REFRESHER M1 MSIAM

C. DAPOGNY* AND S. MEIGNEN*

* *Laboratoire Jean Kuntzmann, Université Grenoble-Alpes*

Contents

Introduction

This course is a refresher of a few selected topics in differential calculus, which are fundamental prerequisites for graduate analysis courses. It is mainly oriented towards calculus and applications; in particular, it does not cover more theoretical material such as the Banach theorems at the principle of functional analysis. On the other hand, whenever possible, concrete applications and algorithmic principles are extracted from the results. We have focused on intuitive (albeit rigorous) presentation of the tackled topics, with the hope that it be more easily understandable by the reader. As a result, the statements may not be "optimal" in terms of assumptions, and the proofs may sometimes be lengthy. We hope, however, that they allow to exemplify important techniques or interesting applications of the fundamental concepts. At the end of each part, some exercises are proposed, ranging from very simple verifications of properties from the main text, to more advanced and conceptual ones, passing through practical calculations.

The central notion of the course is that of differential, as a suitable generalization to abstract spaces of the derivative of functions on the real line. We heavily emphasize on the basic concepts and facts in Section 2, before turning to more advanced topics (such as the mean value theorem, higher-order derivatives and the Taylor formulas) in Section 4. Meanwhile, we recall in Section 3 a few fundamental facts from the theory of Lebesgue integration. The next sections are devoted to slightly more advanced topics, which are also good opportunities to apply the concepts of the first sections: the fixed point theorem is presented in Section 5, together with a few interesting applications. The next Section 6 is devoted to the fundamental implicit function theorem, and illustrations of some of its multiple applications. Then, we briefly broach the subject of mathematical optimization in Section 7, mainly dealing with the statement of necessary conditions for local optimality, before finally turning to differential calculus in regular domains of $\mathbb{R}^d$ in Section 8.

Let us warn the reader that the material presented in these notes goes way beyond the contents taught during the lectures. It is our hope that the reader will find all the mathematical details of the sometimes too fast derivations, as well as some interesting complements. Comments, suggestions, and reports about mistakes will be gratefully received at the following email address:

`charles.dapogny[AT]univ-grenoble-alpes.fr`

1. Reminders: normed vector spaces, convergence and Banach spaces

This first section gathers haphazardly a few facts from undergraduate analysis; it is not meant to be read in linear fashion, but rather to be consulted as support if the need becomes apparent in the study of the subsequent sections.

1.1. General notation

The following notation and conventions are used throughout the text:

- $\mathbb{R}$ denotes the set of real numbers and $\mathbb{R}_+$ is the subset of non negative real numbers.
- For $a, b \in \mathbb{R} \cup \{-\infty, \infty\}$, (a, b) and $[a, b]$ respectively denote the open and closed intervals of $\mathbb{R}$ with ends a and b.
- Unless specified otherwise, all the considered vector spaces in these lecture notes are based on the field $\mathbb{R}$ of real numbers.

- $\mathbb{N}$ is the set of non negative integers: $\mathbb{N} = \{0, 1, \ldots\}$.
- For any integer $d \geq 1$, the canonical basis $\{e_i\}_{i=1,\ldots,d}$ of the Euclidean space $\mathbb{R}^d$ is defined by:

$$e_i = (0, \ldots, \underbrace{1}_{i^{\text{th}} \text{ position}}, \ldots, 0).$$

The canonical inner product $\langle \cdot, \cdot \rangle$ of $\mathbb{R}^d$ is defined by:

$$\forall x = \sum_{i=1}^{d} x_i e_i,\ y = \sum_{i=1}^{d} y_i e_i \in \mathbb{R}^d, \quad \langle x, y \rangle = \sum_{i=1}^{d} x_i y_i,$$

and the corresponding Euclidean norm reads:

$$\forall x == \sum_{i=1}^{d} x_i e_i \in \mathbb{R}^d, \quad |x| := \left(x_1^2 + \ldots + x_d^2\right)^{\frac{1}{2}}.$$

- The vector product $u \wedge v$ of two vectors $u = (u_1, u_2, u_3)$, $v = (v_1, v_2, v_3) \in \mathbb{R}^3$ is the vector in $\mathbb{R}^3$ defined by

$$u \wedge v = (u_2 v_3 - u_3 v_2, u_3 v_1 - u_1 v_3, u_1 v_2 - u_2 v_1).$$

- For any integers $p, q \geq 1$, $M_{p,q}(\mathbb{R})$ is the vector space of $p \times q$ matrices with real entries. An element $M \in M_{p,q}(\mathbb{R})$ may be alternatively denoted via its entries $M = (m_{ij})_{\substack{i=1,\ldots,p \\ j=1,\ldots,q}}$ with $m_{ij} \in \mathbb{R}$. In the latter notation, the reference to the numbers p, q of columns and rows of M is omitted when clear from the context, and we simply write $M = (m_{ij})$.
- For $p, q \geq 1$ and any matrix $M = (m_{ij}) \in M_{p,q}(\mathbb{R})$, we denote by $M^T \in M_{q,p}(\mathbb{R})$ the transpose of M, i.e. the matrix with entries

$$\forall i = 1, \ldots q,\ j = 1, \ldots, p, \quad (M^T)_{ij} = m_{ji}.$$

- When $p = q$, we simply denote by $M_p(\mathbb{R})$ the space of square $p \times p$ matrices.
- For any integer $p \geq 1$, $\mathrm{I}_p \in M_p(\mathbb{R})$ is the identity $p \times p$ matrix.
- The trace of a $p \times p$ matrix $M = (m_{ij})$ is the sum of its diagonal entries:

$$\mathrm{tr}(M) = \sum_{i=1}^{p} m_{ii}.$$

- The determinant of a $p \times p$ matrix $M = (m_{ij})$ is denoted by:

$$\det(M) = \begin{vmatrix} m_{11} & m_{12} & \ldots & m_{1p} \\ \vdots & & & \vdots \\ m_{p1} & m_{p2} & \ldots & m_{pp} \end{vmatrix}.$$

- For any $p \times p$ matrix $M = (m_{ij})$, the associated (i, j) minor Δ_{ij} is the determinant of the $(p-1) \times (p-1)$ matrix obtained from M by deleting the i^{th} row and j^{th} column:

$$\Delta_{ij} = \begin{vmatrix} m_{11} & m_{12} & \ldots & m_{1j-1} & m_{1j+1} & \ldots & m_{1p} \\ \vdots & & & & & & \vdots \\ m_{i-11} & m_{i-12} & \ldots & m_{i-1j-1} & m_{i-1j+1} & \ldots & m_{i-1p} \\ m_{i+11} & m_{i+12} & \ldots & m_{i+1j-1} & m_{i+1j+1} & \ldots & m_{i+1p} \\ \vdots & & & & & & \vdots \\ m_{p1} & m_{p2} & \ldots & m_{pj-1} & m_{pj+1} & \ldots & m_{pp} \end{vmatrix}..$$

- For any square $p \times p$ matrix $M = (m_{ij}) \in M_p(\mathbb{R})$, $\mathrm{com}(M) \in M_p(\mathbb{R})$ is the associated cofactor matrix, that is, the $p \times p$ matrix whose (i, j) entry equals $(-1)^{i+j} \Delta_{ij}$. Let us recall the following fundamental identity from linear algebra:

$$\forall M \in M_p(\mathbb{R}), \quad M^T \mathrm{com}(M) = \det(M) \mathrm{I}_p.$$

- A sequence is an enumeration $\{x_0, x_1, \ldots\}$ of objects that is indexed by $\mathbb{N}$; it is either denoted between braces $\{x_n\}_{n \in \mathbb{N}}$, or most often via its general term x_n.

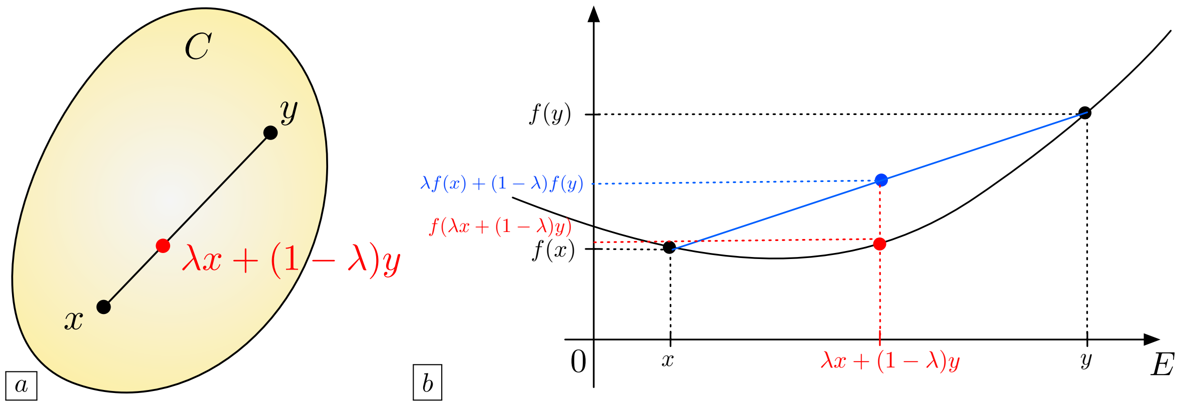

FIGURE 1. *(a) The segment joining two points x, y in a convex subset C of a vector space E completely lies in C; (b) the graph $\{(x, f(x)),\ x \in E\}$ of a convex function lies below any line segment joining two points $(x, f(x))$, $(y, (y))$ of the graph.*

- A subsequence of a sequence $\{x_n\}_{n\in\mathbb{N}}$ is another sequence obtained by retaining only some of the terms among the $x_0, x_1, \dots$. For instance, $\{x_2, x_5, x_8, \dots\}$ is a subsequence of $\{x_n\}_{n\in\mathbb{N}}$. Formally, such a subsequence is defined by means of an "extraction mapping" $\varphi : \mathbb{N} \to \mathbb{N}$, i.e. an increasing mapping identifying the retained indices among $\{0, 1, \dots\}$. The subsequence $\{x_{\varphi(n)}\}_{n\in\mathbb{N}}$ is sometimes denoted by $\{x_{n_k}\}_{k\in\mathbb{N}}$ for short.
- For any subset A of a vector space E, we denote by $\mathbb{1}_A$ the characteristic function of A, that is:

$$\forall x \in E, \quad \mathbb{1}_A(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{otherwise.} \end{cases}$$

1.2. A few facts from real analysis

1.2.1. *Convex sets and functions*

In this short subsection, we recall the notions of convex sets and functions in a vector space, which are ubiquitous in differential calculus.

Definition 1.1. *Let E be a vector space.*

- *A subset $C \subset E$ is called convex if*

$$\forall x, y \in C, \quad \forall \lambda \in [0,1], \ \lambda x + (1-\lambda)y \in C.$$

- *A real-valued function $f : C \to \mathbb{R}$ defined on a convex subset $C \subset E$ is called convex if*

$$\forall x, y \in C, \ \forall \lambda \in [0,1], \quad f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y).$$

- *A real-valued function $f : C \to \mathbb{R}$ defined on a convex subset $C \subset E$ is called strictly convex if*

$$\forall x, y \in C, \ \forall \lambda \in (0,1), \quad f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda) f(y);$$

Intuitively, a convex set C is such that whenever two points x, $y \in E$ belong to C, the whole line segment $\{(1-\lambda)x + \lambda y,\ \lambda \in [0,1]\}$ is included in C. A convex function $f : C \to \mathbb{R}$ is characterized by the fact that its graph $\{(x, f(x)),\ x \in C\} \subset E \times \mathbb{R}$ lies locally below any line segment between two of its points $(x, f(x))$, $(y, f(y))$, see Fig. 1 for an illustration.

1.2.2. *The mean value theorem for real-valued functions on the real line*

We now recall the mean value theorem for a real-valued functions f defined on an interval $[a, b] \subset \mathbb{R}$. The latter is a key tool to express the variations of f over $[a, b]$ as the value of the derivative f' at an intermediate point $a < c < b$, see Fig. 2 for an illustration.

Theorem 1.1 (Mean-value theorem)**.** *Let $a < b$, and let $f : [a, b] \to \mathbb{R}$ be a real-valued function which is continuous on the closed interval $[a, b]$, and differentiable on the open interval (a, b). Then there exists $\theta \in (0, 1)$ such that*

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

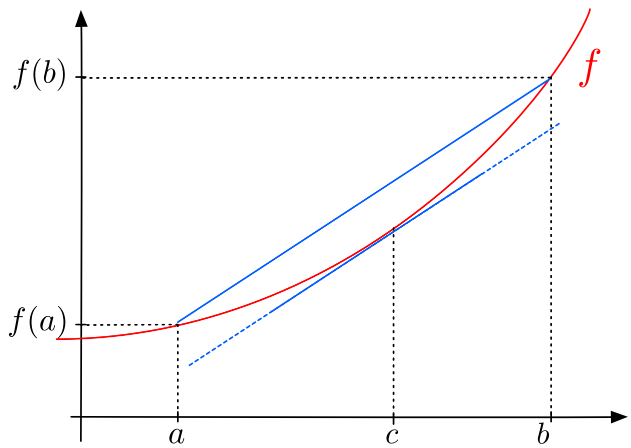

FIGURE 2. *The slope of the solid blue line between the points $(a, f(a))$ and $(b, f(b))$ equals $\frac{f(b)-f(a)}{b-a}$; the mean value Theorem 1.1 asserts that there exists an intermediate point $c := a + \theta(b - a)$, $\theta \in (0, 1)$, such that the derivative $f'(c)$ of f at c is exactly equal to this slope.*

1.3. **Analysis in normed vector spaces**

We now slip into the context of normed vector spaces, in which we investigate convergent sequences. We notably recall the fundamental Cauchy criterion and the related notion of Banach space.

1.3.1. *General facts about normed vector spaces*

Let us start with a few definitions.

Definition 1.2. *Let E be a (real) vector space; a norm $||\cdot||$ on E is a mapping $E \to \mathbb{R}_+$ which satisfies the following three conditions:*

- *(Positive homogeneity): For all $\lambda \in \mathbb{R}$ and $x \in E$, one has $||\lambda x|| = |\lambda| ||x||$;*
- *(Positive definiteness): For all $x \in E$, $||x|| = 0$ if and only if $x = 0$;*
- *(Triangle inequality): For all x, $y \in E$, one has: $||x + y|| \leq ||x|| + ||y||$.*

A vector space E equipped with a norm $||\cdot||$ is called a normed vector space.

The following lemma, sometimes referred to as the “second triangle inequality”, is a straightforward, albeit fundamental consequence of the definition of norm.

Lemma 1.1. *Let $(E, ||\cdot||)$ be a normed vector space; then:*

$$\forall x, y \in E, \quad |||x|| - ||y||| \leq ||x + y||.$$

Proof. For all x, $y \in E$, the triangle inequality yields

$$||x|| = ||x + y - y|| \leq ||x + y|| + ||y||,$$

and so $||x|| - ||y|| \leq ||x + y||$. Applying the same argument after inverting the roles of x and y, we obtain the desired statement. □

One given vector space can generally be endowed with multiple different norms, inducing as many different structures of normed vector space on E. The following notion helps in identifying those norms inducing a “similar” structure on E.

Definition 1.3. *Let E be a vector space, and let $||\cdot||_1$, $||\cdot||_2 : E \to \mathbb{R}_+$ be two norms on E. These norms are called equivalent if there exist positive real numbers $0 < \alpha \leq \beta$ such that*

$$\forall x \in E, \quad \alpha ||x||_2 \leq ||x||_1 \leq \beta ||x||_2.$$

A very particular feature of finite-dimensional spaces is the equivalence of all possible norms, as we recall in the next result.

Proposition 1.1. *Let E be a finite-dimensional space; all norms defined on E are then equivalent.*

Proof. See Exercise 1.1. □

Unfortunately, such a statement is utterly false in an infinite-dimensional space E; its properties, starting from its *topology* (as we shall see in the next sections), depend heavily on the chosen norm.

1.3.2. *Topology on normed vector spaces*

The structure of a norm on a vector space allows to make out different sorts of subsets, a classification which is referred to as the *topology* of the space.

Definition 1.4. *Let $(E, ||\cdot||_E)$ be a normed vector space.*

- *For all $x \in E$, and $r > 0$, we denote by*
$$B(x,r) := \{\ y \in E,\ \ ||y-x|| \ < r\},\ \text{and}\ \overline{B(x,r)} = \{\ y \in E,\ \ ||y-x|| \ \leq r\}$$
the open and closed balls with center x and radius r, respectively.
- *A subset $U \subset E$ is called open if, for all $x \in U$, there exists $r > 0$ such that the ball $B(x,r)$ is enclosed in U.*
- *A subset $F \subset E$ is called closed if its complement $U := E \setminus F$ is open.*
- *Let $A \subset E$ be an arbitrary subset; a neighborhood of A is an open subset U of E such that $A \subset U$.*

Remark 1.1. *The closedness of a subset $F \subset E$ is often used under the following sequential form: F is closed if and only if, for every sequence $x_n \in F$ which converges to some element $x_\infty \in E$, the limit x_∞ actually belongs to F. The verification of the equivalence between this property and Definition 1.4 is left to the reader.*

We next turn to the definition of continuous and uniformly continuous functions on a normed vector space, which are the immediate generalizations of their perhaps more familiar one-dimensional counterparts.

Definition 1.5. *Let $(E, ||\cdot||_E)$ and $(F, ||\cdot||_F)$ be normed vector spaces; let $A \subset E$ and $f : A \to F$ be a function;*

- *The function f is said to be continuous at a particular point $x \in A$ if*
$$\forall \varepsilon > 0,\ \exists \delta > 0,\ \forall y \in A, \quad ||x-y||_E \leq \delta \Rightarrow ||f(x) - f(y)||_F \leq \varepsilon.$$
The function f is called continuous on A if it is continuous at every point $x \in A$.
- *The function f is called uniformly continuous on A if for all $\varepsilon > 0$, there exists $\delta > 0$ such that*
$$\forall x, y \in I, \qquad ||x-y||_E \leq \delta \Rightarrow ||f(x) - f(y)||_F \leq \varepsilon.$$

The difference between the continuous and uniformly continuous functions on $A \subset E$ is a little subtle. The continuity of f on A solely requires that, for each given point $x \in A$, $f(y)$ be as close to $f(x)$ as desired, provided y is close enough from x; formally: for a given error margin $\varepsilon > 0$, $f(y)$ takes values close to $f(x)$ (i.e. within a range ε from $f(x)$) provided y is "close enough" to x (i.e. the distance between x and y is less than δ), but how close y has to be from x for this to happen depends on the considered point x. In particular, one may imagine that f take larger and larger variations as y gets near the border of A. On the contrary, when f is uniformly continuous on A, for each given tolerance $\varepsilon > 0$, there is a distance $\delta > 0$ such that, as soon as the distance between *any* two points x and $y \in A$ is less than δ, the difference between $f(x)$ and $f(y)$ is less than ε. In other terms, f cannot "vary too fast" on A, see Fig. 3 for an illustration in the case where $E = \mathbb{R}$.

Uniformly continuous functions are obviously continuous, but the converse is false in general. However, as we shall recall in Section 1.3.4, when the considered function f is defined on a compact subset K of E, the remarkable Heine's theorem asserts that both notions coincide.

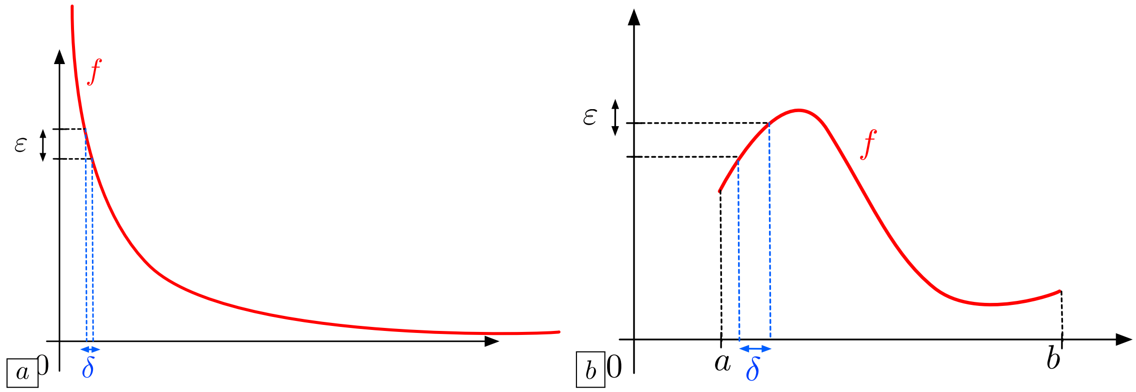

FIGURE 3. *One-dimensional illustrations of the notions of continuity and uniform continuity. (a) The function $f : x \mapsto \frac{1}{x}$ is continuous on $(0, \infty)$, but not uniformly continuous on this interval: as points x, y gets near 0, to get variations between $f(x)$ and $f(y)$ less than a given threshold ε requires x and y to be closer and closer from one another; (b) Heine's Theorem 1.4 states that a continuous function on a compact interval is uniformly continuous, since its variations are controlled.*

One third important notion of continuity is the following.

Definition 1.6. *Let $(E, ||\cdot||_E)$ and $(F, ||\cdot||_F)$ be normed vector spaces and let $A \subset E$. A function $f : A \to F$ is called Lipschitz continuous with ratio $C > 0$ if the following relation holds:*

$$\forall x, y \in A, \quad ||f(x) - f(y)||_F \leq C||x - y||_E.$$

Lipschitz continuity is somehow a quantitative version of uniform continuity, as it gives an explicit measurement of "how much" x and y have to be close from one another so that the distance between $f(x)$ and $f(y)$ be less than a given threshold. Formally,

$$\text{For all } \varepsilon > 0, \quad ||x - y||_E \leq \frac{\varepsilon}{C} \ \Rightarrow \ ||f(x) - f(y)||_E \leq \varepsilon;$$

in particular, a Lipschitz continuous function is uniformly continuous, thus continuous.

1.3.3. *Convergent and Cauchy sequences in normed vector spaces*

We now recall the notions of convergent and Cauchy sequences in a normed vector space.

Definition 1.7. *Let $(E, ||\cdot||_E)$ be a normed vector space, and let x_n be a sequence of elements of E;*

- *The sequence x_n is said to converge to an element $x_\infty \in E$ if,*

$$\forall \varepsilon > 0, \ \exists N \geq 1, \ \forall n \geq N, \quad ||x_n - x_\infty||_E \leq \varepsilon.$$

- *The sequence x_n is called a Cauchy sequence if*

$$\forall \varepsilon > 0, \ \exists N \in \mathbb{N}, \ \forall n, m \geq N, \quad ||x_n - x_m||_E \leq \varepsilon.$$

Intuitively, the elements of a Cauchy sequence tend to "get increasingly closer" from each other as $n \to \infty$, while the elements of a convergent sequence x_n tend to "get ìncreasingly closer" to a fixed element $x_\infty \in E$.

The following proposition sheds some light on the relation between both notions.

Proposition 1.2. *Let $(E, ||\cdot||_E)$ be a normed vector space;*

(i) A Cauchy sequence x_n of elements of E is bounded;
(ii) A sequence x_n of elements of E which converges to some $x_\infty \in E$ is a Cauchy sequence.

Proof. (i): Letting $\varepsilon = 1$ in the Definition 1.7 of a Cauchy sequence, we see that there exists $N \geq 0$ such that

$$\forall n \geq N, \quad ||x_n - x_N||_E \leq 1.$$

Using the triangle inequality, we infer that, for $n \geq N$,

$$||x_n|| - ||x_N||_E \leq 1 \text{ and so } ||x_n||_E \leq 1 + ||x_N||_E,$$

and so:

$$\forall n \geq 0, \quad ||x_n||_E \leq \max(||x_0||_E, ||x_1||_E, \ldots, ||x_{N-1}||_E, 1 + ||x_N||_E),$$

that is, the sequence x_n is bounded.

