ to some map $g : U \to \mathscr{L}(X,Y)$.*

Then there holds:

(i) $(f_n)_{n\in\mathbf{N}}$ *converges pointwisely on* U, *and we denote* $f(x) = \lim_{n\to\infty} f_n(x)$ *for any* $x \in U$;

(ii) *for any* $a \in U$ *there is* $r > 0$ *such that* $(f_n)_{n\in\mathbf{N}}$ *converges uniformly to* f *on* $B_X(a,r)$;

(iii) f *is differentiable on* U *and* $f'(a) = g(a)$ *for any* $a \in U$.

Proof. Let $a \in U$. For any $x \in B_X(a,r)$ and $n, m \in \mathbf{N}$ there holds, thanks to Theorem 9.2,

$$\begin{aligned} \|f_n(x) - f_m(x) - (f_n(a) - f_m(a))\| &\leqslant \|x - a\| \sup_{y\in B_X(a,r)} \|f_n'(y) - f_m'(y)\| \\ &\leqslant r \sup_{y\in B_X(a,r)} \|f_n'(y) - f_m'(y)\|. \end{aligned} \tag{9.2}$$

Therefore, if $(f_n(a))_{n\in\mathbf{N}}$ converges in Y it follows that $(f_n)_{n\in\mathbf{N}}$ is a Cauchy sequence for the uniform norm on $B_X(a,r)$, hence $(f_n)_{n\in\mathbf{N}}$ converges uniformly on $B_X(a,r)$ to some map f which is continuous.

Moreover, let $A = \{x \in U \mid (f_n(x))_{n\in\mathbf{N}} \text{ converges in } Y\}$, which is non-empty by hypothesis. One easily obtains that A is open and closed, hence $A = U$ since U is connected, that is, $\lim_{n\to\infty} f_n(x) = f(x)$ for any $x \in U$.

It remains to prove that, for any $a \in U$, f is differentiable at a and $f'(a) = g(a)$. We write, for any $h \in X$ small enough and $n \in \mathbf{N}$,

$$\begin{aligned} \|f(a+h) - f(a) - g(a)(h)\| &\leqslant \|f(a+h) - f(a) - (f_n(a+h) - f_n(a))\| \\ &\quad + \|f_n(a+h) - f_n(a) - f_n'(a)(h)\| \\ &\quad + \|f_n'(a)(h) - g(a)(h)\|. \end{aligned}$$

Let $\varepsilon > 0$, then there is $N_0 \in \mathbf{N}$ such that for any integer $n, m \geqslant N_0$ one has

$$\sup_{y\in B_X(a,r)} \|f_n'(y) - f_m'(y)\| < \frac{\varepsilon}{3},$$

hence using (9.2) one gets, for any $h \in B_X(0,r)$,

$$\|f_n(a+h) - f_m(a+h) - (f_n(a) - f_m(a))\| < \frac{\varepsilon}{3}\|h\|,$$

and by letting $m \to \infty$ one concludes that

$$\|f(a+h) - f_n(a+h) - (f(a) - f_n(a))\| \leqslant \frac{\varepsilon}{3}\|h\|.$$

Furthermore, there is $N_1 \in \mathbf{N}$ such that for any integer $n \geqslant N_1$ one has

$$\|f_n'(a) - g(a)\| < \frac{\varepsilon}{3}.$$

Let us fix some $n \geqslant \max(N_0, N_1)$. Since f_n is differentiable at a, there is $r' \in (0, r]$ such that for any $h \in B_X(0, r')$ there holds

$$\|f_n(a+h) - f_n(a+h) - f_n'(a)(h)\| \leqslant \frac{\varepsilon}{3}\|h\|.$$

Finally one obtains

$$\|f(a+h) - f(a) - g(a)(h)\| \leqslant \varepsilon\|h\|.$$

□

9.4 High-order derivatives

9.4.1 Second-order derivatives

Let X, Y be Banach spaces and U be an open subset of X.

Definition 9.5. A map $f : U \to Y$ is *twice differentiable at $a \in U$* if f is differentiable on some open neighborhood V of a, and if the map $f' : V \to \mathscr{L}(X, Y)$ is differentiable at a. We then denote

$$\mathrm{d}^2 f(a) \quad \text{or} \quad f''(a) = (f')'(a).$$

Recall from Proposition 7.27 that the spaces $\mathscr{L}(X, \mathscr{L}(X, Y))$ and $\mathscr{L}_2(X, X; Y)$ are isometrically isomorphic. We will therefore always identify $f''(a) \in \mathscr{L}(X, \mathscr{L}(X, Y))$ to the corresponding continuous bilinear map in $\mathscr{L}_2(X, X; Y)$, more precisely to the map

$$\begin{array}{ccc} X \times X & \to & Y \\ (h, k) & \mapsto & f''(a)(h)(k). \end{array}$$

Definition 9.6. Consider a map $f : U \to Y$.

(i) One says that f is *twice differentiable on U* if f is twice differentiable at every point of U or, equivalently, if f is differentiable on U and the map $f' : U \to \mathscr{L}(X, Y)$ is differentiable on U.

(ii) One says that f is *of class $\mathscr{C}^2$ on U* if it is twice differentiable on U and the map $f'' : U \to \mathscr{L}(X, \mathscr{L}(X, Y))$ is continuous.

Proposition 9.7. *If $f : U \to Y$ is twice differentiable at $a \in U$, then $f''(a)$ is a symmetric bilinear continuous map, that is, $f''(a)(h)(k) = f''(a)(k)(h)$ for any $(h, k) \in X \times X$.*

Proof. Let $\varepsilon > 0$, then there is $r > 0$ such that $B(a, r) \subseteq U$ and for any $h, k \in E$ with $\|h\|, \|k\| < r/2$, and any $t \in [0, 1]$, one has

$$\|f'(a + th + k) - f'(a) - f''(a)(th + k)\| \leqslant \varepsilon(\|th + k\|) \leqslant \varepsilon(\|h\| + \|k\|).$$

Define the map $g : t \mapsto f'(a + th + k) - f'(a + th)$ from $[0, 1]$ into Y, then g is differentiable (by composition of differentiable maps) and

$$\begin{aligned} g'(t) &= f'(a + th + k)(h) - f'(a + th)(h) \\ &= \left(f'(a + th + k) - f'(a)\right)(h) - \left(f'(a + th) - f'(a)\right)(h). \end{aligned}$$

Hence one obtains

$$\begin{aligned} \|g'(t) - f''(a)(k, h)\| &= \|g'(t) - f''(a)(th + k, h) + f''(a)(th, h)\| \\ &\leqslant \| \left(f'(a + th + k) - f'(a) - f''(a)(th + k)\right)(h)\| \\ &\quad + \| \left(f'(a + th) - f'(a) - f''(a)(th)\right)(h)\| \\ &\leqslant 2\varepsilon(\|h\| + \|k\|)\|h\| \leqslant 2\varepsilon(\|h\| + \|k\|)^2. \end{aligned}$$

Applying Theorem 9.2 to the function $t \mapsto g(t) - tf''(a)(k, h)$ one gets

$$\|g(1) - g(0) - f''(a)(k, h)\| \leqslant \sup_{0 \leqslant t \leqslant 1} \|g'(t) - f''(a)(k, h)\| \leqslant 2\varepsilon(\|h\| + \|k\|)^2,$$

and $S(h,k) := g(1) - g(0) = f(a+h+k) - f(a+h) - f(a+k) + f(a)$ is symmetric on (h,k). Therefore, for any $h, k \in B(0, r/2)$, it follows that

$$\begin{aligned}\|f''(a)(h,k) - f''(a)(k,h)\| &\leqslant \|f''(a)(h,k) - S(h,k)\| + \|S(k,h) - f''(a)(k,h)\| \\ &\leqslant 4\varepsilon(\|h\| + \|k\|)^2.\end{aligned}$$

By homogeneity, this last inequality holds for any $(h,k) \in X \times X$, and one concludes that $f''(a)(h,k) = f''(a)(k,h)$ since $\varepsilon > 0$ is arbitrary. □

9.4.2 Second-order partial derivatives

Let $X = X_1 \times \cdots \times X_n$ be a finite product of Banach spaces, Y be a Banach space, and U an open subset of X.

Consider a map $f : U \to Y$ twice differentiable at $a \in U$. Then there exists some open neighborhood V of a such that, for any $x \in V$, one has

$$f'(x)(h_1, \dots, h_n) = \sum_{j=1}^{n} \partial_j f(x)(h_j), \quad \text{for all } h = (h_1, \dots, h_n) \in X.$$

Moreover, one also has

$$f''(a)(k_1, \dots, k_n) = \sum_{i=1}^{n} \partial_i f'(a)(k_i), \quad \text{for all } k = (k_1, \dots, k_n) \in X.$$

Combining both identities one finally gets, for any $(h,k) \in X \times X$,

$$f''(a)(k_1, \dots, k_n)(h_1, \dots, h_n) = \sum_{i,j=1}^{n} \partial_i(\partial_j f)(a)(k_i)(h_j),$$

which relates the second-order derivative of f at a with the partial derivatives $\partial_i(\partial_j f)(a)$.

Applying twice (the proof of) Theorem 9.3, one obtains:

Proposition 9.8. *A map $f : U \to Y$ is twice differentiable at $a \in U$ if and only if*

(i) *the partial derivatives $\partial_j f$ exist on U for any $j = 1, \dots, n$ and they are continuous on U;*

(ii) *for any $j = 1, \dots, n$, the partial derivatives $\partial_i(\partial_j f)$ exist on U for any $i = 1, \dots, n$ and they are continuous at a.*

9.4.3 Derivatives of higher order

Let X, Y be Banach spaces and U be an open subset of Y. We denote $f^{(1)} = f'$, $\mathscr{L}_1(X,Y) = \mathscr{L}(X,Y)$ and $\mathscr{L}_n(X,Y) = \mathscr{L}(X, \mathscr{L}_{n-1}(X,Y))$ for any $n \in \mathbf{N}^*$.

Definition 9.9. Consider a map $f : U \to Y$.

(i) One says that f is n-times differentiable at $a \in U$ if f is $(n-1)$-times differentiable on some open neighborhood V of a, and if the map $f^{(n-1)}: V \to \mathscr{L}_{n-1}(X,Y)$ is differentiable at a. We then denote

$$\mathrm{d}^n f(a) \quad \text{or} \quad f^{(n)}(a) = (f^{(n-1)})'(a).$$

(ii) One says that f is n-times differentiable on U if it is n-times differentiable at every point of U.

(iii) One says that f is of class $\mathscr{C}^n$ on U if it is n-times differentiable on U and the map $f^{(n)}: U \to \mathscr{L}_n(X,Y)$ is continuous.

(iv) One says that f is of class $\mathscr{C}^\infty$ on U if it is of class $\mathscr{C}^n$ on U for all $n \in \mathbf{N}$.

As for the case of second-order derivatives, we often identify $f^{(n)}(a) \in \mathscr{L}_n(X,Y)$ to the corresponding element of $\mathscr{L}_n(X,\dots,X;Y)$, the vector space of n-multilinear continuous maps, through the canonical isomorphism

$$\begin{array}{ccc} \mathscr{L}_n(X,Y) & \to & \mathscr{L}_n(X,\dots,X;Y) \\ u & \mapsto & \{(h_1,\dots,h_n) \mapsto u(h_1)(h_2)\cdots(h_n)\}. \end{array}$$

One also obtains an analogous result of Proposition 9.7, which states that if $f: U \to Y$ is n-times differentiable at some point $a \in U$, then its derivative $f^{(n)}(a) \in \mathscr{L}_n(X,Y)$ is a n-multilinear symmetric map, that is, for any $(h_1,\dots,h_n) \in X^n$ and any permutation σ on $\{1,\dots,n\}$, one has

$$f^{(n)}(a)(h_1,\dots,h_n) = f^{(n)}(a)(h_{\sigma(1)},\dots,h_{\sigma(n)}).$$

9.5 Taylor formulas

Consider a map $v: I \to Y$ where $I = (a,b)$ is an open interval of $\mathbf{R}$ and Y a Banach space. If v is $(n+1)$-times differentiable on I, then one has for any $t \in I$

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[v(t) + (1-t)v'(t) + \cdots + \frac{(1-t)^n}{n!}v^{(n)}(t)\right] = \frac{(1-t)^n}{n!}v^{(n+1)}(t).$$

As a consequence on obtains the following particular case of Taylor's[1] formula:

Proposition 9.10. *Let $v: I \to Y$ be $(n+1)$-times differentiable on $I \supseteq [0,1]$.*

(i) *Assume that $v^{(n+1)}$ is continuous on I, then*

$$v(1) - v(0) - v'(0) - \frac{1}{2!}v''(0) - \cdots - \frac{1}{n!}v^{(n)}(0) = \int_0^1 \frac{(1-t)^n}{n!}v^{(n+1)}\,\mathrm{d}t.$$

(ii) *Assume that there is $M > 0$ such that $\|v^{(n+1)}(t)\| \leqslant M$ for any $t \in [0,1]$, then*

$$\left\|v(1) - v(0) - v'(0) - \frac{1}{2!}v''(0) - \cdots - \frac{1}{n!}v^{(n)}(0)\right\| \leqslant \frac{M}{(n+1)!}.$$

[1] Brook Taylor (1685–1731) was an English mathematician who made important contributions to analysis.

Proof. (i) Recall that if $f : I \to Y$ is of class $\mathscr{C}^1$ then $f(1) - f(0) = \int_0^1 f'(t)\,\mathrm{d}t$. The result follows applying this to the function

$$f(t) = v(t) + (1-t)v'(t) + \frac{(1-t)^2}{2!}\,v''(t) + \cdots + \frac{(1-t)^n}{n!}\,v^{(n)}(t).$$

(ii) Consider the function $f(t)$ defined above, and define $g(t) = -M\,\frac{(1-t)^{n+1}}{(n+1)!}$ so that $g'(t) = M\,\frac{(1-t)^n}{n!}$. One has

$$\|f'(t)\| = \left\| \frac{(1-t)^n}{n!}\,v^{(n+1)}(t) \right\| \leqslant \frac{(1-t)^n}{n!}\|v^{(n+1)}(t)\| \leqslant M\frac{(1-t)^n}{n!}.$$

Thanks to Theorem 9.1 one deduces $\|f(1) - f(0)\| \leqslant g(1) - g(0)$, which concludes the proof. □

Consider now X and Y Banach spaces, U an open set in X, and a map $f : U \to Y$. Let $a, a+h \in U$ be points such that the segment $[a, a+h] = \{a + th : t \in [0,1]\}$ is included in U. Define the function $v(t) = f(a+th)$ for $t \in [0,1] \subseteq I$ (where I in an open interval of $\mathbf{R}$). If f is $(n+1)$-times differentiable, then v is also $(n+1)$-times differentiable by composition and, moreover, one has for any $t \in [0,1]$

$$\begin{aligned} v'(t) &= f'(a+th)(h) \\ v''(t) &= f''(a+th)(h,h) \\ &\vdots \\ v^{(n)}(t) &= f^{(n)}(a+th)(h,\ldots,h). \end{aligned}$$

Therefore, applying Proposition 9.10 we obtain the following two Taylor formulas.

Theorem 9.11 (Taylor formula with integral remainder)**.** *Let $f : U \to Y$ be a map of class $\mathscr{C}^{n+1}$ on U. If $[a, a+h] \subseteq U$ then*

$$\begin{aligned} f(a+h) &= f(a) + f'(a)(h) + \frac{1}{2!}f''(a)(h,h) \\ &\quad + \cdots + \frac{1}{n!}f^{(n)}(a)(h,\ldots,h) + \int_0^1 \frac{(1-t)^n}{n!}\,f^{(n+1)}(a+th)(h,\ldots,h)\,\mathrm{d}t. \end{aligned}$$

Theorem 9.12 (Taylor formula with Lagrange remainder)**.** *Let $f : U \to Y$ be $(n+1)$-times differentiable on U and assume that there is $M > 0$ such that $\|f^{(n+1)}(x)\| \leqslant M$ for any $x \in U$. Then*

$$\begin{aligned} &\left\| f(a+h) - f(a) - f'(a)(h) - \frac{1}{2!}f''(a)(h,h) - \cdots - \frac{1}{n!}f^{(n)}(a)(h,\ldots,h) \right\| \\ &\qquad\qquad \leqslant \frac{M}{(n+1)!}\|h\|^{n+1}. \end{aligned}$$

One can actually obtain a third formula that is similar to that obtained in Theorem 9.12, but under weaker hypothesis.

Theorem 9.13. *Let $f : U \to Y$ be $(n-1)$-times differentiable on U, and assume that f is n-times differentiable at $a \in U$. Then for any $\varepsilon > 0$, there is $\delta > 0$ such that for any $h \in E$, $\|h\| < \delta$ implies*

$$\left\| f(a+h) - f(a) - f'(a)(h) - \frac{1}{2!} f''(a)(h,h) - \cdots - \frac{1}{n!} f^{(n)}(a)(h,\ldots,h) \right\| < \varepsilon \|h\|^n.$$

Proof. The case $n = 1$ follows from the definition of differentiability. We shall prove the general case by induction.

Assume that the result holds for some $n-1$ with $n \geqslant 2$. Define the map

$$\varphi(h) = f(a+h) - f(a) - f'(a)(h) - \frac{1}{2!} f''(a)(h,h) - \cdots - \frac{1}{n!} f^{(n)}(a)(h,\ldots,h)$$

which is differentiable. For any $h, k \in X$ one has

$$\begin{aligned}\varphi'(h)(k) = f'(a+h)(k) - f'(a)(k) - \frac{1}{2!}\left\{f''(a)(h,k) + f''(a)(k,h)\right\} - \cdots \\ - \frac{1}{n!}\left\{f^{(n)}(a)(h,\ldots,h,k) + f^{(n)}(a)(h,\ldots,h,k,h) + \cdots + f^{(n)}(a)(k,h,\ldots,h)\right\}.\end{aligned}$$

Using the symmetry of the derivatives $f^{(k)}(a)$ and denoting $(h)^k = (h,\ldots,h) \in E^k$, one gets

$$\varphi'(h) = f'(a+h) - f'(a) - f''(a)(h) - \frac{1}{2!} f^{(3)}(h)^2 - \cdots - \frac{1}{(n-1)!} f^{(n)}(a)(h)^{n-1}.$$

Therefore, by the induction hypothesis applied to f', one obtains that for any $\varepsilon > 0$, there is $\delta > 0$ such that for any $h \in E$, $\|h\| \leqslant \delta$ implies $\|\varphi'(h)\| \leqslant \varepsilon \|h\|^{n-1}$. Hence, thanks to Theorem 9.2, on deduces that $\|\varphi(1) - \varphi(0)\| \leqslant \varepsilon \|h\|^n$ for any $\|h\| \leqslant \delta$, which concludes the proof since $\varphi(0) = 0$. □

Chapter 10

Inverse function and implicit function theorems

10.1 Diffeomorphisms

Let X and Y be Banach spaces.

Definition 10.1. Let V be an open subset of X and W an open subset of Y. A map $f : V \to W$ is a $\mathscr{C}^1$-diffeomorphism if it is bijective, of class $\mathscr{C}^1$ on V, and its inverse map $f^{-1} : W \to V$ is of class $\mathscr{C}^1$ on W.

Remark that a map of class $\mathscr{C}^1$ can be a homeomorphism without being a $\mathscr{C}^1$-diffeomorphism, in other words the inverse homeomorphism need not be of class $\mathscr{C}^1$.