(ii): Let x_n be a sequence of elements of E, converging to some $x_\infty \in E$. By definition, for all $\varepsilon > 0$, there exists $N \geq 0$ such that

$$\forall n \geq N, \quad ||x_n - x_\infty||_E \leq \frac{\varepsilon}{2}.$$

The triangle inequality then yields, for all integers m, $n \geq N$,

$$\begin{aligned} ||x_n - x_m||_E \leq ||x_n - x_\infty||_E + ||x_m - x_\infty||_E &\leq \tfrac{\varepsilon}{2} + \tfrac{\varepsilon}{2} \\ &= \varepsilon, \end{aligned}$$

which proves that x_n is a Cauchy sequence. □

It turns out that the converse property holds true in particularly "nice" vector spaces, which deserve a name of their own.

Definition 1.8. *A normed vector space $(E, ||\cdot||)$ is called complete if all Cauchy sequences of elements of E are convergent. A complete normed vector space is called a Banach space.*

Remark 1.2.

- *All finite-dimensional vector spaces are complete, see Exercise 1.5.*
- *An important example of a non convergent Cauchy sequence in a (necessarily not complete) normed vector space is given in Exercise 1.2.*

1.3.4. *Compact subsets of a normed vector space*

We conclude this section by recalling the key notion of compactness.

Definition 1.9. *Let K be a subset of a normed vector space $(E, ||\cdot||_E)$; K is called compact if, for any sequence $\{x_n\}_{n\in\mathbb{N}}$ of elements of K, there exists a subsequence $\{x_{n_k}\}_{k\in\mathbb{N}}$ which converges to an element $x_\infty \in K$.*

Intuitively, a compact set K is "small" enough so that any sequence of elements $x_n \in K$ must cluster at least at one point in K.

Proposition 1.3. *Let K be a compact subset of a normed vector space $(E, ||\cdot||_E)$. Then K is closed and bounded, that is, there exists $M > 0$ large enough so that $K \subset B(0, M)$.*

Proof. Let us first show the boundedness of K. We argue by contradiction: let x_0 be any point in K. Since K is not bounded, there exists $x_1 \in K$ such that $||x_1||_E \geq ||x_0||_E + 1$. In turn, since K is not bounded, there exists $x_2 \in K$ with $||x_2||_E \geq ||x_1||_E + 1 \geq ||x_0||_E + 2$. By induction, we thus construct a sequence $x_n \in K$ such that

$$\forall n, m \in \mathbb{N}, \;\; n \geq m \Rightarrow ||x_n - x_m||_E \geq ||x_n||_E - ||x_m||_E \geq m.$$

This sequence cannot possess any convergent subsequence, since the previous inequality impedes any subsequence of x_n to have the Cauchy property. This contradicts the compactness of K.

Let us now show that K is closed. To this end, we rely on the sequential definition of the closedness of subsets of E, see Remark 1.1. Let x_n be any sequence of elements of K, converging to some $x_\infty \in E$. By definition, x_n has a convergent subsequence x_{n_k} converging to some element $y \in K$. But since x_{n_k} is a subsequence of x_n, it also converges to x_∞, and by uniqueness of the limit, we obtain $x_\infty = y \in K$, as desired. □

Unfortunately, the converse of the latter property does not hold in general, and compact subsets of a normed vector space may be much more difficult to identify (and actually more scarce) than closed and bounded subsets. The case of finite-dimensional spaces is special in this perspective, as stated by the famous Bolzano-Weierstrass theorem, which we recall without proof.

Theorem 1.2 (Bolzano-Weierstrass theorem)**.** *Let E be a finite-dimensional vector space. A subset $K \subset E$ is compact if and only if it is closed and bounded.*

The compactness property of a subset of a normed vector space is often used via the following equivalent characterization, sometimes referred to as the Borel-Lebesgue property, whose difficult proof is admitted.

Theorem 1.3 (Borel-Lebesgue characterization of compactness)**.** *Let $(E, ||\cdot||_E)$ be a normed vector space, and let K be a subset of E. Then K is compact if and only if, for any collection $\{U_i\}_{i\in I}$ of open subsets of E (where the set of indices I is arbitrary), forming a covering of K, in the sense that*

$$K \subset \bigcup_{i\in I} U_i,$$

there exists a finite subset of indices $\{i_1, \dots, i_N\} \subset I$ such that the collection $\left\{U_{i_j}\right\}_{j=1,\dots,N}$ still forms a covering of K, i.e.

$$K \subset \bigcup_{j=1}^{N} U_{i_j}.$$

To conclude this brief discussion about compact sets, we present one among the many key properties of compact subsets of a normed vector space, that we have already hinted at previously: continuous functions defined on compact sets are uniformly continuous.

Theorem 1.4. *Let $(E, ||\cdot||_E)$ and $(F, ||\cdot||_F)$ be normed vector spaces and let $K \subset E$ be a compact subset. If a function $f : K \to F$ is continuous, it is uniformly continuous.*

Proof. We proceed by contradiction: assume that there exists $\varepsilon > 0$ such that for all $\delta > 0$, one may find two points $x, y \in K$ with

$$||x - y||_E \leq \delta, \text{ and } ||f(x) - f(y)||_F > \varepsilon.$$

In particular, one may define two sequences a_n and b_n of elements in K such that

$$\forall n \in \mathbb{N}, \quad ||a_n - b_n||_E \leq \frac{1}{n} \text{ and } ||f(a_n) - f(b_n)||_F > \varepsilon. \tag{1.1}$$

Since the set K is compact, up to extraction of a subsequence (still indexed by n) the sequences a_n and b_n converge to some elements a_∞ and $b_\infty \in K$, respectively. Passing to the limit in the first inequality of (1.1), we actually see that $a_\infty = b_\infty =: c$.

On another hand, the continuity of f at c reads:

$$\exists \eta > 0, \ ||x - c||_E \leq \eta \ \Rightarrow \ ||f(x) - f(c)||_F \leq \frac{\varepsilon}{2}.$$

In particular, for n larger than some $n_0 \in \mathbb{N}$, we have $||a_n - c||_E \leq \eta$ and $||b_n - c||_F \leq \eta$; this implies:

$$\forall n \geq n_0, \quad ||f(a_n) - f(c)||_F \leq \frac{\varepsilon}{2} \text{ and } ||f(b_n) - f(c)||_F \leq \frac{\varepsilon}{2},$$

and so, by the triangle inequality

$$||f(a_n) - f(b_n)||_F \leq ||f(a_n) - f(c)||_F + ||f(c) - f(b_n)||_F \leq \varepsilon,$$

which contradicts (1.1). This ends the proof. □

Remark 1.3. *The use of Theorem 1.2 yields the perhaps more familiar version of Heine's theorem, which is just a particular case of Theorem 1.4: a function $f : [a, b] \to \mathbb{R}$ defined on a closed bounded interval of $\mathbb{R}$ is uniformly continuous.*

1.4. **Normed vector spaces and linear structures**

In this section, we focus on the linear structure of mappings between normed vector spaces, which shows remarkable properties.

1.4.1. *Linear mappings between normed vector spaces*

Throughout this section, $(E, ||\cdot||_E)$ and $(F, ||\cdot||_F)$ are two normed vector spaces.

Definition 1.10. *A mapping $f : E \to F$ is said to be linear provided*

$$\forall x, y \in E, \ \ \lambda, \mu \in \mathbb{R}, \quad f(\lambda x + \mu y) = \lambda f(x) + \mu f(y).$$

The continuity of linear mappings can be characterized in a very handful way.

Proposition 1.4. *Let $f : E \to F$ be a linear mapping; the following three conditions are equivalent:*

(i) f is continuous on E;

(ii) f is continuous at 0;

(iii) There exists a constant $C > 0$ such that:

$$\forall x \in E, \quad ||f(x)||_F \leq C||x||_E. \tag{1.2}$$

Proof. $(i) \Rightarrow (ii)$ is obvious.

$(ii) \Rightarrow (iii)$: Letting $\varepsilon = 1$ in the definition of the continuity of f at 0 (see Definition 1.5), there exists $\delta > 0$ such that

$$\forall x \in E \text{ s.t. } ||x||_E \leq \delta, \quad ||f(x)||_F \leq 1.$$

Now, let $x \in E$ be given. If $x = 0$, the inequality (1.2) trivially holds true; otherwise, the element $y := \frac{\delta}{||x||_E} x \in E$ is such that $||y||_E = \delta$, and the previous inequality yields:

$$||f(y)||_F \leq 1, \text{ and so } \frac{\delta}{||x||_E}||f(x)||_F \leq 1,$$

where we have used the linearity of f. Since this holds for any $x \in E \setminus \{0\}$, we have proved (1.2) with $C = \frac{1}{\delta}$.

$(iii) \Rightarrow (i)$: Using the inequality (1.2) with x of the form $x = z - y$, for arbitrary $y, z \in E$, we see that f is actually Lipschitz continuous (see Definition 1.6), i.e.

$$\forall y, z \in E, \quad ||f(z) - f(y)||_F \leq C||z - y||_E.$$

The continuity of f on E easily follows from this fact. $\square$

In the following, we denote by $\mathcal{L}(E; F)$ the vector space of continuous linear functions from E to F. We define the operator norm $||f||_{\mathcal{L}(E;F)}$ of f as the best (i.e. the lowest) value of the constant C such that the inequality (1.2) holds, that is:

$$||f||_{\mathcal{L}(E;F)} = \sup_{\substack{x \in E \\ x \neq 0}} \frac{||f(x)||_F}{||x||_E}. \tag{1.3}$$

That this expression indeed defines a norm on $\mathcal{L}(E; F)$ is a simple verification which is left to the reader, see Exercise 1.3.

One interesting property of the space $\mathcal{L}(E; F)$ is that it inherits the completeness of the arrival space F. More precisely, we have the following result, whose proof is important, as it can be adapted to many situations when it comes to showing completeness of a normed vector space.

Proposition 1.5. *Assume that $(F, ||\cdot||_F)$ is a Banach space. Then the vector space $\mathcal{L}(E; F)$ of continuous and linear mappings from E to F, equipped with the norm (1.3), is also a Banach space.*

Proof. Let $f_n \in \mathcal{L}(E; F)$ be a Cauchy sequence of continuous linear mappings from E to F; we wish to show that f_n converges to a certain linear mapping $f_\infty : E \to F$. We proceed within three steps.

Step 1: We find a candidate f_∞ for the limit of f_n. To this end, let us write down the definition of a Cauchy sequence in $\mathcal{L}(E; F)$. For all $\varepsilon > 0$, there exists $N \geq 0$ such that

$$\forall m, n \geq N, \quad ||f_n - f_m||_{\mathcal{L}(E;F)} \leq \varepsilon.$$

Let us fix $x \in E$; using the definition of the operator norm (1.3), we infer that, in particular

$$\forall m, n \geq N, \quad ||f_n(x) - f_m(x)||_F \leq \varepsilon||x||_E.$$

This implies, in particular, that $\{f_n(x)\}_{n\geq 1}$ is a Cauchy sequence of elements of F. Since F is complete, this sequence converges to an element of F, which we denote by $f_\infty(x)$.

We have thus constructed a mapping $E \ni x \mapsto f_\infty(x) \in F$, about which no further information is available at the moment.

Step 2: We prove that the mapping f_∞ belongs to $\mathcal{L}(E; F)$. At first, for any $x, y \in E$, and any $\lambda, \mu \in \mathbb{R}$, we have, on the one hand, by the definition of f_∞:

$$f_n(\lambda x + \mu y) \xrightarrow{n\to 0} f_\infty(\lambda x + \mu y).$$

On the other hand, since each f_n is linear, it holds

$$f_n(\lambda x + \mu y) = \lambda f_n(x) + \mu f_n(y),$$

and the right-hand side of the above expression converges to $\lambda f_\infty(x) + \mu f_\infty(y)$. By uniqueness of the limit of a sequence of elements in F, we obtain that

$$f_\infty(\lambda x + \mu y) = \lambda f_\infty(x) + \mu f_\infty(y),$$

which means that f_∞ is linear.

Let us show that f_∞ is continuous. To this end, we know that f_n is bounded as a Cauchy sequence, see Proposition 1.2. Hence, there exists a constant $C > 0$ such that, for all $n \geq 0$,

$$\forall x \in E, \quad ||f_n(x)||_F \leq C||x||_E.$$

Passing to the limit, it follows that

$$\forall x \in E, \quad ||f_\infty(x)||_F \leq C||x||_E,$$

which means that f_∞ is continuous.

Summarizing, we have proved that $f_\infty \in \mathcal{L}(E; F)$.

Step 3: We prove that $||f_n - f_\infty||_{\mathcal{L}(E;F)} \to 0$ as $n \to \infty$. We return to the fact that f_n is a Cauchy sequence in $\mathcal{L}(E; F)$: for all $\varepsilon > 0$, there exists $N \geq 0$ such that

$$\forall n, m \geq N, \ \forall x \in E, \quad ||f_n(x) - f_m(x)||_F \leq \varepsilon||x||_E.$$

For an arbitrary given $\varepsilon > 0$, keeping n fixed and passing to the limit $m \to \infty$, this yields:

$$\forall n \geq N, \ \forall x \in E, \quad ||f_n(x) - f_\infty(x)||_F \leq \varepsilon||x||_E,$$

i.e.

$$\forall n \geq N, \quad ||f_n - f_\infty||_{\mathcal{L}(E;F)} \leq \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, this expresses the convergence of f_n to f_∞ in $\mathcal{L}(E; F)$. □

Additional properties of interest concerning linear mappings are introduced in the next definition.

Definition 1.11.

- *A linear and continuous mapping $f : E \to F$ is called a linear isomorphism if there exists a continuous, linear mapping $g : F \to E$ such that*

$$g \circ f = \mathrm{id}_E \ \textit{and} \ \ f \circ g = \mathrm{id}_F,$$

  where $\mathrm{id}_E : E \to E$ is the identity mapping of E and $\mathrm{id}_F : F \to F$ is the identity mapping of F.
- *A linear mapping $f : E \to F$ is called an isometry if*

$$\forall x \in E, \quad ||f(x)||_F = ||x||_E.$$

Remark 1.4. *It follows immediately from the definition that an isometry $f : E \to F$ is continuous and injective.*

1.4.2. *Multlinear mappings between normed vector spaces*

The previous definitions and concepts extend almost immediately to the case of multilinear mappings. In this section, we consider $(n+1)$ normed vector spaces $(E_1, ||\cdot||_{E_1}), \dots, (E_n, ||\cdot||_{E_n})$ and $(F, ||\cdot||_F)$. The product space $E_1 \times \dots \times E_n$ is naturally equipped with the norm

$$\forall x = (x_1, \dots, x_n) \in E_1 \times \dots \times E_n, \quad ||x||_{E_1 \times \dots \times E_n} = \sup_{i=1,\dots,n} ||x_i||_{E_i}.$$

Remark 1.5. *From the fact that all norms defined on a finite-dimensional vector space are equivalent (see Proposition 1.1), it is easy to see that the above norm is equivalent to any norm of the form*

$$E_1 \times \dots \times E_n \ni x = (x_1, \dots, x_n) \longmapsto N\Big(||x_1||_{E_1}, \dots, ||x_n||_{E_n}\Big) \in \mathbb{R},$$

where $N(\cdot)$ is an arbitrary norm on $\mathbb{R}^n$.

Definition 1.12. *A mapping $f : E_1 \times \dots \times E_n \to F$ is called multilinear (or n-linear) if for all $k = 1, \dots, n$, for any fixed elements $x_1 \in E_1, \dots, x_{k-1} \in E_{k-1}, x_{k+1} \in E_{k+1}, \dots, x_n \in E_n$, the partial mapping*

$$E_k \ni z \longmapsto f(x_1, \dots, x_{k-1}, z, x_{k+1}, \dots, x_n) \in F$$

is linear.

The following result about the continuity of multilinear mappings is the exact counterpart of Proposition 1.4 in the present context. Its proof is thus left to the reader.

Proposition 1.6. *Let $f : E_1 \times \dots \times E_n \to F$ be a multilinear mapping. Then the following three conditions are equivalent.*

(i) f is continuous on $E_1 \times \dots \times E_n$;
(ii) f is continuous at $(0, \dots, 0) \in E_1 \times \dots \times E_n$;
(iii) There exists a constant $C > 0$ such that:

$$\forall (x_1, \dots, x_n) \in E_1 \times \dots \times E_n, \quad ||f(x_1, \dots, x_n)||_F \leq C ||x_1||_{E_1} \dots ||x_n||_{E_n}. \tag{1.4}$$

We denote by $\mathcal{L}(E_1, \dots, E_n; F)$ the vector space of all n-linear continuous mappings $E_1 \times \dots \times E_n \to F$. The operator norm $||f||_{\mathcal{L}(E_1,\dots,E_n;F)}$ of a continuous multilinear mapping $f : E_1 \times \dots \times E_n \to F$ is defined as the smallest constant C such that the inequality (1.4) holds:

$$\forall f \in \mathcal{L}(E_1, \dots, E_n; F), \quad ||f||_{\mathcal{L}(E_1,\dots,E_n;F)} = \sup_{x_1 \neq 0,\, x_2 \neq 0, \dots, x_n \neq 0} \frac{||f(x_1, \dots, x_n)||_F}{||x_1||_{E_1} \dots ||x_n||_{E_n}}.$$

For later purposes, we eventually introduce the space of symmetric multilinear mappings.

Definition 1.13. *Let $(E, ||\cdot||_E)$ and $(F, ||\cdot||_F)$ be normed vector spaces, and let $n \geq 2$. A multilinear mapping $f : E^n \to F$ is called symmetric if for any permutation $\sigma : \{1, \dots, n\} \to \{1, \dots, n\}$, it holds:*

$$\forall x_1, \dots x_n \in E, \quad f(x_1, \dots, x_n) = f(x_{\sigma(1)}, \dots, x_{\sigma(n)}).$$

The vector space of continuous and symmetric n-linear mappings from E^n into F is denoted by $\mathcal{L}_s(E^n; F)$.

1.4.3. *The dual of a normed vector space*

In the study of a normed vector space E, a central role is played by the associated space of continuous, linear forms on E, i.e. on the continuous linear mappings $E \to \mathbb{R}$.

Definition 1.14. *The dual space of E is the vector space $E^* = \mathcal{L}(E; \mathbb{R})$ of continuous linear forms on E, equipped with the norm*

$$||\ell||_{E^*} = \sup_{\substack{x \in E \\ x \neq 0}} \frac{|\ell(x)|}{||x||_E}.$$

Often, much information can be gleaned about E by looking at the dual space E^*, and the way they interact with one another – a relation which is called *duality* in the literature. Indeed, E^* often turns out to show more pleasant properties than E, in spite of its apparent greater complexity. For instance, it follows from Proposition 1.5 that E^* is always a Banach space, even when E is not.

1.5. **Hilbert spaces**

Among Banach spaces, we make out the subcategory of Hilbert spaces, whose desirable properties stem from the presence of an additional structure, that of inner product.

Definition 1.15. *Let E be a (real) vector space; an inner product on E is a mapping $\langle \cdot, \cdot \rangle : E \times E \to \mathbb{R}$ which is*

(i) Bilinear: For all x, y and $z \in E$ and all $\lambda, \mu \in \mathbb{R}$, there holds:

$$\langle \lambda x + \mu y, z \rangle = \lambda \langle x, z \rangle + \mu \langle y, z \rangle, \text{ and } \langle z, \lambda x + \mu y \rangle = \lambda \langle z, x \rangle + \mu \langle z, y \rangle.$$

(ii) Symmetric: For all x, $y \in E$, one has $\langle x, y \rangle = \langle y, x \rangle$.

(iii) Positive: For all $x \in E$, $\langle x, x \rangle \geq 0$.

(iv) Positive definite: For all $x \in E$, $\langle x, x \rangle = 0$ if and only if $x = 0$.

One fundamental property related to inner products is the well-known Cauchy-Schwarz inequality.

Lemma 1.2 (Cauchy-Schwarz inequality)**.** *Let E be a (real) vector space equipped with an inner product $\langle \cdot, \cdot \rangle$. Then, it holds:*

$$\forall x, y \in E, \quad \langle x, y \rangle^2 \leq \langle x, x \rangle \langle y, y \rangle. \tag{1.5}$$

Proof. Let x, y be two given elements in E; we define the function $p : \mathbb{R} \to \mathbb{R}$ by

$$p(t) = \langle x + ty, x + ty \rangle = \langle x, x \rangle + 2t \langle x, y \rangle + t^2 \langle y, y \rangle.$$

Hence, $p(t)$ is a second-order polynomial function, which taking only non negative values owing to the positivity of the inner product $\langle \cdot, \cdot \rangle$. Its discriminant must therefore be negative, which reads

$$4 \langle x, y \rangle^2 - 4 \langle x, x \rangle \langle y, y \rangle \leq 0,$$

Re-arranging the latter expression yields the desired inequality (1.5). □

As a consequence, it is easily verified that:

- The mapping $|| \cdot || : E \to \mathbb{R}_+$ given by

$$||x|| := \sqrt{\langle x, x \rangle} \tag{1.6}$$

  defines a norm on E.
- For any given $y \in H$, the linear form

$$E \ni x \mapsto \langle x, y \rangle \in \mathbb{R}$$

  is continuous when E is equipped with the norm (1.6).

Definition 1.16. *A vector space E equipped with an inner product $\langle \cdot, \cdot \rangle$ (and the induced norm (1.6)) is called a pre-Hilbert space. If, in addition, E is complete for the norm (1.6), it is called a Hilbert space.*

The following theorem is at the basis of many crucial properties of a Hilbert space H. Roughly, it expresses that when $C \subset H$ is a closed and convex subset of H, one may define a projection mapping $p_C : H \to C$ which to each point $x \in H$ associates the closest point p to x in C, see Fig. 4.

Theorem 1.5. *Let $(H, \langle \cdot, \cdot \rangle)$ be a Hilbert space, $|| \cdot ||$ be the associated norm, and let $C \subset H$ be a closed and convex subset. Then for all $x \in H$, there exists a unique point $p \in C$ such that*

$$||x - p||^2 = \min_{y \in C} ||x - y||^2. \tag{1.7}$$

This point is called the projection of x onto C and it is denoted by $p_C(x)$.

The point $p_C(x)$ is also characterized by the following fact:

$$\forall z \in H, \;\; z = p_C(x) \Leftrightarrow \left\{ \begin{array}{l} z \in C, \\ \forall y \in C, \;\; \langle z - x, y - z \rangle \geq 0. \end{array} \right. \tag{1.8}$$

The so-defined mapping $p_C : H \to C$ is continuous.