Proposition 10.2. *Let V be an open subset of X and W an open subset of Y. Let $f : V \to W$ be a homeomorphism of class $\mathscr{C}^1$. Then f is a $\mathscr{C}^1$-diffeomorphism from V onto W if and only if $f'(a)$ is an isomorphism from X onto Y for all $a \in V$.*

Proof. Assume that f is a $\mathscr{C}^1$-diffeomorphism from V onto W. For any $a \in V$, let $b = f(a) \in W$ and denote by $g = f^{-1} : W \to V$ the inverse map, which is of class $\mathscr{C}^1$. By the composition of differentiable maps one has

$$g'(b) \circ f'(a) = \mathrm{id}_X \quad \text{and} \quad f'(a) \circ g'(b) = \mathrm{id}_Y,$$

therefore $f'(a)$ is isomorphism from X onto Y.

Conversely, assume that $f'(a)$ is an isomorphism from X onto Y for all $a \in V$. Thanks to Proposition 8.10 the inverse map $g = f^{-1} : W \to V$ is differentiable on W with

$$g'(b) = (f'(g(b)))^{-1}, \quad \text{for all } b \in W.$$

It remains to check that g is of class $\mathscr{C}^1$ on W, that is, the map $g' : W \to \mathscr{L}(Y, X)$ is continuous. But the map g' can be written as the composition of three continuous maps, more precisely, $g' = \mathrm{inv} \circ f' \circ g$, where, denoting by $\mathrm{Isom}(X, Y)$ the set of isomorphisms from X into Y, we denote

$$\begin{array}{rccc} \mathrm{inv}: & \mathrm{Isom}(X,Y) & \to & \mathrm{Isom}(Y,X) \\ & u & \mapsto & u^{-1} \end{array}$$

which is continuous thanks to Proposition 7.24. Therefore g is of class $\mathscr{C}^1$ on W, and thus $f : V \to W$ is a $\mathscr{C}^1$-diffeomorphism. □

10.2 Inverse function theorem

Let X and Y be Banach spaces, and U be an open subset of X.

Proposition 10.3. *Consider two maps $g, \phi : U \to X$ such that $g(x) = x + \phi(x)$ and ϕ is k-Lipschitz with $k \in (0,1)$. Then g is a homeomorphism from U onto an open subset W of X.*

Proof. We first prove that g is injective. For any $x, y \in U$ we have

$$\begin{aligned}\|g(x) - g(y)\| &= \|(x-y) + (\phi(x) - \phi(y)\| \\ &\geqslant \|x-y\| - \|\phi(x) - \phi(y)\| \\ &\geqslant (1-k)\|x-y\|,\end{aligned}$$

thus if $g(x) = g(y)$ we deduce that $x = y$. Therefore g is bijective from U onto its image $W := g(U)$.

For any $a, b \in W$, if $x = g^{-1}(a)$ and $y = g^{-1}(b)$ then

$$\begin{aligned}\|g^{-1}(a) - g^{-1}(b)\| &= \|x-y\| \\ &\geqslant \frac{1}{1-k}\|g(x) - g(y)\| \\ &= \frac{1}{1-k}\|a-b\|,\end{aligned}$$

thus g^{-1} is $\frac{1}{1-k}$-Lipschitz.

We now prove that $W = g(U)$ is open in X. Let $y_0 \in W$ then consider $x_0 \in U$ such that $y_0 = g(x_0)$ and $r > 0$ such that the closed ball $\overline{B}(x_0, r) \subseteq U$. For any $y \in B(y_0, (1-k)r)$ and any $x \in \overline{B}(x_0, r)$ we have

$$\begin{aligned}\|[y - \phi(x)] - x_0\| &\leqslant \|y - y_0\| + \|y_0 - x_0 - \phi(x)\| \\ &= \|y - y_0\| + \|\phi(x_0) - \phi(x)\| \\ &\leqslant \|y - y_0\| + k\|x - x_0\| \\ &\leqslant (1-k)r + kr = r,\end{aligned}$$

which implies that $y - \phi(x) \in \overline{B}(x_0, r)$. Therefore the map $h : x \mapsto y - \phi(x)$ is a contraction, since it clearly is k-Lipschitz, and verifies $h(\overline{B}(x_0, r)) \subseteq \overline{B}(x_0, r)$. Since $\overline{B}(x_0, r)$ is complete, it follows from the Banach-Picard fixed point theorem (Theorem 3.11) that there is a unique $\bar{x} \in \overline{B}(x_0, r)$ such that $\bar{x} = h(\bar{x}) = y - \phi(\bar{x})$, namely $y = g(\bar{x})$. This implies that $y \in g(\overline{B}(x_0, r)) \subseteq g(U) = W$, thus W is a neighborhood of y_0 and hence W is open in X. □

Theorem 10.4 (Inverse function theorem)**.** *Let $f : U \to Y$ be a map of class $\mathscr{C}^1$ on U and assume that there is $a \in U$ such that $f'(a)$ is an isomorphism from X into Y. Then there is an open neighborhood $V \subseteq U$ of a in X and an open neighborhood W of $f(a)$ in Y such that the restriction of f to V is a $\mathscr{C}^1$-diffeomorphism from V onto W.*

Proof. Consider the translation maps $\tau : x \mapsto x + a$ from X into X and $\bar{\tau} : y \mapsto y - f(a)$ from Y into Y, which clearly are $\mathscr{C}^1$-diffeomorphisms. Consider also the set $U_0 := U - a = \{x - a : x \in U\}$, which is open in X and contains 0, and define the map

$$\begin{array}{rccl} g: & U_0 & \longrightarrow & X \\ & x & \longmapsto & (f'(a))^{-1}(f(x+a) - f(a)) \end{array} .$$

In other words, g can be written as the following composition

$$g = (f'(a))^{-1} \circ \bar{\tau} \circ f \circ \tau,$$

and moreover

$$f = (\bar{\tau})^{-1} \circ f'(a) \circ g \circ \tau^{-1}. \tag{10.1}$$

The map g is of class $\mathscr{C}^1$ on U_0 with $g'(x) = (f'(a))^{-1} \circ f'(x + a)$, so in particular it verifies

$$g(0) = 0 \quad \text{and} \quad g'(0) = \mathrm{id}_X \in \mathrm{Isom}(X, X).$$

We shall prove below that g is a $\mathscr{C}^1$-diffeomorphism from some open neighborhood V' of 0 in X onto some open neighborhood W' of 0 in X. This will imply the desired result for f since f will be the composition of $\mathscr{C}^1$-diffeomorphisms thanks to (10.1).

Define the map $\phi : x \mapsto g(x) - x$ from U_0 into X, which verifies $\phi(0) = 0$ and is of class $\mathscr{C}^1$ with

$$\phi'(x) = g'(x) - \mathrm{id}_X = g'(x) - g'(0).$$

Since $\mathrm{Isom}(X, X)$ is open in $\mathscr{L}(X, X)$ from Proposition 7.24 and g' is continuous, there is $r > 0$ small enough such that $B(0, r) = \{x \in X : \|x\| < r\} \subseteq U_0$ and for all $x \in B(0, r)$ one has

$$g'(x) \in \mathrm{Isom}(X, X) \quad \text{and} \quad \|\phi'(x)\| \leqslant \frac{1}{2}. \tag{10.2}$$

Applying Theorem 9.2 one deduces, for all $x, y \in B(0, r)$,

$$\|\phi(x) - \phi(y)\| \leqslant \frac{1}{2}\|x - y\|,$$

that is, ϕ is 1/2-Lipschitz on $B(0, r)$. From Proposition 10.3 it follows that g is a homeomorphism from the open set $V' := B(0, r)$ onto the open set $W' := g(B(0, r))$.

Therefore g is a homeomorphism of class $\mathscr{C}^1$ from V' onto W' and $g'(x) \in \mathrm{Isom}(X, X)$ for any $x \in V'$ from (10.2), which implies that $g : V' \to W'$ is a $\mathscr{C}^1$-diffeomorphism thanks to Proposition 10.2. □

10.3 Implicit function theorem

Let X, Y, Z be Banach spaces and $U \subseteq X \times Y$ be open.

Theorem 10.5 (Implicit function theorem)**.** *Consider a map $f : U \to Z$ that is of class $\mathscr{C}^1$ on U. Let $(a,b) \in U$ be such that $f(a,b) = 0$, and assume that $\partial_2 f(a,b) \in \mathscr{L}(Y,Z)$ is an isomorphism from Y onto Z.*

Then there exist an open neighborhood $V \subseteq U$ of (a,b), an open neighborhood W of a, and a map $g : W \to Y$ of class $\mathscr{C}^1$ on W such that

$$(x,y) \in V \quad \text{and} \quad f(x,y) = 0 \quad \Longleftrightarrow \quad x \in W \quad \text{and} \quad g(x) = y.$$

In particular $g(a) = b$ and

$$g'(a) = -(\partial_2 f(a,b))^{-1} \circ \partial_1 f(a,b).$$

Proof. Define the map

$$\begin{array}{cccc} \varphi : & U & \to & X \times Z \\ & (x,y) & \mapsto & (x, f(x,y)) \end{array}$$

which is of class $\mathscr{C}^1$ on U since each component is $\mathscr{C}^1$ on U. Therefore one has

$$\begin{array}{cccc} \varphi'(a,b) : & X \times Y & \to & X \times Z \\ & (h,k) & \mapsto & \big(h, \partial_1 f(a,b)(h) + \partial_2 f(a,b)(k)\big) . \end{array}$$

Since $\partial_2 f(a,b)$ is an isomorphism from Y onto Z, one gets that $\varphi'(a,b)$ is an isomorphism from $X \times Y$ onto $X \times Z$ with inverse map given by

$$\begin{array}{ccc} X \times Z & \to & X \times Y \\ (h',k') & \mapsto & \big(h', (\partial_2 f(a,b))^{-1}(k') - (\partial_2 f(a,b))^{-1} \circ \partial_1 f(a,b)(h')\big) . \end{array}$$

Applying Theorem 10.4 to φ, one gets that there is an open neighborhood $V \subseteq U$ of (a,b) in $X \times Y$ and an open neighborhood W_1 of $\varphi(a,b) = (a,0)$ in $X \times Z$ such that $\varphi : V \to W_1$ is a $\mathscr{C}^1$-diffeomorphism from V to W_1.