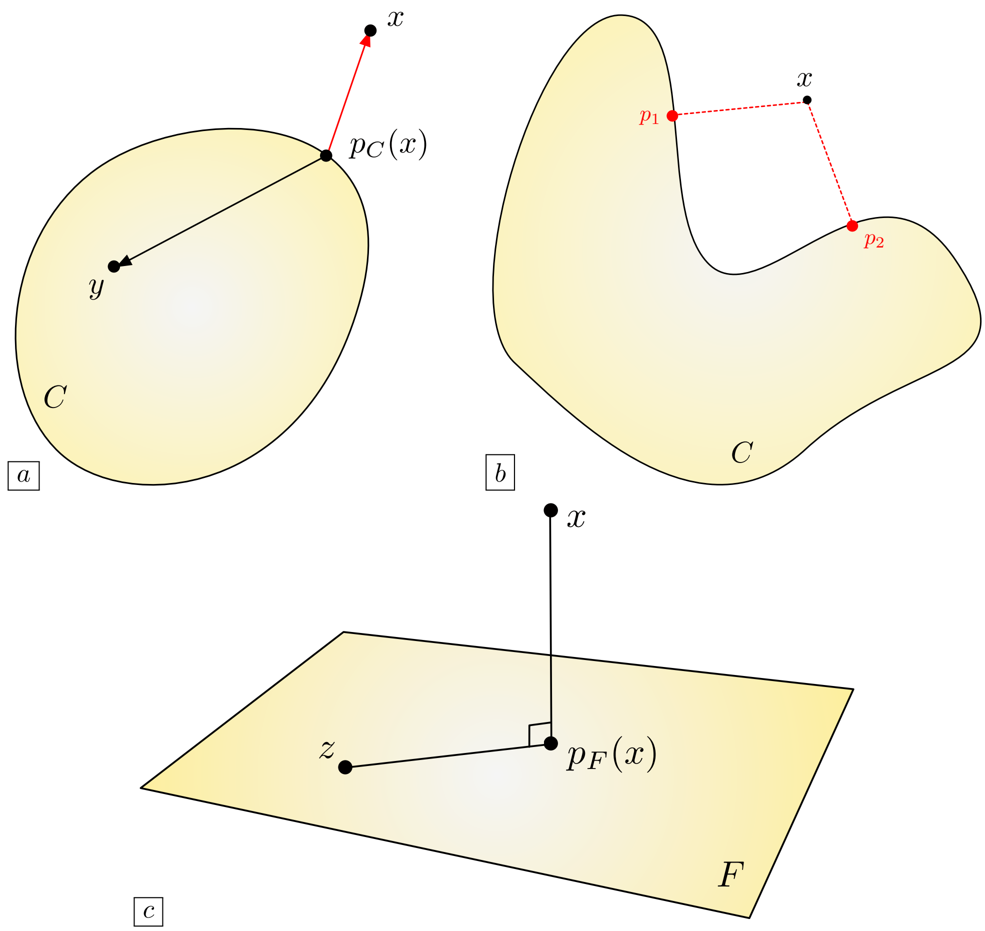

FIGURE 4. *(a) Projection onto a closed convex set C in a Hilbert space; (b) When C is not convex, Theorem 1.5 fails since one point $x \in H$ may have two closest points p_1, $p_2 \in C$; (c) Projection onto a closed vector subspace F.*

Proof. Let us denote by

$$\delta = \inf_{p \in C} ||x - p||.$$

Note that for the moment, we do not know whether this infimum is attained. However, by definition of the infimum, there exists a minimizing sequence for this problem, i.e. a sequence y_n of elements in C such that

$$\lim_{n\to\infty} ||y_n - x|| = \delta.$$

We now show that this sequence actually converges to an element in C, relying on the Cauchy criterion. To this end, we use the parallelogram identity in the Hilbert space H; see Exercise 1.9. For all $n, m \in \mathbb{N}$, we have

$$\begin{aligned}
\tfrac{1}{2}||y_n - y_m||^2 &= \tfrac{1}{2}||y_n - x - (y_m - x)||^2 \\
&= ||y_n - x||^2 + ||y_m - x||^2 - \tfrac{1}{2}||y_n + y_m - x||^2 \\
&= ||y_n - x||^2 + ||y_m - x||^2 - 2\left|\left|x - \tfrac{1}{2}(y_n + y_m)\right|\right|^2 \\
&\leq ||y_n - x||^2 + ||y_m - x||^2 - 2\delta^2,
\end{aligned}$$

where we have used the definition of δ in the last line, together with the fact that $\frac{1}{2}(y_n + y_m)$ belongs to C since this set is convex. Now, since $||y_n - x||$ converges to 0 as $n \to \infty$, the right-hand side of the above inequality converges to δ as $m, n \to \infty$, which proves that y_n is a Cauchy sequence in H. It therefore converges to some $p \in H$, and p belongs to C since this set is closed. At this point, we have thus proved that there exists at least one point p satisfying (1.7).

Let now $p \in C$ be any point satisfying (1.7), and let us prove that p also satisfies (1.8). For any point $y \in C$, and for any $t \in (0,1]$, the point $(1-t)p + ty$ belongs to C, and so

$$||x - p||^2 \le ||x - (1-t)p - ty||^2.$$

Expanding the right-hand side, we obtain that

$$||x - p||^2 \le ||x - p||^2 - 2t\langle x - p, y - p\rangle + t^2||y - p||^2.$$

Canceling the common term $||x-p||^2$ in both side, dividing by t and then letting t tend to 0, we thus obtain the desired inequality

$$\langle x - p, y - p\rangle \le 0.$$

At this point, we have thus proved that there exists at least one solution to (1.7), and that it satisfies (1.8).

Conversely, let $z \in H$ be any point satisfying (1.8), that is:

$$z \in C \text{ and } \forall y \in C, \quad \langle z - x, y - x\rangle \ge 0.$$

Then, we have, for all $y \in C$,

$$\begin{array}{rcl} ||x - y||^2 & = & ||x - z||^2 + 2\langle x - z, z - y\rangle + ||z - y||^2 \\ & \ge & ||x - z||^2 + ||z - y||^2. \end{array}$$

Hence, it is clear that

$$\min_{y \in C} ||x - y||^2 \ge ||z - x||^2,$$

and equality holds if and only if $y = z$. Hence, z is the unique solution to the minimization problem (1.7). Since this argument holds for any point z satisfying (1.8), this shows in the meantime that there exists a unique such point $z \in H$.

We have thus proved that there exists a unique solution $p = p_C(x)$ to (1.7), which is equivalently characterized by (1.8). Eventually, we turn to the continuity of the (non linear) mapping p_C. We actually prove that p_C is 1-Lipschitz. For any two points $x, y \in H$, let us write:

$$\begin{array}{rcl} ||p_C(x) - p_C(y)||^2 & = & \langle p_C(x) - p_C(y), p_C(x) - p_C(y)\rangle \\ & = & \langle p_C(x) - x, p_C(x) - p_C(y)\rangle + \langle x - y, p_C(x) - p_C(y)\rangle + \langle y - p_C(y), p_C(x) - p_C(y)\rangle \\ & \le & \langle x - y, p_C(x) - p_C(y)\rangle, \end{array}$$

where we have used the inequality (1.8) to pass from the second to the third line. Now, by the Cauchy-Schwarz inequality, we obtain:

$$||p_C(x) - p_C(y)||^2 \le ||x - y|| ||p_C(x) - p_C(y)||,$$

which readily yields

$$||p_C(x) - p_C(y)|| \le ||x - y||.$$

This shows the continuity of p_C and the proof of Theorem 1.5 is now complete. □

This fundamental result has even more remarkable implications in the case where the closed convex set C of interest is a closed vector subspace of H. Before studying these, a few definitions are in order.

Definition 1.17. *Let $(H, \langle\cdot,\cdot\rangle)$ be a Hilbert space.*

- *Two elements $a, b \in H$ are called orthogonal when $\langle a, b\rangle = 0$.*
- *The orthogonal of a subset $A \subset H$ is the closed vector space $A^\perp \subset H$ defined by:*

$$A^\perp := \{x \in H,\ \forall a \in A,\ \langle a, x\rangle = 0\}.$$

We are now in position to state and prove the result of interest, see Fig. 4 (b) for an illustration.

Theorem 1.6. *Let F be a closed vector subspace of H. Then,* *(i) The projection mapping $p_F : H \to F$ supplied by Theorem 1.5 is a linear and continuous mapping which is characterized by the identity:*

$$p = p_F(x) \Leftrightarrow \begin{cases} p \in F, \\ \forall z \in F, \ \ \langle x - p, z\rangle = 0. \end{cases} \tag{1.9}$$

(ii) The following decomposition of H holds

$$H = F \oplus F^{\perp}.$$

Proof. (i): Let $x \in H$ be given; Theorem 1.5 ensures that the projection point $p_F(x)$ is characterized by:

$$p_F(x) \in F \text{ and } \forall z \in F, \quad \langle x - p_F(x), p_F(x) - z\rangle \geq 0.$$

Since F is a vector space, this is equivalent to

$$p_F(x) \in F \text{ and } \forall z \in F, \quad \langle x - p_F(x), z\rangle \geq 0.$$

Finally, $p_F(x)$ is the unique point in F such that:

$$p_F(x) \in F \text{ and } \forall z \in F, \quad \langle x - p_F(x), z\rangle = 0.$$

It follows easily from this characterization that the mapping $x \mapsto p_F(x)$ is linear from H to F. Since we already know from Theorem 1.5 that $x \mapsto p_F(x)$ is continuous, it follows that $p_F \in \mathcal{L}(H; F)$, as desired.

(ii): Any point $x \in H$ can be decomposed as

$$x = (x - p_F(x)) + p_F(x),$$

where by definition, $p_F(x) \in F$, and $(x - p_F(x)) \in F^{\perp}$ because of (1.9). Since, obviously, $F \cap F^{\perp} = \{0\}$, the desired result is proved. ☐

We end this short tour of Hilbert spaces with a study of their dual space. On the one hand, it is clear that, for each $u \in H$, the mapping $\ell : H \to \mathbb{R}$ defined by

$$\ell(x) := \langle u, x\rangle$$

is a continuous linear form on H. The following theorem establishes that the converse holds true, i.e. that any continuous linear form $\ell \in H^*$ can be represented by an element $u \in H$.

Theorem 1.7 (Riesz representation theorem)**.** *Let $(H, \langle, \cdot, \cdot\rangle)$ be a Hilbert space, and let $\ell : H \to R$ be a continuous linear form. Then there exists a unique element $u \in H$ such that*

$$\forall x \in H, \quad \ell(x) = \langle u, x\rangle. \tag{1.10}$$

In addition,

$$||\ell||_{H^*} = ||u||_H.$$

Proof. The idea of the proof is that the kernel of a continuous linear form $\ell \in H^*$ is a closed hyperplane F of H; hence, ℓ is characterized by the orthogonal complement $F^{\perp}$ of F, which is a one-dimensional space, and the sought Riesz representative u of ℓ in (1.10) is then a suitably scaled generator of $F^{\perp}$.

The result is trivial is $\ell = 0$, and we thus assume that $\ell \neq 0$. Let $F = \text{Ker}(\ell)$; since ℓ is a continuous linear form which does not vanish identically on H, F is a strict, closed vector subspace of H, and by virtue of Theorem 1.6, we may write

$$H = F \oplus F^{\perp}.$$

As $F^{\perp} \neq \{0\}$, let $u_0 \in F^{\perp}$ be different from 0; in particular, $\ell(u_0) \neq 0$. Then, for all $v \in F^{\perp}$, he have:

$$\ell\left(v - \frac{\ell(v)}{\ell(u_0)}u_0\right) = \ell(v) - \ell(v) = 0,$$

and so, $\left(v - \frac{\ell(v)}{\ell(u_0)}u_0\right)$ belongs at the same time to to $F^{\perp}$ (by definition), and to $F = \text{Ker}(\ell)$. Hence,

$$v = \frac{\ell(v)}{\ell(u_0)}u_0,$$

which proves that $F^{\perp}$ is a vector space with dimension 1, spanned by u_0.

Now, let us consider the linear form $\ell_0 : H \ni v \mapsto \langle u_0, v\rangle$. Then $\ell_0(v) = \ell(v) = 0$ for all $v \in F$, and

$$\forall v = \alpha u_0 \in F^\perp, \quad \ell_0(v) = \alpha ||u_0||^2, \text{ and } \ell(v) = \alpha \ell(u_0).$$

It follows that

$$\forall v \in H, \quad \ell(v) = \frac{\ell(u_0)}{||u_0||^2} \langle u_0, v\rangle;$$

setting $u := \frac{\ell(u_0)}{||u_0||^2} u_0 \in H$, we have proved that there exists one Riesz representative u for ℓ, as in (1.10).

Let us now prove the uniqueness of such a representative: if there exist u_1, $u_2 \in H$ such that

$$\ell(v) = \langle u_1, v\rangle = \langle u_2, v\rangle,$$

then $(u_1 - u_2)$ is orthogonal to all elements in H, and so $u_1 - u_2 = 0$.

Finally, it follows immediately from the Cauchy-Schwarz inequality that if ℓ is represented by $u \in H$, then $||\ell||_{H^*} = ||u||$. □

Remark 1.6. *The Riesz representation Theorem 1.7 supplies an isometric, bijective mapping $H^* \to H$, which allows to identify the dual space H^* with H itself: a continuous linear form $\ell \in H^*$ is often seen as the corresponding Riesz representative $u \in H$.*

1.6. **Exercises**

Exercise 1.1 (Equivalence of norms in finite-dimensional vector spaces). The purpose of this exercise is to prove the well-known result that all norms in a finite-dimensional vector space are equivalent. Let E be a vector space with finite dimension d, equipped with a basis $(e_1, \dots, e_d)$. We define the supremum norm $||\cdot||_\infty$ on E with respect to this basis by

$$\forall x = \sum_{i=1}^{d} x_i e_i \in E, \quad ||x||_\infty := \sup_{i=1,\dots,d} |x_i|.$$

We consider another norm $m(x)$ on E, and we aim to prove that m and $||\cdot||_\infty$ are equivalent.

(i) Prove that there exists a constant $C > 0$ such that:

$$\forall x \in E, \quad m(x) \leq C||x||_\infty.$$

(ii) Infer that the mapping $x \mapsto m(x)$ is continuous from $(E, ||\cdot||_\infty)$ into $\mathbb{R}$.

(iii) Prove that the unit sphere $S := \{x \in E, \ \ ||x||_\infty = 1\}$ is a compact subset of E.

(iv) By using the fact that a continuous function on a compact metric space has a lower bound and an upper bound which are attained, conclude that the norms m and $||\cdot||_\infty$ are equivalent.

Exercise 1.2. Let P be the vector space of real-valued polynomial functions on the interval $[0, 1] \subset \mathbb{R}$, equipped with the supremum norm:

$$\forall p \in P, \quad ||p||_P := \sup_{x\in[0,1]} |p(x)|.$$

Let the sequence p_n of polynomials in P be defined by:

$$p_n(x) := \sum_{k=0}^{n} \frac{x^k}{k!}.$$

(i) Prove that p_n is a Cauchy sequence in P.

(ii) Show that if p_n converges to some polynomial function $f \in P$, then necessarily $p_n(x) \to f(x)$ for all points $x \in [0, 1]$.

(iii) Infer from the previous question that if p_n converges to $f \in P$, then f is necessarily the exponential function $f(x) = e^x$.

(iv) Show that the function $x \mapsto e^x$ is not a polynomial.

(v) Conclude that the Cauchy sequence p_n does not converge to an element of P.

Exercise 1.3. Let $(E, ||\cdot||_E)$ and $(F, ||\cdot||_F)$ be two normed vector spaces; recall that for all linear and continuous mapping $\ell : E \to F$, the quantity $||\ell||_{\mathcal{L}(E;F)}$ is defined by the formula (1.3).

(i) Show that the mapping $\ell \mapsto ||\ell||_{\mathcal{L}(E;F)}$ is a norm on $\mathcal{L}(E; F)$.

(ii) Let $(G, ||\cdot||_G)$ be another normed vector space, and let $\ell_1 : E \to F$, $\ell_2 : F \to G$ be continuous linear mappings; show that

$$||\ell_2 \circ \ell_1||_{\mathcal{L}(E;G)} \leq ||\ell_2||_{\mathcal{L}(F;G)} ||\ell_1||_{\mathcal{L}(E;F)}.$$

Exercise 1.4 (Extension by uniform continuity). Let $(E, ||\cdot||_E)$ be a normed vector space, and $(F, ||\cdot||_F)$ be a Banach space; let $U \subset E$ be a dense subset. Let $f : U \to F$ be a uniformly continuous function. Show that there exists a unique uniformly continuous function $g : E \to F$ such that $f = g|_U$.

Exercise 1.5. The goal of this exercise is to prove that any finite-dimensional vector space is complete, admitting the Bolzano-Weierstrass Theorem 1.2.

(i) Let $(E, ||\cdot||_E)$ be an arbitrary normed vector space. Show that if a Cauchy sequence x_n of elements of E has a convergent subsequence, then x_n itself is a convergent sequence.
(ii) Show that, if E has finite dimension, any Cauchy sequence in E has a convergent subsequence.
(iii) Conclude.

Exercise 1.6 (The adjoint mapping). Let $(E, ||\cdot||_E)$ and $(F, ||\cdot||_F)$ be two normed vector spaces, and let $T : E \to F$ be a linear, continuous mapping. We define the adjoint mapping $T^* : F^* \to E^*$ by the formula

$$\forall \ell \in F^*, \quad T^*\ell = \ell \circ T.$$

(i) Show that T^* is a linear and continuous mapping from F^* into E^* and that its operator norm satisfies:

$$||T^*||_{\mathcal{L}(F^*;E^*)} \leq ||T||_{\mathcal{L}(E;F)}.$$

[Remark: It can be shown that, actually, equality holds in the previous inequality, but this is more difficult to prove.]

(ii) Show that the kernel $\text{Ker}(T^*)$ of T^* coincides with $\overset{\circ}{\text{Ran}}(T)$, where the range $\text{Ran}(T)$ of T is defined by

$$\text{Ran}(T) = \{Tx \in F, \ x \in E\}$$

and where the polar set $\overset{\circ}{V} \subset F^*$ of an arbitrary subset $V \subset F$ is given by:

$$\overset{\circ}{V} = \{\ell \in F^*, \ \langle \ell, y \rangle = 0 \text{ for all } y \in V\}.$$

(iii) We now assume that E and F are Hilbert spaces, denoting by $\langle \cdot, \cdot \rangle_E$ and $\langle \cdot, \cdot \rangle_F$ the associated inner products. We identify the dual spaces E^* and F^* with E and F respectively, according to Remark 1.6. Show that the adjoint mapping T^*, which is a continuous linear mapping $F \to E$ once this identification is made, satisfies:

$$\forall u \in E, v \in F, \quad \langle T^*v, u \rangle_E = \langle Tu, v \rangle_F.$$

(iv) We now assume that $E = \mathbb{R}^d$ and $F = \mathbb{R}^m$ for some $d, m \geq 1$, equipped with their canonical bases and inner products. Let M be the matrix associated to the mapping $T : E \to F$; show that the matrix of $T^* : F \to E$ is M^T.

Exercise 1.7 (*The spaces of bounded continuous functions and functions of class $\mathcal{C}^1$*).

(i) Let E be the vector space of continuous functions on the interval $[0,1]$ equipped with the norm

$$||f||_\infty := \sup_{x \in [0,1]} |f(x)|.$$

Show that E is a Banach space.

(Hint: the proof can be conducted in a quite similar way to that of Proposition 1.5.)

(ii) Let F be the vector space of functions of class $\mathcal{C}^1$ on $[0,1]$ (that is, $f \in F$ if and only if it is the restriction to $[0,1]$ of a function of class $\mathcal{C}^1$ on an open neighborhood of $[0,1]$). Show that

$$u \in F \Leftrightarrow u \in E, \text{ and } \exists v \in E \text{ s.t. } \forall t \in [0,1], \ u(t) = u\left(\frac{1}{2}\right) + \int_{\frac{1}{2}}^t v(s)\,\mathrm{d}s.$$

(iii) We now equip F with the following norm:

$$||u||_F := \sup\left(\sup_{x \in [0,1]} |u(x)|, \sup_{x \in [0,1]} |u'(x)| \right).$$

Infer from the previous question that $(F, ||\cdot||_F)$ is a Banach space.

Exercise 1.8 (Neumann series). Let $(E, ||\cdot||)$ be a Banach space.

(i) Let $T : E \to E$ be a linear operator with norm $||T||_{\mathcal{L}(E)} < 1$. Show that the series

$$U = \sum_{n=0}^{\infty} T^n$$

converges in $\mathcal{L}(E)$.

(ii) Prove that $(\mathrm{Id} - T)$ is invertible, with inverse $(\mathrm{Id} - T)^{-1} = U$.

(iii) Show that the subset $\mathrm{Inv}(E)$ of $\mathcal{L}(E)$ made of invertible operators is open in $\mathcal{L}(E)$.

(iv) Prove that the (non linear) mapping

$$\mathrm{Inv}(E) \ni T \longmapsto T^{-1} \in \mathrm{Inv}(E)$$

is continuous.

(v) Let $(F, ||\cdot||_F)$ be another normed vector space. Show that the subset $\mathrm{Inv}(E; F)$ of $\mathcal{L}(E; F)$ made of invertible mappings is open in $\mathcal{L}(E; F)$.

Exercise 1.9 (The parallelogram identity).

(i) Let $(H, \langle\cdot,\cdot\rangle)$ be a pre-Hilbert space. Show that the associated norm $||\cdot|| = \sqrt{\langle\cdot,\cdot\rangle}$ satisfies the parallelogram identity:

$$\forall x, y \in H, \quad ||x+y||^2 + ||x-y||^2 = 2||x||^2 + 2||y||^2, \tag{1.11}$$

see Fig. 5 for an illustration.

Let now $(E, ||\cdot||)$ be a normed vector space, whose norm satisfies the parallelogram identity (1.11). Our aim is to prove that this norm is actually induced by an inner product on E. More precisely, we define

$$\forall x, y \in E, \quad \langle x, y\rangle := \frac{1}{4}\Big(||x+y||^2 - ||x-y||^2\Big),$$

and we prove that this mapping satisfies the axiom of an inner product.

(ii) Prove that $\langle\cdot,\cdot\rangle$ satisfies the symmetry, positiveness and positive definiteness axioms (ii) (iii) and (iv) of Definition 1.15.

(iii) Prove that, for any fixed $y \in E$, the mapping $x \mapsto \langle x, y\rangle$ is continuous.

(iv) Prove that, for any x, y, $z \in E$, the following identity holds:

$$||x+y+z||^2 = ||x||^2 + ||y||^2 + ||x+z||^2 + ||y+z||^2 - \frac{1}{2}||x-y+z||^2 - \frac{1}{2}||y-x+z||^2.$$

(v) Deduce that it holds:

$$\forall x, y, z \in E, \quad \langle x+y, z\rangle = \langle x, z\rangle + \langle y, z\rangle.$$

(vi) Show that, for any integer $\lambda \in \mathbb{N}$, one has:

$$\forall x, y \in E, \quad \langle \lambda x, y\rangle = \lambda\langle x, y\rangle. \tag{1.12}$$

(vii) Show that (1.12) actually holds when the integer n is replaced with an arbitrary rational number $\lambda \in \mathbb{Q}$.

(viii) Show that (1.12) holds for all $\lambda \in \mathbb{R}$ and conclude.

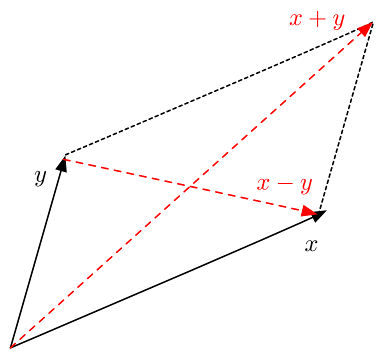

FIGURE 5. *Illustration of the parallelogram identity.*

Exercise 1.10 (The Lax-Milgram theorem). Let $(H, \langle \cdot, \cdot \rangle)$ be a Hilbert space, $a : H \times H \to \mathbb{R}$ be a continuous bilinear form which is elliptic, that is, there exists $\alpha > 0$ such that

$$\forall u \in H, \quad \alpha ||u||^2 \leq a(u, u).$$

(i) Show that there exists a continuous linear mapping $A : H \to H$ such that:

$$\forall u, v \in H, \quad a(u, v) = \langle A(u), v \rangle.$$

(ii) Show that for all linear form $\ell : H \to \mathbb{R}$, there exists a unique $u \in H$ such that

$$\forall v \in H, \quad a(u, v) = \ell(v),$$

and that

$$||u||_H \leq \frac{1}{\alpha} ||\ell||_{H^*}.$$

2. Differential calculus I: elementary facts

This section revolves around the notion of differentiability. Before delving into rigorous mathematical developments, let us first present the main ideas in an informal manner.

The differential is the generalization to arbitrary normed vector spaces of the familiar derivative of a function $f : \mathbb{R} \to \mathbb{R}$. For.a given point $x \in \mathbb{R}$, the derivative $f'(x)$ is defined as the following limit, when it exists:

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

visually, $f'(x)$ is the slope of the tangent line to the graph of f at x. This expression can be re-arranged into an approximation formula for f near x:

$$f(x+h) = f(x) + f'(x)h + \mathrm{o}(h), \tag{2.1}$$

where the notation $\mathrm{o}(h)$ stands for an unspecified function of h which tends to 0 faster than h as $h \to 0$: $\lim_{h \to 0} \frac{\mathrm{o}(h)}{h} = 0$. Roughly speaking, when "small" perturbations $(x + h)$, $h \ll 1$, are considered around x, a coarse approximation of the corresponding values $f(x + h)$ is the (h-independent) value $f(x)$. A more accurate "first-order" approximation of $h \mapsto f(x + h)$ (up to a remainder of the order $\mathrm{o}(h)$) is the affine function $h \mapsto f(x) + f'(x)h$, see Fig. 6.

The benefits of this type of approximation of $f(x + h)$ are manifold: much information about the local behavior of f near x can be gleaned from the knowledge of the first-order approximation $h \mapsto f(x) + f'(x)h$. For instance, just by looking at the sign of $f'(x)$, it is possible to determine whether f is increasing or decreasing in the neighborhood of x.

The concept of differential generalizes these fundamental ideas to the case of functions $f : E \to F$ between arbitrary normed vector spaces E and F. For a given point $x \in E$, we shall approximate (up to a remainder of the order $\mathrm{o}(h)$) the perturbed values $f(x + h)$ of f near x by an affine function $h \mapsto f(x) + \mathrm{d}f_x(h)$, where $\mathrm{d}f_x : E \to F$ is now a continuous, linear mapping from E to F. As we shall see, a lot of information about

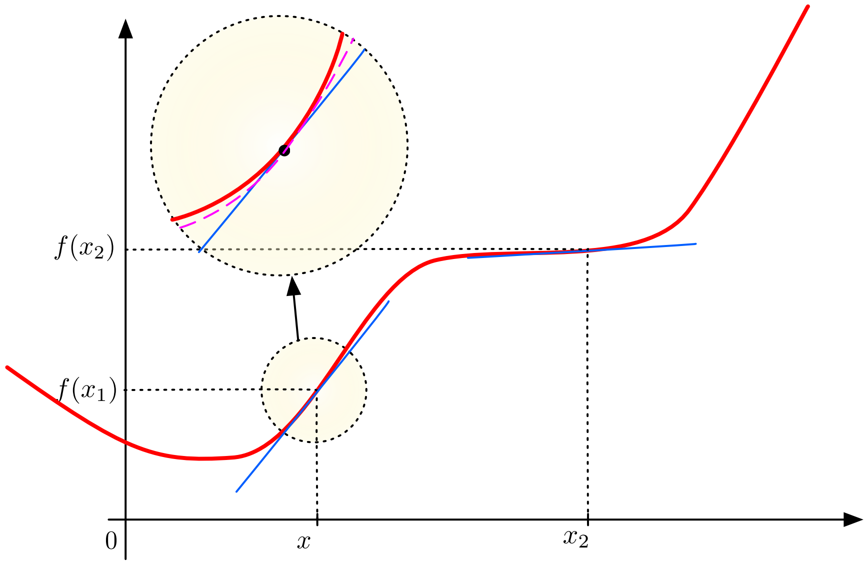

FIGURE 6. *Interpretation of the derivative of a function $f : \mathbb{R} \to \mathbb{R}$: when "zooming" on a point $x \in \mathbb{R}$, the function $h \mapsto f(x+h)$ resembles the constant value $f(x)$; a more precise approximation is given by the affine function $h \mapsto f(x) + f'(x)(h)$ (whose graph is the blue line), and an even more precise approximation is given by the second-order expansion $h \mapsto f(x) + f'(x)h + \frac{1}{2}f''(x)h^2$ (in purple).*

the local behavior of f near x is encoded in this differential (e.g. its local invertibility, as appraised by the local inverse Theorem 6.2 below).

Remark 2.1. *The asymptotic expansion (2.1) of a function $f : \mathbb{R} \to \mathbb{R}$ near a point $x \in \mathbb{R}$ can be pursued to higher orders; the second-order Taylor expansion of f near x indeed yields:*

$$f(x+h) = f(x) + f'(x)h + \frac{1}{2}f''(x)h^2 + \mathrm{o}(h^2),$$

where f is now approximated up to second-order by the quadratic function $h \mapsto f(x) + f'(x)h + \frac{1}{2}f''(x)h^2$. This process will also be extended to functions between general normed vector spaces, see Section 4 about the notion of higher-order differentials and general Taylor's formulas.

2.1. **Fréchet derivatives in normed vector spaces**

Let us start by defining differentiable mappings between normed vector spaces.

Proposition-Definition 2.1. *Let $(E, ||\cdot||_E)$ and $(F, ||\cdot||_F)$ be two normed vector spaces, and let U be an open subset of E. One function $f : U \to F$ is said to be differentiable in the sense of Fréchet at some point $x \in U$ if there exists a linear, continuous mapping $L : E \to F$ such that*

$$\forall \varepsilon > 0, \ \exists \delta > 0, \forall h \in E, \quad ||h||_E \leq \delta \ \Rightarrow \ \frac{||f(x+h) - f(x) - L(h)||_F}{||h||_E} \leq \varepsilon. \tag{2.2}$$

In such case, the linear mapping L satisfying (2.2) is unique, and it is called the differential of f at x, or the Fréchet derivative of f at x.

The function f is called differentiable on U if it is differentiable at every point $x \in U$.

Notation. Several notations exist in the literature for the above differential $L(h)$ of f at $x \in U$ in a direction $h \in E$. Depending on the authors, this quantity may be denoted by $\mathrm{d}f_x(h)$, $Df(x)(h)$ or again $f'(x)(h)$...

Remark 2.2. *The fact that the domain of definition U of f be open in the above definition is essential to guarantee that for $x \in U$ and for $h \in E$ "small enough", the perturbed $(x+h)$ also belongs to U, so that the perturbed quantity $f(x+h)$ is well-defined.*

It is customary to call differentiable a function $f : A \to F$ defined on a possibly non open subset $A \subset E$ if it is the restriction $g|_A$ to A of a differentiable function $g : U \to F$ on an open set $U \subset E$ containing A.

Remark 2.3. *The relation (2.2) is often expressed in alternative forms. In particular, it is equivalent to the existence of a function $r : E \to F$ such that $r(h) \to 0$ as $h \to 0$ and*

$$\text{For small enough } h \in E, \quad f(x+h) = f(x) + L(h) + ||h||_E r(h).$$

Yet another equivalent statement to (2.2) is the following:

$$f(x+h) = f(x) + L(h) + \mathrm{o}(h),$$

where $\mathrm{o}(h)$ stands for an unspecified function of h which tends to 0 faster than h as $h \to 0$, that is

$$\lim_{h \to 0} \frac{\mathrm{o}(h)}{||h||_E} = 0.$$

Let us warn the reader that this last formulation, however widespread and seemingly handful, is dangerous: for instance, when f also depends on other variables and parameters than x, the notation $\mathrm{o}(h)$ hides how the remainder depends on these parameters.

Proof of Proposition-Definition 2.1. Let L_1, L_2 be two linear and continuous mappings from E into F satisfying (2.2). Then, there exist $\delta > 0$ and two functions $r_1, r_2 : E \to \mathbb{R}$ such that $r_1(h) \to 0$ and $r_2(h) \to 0$ as $h \to 0$ and

$$\forall ||h||_E \leq \delta, \quad f(x+h) = f(x) + L_1(h) + ||h||_E r_1(h), \text{ and } f(x+h) = f(x) + L_2(h) + ||h||_E r_2(h).$$

Taking the difference and dividing both sides by $||h||_E$, we see that

$$\forall ||h||_E \leq \delta, \; h \neq 0, \; \frac{L_2(h) - L_1(h)}{||h||_E} = -(r_2(h) - r_1(h)).$$

Let us now fix the direction $h \neq 0$; we have, for $t > 0$ small enough:

$$\text{For } t \text{ small enough,} \quad \frac{L_2(h) - L_1(h)}{||h||_E} = -(r_2(th) - r_1(th)),$$

and since the term in the above right-hand side tends to 0 as $t \to 0$, we arrive at:

$$\frac{L_2(h) - L_1(h)}{||h||_E} = 0.$$

As $h \neq 0$ is arbitrary, it follows that $L_1 = L_2$. □

Let us recall from the introduction of this section that the definition of the differential of f can be thought of as a means to approximate f near a considered point x: roughly, the quantity $f(x+h)$ is approximated at 0^{th} order (i.e. very crudely) by the value $f(x)$ of f at x when h is "small enough"; a more precise, first-order approximation is achieved by the affine function $h \mapsto f(x) + \mathrm{d}f_x(h)$, see Fig. 7.

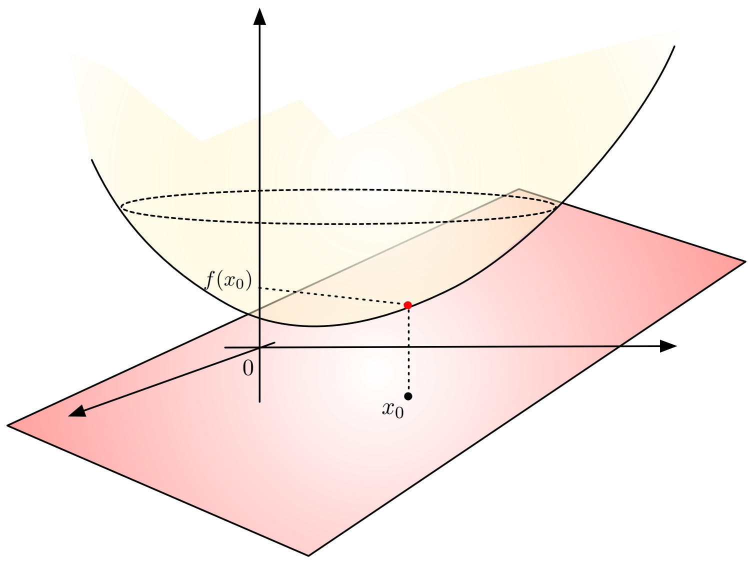

FIGURE 7. *Near the point x_0, the graph of a function $f : \mathbb{R}^2 \to \mathbb{R}$ (in yellow) is approximated by the tangent plane $\left\{f(x_0) + \mathrm{d}f_{x_0}(h),\ \ h \in \mathbb{R}^2\right\}$.*

Let us provide a few elementary examples of Fréchet derivatives. The reader is strongly encouraged to check these simple results by himself to appraise his understanding of the basic notions.

Example 2.1.

(i) Let $f : \mathbb{R} \to \mathbb{R}$ be a function which has a derivative at some point $x_0 \in \mathbb{R}$ in the usual sense; then f is differentiable at x_0 in the sense of Proposition-Definition 2.1, and its Fréchet derivative at x_0 is the continuous linear mapping $\mathrm{d}f_{x_0} : \mathbb{R} \to \mathbb{R}$ defined by:

$$\forall h \in \mathbb{R}, \quad \mathrm{d}f_{x_0}\ (h) = f'(x_0)h.$$

In this situation, by a slight abuse of notations, one usually identifies the differential $\mathrm{d}f_x$ (which is a linear mapping from $\mathbb{R}$ into itself) with the derivative $f'(x)$ (which is a real number).

(ii) Let $f : I \to F$ be a function from an open interval $I \subset \mathbb{R}$ into the normed vector space F. It follows from the definition that f is Fréchet differentiable at some point $t \in I$ if and only if the limit

$$f'(t) := \lim_{h \to 0} \frac{f(t+h) - f(t)}{h}$$

exists in E. The Fréchet differential $\mathrm{d}f_t \in \mathcal{L}(R; F)$ is the defined by

$$\forall h \in \mathbb{R}, \quad \mathrm{d}f_t(h) = hf'(t).$$

(iii) Let $f : E \to F$ be a constant function on E; then $\mathrm{d}f_x = 0$ at all points $x \in U$.

(iv) Let $f : E \to F$ be a continuous linear mapping; then f is differentiable at all points $x \in E$ and

$$\forall h \in E, \quad \mathrm{d}f_x(h) = f(h).$$

(v) Let $m : E^n \to F$ be a continuous, n-linear mapping from E^n to F. We define the function $f : E \to F$ by $f(x) = m(x, \ldots, x)$. Then, f is differentiable at every point $x \in E$ and its derivative reads:

$$\forall h \in E, \quad \mathrm{d}f_x(h) = m(h, x, \ldots, x) + m(x, h, x, \ldots, x) + \ldots + m(x, \ldots, x, h).$$

(vi) Let $(E_1, ||\cdot||_{E_1})$, $(E_2, ||\cdot||_{E_2})$ and $(F, ||\cdot||_F)$ be three normed vector space, and let $b : E_1 \times E_2 \to F$ be a continuous bilinear mapping. Then the mapping b is Fréchet differentiable at any point $(x_1, x_2) \in E_1 \times E_2$, and its Fréchet derivative reads:

$$\forall (h_1, h_2) \times E_1 \times E_2, \quad \mathrm{d}b_{(x_1,x_2)}(h_1, h_2) = b(x_1, h_2) + b(h_1, x_2).$$

Remark 2.4. *In practice, recipes are scarce when it comes to proving the differentiability of a function $f : E \to F$ between infinite-dimensional normed vector spaces. Often, the only available option is to return to the definition, write down the definition of $f(x+h) - f(x)$, try to force out a linear term with respect to h and then estimate the remainder.*

We next turn to investigate how the differential mapping $x \mapsto \mathrm{d}f_x$ attached to a function f (taking values in the space of linear continuous mappings from E to F) depends on the point $x \in U$ where the derivative is calculated. This leads us to the notion of function of class $\mathcal{C}^1$.

Definition 2.1. *The function f is said to be of class $\mathcal{C}^1$ on U if the mapping*

$$\mathrm{d}f : U \ni x \mapsto \mathrm{d}f_x \in \mathcal{L}(E; F)$$

is continuous.

Remark 2.5. *Let us warn the reader about the common confusion between the continuity of the linear mapping*

$$E \ni h \mapsto \mathrm{d}f_x(h) \in F,$$

which stems directly from the definition of the differential $\mathrm{d}f_x$ of f at a given point $x \in U$, and the continuity of the (non linear) differential mapping

$$U \ni x \mapsto \mathrm{d}f_x \in \mathcal{L}(E; F).$$

The latter appraises the dependence of the differential with respect to the base point, and it is precisely the feature which is evaluated by the $\mathcal{C}^1$ character of f.

Another category of function, which will be involved repeatedly in the sequel is introduced in the next definition.

Definition 2.2. *Let $(E, ||\cdot||_E)$ and $(F, ||\cdot||_F)$ be two normed vector spaces, and let $U \subset E$ and $V \subset F$ be open. One function $f : U \to V$ is called a diffeomorphism of class $\mathcal{C}^1$ between U and V if*

- *f is a function of class $\mathcal{C}^1$;*
- *f is bijective;*
- *The inverse mapping $f^{-1} : V \to U$ is also of class $\mathcal{C}^1$.*

Remark 2.6. *As we have seen, the Fréchet derivative of a real-valued function $f : E \to \mathbb{R}$ is a continuous linear mapping on E. It turns out that this derivative can be identified with an element in E – the gradient of f at x – in the particular case where E is a Hilbert space.*

More precisely, when H is a Hilbert space, and $f : A \to \mathbb{R}$ is a Fréchet differentiable function on an open subset $A \subset H$, for any $x \in A$, the continuous linear form $\mathrm{d}f_x : H \to \mathbb{R}$ can be represented by a unique vector $\nabla f(x) \in H$ via the Riesz representation Theorem 1.7:

$$\forall x \in H, \quad \mathrm{d}f_x(h) = \langle \nabla f(x), h \rangle.$$

The vector $\nabla f(x) \in H$ is called the gradient of f at x.

2.2. Operations with Fréchet derivatives

Let us start with a few elementary properties of Fréchet derivatives, whose proofs offer nice opportunities to handle the notion.

Proposition 2.1. *Let $(E, ||\cdot||_E)$ and $(F, ||\cdot||_F)$ be two normed vector spaces and let $U \subset E$ be open.*

(i) Let $f : U \to F$ be a differentiable function at some point $x \in U$; then f is continuous at x.

(ii) Let f and $g : U \to F$ be two differentiable functions at some point $x \in U$. Then $(f+g)$ is differentiable at x and:

$$\forall h \in E, \quad \mathrm{d}(f+g)_x(h) = \mathrm{d}f_x(h) + \mathrm{d}g_x(h).$$

(iii) Let $f : U \to F$ be differentiable at $x \in U$ and let $\lambda \in \mathbb{R}$. Then λf is differentiable at x and:

$$\forall h \in E, \quad \mathrm{d}(\lambda f)_x(h) = \lambda \mathrm{d}f_x(h).$$

(iv) More generally, let $f : U \to F$ and $m : U \to \mathbb{R}$ be two Fréchet differentiable functions at $x \in U$. Then the product $mf : U \to \mathbb{R}$ is also differentiable at x and it holds

$$\forall h \in E, \quad \mathrm{d}(mf)_x(h) = m(x)\mathrm{d}f_x(h) + (\mathrm{d}m_x(h))f(x).$$

Proof. *(i):* Let $\varepsilon > 0$ be given; by the definition of differentiability, there exists $\delta > 0$ such that, for all $h \in E$ with $||h||_E \leq \delta$,

$$||f(x+) - f(x) - \mathrm{d}f_x(h)||_F < \varepsilon ||h||_E.$$

Now, using the triangle inequality and the continuity of the linear mapping $\mathrm{d}f_x \in \mathcal{L}(E;F)$, we obtain, for $||h||_E \leq \delta$,

$$||f(x+h) - f(x)||_F \leq \Big(\varepsilon + ||\mathrm{d}f_x||_{\mathcal{L}(E;F)}\Big)||h||_E.$$

This immediately implies that f is continuous at x.

(ii): Since f is differentiable at x, there exist $\delta > 0$ and a function $r_f : E \to F$ such that $r_f(h) \to 0$ as $h \to 0$ and

$$\forall h \in E,\ ||h||_E \leq \delta \Rightarrow f(x+h) = f(x) + \mathrm{d}f_x(h) + ||h||_E r_f(h).$$

Likewise, the differentiability of g at x implies the existence of a function $r_g : E \to F$ such that $r_g(h) \to 0$ as $h \to 0$ and (up to decreasing the value of δ):

$$\forall h \in E,\ ||h||_E \leq \delta \Rightarrow g(x+h) = g(x) + \mathrm{d}g_x(h) + ||h||_E r_g(h).$$

Adding both identities, we obtain that, for $||h||_E \leq \delta$,

$$(f+g)(x+h) = (f+g)(x) + \Big(\mathrm{d}f_x(h) + \mathrm{d}g_x(h)\Big) + ||h||_E \Big(r_f(h) + r_g(h)\Big),$$

and since $r_f(h) + r_g(h) \to 0$ as $h \to 0$, Proposition-Definition 2.1 shows that $(f+g)$ is differentiable at x with differential $\mathrm{d}(f+g)_x = \mathrm{d}f_x + \mathrm{d}g_x$.

(iii) is on any point simpler than (iv) below, and it is left to the reader.

(iv): Let $\delta > 0$ be such that $\overline{B(x,\delta)} \subset U$; let us write, for $||h||_E \leq \delta$,

(2.3)

$$\begin{aligned} m(x+h)f(x+h) &= \Big(m(x+h) - m(x)\Big)f(x+h) + m(x)f(x+h) \\ &= \Big(m(x+h) - m(x)\Big)f(x+h) + m(x)\Big(f(x+h) - f(x)\Big) + m(x)f(x) \\ &= m(x)f(x) + \Big(m(x+h) - m(x)\Big)f(x) + m(x)\Big(f(x+h) - f(x)\Big) \\ &\qquad + \Big(m(x+h) - m(x)\Big)\Big(f(x+h) - f(x)\Big). \end{aligned}$$

Now, since m is differentiable at x, there exists a function $r_m : E \to \mathbb{R}$ such that (up to decreasing the value of δ):

$$r_m(h) \to 0 \text{ as } h \to 0, \text{ and for } ||h||_E \leq \delta, \quad m(x+h) = m(x) + \mathrm{d}m_x(h) + ||h||_E r_m(h),$$

and likewise, there exists $r_f : E \to F$ such that (up to decreasing again δ):

$$r_f(h) \to 0 \text{ as } h \to 0, \text{ and for } ||h||_E \leq \delta, \quad f(x+h) = f(x) + \mathrm{d}f_x(h) + ||h||_E r_f(h).$$

Hence, (2.3) rewrites, for all $h \in E$ such that $||h||_E \leq \delta$:

$$m(x+h)f(x+h) = m(x)f(x) + \mathrm{d}m_x(h)f(x) + m(x)\mathrm{d}f_x(h) + ||h||_E \widetilde{r}(h),$$

where we have set

$$\widetilde{r}(h) := \begin{cases} 0 & \text{if } h = 0, \\ r_m(h)f(x) + m(x)r_f(h) + ||h||_E \left(\mathrm{d}m_x\left(\frac{h}{||h||_E}\right) + r_m(h)\right)\left(\mathrm{d}f_x\left(\frac{h}{||h||_E}\right) + r_f(h)\right) & \text{otherwise.} \end{cases}$$

Finally, since $\mathrm{d}m_x$ and $\mathrm{d}f_x$ are continuous linear mappings, there exists a constant $C > 0$ such that

$$\forall h \in E \setminus \{0\}, \quad \left|\left|\mathrm{d}f_x\left(\frac{h}{||h||_E}\right)\right|\right|_F \leq C, \text{ and } \left|\mathrm{d}m_x\left(\frac{h}{||h||_E}\right)\right| \leq C.$$

This shows that $\widetilde{r}(h) \to 0$ when $h \to 0$, thus completing the proof. □

One crucial property of Fréchet derivatives is related to their behavior with respect to the composition of functions.