The inverse map of φ, that we denote by $\psi = \varphi^{-1}$, is of the form

$$\psi(x,z) = (x, \bar{g}(x,z)), \quad \text{for all } (x,z) \in W_1,$$

for some function $\bar{g} : W_1 \to Y$ of class $\mathscr{C}^1$ on W_1. Therefore one has

$$(x,y) \in V \quad \text{and} \quad f(x,y) = z \quad \Longleftrightarrow \quad (x,z) \in W_1 \quad \text{and} \quad \bar{g}(x,z) = y.$$

Taking $z = 0$ above one gets

$$(x,y) \in V \quad \text{and} \quad f(x,y) = 0 \quad \Longleftrightarrow \quad (x,0) \in W_1 \quad \text{and} \quad \bar{g}(x,0) = y.$$

Define $W = \{x \in X \mid (x,0) \in W_1\}$ which is open in X and contains a. Moreover, define the function $g(x) = \bar{g}(x,0)$ from W to Y so that g is of class $\mathscr{C}^1$ on W. One finally obtains

$$(x,y) \in V \quad \text{and} \quad f(x,y) = 0 \quad \Leftrightarrow \quad x \in W \quad \text{and} \quad g(x) = y.$$

Let us now check the last statement. The map $u : x \mapsto f(x, g(x))$ is constant on W, thus $u'(x) = 0$ for any $x \in W$, which implies that

$$\partial_1 f(x, g(x)) + \partial_2 f(x, g(x)) \circ g'(x) = 0, \quad \text{for all } x \in W.$$

We complete the proof by taking $x = a$. □

Chapter 11

Optimization

In this chapter we shall consider maps $f : U \to \mathbf{R}$ with U an open subset of a Banach space X.

11.1 Extremum points

Definition 11.1. Let $f : U \to \mathbf{R}$ be a map and $a \in U$. One says that:

(i) f has a *(global) minimum at a* if $f(a) \leqslant f(x)$ for any $x \in U$;

(ii) f has a *strict (global) minimum at a* if $f(a) < f(x)$ for any $x \in U$ with $x \neq a$;

(iii) f has a *local minimum at a* if there is some neighborhood V of a such that $f(a) \leqslant f(x)$ for any $x \in V$;

(iv) f has a *strict local minimum at a* if there is some neighborhood V of a such that $f(a) < f(x)$ for any $x \in V$ with $x \neq a$;

(v) f has a *(global) maximum at a* if $f(a) \geqslant f(x)$ for any $x \in U$;

(vi) f has a *strict (global) maximum at a* if $f(a) > f(x)$ for any $x \in U$ with $x \neq a$;

(vii) f has a *local maximum at a* if there is some neighborhood V of a such that $f(a) \geqslant f(x)$ for any $x \in V$;

(viii) f has a *strict local maximum at a* if there is some neighborhood V of a such that $f(a) > f(x)$ for any $x \in V$ with $x \neq a$.

One says that f has an (strict local/local/strict) extremum at a if f has a (strict local/local/strict) minimum or maximum at a. Moreover, if V is a subset of U, one says that f has an extremum on V if the restriction map $f_{|V} : V \to \mathbf{R}$ of f to V has an extremum.

11.2 First-order conditions

Our first result is a necessary condition for a differentiable function to have a extremum at some point.

Proposition 11.2. *Let $f : U \to \mathbf{R}$ be differentiable at $a \in U$. If f has a local extremum at a then $f'(a) = 0$.*

Proof. Assume that f has a local minimum at a (the proof of the other case being similar). Then there is some neighborhood V_1 of 0 in X such that $f(a+h) - f(a) \geqslant 0$ for any $h \in V_1$. Moreover, since f is differentiable at a, there is a neighborhood V_2 of 0 in X such that $f(a+h) - f(a) = f'(a)(h) + \|h\|\,\varepsilon(h)$, where $\varepsilon : V_2 \to \mathbf{R}$ satisfies $\lim_{h\to 0} \varepsilon(h) = 0$.

For any $h \in X$ and $t > 0$ small enough so that $th \in V := V_1 \cap V_2$, one obtains

$$f(a+th) - f(a) = f'(a)(th) + \|th\|\,\varepsilon(th) \geqslant 0$$

hence

$$f'(a)(h) + \|h\|\,\varepsilon(th) \geqslant 0,$$

and taking the limit $t \to 0$ one gets $f'(a)(h) \geqslant 0$. Applying the same argument with $-h$ instead of h one obtains $f'(a)(h) \leqslant 0$, which concludes the proof. □

Definition 11.3. Let $f : U \to \mathbf{R}$ be a differentiable function. We call *critical points (or stationary points)* of f all points $x \in U$ such that $f'(x) = 0$.

Theorem 11.4. *Let $f : U \to \mathbf{R}$ be a differentiable map. Assume that U is a convex open subset of X, and that f is a convex map. Let $a \in U$ be such that $f'(a) = 0$. Then f has a minimum at a.*

Proof. By way of contradiction, assume that there is some $b \in U \setminus \{a\}$ such that $f(b) < f(a)$. Since U is convex one has $tb + (1-t)a \in U$ for all $t \in [0,1]$ and

$$f(tb + (1-t)a) \leqslant tf(b) + (1-t)f(a) = f(a) + t(f(b) - f(a)).$$

Define the function $g : t \mapsto f(tb + (1-t)a)$ from J into $\mathbf{R}$, where J is some open interval containing $[0,1]$. It is clear that g is differentiable and

$$g(t) \leqslant g(0) + t(f(b) - f(a)).$$

On the one hand, one computes

$$g'(0) = \lim_{t\to 0,\, t>0} \frac{g(t) - g(0)}{t} \leqslant f(b) - f(a) < 0.$$

On the other hand one has $g'(0) = f'(a)(b-a)$, thus $f'(a)(b-a) < 0$, which is a contradiction with $f'(a) = 0$. □

11.3 Optimization with constraints

Theorem 11.5 (Lagrange multipliers' theorem)**.** *Let $f : U \to \mathbf{R}$ be a map of class $\mathscr{C}^1$ on U. Consider a constraint function $\varphi = (\varphi_1, \ldots, \varphi_p) : U \to \mathbf{R}^p$ of class $\mathscr{C}^1$ on U, and define the set $S := \{x \in U \mid \varphi(x) = 0\}$. Assume that f has a local extremum on S at a*

point $a \in S$ and that $\varphi'(a)$ is a surjective map. Then there exist $\lambda_1, \cdots, \lambda_p \in \mathbf{R}$ (called Lagrange[1] multipliers*) such that*

$$f'(a) = \sum_{i=1}^{p} \lambda_i \varphi_j'(a).$$

Proof. Let $(e_1, \ldots, e_p)$ be a basis of $\mathbf{R}^p$. Since $\varphi'(a)(X) = \mathbf{R}^p$, there is $u_1, \ldots, u_p \in X$ such that $\varphi'(a)(u_i) = e_i$ for any $i = 1, \ldots, p$. Define $V = \operatorname{span}(u_1, \ldots, u_p)$ a vector subspace of X.

Consider the set $\Omega = \{(x, t) \in X \times \mathbf{R}^p \mid x + t_1 u_1 + \cdots + t_p u_p \in U\}$ where we denote $t = (t_1, \ldots, t_p) \in \mathbf{R}^p$, and define the function $\Phi : \Omega \to \mathbf{R}^p$ By

$$\Phi(x, t) = \varphi(x + t_1 u_1 + \cdots + t_p u_p).$$

One has $(a, 0) \in \Omega$, $\Phi(a, 0) = \varphi(a) = 0$ and

$$\partial_{t_i} \Phi(a, 0) = \varphi'(a) u_i = e_i, \quad \text{for all } i = 1, \ldots, p.$$

In particular, one has $\partial_t \Phi(a, 0) = \mathrm{id}_{\mathbf{R}^p} \in \mathscr{L}(\mathbf{R}^p, \mathbf{R}^p)$. Hence by the Implicit Function Theorem 10.5 there is some open neighborhood W of a in X and some function $g = (g_1, \ldots, g_p) = W \to \mathbf{R}^p$ of class $\mathscr{C}^1$ such that $g(a) = 0$ and $\Phi(x, g(x)) = 0$ for any $x \in W$, that is

$$x + g_1(x) u_1 + \cdots g_p(x) u_p \in S = \Phi^{-1}(\{0\}),$$

and moreover

$$g'(a)(h) = -(\partial_t \Phi(a, 0))^{-1} \circ \partial_x \Phi(a, 0)(h),$$

hence

$$g_i'(a)(h) = -\varphi_i'(a)(h), \quad \text{for all } i = 1, \ldots, p.$$

Assume now that f has a local maximum on S at the point a (the other case can be treated in a similar fashion). For x close enough to a in W one has

$$f(x + g_1(x) u_1 + \cdots g_p(x) u_p) \leqslant f(a) = f(a + g_1(a)(u_1) + \cdots + g_p(a) u_p).$$

Define the map $\psi : x \mapsto f(x + g_1(x) u_1 + \cdots g_p(x) u_p)$ from W into $\mathbf{R}$, so that ψ has a local maximum at a by above, therefore for any $h \in X$ one has

$$\psi'(a)(h) = f'(a) \Big(h + g_1'(a)(h) u_1 + \cdots + g_p'(a)(h) u_p \Big) = 0.$$

It follows that

$$f'(a)(h) = \sum_{i=1}^{p} (-g_i'(a)(h)) f'(a)(u_i)$$

which implies

$$f'(a)(h) = \sum_{i=1}^{p} (f'(a)(u_i)) \varphi_i'(a)(h),$$

which concludes the proof by taking $\lambda_i = f'(a)(u_i)$ for any $i = 1, \ldots, p$. □

[1] Joseph-Louis Lagrange (1736–1813) was an Italian mathematician (later naturalized French) who made fundamental contributions to analysis, number theory, and mechanics. He was a professor at École polytechnique.