Theorem 2.1 (Chain rule)**.** *Let $(E, ||\cdot||_E)$, $(F, ||\cdot||_F)$ and $(G, ||\cdot||_G)$ be three normed vector spaces, and let $U \subset E$ and $V \subset F$ be open subsets. Let $f : U \to F$ and $g : V \to G$ be two functions; we assume that $f(U) \subset V$ so that the composite mapping $g \circ f : U \to G$ is well-defined. Let $x \in U$ be a point such that f is differentiable at x and g is differentiable at $f(x)$. Then, $g \circ f : U \to G$ is differentiable at x and its Fréchet derivative reads:*

$$\forall h \in E, \quad \mathrm{d}(g \circ f)_x(h) = \mathrm{d}g_{f(x)}\left(\mathrm{d}f_x(h)\right).$$

Proof. Let us set $y = f(x) \in F$; from the definition of the differentiability of f at x, there exist $\delta_1 > 0$ and a function $r_f : E \to F$ such that

$$\text{For } ||h||_E \leq \delta_1, \quad f(x+h) = f(x) + \mathrm{d}f_x(h) + ||h||_E r_f(h), \text{ and } \lim_{h\to 0} r_f(h) = 0. \tag{2.4}$$

Likewise, there exist $\delta_2 > 0$ and a function $r_g : F \to G$ with

$$\text{For } ||k||_F \leq \delta_2, \quad g(y+k) = g(y) + \mathrm{d}g_y(k) + ||k||_F r_g(k), \text{ and } \lim_{k\to 0} r_g(k) = 0. \tag{2.5}$$

Let us now write, for $||h||_E \leq \delta_1$:

$$g \circ f(x+h) = g(f(x)+k), \text{ where we have set } k := \mathrm{d}f_x(h) + ||h||_E r_f(h) \in F.$$

The continuity of $\mathrm{d}f_x \in \mathcal{L}(E;F)$ and the fact that $r_f(h) \to 0$ as $h \to 0$ imply that, up to decreasing the value of δ_1, it holds $||k||_F \leq \delta_2$. Therefore, the expansion (2.5) yields, for $h \in E$ such that $||h||_E \leq \delta_1$:

$$\begin{aligned} g \circ f(x+h) &= g(y) + \mathrm{d}g_y(k) + ||k||_F r_g(k) \\ &= g \circ f(x) + \mathrm{d}g_y\Big(\mathrm{d}f_x(h) + ||h||_E r_f(h)\Big) + ||\mathrm{d}f_x(h) + ||h||_E r_f(h)||_F \, r_g\Big(\mathrm{d}f_x(h) + ||h||_E r_f(h)\Big) \\ &= g \circ f(x) + \mathrm{d}g_y\Big(\mathrm{d}f_x(h)\Big) + ||h||_E \, \mathrm{d}g_y\Big(r_f(h)\Big) + ||\mathrm{d}f_x(h) + ||h||_E r_f(h)||_F \, r_g\Big(\mathrm{d}f_x(h) + ||h||_E r_f(h)\Big) \\ &= g \circ f(x) + \mathrm{d}g_{f(x)}\Big(\mathrm{d}f_x(h)\Big) + ||h||_E \widetilde{r_f}(h), \end{aligned}$$

where we have defined the remainder:

$$\widetilde{r_f}(h) = \begin{cases} 0 & \text{if } \; h = 0, \\ \mathrm{d}g_y(r_f(h)) + \left|\left|\mathrm{d}f_x\left(\frac{h}{||h||_E}\right) + r_f(h)\right|\right|_F r_g\Big(\mathrm{d}f_x(h) + ||h||_E r_f(h)\Big) & \text{otherwise.} \end{cases}$$

To complete the proof of the theorem, we simply need to show that the remainder $\tilde{r}_f(h)$ tends to 0 in G as $h \to 0$. The definition of this quantity features a sum of two contributions, which we analyze separately:

- Since $r_f(h) \to 0$ as $h \to 0$ and $\mathrm{d}g_y$ is a continuous mapping from F to $\mathbb{R}$, we have:

$$\mathrm{d}g_y(r_f(h)) \to 0 \text{ as } \; h \to 0.$$

- Since $\mathrm{d}f_x : E \to \mathbb{R}$ is a continuous linear mapping, there exists a constant $C > 0$ independent of h such that

$$\forall h \in E \setminus \{0\}, \quad \left|\left|\mathrm{d}f_x\left(\frac{h}{||h||_E}\right)\right|\right|_F \leq C.$$

  Besides, since $r_f(h) \to 0$ as $h \to 0$, there exists $\delta_3 > 0$ such that r_f is bounded on $\overline{B(0,\delta)}$ (by a constant $\widetilde{C}$):

$$\forall ||h||_E \leq \delta, \quad ||r_f(h)||_F \leq \widetilde{C}.$$

  It follows from the triangle inequality that:

$$\forall h \in E \setminus \{0\}, \; ||h||_E \leq \delta_3 \Rightarrow \; \begin{aligned} \left|\left|\mathrm{d}f_x\left(\tfrac{h}{||h||_E}\right) + r_f(h)\right|\right|_F &\leq \left|\left|\mathrm{d}f_x\left(\tfrac{h}{||h||_E}\right)\right|\right|_F + ||r_f(h)||_F \\ &\leq C + \widetilde{C} \end{aligned}$$

  Finally, since $\mathrm{d}f_x(h) \to 0$ and $||h||_E r_f(h) \to 0$ as $h \to 0$ in E, and since $r_g(k) \to 0$ as $k \to 0$ in F, we obtain by composition of limits that

$$r_g\Big(\mathrm{d}f_x(h) + ||h||_E r_f(h)\Big) \to 0 \text{ as } \; h \to 0.$$

Hence,
$$\left|\left|\mathrm{d}f_x\left(\frac{h}{||h||_E}\right)+r_f(h)\right|\right|_F r_g\Big(\mathrm{d}f_x(h)+||h||_E r_f(h)\Big)\xrightarrow{h\to 0}0,$$
Summarizing, we have proved that $\widetilde{r}_f(h)\to 0$ as $h\to 0$, and the proof is complete. □

The following example is fundamental, as it often allows to reduce the study of a function f from an open subset U of a normed vector space E into another normed vector space F to that of a function g from an interval $I\subset\mathbb{R}$ into F.

Example 2.2. Let $(E,||\cdot||_E)$ and $(F,||\cdot||_F)$ be two normed vector spaces, and let $U\subset E$ be open. Let $f:U\to F$ be a function which is n times differentiable on U, and let $h\in E$ be such that the whole segment $\{x+th,\ t\in[0,1]\}$ is contained in U. Let us define the composite function $g:[0,1]\to F$ by
$$\forall t\in[0,1],\quad g(t)=f(x+th),$$
that is, $g=f\circ\ell$, where $\ell:[0,1]\to U$ is given by $\ell(t)=x+th$. The chain rule Theorem 2.1 allows to conclude that g is differentiable on $(0,1)$ and that its derivative reads:
$$\forall t\in(0,1),\quad g'(t)=\mathrm{d}f_{x+th}(h).$$

2.3. Partial Fréchet derivatives

Sometimes, the function f under scrutiny depends on two (or more) variables belonging to different normed vector spaces, and it is desirable to take derivatives with respect to only one of them.

Definition 2.3. *Let $(E_1,||\cdot||_{E_1})$, $(E_2,||\cdot||_{E_2})$ and $(F,||\cdot||_F)$ be three normed vector spaces, and let $U\subset E_1$, $V\subset E_2$ be open subsets. A function $f:U\times V\to F$ is said to possess a partial differential (or partial Fréchet derivative) with respect to the first variable at some point $(x,y)\in U\times V$ if the partial mapping $U\ni z\mapsto f(z,y)\in F$ is Fréchet differentiable at $z=x$, i.e. if there exists a continuous linear mapping $L_1:E_1\to F$ such that*
$$\forall\varepsilon>0,\ \exists\delta>0,\ ||h||_{E_1}\leq\delta\ \Rightarrow\ \frac{||f(x+h,y)-f(x,y)-L_1(h)||_F}{||h||_{E_1}}\leq\varepsilon.$$
This mapping L_1 is denoted by $\frac{\partial f}{\partial x}(x,y)\in\mathcal{L}(E_1;F)$.

Likewise, one defines the partial derivative $\frac{\partial f}{\partial y}(x,y)\in\mathcal{L}(E_2;F)$ on f with respect to the second variable.

In the above context, it is quite simple to verify that, if $f:U\times V\to F$ is Fréchet differentiable at some point $(x,y)\in U\times V$, then it has partial differentials $\frac{\partial f}{\partial y}(x,y)$ and $\frac{\partial f}{\partial y}(x,y)$ with respect to both variables. The converse statement does not hold true: f may very well possess partial differentials $\frac{\partial f}{\partial y}(x,y)$ and $\frac{\partial f}{\partial y}(x,y)$ without being differentiable with respect to the pair (x,y), see Exercise 2.1. Stronger assumptions are actually required for partial differentiability to provide information about the total differentiability of a function.

The next proposition makes the previous discussion rigorous.

Proposition 2.2. *Let $(E_1,||\cdot||_{E_1})$, $(E_2,||\cdot||_{E_2})$ and $(F,||\cdot||_F)$ be three normed vector spaces, and let $U\subset E_1$, $V\subset E_2$ be open subsets. Let $f:U\times V\to F$ be a function and let $(x_0,y_0)\in U\times V$.*

(i) If f is Fréchet differentiable at (x_0,y_0), then it admits partial Fréchet derivatives $\frac{\partial f}{\partial y}(x,y)$ and $\frac{\partial f}{\partial y}(x,y)$, which are respectively given by:
$$\forall h\in E_1,\quad \frac{\partial f}{\partial x}(x_0,y_0)(h)=\mathrm{d}f_{(x_0,y_0)}(h,0),\ \textit{and}\ \forall k\in E_2,\quad \frac{\partial f}{\partial y}(x_0,y_0)(h)=\mathrm{d}f_{(x_0,y_0)}(0,k).$$

(ii) Assume that there exist open neighborhoods $U'\subset U$ and $V'\subset V$ of x_0 and y_0 respectively such that the partial derivatives
$$U'\times V'\ni(x,y)\mapsto\frac{\partial f}{\partial x}(x,y)\in\mathcal{L}(E_1;F)\ \textit{and}\ U'\times V'\ni(x,y)\mapsto\frac{\partial f}{\partial y}(x,y)\in\mathcal{L}(E_2;F)$$
exist and define continuous mappings. Then f is Fréchet differentiable at (x_0,y_0) and its derivative reads:
$$\forall(h,k)\in E_1\times E_2,\quad \mathrm{d}f_{(x_0,y_0)}(h,k)=\frac{\partial f}{\partial x}(x_0,y_0)(h)+\frac{\partial f}{\partial y}(x_0,y_0)(k).$$

Proof. *(i):* From the definition of the differentiability of f at (x_0, y_0), for all $\varepsilon > 0$, there exists a number $\delta > 0$ such that

$$\forall h \in E_1,\ k \in E_2,\ \ ||h||_{E_1} + ||k||_{E_2} \leq \delta\ \Rightarrow \left|\left|\frac{f(x_0+h, y_0+k) - f(x_0, y_0) - \mathrm{d}f_{(x_0,y_0)}(h,k)}{||h||_{E_1} + ||k||_{E_2}}\right|\right| \leq \varepsilon.$$

In particular, this implies that for all $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\forall h \in E_1,\ \ ||h||_{E_1} \leq \delta\ \Rightarrow \left|\left|\frac{f(x_0+h, y_0) - f(x_0, y_0) - \mathrm{d}f_{(x_0,y_0)}(h,0)}{||h||_{E_1} + ||k||_{E_2}}\right|\right| \leq \varepsilon,$$

which proves that the partial mapping $x \mapsto f(x, y_0)$ is Fréchet differentiable at $x = x_0$, with Fréchet derivative $E_1 \ni h \mapsto \mathrm{d}f_{(x_0,y_0)}(h, 0)$. The corresponding statement related to the partial differentiation of f with respect to the second variable follows in a similar way.

(ii): This point is more delicate, and it relies on the generalized version of the Mean Value Theorem 1.1 to the case of functions taking values in an arbitrary normed vector space. It is proposed as Exercise 4.1. □

Remark 2.7. *The above definition and proposition are easily generalized to the case of a function $f : E_1 \times \ldots E_n \to F$ of $n \geq 3$ variables in different normed vector spaces $(E_1, ||\cdot||_{E_1}), \ldots (E_n, ||\cdot||_{E_n})$.*

2.4. **Other notions of derivatives**

The notion of Fréchet derivative that we have been discussing in the previous two sections is only one among the various means to appraise the first-order behavior of a function between normed vector spaces. Other, weaker notions of derivatives are available, that we now present, together with their connections with Fréchet derivatives.

Definition 2.4. *Let $(E, ||\cdot||_E)$ and $(F, ||\cdot||_F)$ be two normed vector spaces, and $U \subset E$ be open. Let $f : U \to F$ be a function.*

(i) The function f has a one-sided directional derivative at some point $x \in U$ and in a direction $h \in E$ if the limit

$$f'(x; h) := \lim_{\substack{t \to 0 \\ t > 0}} \frac{f(x + th) - f(x)}{t}$$

exists and is finite.

(ii) The function f is called Gateaux-differentiable at $x \in U$ if it has a one-sided directional derivative at x in all the directions $h \in E$, and if the mapping

$$E \ni h \longmapsto f'(x; h) \in F$$

is a linear continuous mapping. The Gateaux derivative $f'_G(x) \in \mathcal{L}(E; F)$ is then defined by

$$\forall h \in E, \quad f'_G(x)(h) = f'(x; h).$$

The difference between the notions of (one-sided) directional differentiability and Gateaux differentiability is fairly simple to appraise. Directional differentiability only evaluates how the function f behaves in all directions of space, when the latter are considered independently from one another. Gateaux differentiability additionally demands that these behaviors somehow "agree with each other" (i.e. the mapping $h \mapsto f(x; h)$ should be linear and continuous). One should bear in mind that a function can possess directional derivatives in all directions without being Gateaux differentiable, as is revealed in the following example.

Example 2.3. The function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = |x|$ is differentiable at $x = 0$ in all directions since:

$$\lim_{\substack{t \to 0 \\ t > 0}} \frac{f(0 + th) - f(0)}{t} = \begin{cases} h & \text{if } \ v \geq 0, \\ -h & \text{otherwise.} \end{cases}$$

and this calculation yields $f'(0; h) = |h|$. Since the mapping $h \mapsto f(x; h)$ is not linear, f is not Gateaux differentiable at $x = 0$.

The definitions of Fréchet and Gateaux differentiability admittedly look quite similar, but they actually differ from their assumptions about the remainder term $o(h)$. Both notions require f to be differentiable at x in all directions $h \in E$, and that the resulting derivative be a linear and continuous mapping of h. When f is only required to be Gateaux differentiable, the rate at which $f(x+th) = f(x) + tf'(x:h) + o(t)$ is approximated by its first-order expansion $f(x) + tf'(x;h)$ (i.e. the size of the remainder $o(t)$) depends on the direction h, whereas when f is Fréchet differentiable, this rate is uniform in all directions $h \in E$.

Proposition 2.3. *Let $(E, ||\cdot||_E)$ and $(F, ||\cdot||_F)$ be two normed vector spaces, and $U \subset E$ be open. If a function $f: U \to F$ is Fréchet differentiable at some point $x \in U$, then it is also Gateaux differentiable at x.*

Proof. Since f is Fréchet differentiable at x, there exist $\delta > 0$ and a function $r_f : E \to F$ such that $\lim_{h\to 0} r_f(h) = 0$ and

$$\forall h \in E,\ ||h||_E \leq \delta \Rightarrow f(x+h) = f(x) + \mathrm{d}f_x(h) + ||h||_E r_f(h).$$

Hence, for any fixed direction $h \in E$, and for $t > 0$ small enough, it holds:

$$\left| \frac{f(x+th) - f(x)}{t} - \mathrm{d}f_x(h) \right| = |r_f(th)| \xrightarrow{t\to 0} 0,$$

so that f has directional derivatives in all directions and $f'(x;h) = \mathrm{d}f_x(h)$. Clearly, this implies that f is Gateaux differentiable at x, with Gateaux derivative $f'_G(x) = \mathrm{d}f_x$. □

The fact that not all Gateaux differentiable functions are Fréchet differentiable is illustrated in the next example.

Example 2.4 (One function which is Gateaux differentiable without being Fréchet differentiable)**.** Let us consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by the following formula, see Fig. 8:

$$f(x_1, x_2) = \begin{cases} 1 & \text{if } (x_1; x_2) \neq (0,0) \text{ and } x_2 = x_1^2, \\ 0 & \text{otherwise.} \end{cases}$$

For any given direction $h = (h_1, h_2) \in \mathbb{R}^2$, and for $t > 0$ small enough, it is clear that $t|h_2| > t^2 h_1^2$, and so, for t small enough, $f(0 + th) = 0$. As a result, f is differentiable in the direction h, with one-sided derivative $f'(0;h) = 0$. It follows immediately that f is Gateaux-differentiable at $x = 0$, and that $f'_G(0) = 0$. However, f is not Fréchet differentiable at 0 since it is not even continuous at this point.

Intuitively, the expansion $f(0+h) = f(0) + tf'(0;h) + o(t)$ holds for any fixed direction $h \in \mathbb{R}^2$, but $o(t)$ depends on h, and it converges to 0 slower and slower as h tends to a horizontal direction $h = (h_1, 0)$.

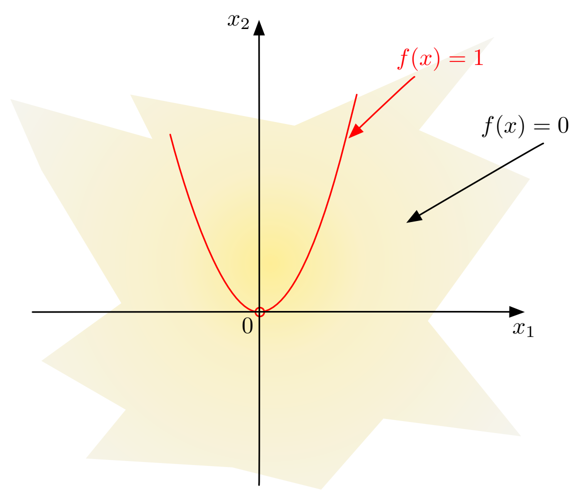

Figure 8. *Illustration of the Gateaux differentiable function which is not Fréchet differentiable presented in Example 2.4.*

Some sort of a converse to Proposition 2.3 holds however, with slightly stronger assumptions. The proof requires some of the tools developed in Section 3, and it is therefore proposed as Exercise 3.6.

Proposition 2.4. *Let $(E, ||\cdot||_E)$ be a normed vector space, $U \subset E$ be open, and let $(F, ||\cdot||_F)$ be a Banach space. Assume that $f : U \to F$ is Gateaux differentiable in the neighborhood V of a given point $x \in U$, and that the mapping induced by the Gateaux derivative $V \ni z \mapsto f'_G(z) \in \mathcal{L}(E;F)$ is continuous. Then f is Fréchet differentiable at x, and $\mathrm{d}f_x = f'_G(x)$.*

Remark 2.8. *When f is a convex function, it is possible to define yet another notion of derivative which is very useful in convex analysis: the notion of subdifferential.*

2.5. **The finite-dimensional case**

We now turn to specialize the previous concepts and results to the particular case where the considered normed vector spaces are finite-dimensional. Throughout this section, we shall consider functions f from an open subset U of $E = \mathbb{R}^d$, taking values in $F = \mathbb{R}^n$, both spaces being equipped with their canonical basis, the associated canonical inner product $\langle \cdot, \cdot \rangle$ and the induced norm $|\cdot|$. More precisely, we shall write

$$\forall x = (x_1, \dots, x_d) \in \mathbb{R}^d, \quad f(x) = f(x_1, \dots, x_d) = (f_1(x_1, \dots, x_d), \dots, f_n(x_1, \dots, x_d)),$$

where $f_1, \dots, f_n : U \to \mathbb{R}$ are the components of the $\mathbb{R}^n$-valued function f.

Definition 2.5. *Let $U \subset \mathbb{R}^d$ be an open subset, and let $f : U \to \mathbb{R}^n$ be a function. For $x = (x_1, \dots, x_d) \in U$ and $i = 1, \dots, d$, f is said to have a i^{th} partial derivative at x if the partial mapping*

$$\mathbb{R} \ni t \longmapsto f(x_1, \dots, x_{i-1}, x_i + t, x_{i+1}, \dots, x_d) \in \mathbb{R}^n$$

is differentiable at $t = 0$ in the usual sense, that is, if each of the component functions $t \mapsto f_p(x_1, \dots, x_{i-1}, x_i + t, x_{i+1}, \dots, x_d)$ is differentiable in the usual sense, $p = 1, \dots, n$.

Remark 2.9. *This definition of partial derivatives extends mutatis mutandis to the case where f takes values in an (infinite-dimensional) arbitrary normed vector space F.*

The relation between the partial derivatives of a function f as above and its differential is supplied by Proposition 2.2, whose finite-dimensional instance is provided below.

Proposition 2.5. *Let $U \subset \mathbb{R}^d$ be an open subset, and let $f : U \to \mathbb{R}^n$ be a function. Then,*

(i) If f is Fréchet differentiable at some point $x = (x_1, \dots, x_d) \in U$, it admits partial derivatives $\frac{\partial f}{\partial x_i}(x)$, $i = 1, \dots, d$, which are given by:

$$\frac{\partial f}{\partial x_i}(x) = \mathrm{d}f_x(e_i), \text{ where } e_i \text{ is the } i^{th} \text{ vector in the canonical basis of } \mathbb{R}^d.$$

(ii) Let $x \in U$ be a point which has an open neighborhood $U' \subset U$ such that the partial derivatives

$$U' \ni x \mapsto \frac{\partial f}{\partial x_i}(x) \in \mathbb{R}^n,$$

exist and define continuous mappings for $i = 1, \dots, d$. Then f is Fréchet differentiable at x and its differential reads:

$$\forall h = (h_1, \dots, h_d) \in \mathbb{R}^d, \quad \mathrm{d}f_x(h) = \sum_{i=1}^d \frac{\partial f}{\partial x_i}(x) h_i.$$

Remark 2.10. *In the physics literature, it is customary to denote by $\{\mathrm{d}x_i\}_{i=1,\dots,d}$ the dual basis of the canonical basis $\{e_i\}_{i=1,\dots,d}$ of $\mathbb{R}^d$, that is, $\{\mathrm{d}x_i\}$ is the basis of the dual space $(\mathbb{R}^d)^*$ defined by:*

$$\forall h = (h_1, \dots, h_d) \in \mathbb{R}^d, \quad \mathrm{d}x_i(h) = h_i,$$

so that $\mathrm{d}x_i(e_j) = 1$ if $i = j$ and 0 otherwise. With this notation, the differential $\mathrm{d}f_x$ of a differentiable function $f : U \subset \mathbb{R}^d \to \mathbb{R}$ at some point $x \in U$ can be rewritten:

$$\mathrm{d}f_x = \sum_{i=1}^d \frac{\partial f}{\partial x_i}(x)\, \mathrm{d}x_i.$$

Let us now specialize some of the previous general concepts and results to the case of finite dimensional spaces.

Definition 2.6. *Let U be an open subset of the Euclidean space $\mathbb{R}^d$.*

- *Let $\varphi : U \to \mathbb{R}$ be a scalar-valued, differentiable function at some point $x \in U$. The gradient of φ at x is the vector $\nabla\varphi(x) \in \mathbb{R}^d$ defined by*
$$\nabla\varphi(x) = \left(\frac{\partial\varphi}{\partial x_1}(x), \ldots, \frac{\partial\varphi}{\partial x_d}(x)\right).$$
- *Let $f : U \to \mathbb{R}^n$ be a vector-valued, differentiable function at some point $x \in U$. The Jacobian matrix $\nabla f(x)$ of f at x is the $n \times d$ matrix defined by*
$$\nabla f(x) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x) & \ldots & \frac{\partial f_1}{\partial x_d}(x) \\ \vdots & & \vdots \\ \frac{\partial f_n}{\partial x_1}(x) & \ldots & \frac{\partial f_n}{\partial x_d}(x) \end{pmatrix}.$$
In other terms, for $i = 1, \ldots, n$ and $j = 1, \ldots, d$, the (i, j) entry of the matrix $\nabla f(x)$ equals $\frac{\partial f_i}{\partial x_j}(x)$.
- *Let $f : U \to \mathbb{R}^d$ be a vector field which is differentiable at some point $x \in U$. The divergence of f is the number $\operatorname{div}(f)(x)$ defined by*
$$\operatorname{div}(f)(x) = \frac{\partial f_1}{\partial x_1}(x) + \ldots + \frac{\partial f_1}{\partial x_d}(x).$$
From the physical point of view, $\operatorname{div}(f)$ is a measure of the local concentration or dilation induced by the vector field f; see Section 8 for more details on the subject.

With these notations,

- Let $\varphi : U \to \mathbb{R}$ be a scalar-valued function which is differentiable at $x \in U$; the differential $d\varphi_x$ of φ at x reads:
$$\forall h \in \mathbb{R}^d, \quad d\varphi_x(h) = \langle \nabla\varphi(x), h \rangle.$$
- Let $f : U \to \mathbb{R}^n$ be a vector-valued function which is differentiable at $x \in U$; the differential df_x of f at x satisfies:
$$\forall h \in \mathbb{R}^d, \quad \underbrace{df_x(h)}_{\text{vector in } \mathbb{R}^n} = \underbrace{\nabla f(x)}_{n\times d \text{ matrix}} \underbrace{h}_{\text{vector in } \mathbb{R}^d}.$$

Remark 2.11. *However ubiquitous in the literature, there is an unfortunate ambiguity in these notations. When $\varphi : U \to \mathbb{R}$ is a real-valued function defined on an open subset $U \subset \mathbb{R}^d$, $\nabla\varphi(x)$ may either denote the vector $\nabla\varphi(x) \in \mathbb{R}^d$, or the $1 \times d$ Jacobian matrix with entries $(\nabla\varphi(x))_{1j} = \frac{\partial\varphi}{\partial x_j}(x)$. Usually, unless stated otherwise, the former meaning is often retained.*

Remark 2.12. *Other, important differential operators can be devised from the derivatives of a vector field on $\mathbb{R}^d$, see notably Exercises 2.9 and 4.6 about the strain tensor and the curl operator.*

Let us finally reformulate the chain rule in Theorem 2.1 in this particular case.

Theorem 2.2 (Chain rule in finite-dimensional spaces)**.** *Let d, n and p be integers, and $U \subset \mathbb{R}^d$, $V \subset \mathbb{R}^n$ be two open subsets. Let $f : U \to \mathbb{R}^n$ and $g : V \to \mathbb{R}^p$ be two functions such that $f(U) \subset V$, and let $x \in U$ be such that f is differentiable at x and g is differentiable at $f(x)$. Then the composite mapping $g \circ f : U \to \mathbb{R}^p$ is differentiable at x and its Jacobian matrix $\nabla(g \circ f)(x)$ reads*
$$\underbrace{\nabla(g \circ f)(x)}_{p\times d \text{ matrix}} = \underbrace{\nabla g(f(x))}_{p\times n \text{ matrix}} \underbrace{\nabla f(x)}_{n\times d \text{ matrix}}.$$

A handful corollary of this result is the following, whose proof is left as an exercise for the reader:

Corollary 2.1. *Let U be an open subset of $\mathbb{R}^d$, and let $\varphi : U \to \mathbb{R}$ be a differentiable function. Let $f : I \to \mathbb{R}^d$ be a function which is differentiable on an interval I of $\mathbb{R}$, such that $f(I) \subset U$. Then the composite mapping $g := \varphi \circ f$ is a differentiable function $I \to \mathbb{R}$, whose derivative reads:*

$$\forall t \in I, \quad g'(t) = \frac{\partial \varphi}{\partial x_1}(f(t))f_1'(t) + \ldots + \frac{\partial \varphi}{\partial x_d}(f(t))f_d'(t).$$

Example 2.5 (Chain rule and backpropagation)**.** Let us consider the following situation. We introduce differentiable functions $f_1 : \mathbb{R}^{d_1} \to \mathbb{R}^{d_2}$, $f_2 : \mathbb{R}^{d_2} \to \mathbb{R}^{d_3}$, …, $f_n : \mathbb{R}^{d_n} \to \mathbb{R}^{d_{n+1}}$, and a final scalar-valued differentiable function $\varphi : \mathbb{R}^{d_{n+1}} \to \mathbb{R}$. We aim to calculate the gradient (i.e. the collection of all partial derivatives) of the composite function $f = \varphi \circ f_n \circ \ldots \circ f_1 : \mathbb{R}^{d_1} \to \mathbb{R}$.

Such a situation typically models a neural network with $(n+1)$ layers: for each $i = 1, \ldots n+1$, d_i is the number of neurons of the i^{th} layer; φ is the activation function, delivering the output of the network f, which approximates a quantity of interest. One then aims to calculate the gradient of f in order to optimize the response of this model.

From the chain rule of Theorem 2.1, the differential of f at a point $x \in \mathbb{R}^{d_1}$ reads:

$$\forall h \in \mathbb{R}^{d_1}, \quad \mathrm{d}f_x = \mathrm{d}\varphi_{f_n \circ \ldots \circ f_1(x)} \circ \mathrm{d}f_{n\, f_{n-1} \circ \ldots \circ f_1(x)} \circ \ldots \circ \mathrm{d}f_{1x}.$$

In particular, for each $i = 1, \ldots, d_1$, the partial derivative $\frac{\partial f}{\partial x_i}(x)$ equals:

$$\frac{\partial f}{\partial x_i}(x) = \Big(\mathrm{d}\varphi_{f_n \circ \ldots \circ f_1(x)} \circ \mathrm{d}f_{n\, f_{n-1} \circ \ldots \circ f_1(x)} \circ \ldots \circ \mathrm{d}f_{1x} \Big)(e_i). \tag{2.6}$$

The calculation of any of these partial derivatives requires first to multiply the Jacobian matrix $\nabla f_1(x)$ with the vector $e_i \in \mathbb{R}^{d_1}$, then to multiply the Jacobian matrix $\nabla f_2(f_1(x))$ with the result, etc. The calculation of the gradient $\nabla f(x)$ is then very large, it demands that this procedure be repeated d_1 times.

A more efficient calculation can be conducted by taking advantage of the fact that φ (and thus f) is a scalar-valued function. One easily verifies that the gradient of f reads:

$$\underbrace{\nabla f(x)}_{\text{vector with size } d_1} \quad = \quad \underbrace{\nabla f_1^T(x)}_{d_1 \times d_2 \text{ matrix}} \ \nabla f_2(f_1(x))^T \ldots \underbrace{\nabla f_n(f_{n-1} \circ \ldots f_1(x))^T}_{d_n \times d_{n+1} \text{ matrix}} \underbrace{\nabla \varphi(f_n \circ \ldots f_1(x))}_{\text{vector with size } d_{n+1}},$$

see for instance Exercise 2.4. The calculation of the full gradient $\nabla f(x)$ by this method, called backpropagation, is more efficient, since it has the same cost as that of the calculation of one single partial derivative of f with the naive formula (2.6).

2.6. **Exercises**

Exercise 2.1 (The existence of partial derivatives at one point without any additional assumption about their continuity is very weak). Let the function $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by

$$\forall x(x_1, x_2) \in \mathbb{R}^2, \quad f(x_1, x_2) = \begin{cases} \frac{x_2^2}{x_1} & \text{if } x_1 \neq 0 \\ 0 & \text{if } \ x_1 = 0. \end{cases}$$

(i) Show that the partial derivatives $\frac{\partial f}{\partial x_1}(0,0)$ and $\frac{\partial f}{\partial x_2}(0,0)$ exist and calculate them.
(ii) Show that f is not continuous at $(0,0)$ (in particular, f is not Fréchet differentiable at $(0,0)$).

Exercise 2.2. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a function of class $\mathcal{C}^1$.

(i) Let $g : \mathbb{R} \to \mathbb{R}$ be defined by $g(t) = f(3-2t, t^2)$; show that g is of class $\mathcal{C}^1$ on $\mathbb{R}$ and calculate its derivative $g'(t)$.
(ii) Let $h : \mathbb{R}^2 \to \mathbb{R}$ be defined by $h(x_1, x_2) = f(x_1^2 + x_2, x_1 x_2)$. Show that h is of class $\mathcal{C}^1$ on $\mathbb{R}^2$ and calculate its gradient $\nabla h(x)$.

Exercise 2.3.

(i) Show that the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by $f(x_1, x_2) = e^{x_1 x_2}(x_1 + x_2)$ is differentiable on $\mathbb{R}^2$ and calculate its gradient.
(ii) Show that the function $f : \mathbb{R}^3 \to \mathbb{R}^2$ defined by $f(x_1, x_2, x_3) = (x_1^2 - \frac{1}{2}x_3^2, \sin x_1 \cos x_2)$ is differentiable on $\mathbb{R}^3$ and calculate its Jacobian matrix.

(iii) Show that the function $f : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $f(x_1, x_2) = (\log(1 + x_2^2), x_1 x_2)$ is differentiable on $\mathbb{R}^2$ and calculate its Jacobian matrix.

Exercise 2.4. Let d and n be two integers ≥ 1, and let $f : \mathbb{R}^d \to \mathbb{R}^n$ and $\varphi : \mathbb{R}^n \to \mathbb{R}$ be differentiable functions. Show that the composite mapping $g := \varphi \circ f : \mathbb{R}^d \to \mathbb{R}$ is differentiable, and that its gradients reads:

$$\forall x \in \mathbb{R}^d, \quad \nabla g(x) = \nabla f(x)^T \nabla \varphi(x).$$

Exercise 2.5. Let $(E, ||\cdot||_E)$, $(F, ||\cdot||_F)$, $(G, ||\cdot||_G)$ and $(H, ||\cdot||_H)$ be normed vector spaces, and let U be an open subset of U. We introduce functions $u : U \to F$ and $v : U \to G$ which are differentiable at some point $x_0 \in U$, as well as a bilinear continuous mapping $b : F \times G \to H$. Show that the function $w : U \to H$ defined by

$$\forall x \in U, \quad w(x) = b(u(x), v(x))$$

is differentiable at x_0, and calculate its derivative.

Exercise 2.6. Let $(E, ||\cdot||_E)$ be a normed vector space and let $d \geq 1$. We consider the mapping $f : \mathbb{R}^d \times E \to E$ defined by:

$$\forall x = (x_1, \dots, x_d) \in \mathbb{R}^d, \; u \in E, \quad f(x, u) = |x|^2 u.$$

Show that f is differentiable on $\mathbb{R}^d \times E$ and calculate its derivative.

Exercise 2.7 (Homogeneous functions and Euler's identity). Let $(E, ||\cdot||_E)$ and $(F, ||\cdot||_F)$ be normed vector spaces, and let $U := E \setminus \{0\}$. One function $f : U \to F$ is called homogeneous of degree $\alpha \in \mathbb{R}$ if the following identity holds:

$$\forall x \in U, \; t > 0, \quad f(tx) = t^\alpha f(x).$$

(i) Show that if $f : U \to \mathbb{R}$ is differentiable and homogeneous of degree α, it holds:

$$\forall x \in U, \quad \mathrm{d}f_x(x) = \alpha f(x).$$

(ii) Conversely, let $f : U \to \mathbb{R}$ be a differentiable function such that $\mathrm{d}f_x(x) = \alpha f(x)$ for all $x \in U$. Show that f is homogeneous with degree α.

[Hint: For an arbitrary given point $x \in U$, define $g : \mathbb{R} \to F$ by $g(s) = f(sx)$ and find out an ordinary differential equation satisfied by g.]

Exercise 2.8 (The differential of the determinant). The purpose of this exercise is to prove that the differential of the mapping

$$\det : M_d(\mathbb{R}) \ni M \mapsto \det(M) \in \mathbb{R}$$

at any point $A \in M_d(\mathbb{R})$ is

$$\mathrm{d}(\det)_A(H) = \mathrm{tr}\Big(\mathrm{com}(A)^T H\Big). \tag{2.7}$$

(i) Show that the subset of $M_d(\mathbb{R})$ composed of invertible matrices is dense in $M_d(\mathbb{R})$.
(ii) Show that the mapping det is of class $\mathcal{C}^1$ on $M_d(\mathbb{R})$.
(iii) Calculate the differential of $M \mapsto \det(M)$ at the point $M = \mathrm{I}_d$.
(iv) Deduce from the previous question that the differential of det at any invertible matrix A is given by (2.7).
(v) Use the results of (i) and (ii) to conclude that the above expression actually holds at all matrices $A \in M_d(\mathbb{R})$.

Exercise 2.9 (About strain and curl). The purpose of this exercise is to introduce and provide a physical intuition of the strain and curl operators for vector fields in the three-dimensional space. Let $u : \mathbb{R}^3 \to \mathbb{R}^3$ be a vector field of class $\mathcal{C}^1$;

- The strain tensor $e(u)$ associated to u is the 3×3 matrix field defined by:

$$\forall x \in \mathbb{R}^3, \quad e(u)(x) = \frac{1}{2}\left(\nabla u(x) + \nabla u(x)^T\right).$$

- The curl of u is the vector field $\mathrm{curl}(u)$ defined by:

$$\forall x \in \mathbb{R}^3, \quad \mathrm{curl}(u)(x) = \left| \begin{array}{c} \frac{\partial}{\partial x_1} \\ \frac{\partial}{\partial x_2} \\ \frac{\partial}{\partial x_3} \end{array} \right. \wedge \left| \begin{array}{c} u_1(x) \\ u_2(x) \\ u_3(x) \end{array} \right. = \begin{pmatrix} \frac{\partial u_3}{\partial x_2}(x) - \frac{\partial u_2}{\partial x_3}(x) \\ \frac{\partial u_1}{\partial x_3}(x) - \frac{\partial u_3}{\partial x_1}(x) \\ \frac{\partial u_2}{\partial x_1}(x) - \frac{\partial u_1}{\partial x_2}(x) \end{pmatrix}.$$

(i) Show that for all point $x \in \mathbb{R}^3$, the following expansion holds:

$$u(x+h) = u(x) + e(u)(x)h + \frac{1}{2}\mathrm{curl}(u)(x) \wedge h + \mathrm{o}(h).$$

[Remark: Roughly speaking, $e(u)$ and $\mathrm{curl}(u)$ can be interpreted as follows: the eigenvalues of the symmetric matrix $e(u)$ quantify the amount of compression or stretching induced by u in the associated principal directions, while $\mathrm{curl}(u)$ is a vector oriented in the direction of the rotation induced by u, with modulus proportional to the rotation angle, see §1.2 in [5] for more details.]

(ii) Show that $\mathrm{div}(\mathrm{curl}(u)) = 0$.

Exercise 2.10. Let $(E, ||\cdot||)_E$ and $(F_1, ||\cdot||_{F_1})$, ..., $(F_n, ||\cdot||_{F_n})$ be normed vector spaces. Let $f : E \to F_1 \times F_n$ be a mapping, whose components are denoted by

$$\forall x \in E, \quad f(x) = (f_1(x), \dots, f_n(x)).$$

Show that f is Fréchet differentiable at a point x_0 in E if and only if each component function f_i is differentiable at x_0, and then

$$\forall h \in E, \quad \mathrm{d}f_x(h) = (\mathrm{d}f_{1,x}(h), \dots, \mathrm{d}f_{n,x}(h)).$$

3. A reminder of integral calculus

In this section, we sketch the main results attached to the Lebesgue integration theory. After a brief overview of the construction of the Lebesgue integral in Section 3.1, we recall the dominated convergence theorem, the Fubini theorem, and the change of variables theorem. Eventually, we detail the construction of a Riemann-like integral for functions taking values in a Banach space.

3.1. A few reminders about the Lebesgue measure and the Lebesgue integral

Let us start this section with a few concepts from abstract measure theory.

Definition 3.1. *Let X be a non empty set. A σ-algebra in X is a collection $\mathcal{E}$ of subsets of X such that*

- *$\mathcal{E}$ contains the empty set $\emptyset$ and X itself;*
- *For all sets E_1, $E_2 \in \mathcal{E}$, the reunion $E_1 \cup E_2$, the intersection $E_1 \cap E_2$ and the complement $X \setminus E_1$ belong to $\mathcal{E}$.*
- *For any countable collection $E_n \in \mathcal{E}$, $n = 0, \dots$, the reunion $\cup_{n=0}^{\infty} E_n$ also belongs to $\mathcal{E}$.*

If $\mathcal{E}$ is a σ-algebra in X, the pair $(X, \mathcal{E})$ is called a measure space.

Proposition-Definition 3.1. *Let X be a non empty set, and let $\mathcal{X}$ be any collection of subsets of X; there exists a σ-algebra in X which is the smallest σ-algebra containing $\mathcal{X}$ (in the sense that it is included in any σ-algebra containing $\mathcal{X}$); it is called the σ-algebra generated by $\mathcal{X}$.*

Sketch of proof. One verifies that for any (possibly uncountable) collection $\{\mathcal{E}_i\}_{i \in I}$ of σ-algebras in X, the intersection $\cap_{i \in I} \mathcal{E}_i$ is still a σ-algebra in X. Hence, the intersection of all the σ-algebras containing $\mathcal{X}$ is clearly that with the desired properties. □

Definition 3.2. *Let $(X, \mathcal{E})$ be a measure space. A positive measure defined on $\mathcal{E}$ is a function $\mu : \mathcal{E} \to [0, \infty]$ (possibly taking infinite values) such that $\mu(\emptyset) = 0$ and μ is σ-additive, i.e. for all sequences E_n of pairwise disjoint elements of $\mathcal{E}$, one has:*

$$\mu\left(\bigcup_{n=0}^{\infty} E_n\right) = \sum_{n=0}^{\infty} \mu(E_n).$$

The rest of this section unfolds in the Euclidean space $\mathbb{R}^d$, which we equip with a particular σ-algebra.

Definition 3.3. *The σ-algebra generated by the collection of open subsets of $\mathbb{R}^d$ is called the Borel σ-algebra of $\mathbb{R}^d$, and is denoted by $\mathcal{B}(\mathbb{R}^d)$. The sets $E \in \mathcal{B}(\mathbb{R}^d)$ are called Borel subsets of $\mathbb{R}^d$.*

Let us now proceed with the definition of the Lebesgue measure on (Borel subsets of) $\mathbb{R}^d$. One classical construction hinges on the following rationale: we have a good intuition of what should be the measure of a product of intervals, of the form $A = \prod_{i=1}^n [a_i, b_i] \subset \mathbb{R}^d$; a difficult result allows then to see that there is only one positive measure on the Borel subsets of $\mathbb{R}^d$ which has this property. Precisely, we shall admit the following theorem.

Theorem 3.1. *There exists a unique positive measure μ on $\mathcal{B}(\mathbb{R}^d)$ such that*

$$\text{For all hypercube } A = \prod_{i=1}^d [a_i, b_i] \subset \mathbb{R}^d, \quad \mu(A) = \prod_{i=1}^d (b_i - a_i).$$

This measure is called the Lebesgue measure on $\mathbb{R}^d$. In the following, we denote by $|E|$ the Lebesgue measure of a Borel subset $E \in \mathcal{B}(\mathbb{R}^d)$.

We now come to the definition of negligible subsets and of Lebesgue measurable subsets of $\mathbb{R}^d$.

Definition 3.4.

- *A subset $N \subset \mathbb{R}^d$ is called negligible if there exists a Borel subset $E \subset \mathbb{R}^d$ with $|E| = 0$ such that $N \subset E$.*
- *A subset $A \subset \mathbb{R}^d$ is called Lebesgue measurable if there exist a Borel subset $E \in \mathcal{B}(\mathbb{R}^d)$ and a negligible set N such that $A = E \cup N$. The measure of such a set A is then defined to be $|E|$, and one easily verifies that this value does not depend on the choice of E and N such that $A = E \cup N$.*
- *One property is said to hold almost everywhere if there exists a negligible set $N \subset \mathbb{R}^d$ such that it holds for all points $x \in \mathbb{R}^d \setminus N$.*
- *One function $f : \mathbb{R}^d \to \mathbb{R}$ is called Lebesgue measurable if for every open subset $U \subset \mathbb{R}$, the set $f^{-1}(U) \subset \mathbb{R}^d$ is Lebesgue measurable.*

We next turn to the definition of the integral with respect to the Lebesgue measure.

Definition 3.5.

- *A simple function $f : \mathbb{R}^d \to \mathbb{R}$ is a function of the form*

$$f = \sum_{n=0}^N \alpha_n \mathbb{1}_{A_n},$$

where the A_n are Lebesgue measurable subsets of $\mathbb{R}^d$, $n = 0, \dots, N$.

- *Let $f : \mathbb{R}^d \to [0, \infty)$ be a non negative, simple function. The integral of f is defined by:*

$$\int_{\mathbb{R}^d} f(x)\, \mathrm{d}x := \sum_{n=0}^N \alpha_n |A_n|.$$

- *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a non negative function. The integral of f is the (possibly infinite) real value defined by*

$$\int_{\mathbb{R}^d} f(x)\, \mathrm{d}x := \sup \left\{ \int_{\mathbb{R}^d} v(x)\, \mathrm{d}x, \ v \text{ is a non negative simple function, } v \leq u \right\}.$$

- *One measurable function $f : \mathbb{R}^d \to \mathbb{R}$ is said to be integrable if*

$$\int_{\mathbb{R}^d} |f(x)|\, \mathrm{d}x < \infty.$$

In this case, the integral of f is defined by:

$$\int_{\mathbb{R}^d} f(x)\, \mathrm{d}x = \int_{\mathbb{R}^d} f^+(x)\, \mathrm{d}x - \int_{\mathbb{R}^d} (-f^-(x))\, \mathrm{d}x,$$

where $f^+ = \max(0, f)$ is the positive part of f and $f^- = \min(0, f)$ is its negative part.

- *Let $A \subset \mathbb{R}^d$ be a measurable subset. The integral of a non negative, measurable function $f : A \to [0, \infty)$ is defined by*
$$\int_A f(x)\,\mathrm{d}x = \int_{\mathbb{R}^d} \mathbb{1}_A(x) f(x)\,\mathrm{d}x.$$
Likewise, a measurable function $f : A \to [0, \infty)$ is called integrable if $\mathbb{1}_A f$ is integrable, and we set
$$\int_A f(x)\,\mathrm{d}x = \int_{\mathbb{R}^d} \mathbb{1}_A(x) f(x)\,\mathrm{d}x.$$
The vector space of integrable functions on A is denoted by $L^1(A)$.

Remark 3.1. *The above concepts are easily extended to the case of $\mathbb{R}^n$-valued functions $f : A \to \mathbb{R}^n$, $n \geq 2$. Such a function $f = (f_i)_{i=1,\dots,n}$ is called integrable if each component f_i is, and the integral $\int_A f\,\mathrm{d}x$ is defined as the vector in $\mathbb{R}^n$ with components*
$$\left(\int_A f(x)\,\mathrm{d}x\right)_i = \int_A f_i(x)\,\mathrm{d}x.$$

Remark 3.2. *Abstract measure and integration theory play a crucial role in the field of probability. In a nutshell, a probability measure on $\mathbb{R}^d$ is a positive measure $\mathbb{P}$ on $\mathbb{R}^d$ with total mass $\mathbb{P}(\mathbb{R}^d) = 1$. A subset $A \in \mathbb{R}^d$ is called an event (as a collection of outcomes $x \in \mathbb{R}^d$), and its probability equals $\mathbb{P}(A)$. When $f : \mathbb{R}^d \to \mathbb{R}$ is a quantity of interest, the integral $\int_{\mathbb{R}^d} f\,\mathrm{d}\mathbb{P}$ is the expectation of f.*

3.2. The Lebesgue dominated convergence theorem and some of its avatars

The perhaps most fundamental result in integration theory is the celebrated Lebesgue dominated convergence theorem, which conveniently allows to intertwine limits and integrals.