11.4 Second-order conditions

Proposition 11.6. *Let $f : U \longrightarrow \mathbf{R}$ be of class $\mathscr{C}^2$ on U. If f has a local minimum (resp. local maximum) at $a \in U$, then $f''(a)(h, h) \geqslant 0$ (resp. $f''(a)(h, h) \leqslant 0$) for any $h \in X$.*

Proof. If f has a local minimum at a one knows that $f'(a) = 0$ and that there is some neighborhood V of 0 in X such that $f(a + h) \geqslant f(a)$ for any $h \in V$.

Let $h \in X$. For $t \in \mathbf{R}$ small enough one has $f(a + th) \geqslant f(a)$. Define the function $g : t \mapsto f(a + th)$ which also is of class $\mathscr{C}^2$. Applying Taylor formula (Theorem 9.13) one obtains

$$g(t) = g(0) + tg'(0) + \frac{t^2}{2} g''(0) + o(t^2).$$

This implies that

$$0 \leqslant \frac{2}{t^2}[f(a + th) - f(a)] = f''(a)(h, h) + o(1)$$

and taking the limit $t \longrightarrow 0$ one concludes that $f''(a)(h, h) \geqslant 0$. □

Definition 11.7. Let $f : U \longrightarrow \mathbf{R}$ be of class $\mathscr{C}^2$ on U. A point $a \in U$ is a *saddle-point* if it is a critical point and if for every neighborhood of a the function f attains values strictly greater and strictly smaller than $f(a)$.

Theorem 11.8. *Let $f : U \longrightarrow \mathbf{R}$ be of class $\mathscr{C}^2$ on U. Suppose that $a \in U$ is a critical point of f and that there is $\alpha > 0$ such that $f''(a)(h, h) \geqslant \alpha \|h\|^2$ for any $h \in X$. Then f has a strict local minimum at a.*

Proof. Since f'' is continuous there is $r > 0$ such that $\|f''(a + h) - f''(a)\| \leqslant \frac{\alpha}{2}$ for any $h \in B_X(0, r)$.

Define the function $g : x \mapsto f(a + x) - \frac{1}{2} f''(a)(x, x)$ from $B_X(0, r)$ into $\mathbf{R}$. The map g is also of class $\mathscr{C}^2$ and one has

$$g'(x)(h) = f'(a + x)(h) - f''(a)(x, h), \quad g''(x)(h, k) = f'(a + x)(h, k) - f''(a)(h, k)$$

and $g'(0) = f'(a) = 0$. Hence $\|g''(x)\| \leqslant \frac{\alpha}{2}$ for any $x \in B_X(0, r)$ and by Taylor's formula (Theorem 9.12) one gets

$$|g(x) - g(0) - g'(0)(x)| \leqslant \frac{1}{2} \sup_{y \in B_X(0,r)} \|g''(y)(x, x)\| \leqslant \frac{\alpha}{4} \|x\|^2,$$

which implies that

$$|f(a + x) - f(a) - \frac{1}{2} f''(a)(x, x)| \leqslant \frac{\alpha}{4} \|x\|^2.$$

Therefore, for any $x \in B_X(0, r)$ such that $x \neq 0$,

$$f(a + x) \geqslant f(a) + \frac{1}{2} f''(a)(x, x) - \frac{\alpha}{4} \|x\|^2 \geqslant f(a) + \frac{\alpha}{4} \|x\|^2 > f(a),$$

which completes the proof. □

Appendix A

Set theory

We adopt the naive point of view regarding set theory, and assume that the intuitive notion of a set, as being a collection of objects, is clear.

The fundamental notion is that of membership. If an object x belongs to a set A we shall write

$$x \in A,$$

in that case we also say that x is a member of A or that x is an element of A. Otherwise, if x does not belong to A we then write

$$x \notin A.$$

Two sets A and B are equal if and only if they have the same elements, which we denote by

$$A = B.$$

There exists a unique set which has no elements; we call it the *empty set* and denote

$$\varnothing.$$

Let A and B be sets. We say that A is a *subset* of B if every element of A is also an element of B, which we denote

$$A \subseteq B.$$

In that case we also say that B contains A, which is written $B \supseteq A$. If $A \subseteq B$ and $A \neq B$, we say that A is a *proper subset* of B and denote

$$A \subsetneq B.$$

In particular, one easily observes that $A = B$ if and only if $A \subseteq B$ and $B \subseteq A$.

Given a set A, one can form the set of all subsets of A, denoted by $\mathscr{P}(A)$ and called the *powerset* of A, namely

$$\mathscr{P}(A) = \{B : B \subseteq A\}.$$

A.1 Intersection, union, and complement

Let A and B be two given sets. We can form a new set consisting in all elements of A toghether with all elements of B, called the *union* of A and B and denoted by $A \cup B$. More precisely, the union of A and B is given by

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

We remind that "$x \in A$ or $x \in B$" means that x belongs to A, or to B, or to both of them.

We can also construct a new set consisting in all elements belonging to both A and B, called the *intersection* of A and B and denoted by $A \cap B$. More precisely, the intersection of A and B is given by

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

If A and B do not have elements in common, then their intersection is the empty set $\varnothing$ and we say that these sets are *disjoint*.

We easily obtain from the definition that

$$A \cup \varnothing = A \quad \text{and} \quad A \cap \varnothing = \varnothing$$

for every set A, as well as

$$A \cup B = B \cup A \quad \text{and} \quad A \cap B = B \cap A$$

for every sets A and B. Moreover, for any sets A, B and C, there holds

$$A \cup (B \cup C) = (A \cup B) \cup C \quad \text{and} \quad A \cap (B \cap C) = (A \cap B) \cap C.$$

Finally, we can define the *difference* of A with respect to B (or the *complement* of B relative to A) as the set consisting of all elements of A that do not belong to B, which is denoted by $A \setminus B$. Namely

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

In the particular case where B is a subset of A, then $A \setminus B = \{x \in A \mid x \notin B\}$ is simply called the *complement set* of B.

From the definition of the complement set, for any subsets A, B and C of a given set X the following hold

$$A \setminus B = A \cap (X \setminus B), \quad A \setminus (A \setminus B) = A \cap B,$$

and morever

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

A.2 Family of sets

Let $\mathscr{A}$ be a non-empty family of sets. We define the the intersection of $\mathscr{A}$ by

$$\bigcap \mathscr{A} = \bigcap_{A \in \mathscr{A}} A = \bigcap \{A : A \in \mathscr{A}\} = \{x \mid \forall A \in \mathscr{A},\ x \in A\},$$

and the union of the family $\mathscr{A}$ by

$$\bigcup \mathscr{A} = \bigcup_{A \in \mathscr{A}} A = \bigcup \{A : A \in \mathscr{A}\} = \{x \mid \exists A \in \mathscr{A},\ x \in A\}.$$

When dealing with collection of sets, we shall often use indexed family of sets. More precisely, consider a non-empty family $\mathscr{A}$ of sets. An index function for $\mathscr{A}$ is a surjective map φ from some set I (called the index set) to $\mathscr{A}$. We then denote $\mathscr{A} = \{A_i\}_{i \in I}$, where $A_i = \varphi(i)$ for all $i \in I$. From above we deduce that the intersection of the (indexed) family of subsets $\{A_i\}_{i \in I}$ is given by

$$\bigcap_{i \in I} A_i = \{x \mid \forall i \in I,\ x \in A_i\},$$

and its union by

$$\bigcup_{i \in I} A_i = \{x \mid \exists i \in I,\ x \in A_i\}.$$

Remark A.1. Given a collection of sets $\mathscr{A}$, it is always possible to write it as an indexed family of sets. Indeed one can take $\mathscr{A}$ itself as the index set and use the identity map on $\mathscr{A}$ to do the indexing.

Remark A.2. We shall adopt the convention that the union of an empty family of subsets of a given set X is equal to the empty set $\varnothing$; and that the intersection of an empty family of subsets of a given set X is equal to X itself.

Proposition A.3 (De Morgan's laws). *Let X be a set, $\{A_i\}_{i \in I}$ be a family of subsets of X. Then*

$$X \setminus \bigcap_{i \in I} A_i = \bigcup_{i \in I} (X \setminus A_i) \quad \text{and} \quad X \setminus \bigcup_{i \in I} A_i = \bigcap_{i \in I} (X \setminus A_i).$$

Proof. Let $x \in X$, then

$$\begin{aligned} x \in X \setminus \bigcap_{i \in I} A_i &\iff x \notin \bigcap_{i \in I} A_i \iff \exists i \in I, x \notin A_i \\ &\iff \exists i \in I, x \in X \setminus A_i \iff x \in \bigcup_{i \in I} (X \setminus A_i), \end{aligned}$$

which proves the first assertion. Moreover we have

$$\begin{aligned} x \in X \setminus \bigcup_{i \in I} A_i &\iff x \notin \bigcup_{i \in I} A_i \iff \forall i \in I, x \notin A_i \\ &\iff \forall i \in I, x \in X \setminus A_i \iff x \in \bigcap_{i \in I} (X \setminus A_i), \end{aligned}$$

which concludes the proof. ◻

Proposition A.4. *Let X be a set, $\{A_i\}_{i\in I}$ be a family of subsets of X, and B be a subset of X. Then*

$$(\bigcup_{i\in I} A_i) \cap B = \bigcup_{i\in I}(A_i \cap B) \quad \textit{and} \quad (\bigcap_{i\in I} A_i) \cup B = \bigcap_{i\in I}(A_i \cup B).$$