Theorem 3.2 (Dominated convergence theorem)**.** *Let $f_n : \mathbb{R}^d \to \mathbb{R}$ be a sequence of measurable functions which converges almost everywhere to a function $f : \mathbb{R}^d \to \mathbb{R}$. We assume that there exists a non negative, integrable function $h : \mathbb{R}^d \to \mathbb{R}$ such that*
$$\forall n \geq 0, \text{ for a.e. } x \in \mathbb{R}^d, \quad |f_n(x)| \leq h(x),$$
so that in particular, all the functions f_n are integrable. Then, f is also integrable, and one has:
$$\int_{\mathbb{R}^d} |f_n(x) - f(x)|\,\mathrm{d}x \xrightarrow{n\to\infty} 0, \text{ and } \int_{\mathbb{R}^d} f_n(x)\,\mathrm{d}x \xrightarrow{n\to\infty} \int_{\mathbb{R}^d} f(x)\,\mathrm{d}x.$$

Remark 3.3.

- *Let $A \subset \mathbb{R}^d$ be an arbitrary Lebesgue measurable subset; by applying the dominated convergence theorem with $\chi_A f_n$ (resp. $\chi_A f$) instead of f_n (resp. f), the same conclusion as in Theorem 3.2 obviously holds with A as integration domain.*
- *A similar statement to that of Theorem 3.2 holds when the considered sequence of functions is indexed with a real parameter $\varepsilon > 0$ tending to 0 instead of an integer $n \to \infty$.*

The next result allows to take derivatives under the integral sign. It is a fairly simple consequence of the Lebesgue dominated convergence theorem, and we present its proof as an instructive exercise We refer to Exercise 3.5 about an interesting situation where the conclusion of the theorem fails when one of its subtle assumptions is violated.

Theorem 3.3 (Differentiation under the integral sign)**.** *Let $A \subset \mathbb{R}^d$ be a measurable subset, and let $I \subset \mathbb{R}$ be an open interval. Let $f : I \times A$ be a function; we assume that there exists a negligible subset $N \subset A$ such that the following hypotheses are satisfied:*

(i) For every $t \in I$, the partial mapping $x \mapsto f(t, x)$ is integrable on A;
(ii) The partial derivative $\frac{\partial f}{\partial t}(t, x)$ exists at every point $(t, x) \in I \times (A \setminus N)$;
(iii) There exists a positive, integrable function $h : A \to \mathbb{R}$ such that
$$\forall t \in I, \ \forall x \in (A \setminus N), \quad \left|\frac{\partial f}{\partial t}(t, x)\right| \leq h(x).$$

Then the function $F : I \to \mathbb{R}$ defined by

$$F(t) := \int_A f(t,x)\,\mathrm{d}x$$

is differentiable on I and its derivative reads:

$$F'(t) = \int_A \frac{\partial f}{\partial t}(t,x)\,\mathrm{d}x.$$

Proof. Let $t \in I$ be fixed, and let s_n be an arbitrary sequence of real numbers such that $s_n \to 0$ as $n \to \infty$. Then, a simple calculation yields:

$$\frac{F(t+s_n) - F(t)}{s_n} = \int_A g_n(x)\,\mathrm{d}x, \text{ where } g_n(x) := \frac{f(t+s_n,x) - f(t,x)}{s_n}.$$

We aim to apply the Lebesgue dominated convergence Theorem 3.2 to calculate the limit of the above expression. To achieve this, we first note that, by assumption (ii), for almost every point $x \in A \setminus N$, it holds:

$$g_n(x) \xrightarrow{n\to\infty} \frac{\partial f}{\partial t}(t,x);$$

in particular, the sequence of functions g_n converges a.e. to the function $\frac{\partial f}{\partial t}(t,\cdot)$. Let us now verify the second assumption of Theorem 3.2. For all $n \geq 0$ and a.e. $x \in A \setminus N$, the Mean Value Theorem 1.1 yields the existence of a number $\theta_{n,x} \in (0,1)$ such that

$$g_n(x) = \frac{f(t+s_n,x) - f(t,x)}{s_n} = \frac{\partial f}{\partial t}(t + \theta_{n,x}s_n, x), \text{ and so } |g_n(x)| \leq h(x),$$

where we have used assumption (iii).

The assumptions of Theorem 3.2 are then fulfilled, and its application allows to conclude. $\square$

3.3. The Fubini theorem

In this section, we briefly discuss multiple integrals. Let $p, q \geq 1$ and let us consider a function $f : (x,y) \in \mathbb{R}^p \times \mathbb{R}^q \mapsto f(x,y) \in \mathbb{R}$. The three Euclidean spaces $\mathbb{R}^p$, $\mathbb{R}^q$ and $\mathbb{R}^p \times \mathbb{R}^q \approx \mathbb{R}^{p+q}$ are equipped with the Lebesgue measure; the integration of f over the product space $\mathbb{R}^p \times \mathbb{R}^q$ can a priori be realized in three different ways:

- One integrates first in the y variable for fixed $x \in \mathbb{R}^p$, thus defining the function $\mathbb{R}^p \ni x \mapsto \int_{\mathbb{R}^q} f(x,y)\,\mathrm{d}y \in \mathbb{R}$, and then integrates the latter in the x variable; this leads to the quantity
$$\int_{\mathbb{R}^p} \left(\int_{\mathbb{R}^q} f(x,y)\,\mathrm{d}y \right) \mathrm{d}x.$$
- Symmetrically, one integrates first in the x variable for fixed $y \in \mathbb{R}^q$, and then integrates the resulting function $\mathbb{R}^q \ni y \mapsto \int_{\mathbb{R}^q} f(x,y)\,\mathrm{d}x \in \mathbb{R}$ in the y variable; this leads to the quantity
$$\int_{\mathbb{R}^q} \left(\int_{\mathbb{R}^p} f(x,y)\,\mathrm{d}x \right) \mathrm{d}y.$$
- One integrates directly f with respect to the couple (x,y) over the product space $\mathbb{R}^p \times \mathbb{R}^q$, which results in the quantity
$$\int_{\mathbb{R}^p \times \mathbb{R}^q} f(x,y)\,\mathrm{d}x\mathrm{d}y.$$

The Fubini theorem states that, under mild assumptions, the three procedures lead to the same result.

Theorem 3.4 (Fubini theorem). *Let $f(x,y)$ be a measurable function defined on the product space $\mathbb{R}^p \times \mathbb{R}^q$.*

(i) If f is non negative, then for almost all $x \in \mathbb{R}^p$, the partial mapping $\mathbb{R}^q \ni y \mapsto f(x,y)$ is measurable. The induced mapping $\mathbb{R}^p \ni x \mapsto \int_{\mathbb{R}^q} f(x,y)\,\mathrm{d}y \in \mathbb{R} \cup \{\infty\}$ is well-defined for a.e. $x \in \mathbb{R}^p$ and it induces a non negative mapping which is also measurable (possibly taking infinite values). Likewise, the mapping $\mathbb{R}^q \ni y \mapsto \int_{\mathbb{R}^p} f(x,y)\,\mathrm{d}x \in \mathbb{R} \cup \{\infty\}$ is well-defined and measurable. Then,

$$\int_{\mathbb{R}^p} \left(\int_{\mathbb{R}^q} f(x,y)\,\mathrm{d}y \right) \mathrm{d}x = \int_{\mathbb{R}^q} \left(\int_{\mathbb{R}^p} f(x,y)\,\mathrm{d}x \right) \mathrm{d}y = \int_{\mathbb{R}^p \times \mathbb{R}^q} f(x,y)\,\mathrm{d}x\mathrm{d}y,$$

in the sense that one of the above quantities is finite if and only if the other two are, and their values are then equal.

(ii) If $f \in L^1(\mathbb{R}^p \times \mathbb{R}^q)$, then the partial mappings $y \mapsto f(x,y)$ and $x \mapsto f(x,y)$ are in $L^1(\mathbb{R}^q)$ and $L^1(\mathbb{R}^p)$, respectively. The mappings $x \mapsto \int_{\mathbb{R}^q} f(x,y)\,\mathrm{d}y$ and $y \mapsto \int_{\mathbb{R}^p} f(x,y)\,\mathrm{d}x$ are in $L^1(\mathbb{R}^p)$ and $L^1(\mathbb{R}^q)$, respectively and:

$$\int_{\mathbb{R}^p} \left(\int_{\mathbb{R}^q} f(x,y)\,\mathrm{d}y \right) \mathrm{d}x = \int_{\mathbb{R}^q} \left(\int_{\mathbb{R}^p} f(x,y)\,\mathrm{d}x \right) \mathrm{d}y = \int_{\mathbb{R}^p \times \mathbb{R}^q} f(x,y)\,\mathrm{d}x\mathrm{d}y,$$

Remark 3.4. *The first part of this statement is referred to as Tonelli's theorem. It is often the most flexible form of the Fubini theorem in applications.*

3.4. The change of variable formula

We eventually recall one last fundamental result in integration theory, namely the change of variables theorem, as a key tool for simplifying either the integrand or the domain of an integral.

Theorem 3.5 (Change of variables in integrals)**.** *Let U be a Borel subset of $\mathbb{R}^d$, and let $\varphi : U \to \varphi(U)$ be a diffeomorphism of class $\mathcal{C}^1$. Then, a function $f : \varphi(U) \to \mathbb{R}$ belongs to $L^1(\varphi(U))$ if and only if $f \circ \varphi$ belongs to $L^1(U)$, and it holds:*

$$\int_{\varphi(U)} f\,\mathrm{d}x = \int_U (f \circ \varphi) |\det \nabla\varphi|\,\mathrm{d}x. \tag{3.1}$$

This formula has an intuitive interpretation, which is illustrated on Fig. 9. Let us recall that any invertible $d \times d$ matrix M admits a polar decomposition:

$$M = OP,$$

where

- O is a unitary matrix, i.e. $O^T O = \mathrm{I}_d$; a standard result from linear algebra describes the structure of such matrices: O is a composition of rotations and reflections in an appropriate coordinate system.
- P is a symmetric positive definite matrix. Writing P under diagonal form,

$$P = \tilde{O} \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_d \end{pmatrix} \tilde{O}^T, \text{ with } \tilde{O}\tilde{O}^T = \mathrm{I}_d \text{ and } \lambda_i \geq 0,$$

  we see that P expresses compressions or stretchings in the orthonormal directions of the column vectors of $\tilde{O}$, with amplitudes λ_i.

It follows that $|\det M| = \det P = \lambda_1 \dots \lambda_d$, i.e. the change of volume entailed by the linear mapping M is that induced by the compression-stretching expressed by P.

Returning to the statement of Theorem 3.5, let $\nabla\varphi(x) = O(x)P(x)$ be the polar decomposition of the invertible matrix $\nabla\varphi(x)$. Then $P(x)$ is the matrix which accounts for the local stretching induced by $\nabla\varphi(x)$, and $|\det \nabla\varphi(x)| = \det P(x)$.

3.5. Integration of vector-valued functions

We have hitherto been dealing with the integration of real-valued functions on the Borel subsets of $\mathbb{R}^d$. This theory can be adapted in a straightforward way to handle integrals of $\mathbb{R}^n$-valued functions defined on Borel subsets in $\mathbb{R}^d$, working component-wise, see Remark 3.1.

However, for a variety of purposes, it is of great interest to be able to integrate so-called *vector-valued functions*, taking values in a Banach space $(E, ||\cdot||_E)$. In this section, we provide one possible construction of such integrals, which is described in details in Chap. 2 in [9]; it mimicks the Riemann construction of the integral of a real-valued function, and is thereby not the most general construction, but it is enough to deal with continuous functions from an interval $[a,b] \subset \mathbb{R}$ into E. The construction can be omitted on first reading; all the reader needs to keep in mind is that "everything works as for the integral of real-valued functions", as reflected by Proposition-Definition 3.2.

Let us start by introducing a few notations and definitions.

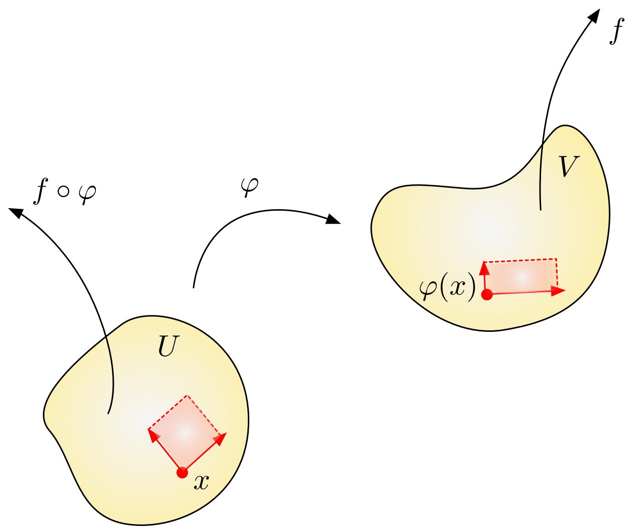

FIGURE 9. *The change of variables Theorem 3.5 in integral calculus features a diffeomorphism $\varphi : U \to V$ of class $\mathcal{C}^1$. The Jacobian term $|\det(\nabla\varphi)|$ featured in the formula (3.1) reflects the local distortion of the volume element in $\mathbb{R}^d$ (in red) caused by φ.*

- Let $[a, b] \subset \mathbb{R}$ be a compact interval of $\mathbb{R}$: $-\infty < a \leq b < \infty$. A partition $\mathcal{P}$ of $[a, b]$ is a collection of points $\{x_0, x_1, \dots, x_I\}$ in $[a, b]$ such that
$$a = x_0 < x_1 < \ldots < x_I = b.$$
- For any such partition, we define $\Delta x_i := |x_i - x_{i-1}|$, $i = 1, \dots, I$; the size of the partition is then $m(\mathcal{P}) := \sup\limits_{i=1,\dots,I} \Delta x_i$.
- Let $\mathcal{P}$ be a partition of $[a, b]$ and let $\xi = \{\xi_1, \dots, \xi_I\}$ be a collection of points such that
$$\xi_i \in [x_{i-1}, x_i] \text{ for } i = 1, \dots, I. \tag{3.2}$$
The Riemann sum $R_{\mathcal{P},\xi}(f)$ associated to a continuous function $f : [a, b] \to E$ is the element of E defined by:
$$R_{\mathcal{P},\xi}(f) = \sum_{i=1}^{I} \Delta x_i f(\xi_i).$$

Definition 3.6. *The Riemann sums $R_{\mathcal{P},\xi}(f)$ of f admit a limit $\ell \in E$ as the size of partitions $m(\mathcal{P})$ tends to 0 if, for all $\varepsilon > 0$, there exists a number $\delta > 0$ such that for all partition $\mathcal{P} = \{x_0, \dots, x_I\}$ with size $m(\mathcal{P})$ and for any choice of points $\xi_i \in [x_{i-1}, x_i]$, the following estimate holds:*
$$||R_{\mathcal{P},\xi}(f) - \ell||_E \leq \varepsilon.$$

The next proposition defines the (Riemann) integral of a continuous function $f : [a, b] \to E$ as the limit of its Riemann sums, and gathers of the main properties of this notion.

Proposition-Definition 3.2. *Let $(E, || \cdot ||_E)$ be a Banach space and let $f : [a, b] \to E$ be a continuous function. Then the Riemann sums $R_{\mathcal{P},\xi}(f)$ have a limit in E as the size $m(\mathcal{P})$ of the partition tends to 0, that we denote by $\int_a^b f(t)\,\mathrm{d}t \in E$. The so-defined Riemann integral of f over $[a, b]$ satisfies the following properties:*

(i) $\left|\left|\int_a^b f(t)\,\mathrm{d}t\right|\right|_E \leq \int_a^b ||f(t)||_E \mathrm{d}t.$

(ii) For all $c \in [a, b]$, it holds:
$$\int_a^b f(t)\,\mathrm{d}t = \int_a^c f(t)\,\mathrm{d}t + \int_c^b f(t)\,\mathrm{d}t.$$

(iii) The function $g : t \mapsto \int_a^t f(s)\,\mathrm{d}s$ is a primitive of f.
(iv) For all continuous linear mapping $\ell \in E^*$, it holds:

$$\left\langle \ell, \int_a^b f(t)\,\mathrm{d}t \right\rangle = \int_a^b \langle \ell, f(t)\rangle\,\mathrm{d}t. \tag{3.3}$$

Proof. The results of the proof rely on an important preliminary calculation. Since $f : [a,b] \to E$ is continuous, by the Heine Theorem 1.4, it is uniformly continuous on this interval. Hence, for each $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\forall s,t \in [a,b],\ \ |t-s| \leq \delta \ \Rightarrow\ ||f(t)-f(s)||_E \leq \varepsilon. \tag{3.4}$$

Let $\mathcal{P} = \{x_0,\dots,x_N\}$ and $\mathcal{Q} = \{y_0,\dots,y_M\}$ be two partitions of f with size $\max(m(\mathcal{P}), m(\mathcal{Q})) \leq \delta$. We introduce the common subdivision $\mathcal{T} = \{z_k\}_{k=1,\dots,K}$ of $\mathcal{P}$ and $\mathcal{Q}$, whose points z_k are the reunion of the x_i and the y_j, arranged in increasing order. For any index $i = 1,\dots,I$, we denote by $K_i \subset \{1,\dots,K\}$ the set of indices $k = 1,\dots,K$ such that x_i is the closest to z_k among the points of $\mathcal{P}$. The sets K_i, $i = 1,\dots,I$, form a partition of $\{1,\dots,K\}$ and for each $k = 1,\dots,K$, we denote by i_k the unique index in $\{1,\dots,I\}$ such that $k \in K_{i_k}$.

Let ξ_i and τ_k be arbitrary points satisfying $\xi_i \in [x_{i-1}, x_i]$, $\tau_k \in [z_k, z_{k+1}]$ for $i = 1,\dots,I$ and $k = 1,\dots,K$. We estimate the difference

$$\begin{aligned}
||R_{\mathcal{P},\xi}(f) - R_{\mathcal{T},\tau}(f)||_E &= \left|\left| \sum_{i=1}^{I} \Delta x_i f(\xi_i) - \sum_{k=1}^{K} \Delta z_k f(\tau_k) \right|\right|_E \\
&= \left|\left| \sum_{i=1}^{I} \sum_{k\in K_i} \Delta z_k f(\xi_i) - \sum_{k=1}^{K} \Delta z_k f(\tau_k) \right|\right|_E \\
&= \left|\left| \sum_{k=1}^{K} \Delta z_k f(\xi_{i_k}) - \sum_{k=1}^{K} \Delta z_k f(\tau_k) \right|\right|_E \\
&\leq \sum_{k=1}^{K} \Delta z_k \, ||f(\xi_{i_k}) - f(\tau_k)||_E \, .
\end{aligned}$$

Combining the facts that for each $k = 1,\dots,K$, we have $|\xi_{i_k} - \tau_k| \leq \delta$, and that $\sum_{k=1}^{K} \Delta z_k = (b-a)$, we obtain

$$||R_{\mathcal{P},\xi}(f) - R_{\mathcal{T},\tau}(f)||_E \leq (b-a)\varepsilon.$$

By the same token, we also have, for any points $\zeta_j \in [y_{j-1}, y_j]$:

$$||R_{\mathcal{Q},\zeta}(f) - R_{\mathcal{T},\tau}(f)||_E \leq (b-a)\varepsilon.$$

As a result, the triangle inequality yields:

$$||R_{\mathcal{P},\xi}(f) - R_{\mathcal{Q},\zeta}(f)||_E \leq ||R_{\mathcal{P},\xi}(f) - R_{\mathcal{T},\tau}(f)||_E + ||R_{\mathcal{T},\tau}(f) - R_{\mathcal{Q},\zeta}(f)||_E \leq 2(b-a)\varepsilon. \tag{3.5}$$

We are now in position to show that the Riemann sums $R_{\mathcal{P},\xi}(f)$ of f have a limit in the sense of Definition 3.6, and to this end, we rely on the Cauchy criterion. Let $\mathcal{P}^n$ be any sequence of partitions of $[a,b]$ with size $m(\mathcal{P}^n)$ tending to 0 as $n \to \infty$, and let ξ^n be any sequence of associated points via (3.2). Let $\varepsilon > 0$ be given, and let $\delta > 0$ be such that (3.4) holds. Then there exists $n_0 \in \mathbb{N}$ such that $m(\mathcal{P}^n) \leq \delta$ for $n \geq n_0$, and so, by the above preliminary calculation:

$$\forall m,n \geq n_0, \quad ||R_{\mathcal{P}^n,\xi^n}(f) - R_{\mathcal{P}^m,\xi^m}(f)||_E \leq 2(b-a)\varepsilon.$$

Thus, the sequence $R_{\mathcal{P}^n,\xi^n}(f)$ is a Cauchy sequence of elements of E, and so it converges to some limit $\ell \in E$. It remains to prove that this limit does not depend on the choice of the particular sequence of partitions $\mathcal{P}^n$ and points ξ^n. Let $\mathcal{P}^n$, ξ^n and $\mathcal{Q}^n$, ζ^n be such that

$$R_{\mathcal{P}^n,\xi^n}(f) \xrightarrow{n\to\infty} \ell_1, \text{ and } R_{\mathcal{Q}^n,\zeta^n}(f) \xrightarrow{n\to\infty} \ell_2.$$

For a given $\varepsilon > 0$, let $\delta > 0$ be such that (3.4) holds. For n large enough, we have $m(\mathcal{P}^n) \leq \delta$ and $m(\mathcal{Q}^n) \leq \delta$, and so, using again our preliminary calculation:

$$||R_{\mathcal{P}^n,\xi^n}(f) - R_{\mathcal{Q}^n,\zeta^n}(f)||_E \leq 2(b-a)\varepsilon.$$

Taking limits as $n \to \infty$, we have $||\ell_1 - \ell_2||_E \le 2(b-a)\varepsilon$, and since ε is arbitrary, it follows that $\ell_1 = \ell_2$.

Let us now prove (i): For any partition $\mathcal{P} = \{x_0, \dots, x_I\}$ of $[a,b]$, and any associated points $\xi_i \in [x_{i-1}, x_i]$, we have by the triangle inequality

$$\left|\left|\sum_{i=1}^{I} \Delta x_i f(\xi_i)\right|\right|_E \le \sum_{i=1}^{I} \Delta x_i \,||f(\xi_i)||_E\,,$$

where the right-hand side is nothing but a Riemann sum associated to the continuous function $\mathbb{R} \ni t \mapsto ||f(t)||_E \in \mathbb{R}$. Passing to the limit as the size of the partition tends to 0 yields the desired result.