Proof. Let $x \in X$. On the one hand we have

$$\begin{aligned} x \in (\bigcup_{i\in I} A_i) \cap B &\iff x \in \bigcup_{i\in I} A_i \text{ and } x \in B \iff (\exists i \in I, x \in A_i) \text{ and } x \in B \\ &\iff \exists i \in I, x \in A_i \cap B \iff x \in \bigcup_{i\in I}(A_i \cap B), \end{aligned}$$

which gives the first property. On the other hand, we have

$$\begin{aligned} x \in (\bigcap_{i\in I} A_i) \cup B &\iff x \in \bigcap_{i\in I} A_i \text{ or } x \in B \iff (\forall i \in I, x \in A_i) \text{ or } x \in B \\ &\iff \forall i \in I, x \in A_i \cup B \iff x \in \bigcap_{i\in I}(A_i \cup B), \end{aligned}$$

which proves the second assertion. □

A.3 Product of sets

Consider two sets X and Y. We define the *(Cartesian) product* $X \times Y$ of the sets X and Y as the set of all ordered pairs (x, y) with $x \in X$ and $y \in Y$, more precisely

$$X \times Y = \{(x, y) \mid x \in X,\ y \in Y\}.$$

Proposition A.5. *If A and B are subsets of X and C and D are subsets of Y, then we have*

$$\begin{aligned} A \times (C \cap D) &= (A \times C) \cap (A \times D), \\ A \times (C \cup D) &= (A \times C) \cup (A \times D), \\ A \times (C \setminus D) &= (A \times C) \setminus (A \times D), \\ (A \cap B) \times (C \cap D) &= (A \times C) \cap (B \times D). \end{aligned}$$

Proof. Let $(x, y) \in X \times Y$. First of all we have

$$\begin{aligned} (x, y) \in A \times (C \cap D) &\iff x \in A, y \in C \cap D \\ &\iff x \in A, y \in C, y \in D \\ &\iff (x, y) \in A \times C,\ (x, y) \in A \times D \\ &\iff (x, y) \in (A \times C) \cap (A \times D), \end{aligned}$$

which proves the first equality above. Moreover

$$\begin{aligned} (x, y) \in A \times (C \cup D) &\iff x \in A, y \in C \cup D \\ &\iff x \in A, y \in C \text{ or } y \in D \\ &\iff (x, y) \in A \times C \text{ or } (x, y) \in A \times D \\ &\iff (x, y) \in (A \times C) \cup (A \times D), \end{aligned}$$

which proves the second assertion. For the third equality we observe that

$$\begin{aligned}(x,y)\in A\times(C\setminus D) &\iff x\in A, y\in C\setminus D\\ &\iff x\in A, y\in C, y\notin D\\ &\iff (x,y)\in A\times C,\ (x,y)\notin A\times D\\ &\iff (x,y)\in (A\times C)\setminus(A\times D).\end{aligned}$$

Finally, we have

$$\begin{aligned}(x,y)\in (A\cap B)\times(C\cap D) &\iff x\in A\cap B, y\in C\cap D\\ &\iff x\in A, x\in B, y\in C, y\in D\\ &\iff (x,y)\in A\times C,\ (x,y)\in B\times D\\ &\iff (x,y)\in (A\times C)\cap(B\times D),\end{aligned}$$

which concludes the proof of the last property. □

Consider now a family $\{X_i\}_{i\in I}$ of sets indexed by a non-empty set I. We then define the (Cartesian) product $\prod_{i\in I} X_i$ of the sets $\{X_i\}_{i\in I}$ by

$$\prod_{i\in I} X_i = \left\{ x : I \to \bigcup_{i\in I} X_i \;\middle|\; \forall i\in I,\ x(i)\in X_i \right\}.$$

We often denote an element $x\in\prod_{i\in I} X_i$ by $x=(x_i)_{i\in I}$ where $x_i = x(i)$. Moreover, for each $j\in I$ we define the canonical projection $\mathrm{pr}_j : \prod_{i\in I} X_i \to X_j$ given by $\mathrm{pr}_j(x) = x_j$.

A.4 Maps

Let X and Y be sets and $f : X\to Y$ a map. The map f is said to be:

- *injective* (or *one-to-one*) provided for any $x, x'\in X$, if $f(x) = f(x')$ then $x = x'$;
- *surjective* (or *onto*) if for any $y\in Y$ there is $x\in X$ such that $y = f(x)$;
- *bijective* (or a *one-to-one correspondence*) if it is both injective and surjective.

Proposition A.6. *Let $f : X\to Y$ and $g : Y\to Z$ be two maps between sets, then there holds:*

(i) *If f and g are injective, then $g\circ f$ is injective.*

(ii) *If f and g are surjective, then $g\circ f$ is surjective.*

(iii) *If $g\circ f$ is injective, then f is injective.*

(iv) *If $g\circ f$ is surjective, then g is surjective.*

As a consequence, taking $Z = X$ above, we deduce that if $g\circ f = \mathrm{id}_X$ and $f\circ g = \mathrm{id}_Y$, then g and f are bijective and $g = f^{-1}$ is the inverse of f.

Proof of Proposition A.6. (i) Let $x, x' \in X$ such that $g \circ f(x) = g \circ f(x')$. This means that $g(f(x)) = g(f(x'))$ and hence, since g is injective, one obtains $f(x) = f(x')$. Since f is injective one deduces that $x = x'$.

(ii) Let $z \in Z$. Since g is surjective there is $y \in Y$ such that $g(y) = z$. Since f is surjective there is $x \in X$ such that $y = f(x)$, therefore $g \circ f(x) = g(y) = z$.

(iii) Let $x, x' \in X$ such that $f(x) = f(x')$. Applying the map g to last equality, one obtains that $g \circ f(x) = g \circ f(x')$ and hence, since $g \circ f$ is injective, one deduces $x = x'$.

(iv) Let $z \in Z$. Since $g \circ f$ is surjective there is $x \in X$ such that $g \circ f(x) = z$. Therefore $y = f(x) \in Y$ and $g(y) = z$. □

A.4.1 Image and preimage

Let X and Y be sets and $f : X \longrightarrow Y$ a map. For any subset A of X, one defines the *image of A under f* as the subset of Y given by

$$f(A) = \{y \in Y \mid \exists x \in A, f(x) = y\} = \{f(x) : x \in A\}.$$

For any subset B of Y, one defines the *preimage* or *inverse image of B under f* as the subset of X given by

$$f^{-1}(B) = \{x \in X \mid f(x) \in B\}.$$

We easily obtain the following properties.

Proposition A.7. *For any subset $A \subseteq X$ one has $A \subseteq f^{-1}(f(A))$. Moreover, the map f is injective if and only if $A = f^{-1}(f(A))$ for any subset $A \subseteq X$.*

Proof. Let $A \subseteq X$ be a subset and $x \in A$, then $f(x) \in f(A)$ and hence $x \in f^{-1}(f(A))$. This proves $A \subseteq f^{-1}(f(A))$.

We give an example in which the previous inclusion is proper: Consider the map $f : \{0,1\} \longrightarrow \{0,1\}$ defined by $0 \mapsto 0$ and $1 \mapsto 0$. Then $\{0\} \subsetneq \{0,1\} = f^{-1}(f(\{0\}))$.

We now prove the second assertion. Suppose f injective and let $A \subseteq X$. Let $x \in f^{-1}(f(A))$, then $f(x) \in f(A)$. Therefore there is $a \in A$ such that $f(x) = f(a)$, which implies that $x = a \in A$, whence $f^{-1}(f(A)) \subseteq A$. From the first assertion one deduces $A = f^{-1}(f(A))$.

Conversely, assume that $A = f^{-1}(f(A))$ for any subset $A \subseteq X$. Let $x, y \in X$ such that $f(x) = f(y)$. Then, by hypothesis, $\{x\} = f^{-1}(f(\{x\})) = f^{-1}(\{f(x)\}) = f^{-1}(\{f(y)\}) = f^{-1}(f(\{y\})) = \{y\}$, therefore $x = y$ and thus f is injective. □

Proposition A.8. *For any subset $B \subseteq Y$ one has $f(f^{-1}(B)) \subseteq B$. Moreover, the map f is surjective if and only if $f(f^{-1}(B)) = B$ for any subset $B \subseteq Y$.*

Proof. Let $B \subseteq Y$ and $y \in f(f^{-1}(B))$, then there is $x \in f^{-1}(B)$ such that $y = f(x)$. Since $f(x) \in B$ we conclude that $y \in B$. This proves $f(f^{-1}(B))$.

We give an example in which the inclusion is proper: Consider the map $f : \{0,1\} \longrightarrow \{0,1\}$ defined by $0 \mapsto 0$ and $1 \mapsto 0$. Then $f(f^{-1}(\{0,1\})) = \{0\} \subsetneq \{0,1\}$.

We now prove the second assertion. Suppose f surjective and let $B \subseteq Y$. Let $y \in B$, then there is $x \in X$ such that $f(x) = y$, that is $x \in f^{-1}(B)$, whence $y \in f(f^{-1}(B))$. From the first assertion one deduces that $f(f^{-1}(B)) = B$.

Conversely, assume that $f(f^{-1}(B)) = B$ for any subset $B \subseteq Y$. Let $y \in Y$, then $f(f^{-1}(\{y\})) = \{y\}$ is non-empty, therefore there is $x \in X$ such that $f(x) = y$, and thus f is surjective. □

Proposition A.9. *For any subset $B \subseteq Y$ one has $X \setminus f^{-1}(B) = f^{-1}(Y \setminus B)$.*

Proof. Let $B \subseteq Y$ and $x \in X$. Then $x \notin f^{-1}(B)$ if and only if $f(x) \notin B$, which means $x \in f^{-1}(Y \setminus B)$. □

Consider a non-empty family $\{A_i\}_{i\in I}$ of subsets of X. We obtain the following properties concerning the image of unions and intersections.

Proposition A.10. *There holds*

$$f(\bigcup_{i\in I} A_i) = \bigcup_{i\in I} f(A_i) \quad \textit{and} \quad f(\bigcap_{i\in I} A_i) \subseteq \bigcap_{i\in I} f(A_i).$$

Moreover, if f is injective then the above inclusion is an equality.

Proof. Let $y \in Y$, then

$$\begin{aligned} y \in f(\bigcup_{i\in I} A_i) &\iff \exists x \in \bigcup_{i\in I} A_i, y = f(x) \iff \exists i \in I, \exists x \in A_i, y = f(x) \\ &\iff \exists i \in I, y \in f(A_i) \iff y \in \bigcup_{i\in I} f(A_i), \end{aligned}$$

which proves the first assertion.

Let $y \in f(\bigcap_{i\in I} A_i)$, then there is $x \in \bigcap_{i\in I} A_i$ such that $y = f(x)$, which means that $x \in A_i$ for any $i \in I$ and $y = f(x)$. This implies that for any $i \in I$ there is a $x \in A_i$ such that $y = f(x)$, thus $y \in f(A_i)$ for any $i \in I$. Therefore $y \in \bigcap_{i\in I} f(A_i)$. This proves the second assertion.

Let us give an example of a proper inclusion: Take $f : x \mapsto x^2$ from $\mathbf{R}$ to $\mathbf{R}$, $A_1 = \{1\}$ and $A_2 = \{-1\}$, then $f(A_1 \cap A_2) = \varnothing$ and $f(A_1) \cap f(A_2) = \{1\}$.

Finally, we assume that f is injective and we prove the inclusion $\bigcap_{i\in I} f(A_i) \subseteq f(\bigcap_{i\in I} A_i)$. Let $y \in \bigcap_{i\in I} f(A_i)$, then $y \in f(A_i)$ for any $i \in I$. This means that, for any $i \in I$, there is $x_i \in A_i$ such that $y = f(x_i)$. Since f is injective we deduce that all the x_i are the same, that is, there is $x \in X$ such that $x = x_i \in A_i$ for any $i \in I$. Therefore we have that $x \in \bigcap_{i\in I} A_i$ and $y = f(x)$, hence $y \in f(\bigcap_{i\in I} A_i)$. □

Finally, we consider a non-empty family $\{B_j\}_{j\in J}$ of subsets of Y and obtain the following properties concerning the preimage of unions and intersections.

Proposition A.11. *There holds*

$$f^{-1}(\bigcup_{j\in J} B_j) = \bigcup_{j\in J} f^{-1}(B_j) \quad \textit{and} \quad f^{-1}(\bigcap_{j\in J} B_j) = \bigcap_{j\in J} f^{-1}(B_j).$$

Proof. Let $x \in X$. On the one hand, we have

$$\begin{aligned} x \in f^{-1}(\bigcup_{j\in J} B_j) &\iff f(x) \in \bigcup_{j\in J} B_j \iff \exists j \in J, f(x) \in B_j \\ &\iff \exists j \in J, x \in f^{-1}(B_j) \iff x \in \bigcup_{j\in J} f^{-1}(B_j), \end{aligned}$$

which proves the first assertion.

On the other hand, there holds

$$\begin{aligned} x \in f^{-1}(\bigcap_{j\in J} B_j) &\iff f(x) \in \bigcap_{j\in J} B_j \iff \forall j \in J, f(x) \in B_j \\ &\iff \forall j \in J, x \in f^{-1}(B_j) \iff x \in \bigcap_{j\in J} f^{-1}(B_j), \end{aligned}$$

whoch concludes the proof of the second assertion. □

A.5 Relations

A *binary relation* on a set X is a subset $\mathscr{R}$ of $X \times X$. For all $x, y \in X$, if $(x, y) \in \mathscr{R}$ we then denote $x\mathscr{R}y$.

Definition A.12. An *equivalence relation* on X is a binary relation $\mathscr{R}$ that is reflexive, symmetric and transitive, i.e. satisfying:

(i) (*reflexive*) $x\mathscr{R}x$ for all $x \in X$;

(ii) (*symmetric*) $x\mathscr{R}y$ if and only if $y\mathscr{R}x$;

(iii) (*transitive*) $x\mathscr{R}y$ and $y\mathscr{R}z$ imply $x\mathscr{R}z$.

A partition of a set X is a collection $\{A_i\}_{i\in I}$ of non-empty subsets of X such that X is the disjoint union of A_i, that is $X = \bigcup_{i\in I} A_i$ and $A_i \cap A_j = \varnothing$ for all $i \neq j$.

Consider an equivalent relation $\mathscr{R}$ on a set X. We denote by $[x]$ the *equivalence class* of $x \in X$ under $\mathscr{R}$, which is defined by

$$[x] := \left\{ y \in X \mid x\mathscr{R}y \right\} .$$

Any element of $[x]$ is called a *representative* of $[x]$.

Proposition A.13. *If $\mathscr{R}$ is an equivalence relation on a set X then, for all $x, y \in X$, $x\mathscr{R}y$ if and only if $[x] = [y]$.*

Proof. Assume that $x\mathscr{R}y$. Let $z \in [x]$, thus $x \mathscr{R} z$ and then, by the transitivity property, we deduce that $y \mathscr{R} z$, that is $z \in [y]$. This proves that $[x] \subseteq [y]$. Arguing similarly, exchanging x and y, we also obtain $[y] \subseteq [x]$ and therefore $[x] = [y]$.

Conversely, assume that $[x] = [y]$. We have $y \in [y]$ since $y\mathscr{R}y$, hence $y \in [x]$ which means that $x\mathscr{R}y$. □

The set

$$X/\mathscr{R} = \{[x] : x \in X\}$$

of equivalence classes of X is called the *quotient set* of X by $\mathscr{R}$. As a consequence of Proposition A.13, $X/\mathscr{R}$ forms a partition of X. There is a canonical projection

$$\begin{array}{rccc} \pi : & X & \to & X/\mathscr{R} \\ & x & \mapsto & [x] \end{array}$$

that maps each element into its equivalence class.

A.6 Orders

Definition A.14. A binary relation $\mathscr{R}$ on a set X is a *partial order* if it is:

(i) (*reflexive*) $x\mathscr{R}x$ for all $x \in X$;

(ii) (*transitive*) $x\mathscr{R}y$ and $y\mathscr{R}z$ imply $x\mathscr{R}z$;

(iii) (*anti-symmetric*) $x\mathscr{R}y$ and $y\mathscr{R}x$ imply $x = y$.

A partial order on a set is usually denoted by $\leqslant$. If $\leqslant$ is a partial order on a set X, we say that $(X, \leqslant)$ is a *partially ordered set*. We then denote $x \geqslant y$ for $y \leqslant x$, $x < y$ for "$x \leqslant y$ and $x \neq y$", and $x > y$ for $y < x$. The partial order $\leqslant$ is said to be *total* if every two elements of X can be compared to each order, namely if for all $x, y \in X$ we have $x \leqslant y$ or $y \leqslant x$.

We now consider a partially ordered set $(X, \leqslant)$ and a subset $Y \subseteq X$. One easily observes that Y can be endowed with a partial order given by the restriction of $\leqslant$ to Y, namely $\leqslant \cap (Y \times Y)$, which we often still denote by $\leqslant$.

An element $m \in Y$ is a *maximal element* of Y if for any $x \in Y$, $m \leqslant x$ implies $m = x$. An element $u \in X$ is an *upper bound* of Y if for any $x \in Y$ one has $x \leqslant u$. If $u \in X$ is an upper bound if Y and $u \in Y$ then one says that u is a *greatest element* of Y.

In a similar fashion, one can define the notions of *minimal element*, *lower bound*, and *least element*. More precisely, $m \in Y$ is a *minimal element* of Y if for any $x \in Y$, $x \leqslant m$ implies $m = x$; $\ell \in X$ is an *lower bound* of Y if for any $x \in Y$ one has $\ell \leqslant x$; $\ell \in X$ is a *least element* of Y if ℓ is a lower bound of Y and $\ell \in Y$.

One says that Y is a *totally ordered subset* (or *chain*) if the restriction of $\leqslant$ to Y is total, i.e. for any $x, y \in Y$ one has $x \leqslant y$ or $y \leqslant x$.

Definition A.15. Let $(X, \leqslant)$ be a a partially ordered set. We say that $(X, \leqslant)$ is *well-ordered*, or that $\leqslant$ is a *well-order* on X, if $\leqslant$ is total and if every non-empty subset of X has a least element.

If C is a chain (totally ordered subset) in a partially ordered set $(X, \leqslant)$ and $x \in C$, we define the set

$$C_{<x} = \{y \in C \mid y < x\},$$

which is called an initial segment in C.

Proposition A.16. *Let $(X, \leqslant)$ be a partially ordered set.*

(i) *If A is a well-ordered subset of X and $x \notin A$ is an upper bound of A in X, then $A \cup \{x\}$ is well-ordered.*

(ii) *If $\mathscr{C}$ is a collection of well-ordered subsets of X such that for all $A, B \in \mathscr{C}$, either A is an initial segment of B or B is an initial segment of A, then the union $\bigcup \mathscr{C} = \bigcup\{C : C \in \mathscr{C}\}$ is well-ordered.*

Proof. (i) It is clear that $A \cup \{x\}$ is totally ordered. Let B be a subset of $A \cup \{x\}$. If $B = \{x\}$ then $x = \min B$; otherwise $\min A \cap B = \min B$.

(ii) Let $a, b \in \bigcup\{C : C \in \mathscr{C}\}$, then there are $A, B \in \mathscr{C}$ such that $a \in A$ and $b \in B$. By hypothesis we have either $A \subseteq B$ or $B \subseteq A$, in any case the elements a anb b belong to the same subset A or B which is totally ordered, hence $a \leqslant b$ or $b \leqslant a$.

Let A be a subset of the union $\bigcup\{C : C \in \mathscr{C}\}$, thus there is $C \in \mathscr{C}$ such that $A \cap C \neq \varnothing$. Let $m = \min A \cap C$, we shall prove that $m = \min A$. Let $x \in A \setminus C$ then there is $C' \in \mathscr{C}$ such that $x \in C'$. If C' is an initial segment of C, then $C' \subseteq C$ and thus $m \leqslant x$. Otherwise C is an initial segment of C', which implies that $x \in C' \setminus C$ and hence $m \leqslant x$ again. □

A.7 Size of sets

Definition A.17. A set is said to be *finite* if it is in bijection with $\{0, 1, \ldots, n-1\}$ for some natural number $n \in \mathbf{N}$. A set is said to be *infinite* if it is not finite.

Definition A.18. A set is said to be *countable* if it is in bijection with some subset of $\mathbf{N}$. A set is said to be *uncountable* if it is not countable.

In other words, a set X is countable if either X is a finite set, or X is an infinite set and there is a bijective map from X onto $\mathbf{N}$. Equivalently, X is countable if there is a injection from X into $\mathbf{N}$.

Proposition A.19. *There is no surjective map from a non-empty set X onto $\mathscr{P}(X)$.*

As a consequence of this result, we deduce that there is no surjection from $\mathbf{N}$ onto $\mathscr{P}(\mathbf{N})$, which is equivalent to say that there is no injection from $\mathscr{P}(\mathbf{N})$ to $\mathbf{N}$, and thus $\mathscr{P}(\mathbf{N})$ is uncountable.

Proof of Proposition A.19. By way of contradiction, assume that $f : X \to \mathscr{P}(X)$ is surjective. Consider the set $A = \{x \in X \mid x \notin f(x)\}$. Since f is surjective, there is $a \in X$ such that $f(a) = A$. On the other hand, by construction we have $a \in A$ if and only if $a \notin f(a) = A$, thus a contradiction. □

Proposition A.20. *Every finite product of countable sets is countable.*

Proof. We only show that the product of two countable sets is countable, the general case being deduced from this by induction. Let X and Y be countable sets. We know that there is an injection $X \to \mathbf{N}$ and an injection $Y \to \mathbf{N}$, thus there exists an injection $\phi : X \times Y \to \mathbf{N} \times \mathbf{N}$. But $\mathbf{N} \times \mathbf{N}$ is countable, since the map $f : \mathbf{N} \times \mathbf{N} \to \mathbf{N}$ given by $f(n, m) = 2^n 3^m$ is injective. Therefore the map $f \circ \phi : X \times Y \to \mathbf{N}$ is injective, thus $X \times Y$ is countable. □

Proposition A.21. *Every countable union of countable sets is countable.*

Proof. Let $\{X_n\}_{n \in \mathbf{N}}$ be a countable family of countable sets and consider their union $X = \bigcup_{n \in \mathbf{N}} X_n$. Define, for any $n \in \mathbf{N}$, the set $\mathcal{I}_n$ of all injections from X_n to $\mathbf{N}$. We remark that $\mathcal{I}_n \neq \varnothing$ since X_n is countable.

Thanks to the Axiom of Choice in Section A.8 below (actually we need only a weaker version of it, called the "Axiom of Countable Choice") there is a sequence $(i_n)_{n \in \mathbf{N}}$ such that $i_n \in \mathcal{I}_n$ for all $n \in \mathbf{N}$. Define the map

$$\begin{array}{rccc} \phi : & X & \to & \mathbf{N} \times \mathbf{N} \\ & x & \mapsto & (n, i_n(x)) \end{array}$$

where n is the smallest natural number verifying $x \in X_n$. By construction ϕ is an injection, since every i_n is an injection. But there is an injection $f : \mathbf{N} \times \mathbf{N} \to \mathbf{N}$ since $\mathbf{N} \times \mathbf{N}$ is countable. Therefore $f \circ \phi : X \to \mathbf{N}$ is a injective map and thus X is countable. □

A.8 Axiom of Choice

We shall accept the Axiom of Choice (see below) and use it freely through these notes either in its original form or in the form of Zorn's[1] Lemma (Theorem A.23).

One should remark that, over the Zermelo[2]-Fraenkel[3] axioms of set theory, the Axiom of Choice is equivalent to Zorn's Lemma and also to Zermelo's Theorem (Theorem A.24). We shall briefly describe these three assertions and prove their equivalence below.

Axiom of Choice. For every set X there is a map $\varphi : \mathscr{P}(X) \setminus \{\varnothing\} \to X$ such that $\varphi(A) \in A$ for all non-empty subset A of X.

Such a map φ is called a *choice function* over X.

Proposition A.22. *The followin are equivalent:*

(i) *Axiom of Choice ;*

(ii) *The product of any non-empty family of non-empty sets is itself non-empty.*

Proof. Assume the Axiom of Choice holds. Let $\{X_i\}_{i \in I}$ be a non-empty collection of non-empy sets. Let $X = \bigcup_{i \in I} X_i$ be their union, then there exists a choice function φ over X, that is, $\varphi(A) \in A$ for every non-empty subset A of X. In particular

[1] Max August Zorn (1906–1993) was a German mathematician (later naturalized American) who contributed to set theory.

[2] Ernst Zermelo (1871–1953) was a German mathematician who made fundamental contributions to the foundations of mathematics. He is best known for the *Zermelo-Fraenkel axioms* of set theory, and for the *Zermelo's theorem* (or *well-ordering theorem*).

[3] Abraham Fraenkel (1891–1965) was a German-Israeli mathematician who made major contributions to set theory. He is best known for the *Zermelo-Fraenkel axioms* of set theory.

one has $\varphi(X_i) \in X_i$ for any $i \in I$, that is $(\varphi(X_i))_{i\in I} \in \prod_{i\in I} X_i$, which implies that $\prod_{i\in I} X_i$ is non-empty.

Conversely, assume that the product of any family of non-empty sets is non-empty. Let X be a set, then the product $\prod_{A\subseteq \mathscr{P}(X)\setminus\{\varnothing\}} A$ is non-empty. Therefore, every element $\varphi = (\varphi(A))_{A\subseteq\mathscr{P}(X)\setminus\{\varnothing\}}$ is a choice function over X. □

Theorem A.23 (Zorn's Lemma)**.** *If X is a partially ordered set with the property that every chain (i.e. totally ordered subset) has an upper bound, then X contains at least one maximal element.*

Proof. We shall prove Zorn's lemma assuming the Axiom of Choice. Let $(X, \leqslant)$ be a partially ordered set such that every chain has an upper bound. By way of contradiction, assume that X does not contain a maximal element. If C is a chain in X then it has an upper bound $m \in X$; thus one can choose $x \in X$ such that $m < x$, which then implies $c < x$ for every $c \in C$. By the axiom of choice one can choose a function φ that maps every chain C in X into a strict upper bound $\varphi(C)$ as explained above.

We say that a subset A of X is a φ-set if

(i) $\leqslant$ is a well-order on A ;

(ii) $x = \varphi(A_{<x})$ for every $x \in A$.

We claim that if A and B are distinct φ-sets of X, then one of these sets is an initial segment of the other. Indeed we may assume that $A \setminus B$ is non-empty and then consider x the least element of $A \setminus B$, thus $A_{<x} = \{y \in A \mid y < x\} \subseteq B$. On the other hand if $B \setminus A_{<x}$ is non empty, let y be its least element. For any $b \in B_{<y}$ and $a \in A$ such that $a < b$, one clearly has $a \in B_{<y}$. Letting z be the least element of $A \setminus B_{<y}$, one obtains $A_{<z} = B_{<y}$ so in particular $z \leqslant x$. Observing that $z = \varphi(A_{<z}) = \varphi(B_{<y}) = y$ and $y \in B$, one cannot have $z = x$ and hence $z < x$. This implies that $y = z \in A_{<x}$, contradicting the choice of y. This proves that $B = A_{<x}$.

Now let U be the union of all φ-sets. We observe that if A is a φ-set of X and $x \in A$, then for every y such that $y < x$ one has either $y \in A$ or $y \notin U$. It follows that U is a φ-set of X, and thus we deduce that $U \cup \{\varphi(U)\}$ is also a φ-set. This implies that $\varphi(U) \in U$, which contradicts the fact that $\varphi(U)$ is a strict upper bound of U. □

As a consequence of Zorn's lemma one can prove Zermelo's theorem.

Theorem A.24 (Zermelo's Theorem or well-ordering theorem)**.** *Every set can be well-ordered.*

Proof. Let X be a non-empty set and denote by $\mathscr{A}$ the set of pairs $(A, \leqslant_A)$ where A is a subset of X and $\leqslant_A$ is a well-order on A. Define a partial order $\preceq$ on $\mathscr{A}$ by $(A, \leqslant_A) \preceq (B, \leqslant_B)$ if and only if A is an initial segment of B and $\leqslant_A$ is the restriction of $\leqslant_B$ to A.

For every chain $\mathscr{C} \subseteq \mathscr{A}$, consider $U = \cup\{A : (A, \leqslant_A) \in \mathscr{C}\}$ and $\leqslant_U = \cup\{\leqslant_A : (A, \leqslant_A) \in \mathscr{C}\}$ so that $(U, \leqslant_U) \in \mathscr{A}$ is an upper bound for $\mathscr{C}$. By Zorn's lemma $\mathscr{A}$ has a maximal element $(M, \leqslant_M)$ and it only remains to check that $M = X$. By way of contradiction, suppose there exists an element $x \in X \setminus M$. We then consider the set $M' = M \cup \{x\}$ and the relation $\leqslant_{M'}$ as the unique partial order on M' such that $\leqslant_{M'} \cap (M \times M) = \leqslant_M$ and x is a strict upper bound of M. Hence we have $(M, \leqslant_M) \prec (M', \leqslant_{M'})$, which contradicts the maximality of $(M, \leqslant_M)$. Therefore $M = X$ and $\leqslant_M$ is a well-order on X. □

From Theorems A.23 and A.24 we have that "Axiom of Choice ⇒ Zorn's Lemma ⇒ Zermelo's Theorem". We shall prove below that "Zermelo's Theorem ⇒ Axiom of Choice", concluding then that all these three assertions are equivalent.

Proof of "Zermelo's Theorem ⇒ Axiom of Choice". If X is well-ordered then $A \mapsto \min A$ is a choice function on X. □