We next prove (ii): Let $n \ge 1$ be given; we introduce the regular partitions $\mathcal{P}_1^n = \{x_0^n, \dots x_n^n\}$ and $\mathcal{P}_2^n = \{y_0^n, \dots y_n^n\}$ of the intervals $[a,c]$ and $[c,b]$:

$$\forall i = 0, \dots, n,\;\; x_i^n = a + i\frac{c-a}{n},\;\text{ and } \forall j = 0, \dots, n,\;\; y_j^n = c + \frac{b-c}{n}.$$

We also introduce the points ξ_i^n and ζ_j^n defined by

$$\forall i = 1, \dots, n,\;\; \xi_i^n = a + \frac{2i-1}{n}(c-a),\;\text{ and } \forall j = 1, \dots, n,\;\; \zeta_j^n = c + \frac{2j-1}{n}(b-c).$$

Then, $\mathcal{P}^n := \{x_0^n, \dots x_n^n, y_1^n, \dots, y_n^n\}$ defines a partition of $[a,b]$ where c explicitly appears. For $k = 1, \dots, 2n$, let us denote $\tau_k^n = x_k^n$ if $k = 1, \dots, n$ and $\tau_k^n = y_{k-n}^n$ if $k = n+1, \dots, 2n$; it follows from the previous definitions that:

$$R_{\mathcal{P}^n, \tau^n}(f) = R_{\mathcal{P}_1^n, \xi^n}(f) + R_{\mathcal{P}_2^n, \zeta^n}(f). \tag{3.6}$$

On another hand, the preliminary calculation reveals that

$$R_{\mathcal{P}_1^n, \xi^n}(f) \xrightarrow{n\to\infty} \int_a^c f(t)\,\mathrm{d}t, \quad R_{\mathcal{P}_2^n, \zeta^n}(f) \xrightarrow{n\to\infty} \int_c^b f(t)\,\mathrm{d}t, \text{ and } R_{\mathcal{P}^n, \tau^n}(f) \xrightarrow{n\to\infty} \int_a^b f(t)\,\mathrm{d}t.$$

The desired identity follows from taking limits in (3.6).

Let us now prove (iii): Since f is uniformly continuous on $[a,b]$, again, for all $\varepsilon > 0$, there exists $\delta > 0$ such that (3.4) holds. Let us then write

$$\frac{g(t+h) - g(t)}{h} = \frac{1}{h}\int_t^{t+h} f(s)\,\mathrm{d}s.$$

Now, if $|h| \le \delta$, then $\mathcal{P} := \{t, t+h\}$ is a partition of $[t, t+h]$ with size $m(\mathcal{P}) \le \delta$, and so, as we have seen in (3.5):

$$\left|\left|\int_t^{t+h} f(s)\,\mathrm{d}s - hf(t)\right|\right|_E \le 2h\varepsilon.$$

Hence,

$$\left|\left|\frac{g(t+h) - g(t)}{h} - f(t)\right|\right|_E \le 2\varepsilon,$$

which is the desired result.

(iv): The proof of this point is very similar to that of the above three points: at first, the desired relation (3.3) is proved when $\int_a^b f(t)\,\mathrm{d}t$ is replaced by any Riemann sum associated to f; then, taking limits as the size of partitions tends to 0 allows to conclude. Details are left to the reader. ☐

Remark 3.5. *Other, more "powerful" notions of integration for functions taking values in Banach spaces can be devised; let us mention, in particular, the theory of Bochner integral, which follows the philosophy of Lebesgue integration and inherits many of its assets with respect to the present construction inspired from Riemann integration, see Appendix E in* [7] *for more details on this subject.*

3.6. **Exercises**

Exercise 3.1.

(i) Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a function of class $\mathcal{C}^k$ for some $k \geq 1$, such that $\varphi(0) = 0$. Show that there exists a function ψ of class $\mathcal{C}^{k-1}$ such that

$$\forall x \in \mathbb{R}, \quad \varphi(x) = x\psi(x).$$

[Hint: Write $\varphi(x) = \varphi(x) - \varphi(0) = \int_0^x \varphi'(t)\,\mathrm{d}t$ *and use a change of variables in the last integral.]*

(ii) More generally, let $\varphi : \mathbb{R}^d \to \mathbb{R}$ be a function of class $\mathcal{C}^k$ for some $k \geq 1$, such that $\varphi(0) = 0$. Show that there exist functions $\psi_i : \mathbb{R}^d \to \mathbb{R}$ of class $\mathcal{C}^{k-1}$ such that

$$\forall x \in \mathbb{R}^d, \quad \varphi(x) = \sum_{i=1}^{d} x_i \psi_i(x).$$

Exercise 3.2 (Polar change of variables and calculation of the integral $I := \int_{\mathbb{R}} e^{-x^2}\,\mathrm{d}x$). (i) Show that the mapping induced by the change from polar to Cartesian coordinates

$$T(r, \theta) = (r\cos\theta, r\sin\theta),$$

from $(0,\infty) \times (-\pi, \pi)$ into $\mathbb{R}^2 \setminus \{(-c, 0),\ c \in [0,\infty)\}$ is a diffeomorphism of class $\mathcal{C}^1$.

(ii) Calculate I^2 by using the Fubini theorem and a change of variables.

(iii) Conclude about the value of I.

Exercise 3.3 (Spherical change of coordinates and calculation of volumes). This exercise unfolds in the setting of the three-dimensional space $\mathbb{R}^3$.

(i) Show that the mapping induced by the change from spherical to Cartesian coordinates

$$T(r, \theta, \varphi) = (r\sin\theta\cos\varphi, r\sin\theta\sin\varphi, r\cos\theta),$$

from $[0,\infty) \times [0,\pi] \times [0, 2\pi)$ is a diffeomorphism of class $\mathcal{C}^1$, and calculate its Jacobian matrix ∇T and the determinant $\det \nabla T$ (ee Fig. 10 for an illustration).

(ii) Infer the following change of variables formula: for any function $f \in L^1(\mathbb{R}^3)$, the function $f \circ T$ also belongs to $L^1(\mathbb{R}^3)$ and

$$\int_{\mathbb{R}^3} f(x)\,\mathrm{d}x = \int_0^\infty \int_0^\pi \int_0^{2\pi} f \circ T\, r^2 \sin\theta\,\mathrm{d}r\mathrm{d}\theta\mathrm{d}\varphi.$$

(iii) Show that the volume of the unit ball $\mathcal{B} := \left\{x = (x_1, x_2, x_3) \in \mathbb{R}^3,\ x_1^2 + x_2^2 + x_3^2 = 1\right\}$

$$\int_{\mathcal{B}} \mathrm{d}x = \frac{4}{3}\pi.$$

(iv) Let $\mathcal{E} := \left\{x = (x_1, x_2, x_3) \in \mathbb{R}^3,\ \frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} + \frac{x_3^2}{c^2} = 1 \in \mathbb{R}^d,\right\}$ be the ellipsoid with semi-axes equal to a, b and c. Show that the volume of $\mathcal{E}$ equals $\frac{4}{3}\pi abc$.

Exercise 3.4 (The Markov identity). Let $f : \mathbb{R}^d \to \mathbb{R}$ be an integrable, non negative function. Show the Markov inequality:

$$\left|\left\{x \in \mathbb{R}^d,\ f(x) \geq \alpha\right\}\right| \leq \frac{1}{\alpha}\int_{\mathbb{R}^d} f(x)\,\mathrm{d}x.$$

Exercise 3.5. The goal of this exercise is to show that the conclusion of Theorem 3.3 fails when one of its assumptions is not satisfied. Let $\varphi : [0,1] \to \mathbb{R}$ be a continuous function, and let $f : [0,1] \to [0,1] \to \mathbb{R}$ be the function defined by:

$$\forall t, x \in [0,1], \quad f(t,x) = \begin{cases} \varphi(x) & \text{if } x \leq t, \\ 0 & \text{if } t > x. \end{cases}$$

In particular, there does not exist a negligible subset $N \subset [0,1]$ such that the mapping $t \mapsto f(t,x)$ is differentiable on $[0,1]$ for all $x \in [0,1] \setminus N$. For all $t \in [0,1]$, the partial derivative $t \mapsto \frac{\partial f}{\partial t}(t,x)$ exists and equals 0, except when $x = t$. Calculate the function $F(t) = \int_0^1 f(t,x)\,\mathrm{d}x$ and its derivative; observe that, in particular, $F'(t)$ may not vanish.

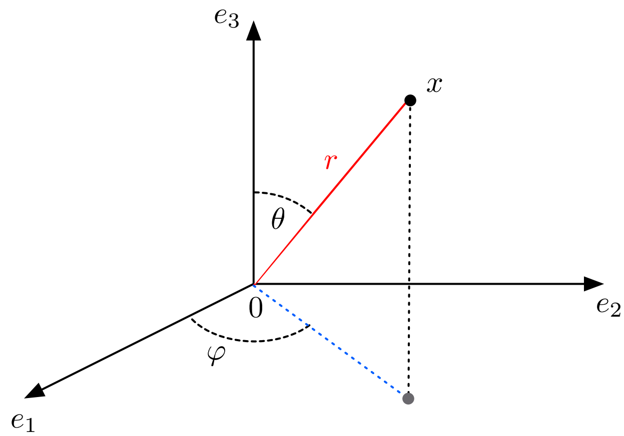

FIGURE 10. *Illustration of the spherical coordinates in* $\mathbb{R}^3$.

Exercise 3.6. Prove Proposition 2.4: let U be an open subset of a normed vector space $(E, ||\cdot||_E)$, $(F, ||\cdot||_F)$ be a Banach space; a function $f : U \to F$ is Gateaux differentiable in the neighborhood V of a point $x \in U$, and we assume that the mapping induced by the Gateaux derivative $V \ni x \mapsto f'_G(x) \in \mathcal{L}(E;F)$ is continuous. Show that f is Fréchet differentiable at x, and $\mathrm{d}f_x = f'_G(x)$.

[Hint: For fixed $x \in U$, $h \in E$, introduce the function $\varphi : [0,1] \to F$ defined by $\varphi(t) = f(x+th)$, then estimate the quantity $\varphi(1) - \varphi(0) - \varphi'(0)$ by expressing $\varphi(1) - \varphi(0) = \int_0^1 \varphi'(t)\,\mathrm{d}t$.]

4. Differential calculus II: some more advanced topics

As we have mentioned in the foreword to Section 2, the differentiability of a function $f : E \to F$ between two normed vector spaces at some point $x_0 \in E$ can be interpreted as the fact that f can be approximated by a first-order polynomial, up to a rest which is "small" when f is only considered near x_0, see (2.2). In this section, we extend this idea to higher-order Taylor formulas, approximating f with increased accuracy by higher-order polynomials, involving higher-order derivatives of f. This raises the need to define such higher-order derivatives for functions between normed vector spaces, which is the topic of Section 4.2. Before entering into the core of the matter, we discuss in Section 4.1 the generalization of the Mean Value Theorem 1.1 to the context of functions between normed vector spaces, as a key technical tool for many purposes.

4.1. The Mean Value theorem

The Mean Value Theorem 1.1, that we have presented in Section 1.2.2 states that the finite difference quotient $\frac{f(b)-f(a)}{b-a}$ of a function $f : [a,b] \to \mathbb{R}$ which is continuous on $[a,b]$ and differentiable on (a,b) exactly equals the value of the derivative $f'(c)$ of f at some intermediate point $c \in (a,b)$. Unfortunately, such a result cannot possibly hold as soon as the arrival space for f is no longer $\mathbb{R}$, as is revealed by the following example.

Example 4.1. Let $f : [0,2\pi] \to \mathbb{C}$ be defined by $f(t) = e^{it}$. Then, obviously, $\frac{f(2\pi)-f(0)}{2\pi-0} = 0$, while the derivative $f'(t) = ie^{it}$ does not vanish on $[0,2\pi]$.

Fortunately, a weaker version of the Mean Value theorem stays valid in the more general context of interest in this section, which is most often sufficient for our purposes. This general result reads as follows:

Theorem 4.1. *Let $(F, ||\cdot||_F)$ be a normed vector space, and let $[a,b] \subset \mathbb{R}$ be a compact interval. Let $f : [a,b] \to F$ be a continuous mapping which is differentiable on (a,b); we assume that there exists a function $\varphi : [a,b] \to \mathbb{R}$ which is continuous on $[a,b]$ and differentiable on (a,b) such that:*

$$\forall t \in (a,b), \quad ||f'(t)||_F \leq \varphi'(t).$$

then, it holds:

$$\forall t, s \in [a,b], \quad ||f(t) - f(s)||_F \leq |\varphi(t) - \varphi(s)|. \tag{4.1}$$

Proof. At first, we note that, without loss of generality, we may limit ourselves to proving (4.1) with $s = a$. In turn, it is actually enough to verify that for all $\varepsilon > 0$,

$$\forall t \in [a,b], \ ||f(t) - f(a)||_F \leq \varphi(t) - \varphi(a) + \varepsilon(t-a) + \varepsilon. \tag{4.2}$$

Indeed, if (4.2) holds true for all $\varepsilon > 0$, then for any given $t \in [a,b]$, letting $\varepsilon > 0$ tend to 0 yields (4.1).

Let us then prove that (4.2) holds. Fixing $\varepsilon > 0$, we consider the set:

$$A := \{t_0 \in (a,b], \ \forall a \leq t \leq t_0, \ ||f(t) - f(a)||_F \leq \varphi(t) - \varphi(a) + \varepsilon(t-a) + \varepsilon\}.$$

Since f and φ are continuous on $[a,b]$, the set A contains the values $a + \eta$ for some small enough $\eta > 0$. Hence, A being non empty and bounded subset of $[a,b]$, it admits a supremum $\theta := \sup A$. By the continuity of f and φ at θ, we see that

$$||f(\theta) - f(a)||_F \leq \varphi(\theta) - \varphi(a) + \varepsilon(\theta - a) + \varepsilon. \tag{4.3}$$

In order to prove (4.2), it is enough to prove that $\theta = b$. To this end, we argue by contradiction, assuming that $\theta < b$. At first, expressing the differentiability of f at θ, we see that there exists $\delta > 0$ such that:

$$\forall t \in [a,b] \text{ s.t. } |t - \theta| \leq \delta, \quad ||f(t) - f(\theta) - f'(\theta)(t-\theta)||_F \leq \frac{\varepsilon}{2}|t - \theta|. \tag{4.4}$$

Likewise, since φ is differentiable at θ, up to decreasing the value of δ, we have:

$$\forall t \in [a,b] \text{ s.t. } |t - \theta| \leq \delta, \quad |\varphi(t) - \varphi(\theta) - \varphi'(\theta)(t-\theta)| \leq \frac{\varepsilon}{2}|t - \theta|. \tag{4.5}$$

Hence, for all $t \in [a,b]$ with $|t - \theta| \leq \delta$, we have by (4.4) and the triangle inequality:

$$\begin{array}{rcl} ||f(t) - f(\theta)||_F & \leq & ||f'(\theta)||_F |t - \theta| + \frac{\varepsilon}{2}|t - \theta| \\ & \leq & \varphi'(\theta)|t - \theta| + \frac{\varepsilon}{2}|t - \theta|, \end{array}$$

and using now (4.5) and the triangle inequality:

$$||f(t) - f(\theta)||_F \leq |\varphi(t) - \varphi(\theta)| + \varepsilon|t - \theta|.$$

Hence using once more the triangle inequality and (4.3), for all $t \in [\theta, \theta + \delta)$, we have, eventually:

$$\begin{array}{rcl} ||f(t) - f(a)||_F & \leq & ||f(t) - f(\theta)||_F + ||f(\theta) - f(a)||_F \\ & \leq & |\varphi(t) - \varphi(\theta)| + \varepsilon|t - \theta| + \varphi(\theta) - \varphi(a) + \varepsilon(\theta - a) + \varepsilon \\ & \leq & \varphi(t) - \varphi(a) + \varepsilon(t - a) + \varepsilon. \end{array}$$

We have thus proved the inclusion $[\theta, \theta + \eta) \subset A$, which contradicts the definition of θ as the supremum of A. Hence, (4.2) holds, and it leads to the desired inequality (4.1). □

Corollary 4.1. *Let $U \subset E$ be a convex open subset of a normed vector space $(E, ||\cdot||_E)$ and let $(F, ||\cdot||_F)$ be another normed vector space. Let $f : U \to F$ be a differentiable mapping such that there exists $k > 0$ with*

$$\forall x \in U, \quad ||\mathrm{d}f_x||_{\mathcal{L}(E;F)} \leq k.$$

Then,

$$\forall x, y \in U, \quad ||f(y) - f(x)||_F \leq k||y - x||_E.$$

Proof. As usual, we reduce to the one-dimensional case. Let $x, y \in U$ be given; we consider the function $g : [0,1] \to F$ defined by

$$g(t) = f(x + t(y - x)),$$

which is possible because U being convex, the whole segment $\{(1-t)x + ty, \ t \in [0,1]\}$ is included in U. The function g is continuous on $[0,1]$, and an application of the chain rule of Theorem 2.1 reveals that it is differentiable on $(0,1)$, with derivative

$$\forall t \in (0,1), \quad g'(t) = \mathrm{d}f_{x+t(y-x)}(y - x).$$

This entails:

$$||g'(t)||_F \leq ||\mathrm{d}f_{x+t(y-x)}||_{\mathcal{L}(E;F)} ||y - x||_E \leq k||y - x||_E.$$

Then, applying Theorem 4.1 with $\varphi(t) = k||x-y||_E\, t$, it follows

$$||g(1) - g(0)||_F = ||f(y) - f(x)||_F \le k||y - x||_E,$$

as desired. □

Corollary 4.2. *Let $(E, ||\cdot||_E)$ and $(F, ||\cdot||_F)$ be two normed vector spaces, and let $U \subset E$ be a connected open subset of E. If $f : U \to F$ is a differentiable mapping such that $\mathrm{d}f_x = 0$ for all $x \in U$, then f is constant on U.*

Proof. Let $x_0 \in U$ be fixed; we introduce the non empty subset

$$S := \{x \in U,\ \ f(x) = f(x_0)\}.$$

Since f is continuous on U, S is a closed subset of U. Let now $x \in S$, and let $r > 0$ be small enough so that the open ball $B := B(x, r)$ is contained in U. Since B is convex and $\mathrm{d}f_x = 0$ for $x \in B$, Corollary 4.1 implies that

$$\forall z \in B, \quad ||f(z) - f(x)||_F = 0.$$

Hence, $f(z) = f(x) = f(x_0)$ on B, so that $B \subset S$. Summarizing, we have proved that S is a non empty subset of the connected set U, which is at the same time closed and open. Hence, $S = U$, thus terminating the proof. □

4.2. **Higher-order derivatives**

We now turn to the definition of the higher-order derivatives of functions between normed vector spaces, starting with that of second-order derivatives.

As in the familiar context of a function f, from an open interval $I \subset \mathbb{R}$ into $\mathbb{R}$, the second-order derivative appraises how the differential mapping $x \mapsto \mathrm{d}f_x$ depends on the base point x. More precisely, let $(E, ||\cdot||_E)$, $(F, ||\cdot||_F)$ be two normed vector spaces, $U \subset E$ be open, and let $f : U \to F$ be a Fréchet differentiable function on U. As we have seen, the Fréchet derivative of f induces a mapping

$$\mathrm{d}f : U \ni x \longmapsto \mathrm{d}f_x \in \mathcal{L}(E; F), \tag{4.6}$$

that is, to each point $x \in U$, $\mathrm{d}f$ associates the linear continuous mapping $\mathrm{d}f_x \in \mathcal{L}(E; F)$ supplied by Proposition-Definition 2.1. Since $\mathrm{d}f$ is a mapping from an open subset of a normed vector space into the normed vector space $\mathcal{L}(E; F)$, one may consider its differentiation.

Definition 4.1. *Let $(E, ||\cdot||_E)$, $(F, ||\cdot||_F)$ be two normed vector spaces, and let $U \subset E$ be open. A differentiable mapping $f : U \to F$ is called twice differentiable at some point $x \in U$ if f is differentiable on a neighborhood $V \subset U$ of x, and if the mapping $\mathrm{d}f : V \to \mathcal{L}(E; F)$ in (4.6) is Fréchet differentiable at x. The associated differential is called the second-order derivative of f at x and is denoted by $\mathrm{d}^2 f_x \in \mathcal{L}(E; \mathcal{L}(E; F))$.*

Let us clarify notations: for all $h, k \in E$, $\mathrm{d}^2 f_x(h) \in \mathcal{L}(E; F)$ is the differential of the mapping $U \ni x \mapsto \mathrm{d}f_x \in \mathcal{L}(E; F)$ in the direction h, and $\mathrm{d}^2 f_x(h)(k)$ is the element of F resulting from the evaluation of this mapping at k.

By nature, the second-order derivative of a function $f : U \to F$ at $x \in U$ belongs to $\mathcal{L}(E; \mathcal{L}(E; F))$, i.e. it is a continuous linear mapping from E into the space $\mathcal{L}(E; F)$ of operators. The following lemma allows to identify it with a continuous bilinear mapping from $E \times E$ into f.

Lemma 4.1. *Let $(E_1, ||\cdot||_{E_1})$, $(E_2, ||\cdot||_{E_2})$ and $(F, ||\cdot||_F)$ be three normed vector spaces. The mapping $\mathcal{I} : \mathcal{L}(E_1; \mathcal{L}(E_2; F)) \to \mathcal{L}(E_1, E_2; F)$ defined by*

$$\forall f \in \mathcal{L}(E_1; \mathcal{L}(E_2; F)), \quad \mathcal{I}(f)(x_1, x_2) = F(x_1)(x_2),$$

is an isometry.

Proof. At first, let us make precise the definition of $\mathcal{I}$: if f is a continuous linear mapping from E_1 into $\mathcal{L}(E_2; F)$, then for all $x_1 \in E_1$, $f(x_1)$ belongs to $\mathcal{L}(E_2; F)$, i.e. $f(x_1)$ is a continuous linear mapping from E_2 into F. Hence, $f(x_1)(x_2)$ defines an element of F, which depends on x_1 and x_2 in a linear way.

Let us now turn to the proof, which is a simple, albeit a little tedious handling of the definition of operator norm. The very definition of the operator norm yields:

$$\begin{aligned}\forall x_1 \in E_1,\ x_2 \in E_2, \quad ||f(x_1)(x_2)||_F &\leq ||f(x_1)||_{\mathcal{L}(E_2;F)}||x_2||_{E_2} \\ &\leq ||f||_{\mathcal{L}(E_1;\mathcal{L}(E_2;F))}||x_1||_{E_1}||x_2||_{E_2},\end{aligned}$$

which proves that

$$||\mathcal{I}(f)||_{\mathcal{L}(E_1,E_2;F)} \leq ||f||_{\mathcal{L}(E_1;\mathcal{L}(E_2;F))}.$$

To prove that $\mathcal{I}$ is an isomorphism, we simply exhibit the inverse mapping. For any bilinear mapping $g \in \mathcal{L}(E_1, E_2; F)$, we define $\mathcal{J}(g) \in \mathcal{L}(E_1; \mathcal{L}(E_2; F))$ by the formula:

$\forall x_1 \in E_1,\ \mathcal{J}(g)(x_1)$ is the continuous linear mapping in $\mathcal{L}(E_2; F)$ defined by

$$\forall x_2 \in E_2,\ \mathcal{J}(g)(x_1)(x_2) = g(x_1, x_2).$$

Again, the definition of the operator norm yields:

$$\forall x_1 \in E_1,\ x_2 \in E_2, \quad ||\mathcal{J}(g)(x_1)(x_2)||_F \leq ||g||_{\mathcal{L}(E_1,E_2;F)}||x_1||_{E_1}||x_2||_{E_2},$$

and so

$$\forall x_1 \in E_1, \quad ||\mathcal{J}(g)(x_1)||_{\mathcal{L}(E_2;F)} \leq ||g||_{\mathcal{L}(E_1,E_2;F)}||x_1||_{E_1},$$

which finally shows that

$$||\mathcal{J}(g)||_{\mathcal{L}(E_1;\mathcal{L}(E_2;F))} \leq ||g||_{\mathcal{L}(E_1,E_2;F)}.$$

This completes the proof. □

We shall thenceforth equivalently see the second-order derivative of a function $f : U \to F$ at $x \in U$ as the operator-valued mapping $\mathrm{d}^2 f_x \in \mathcal{L}(E; \mathcal{L}(E; F))$ produced by the Definition 4.1 of second-order differentiability, or as the bilinear mapping $E \times E \to F$ supplied by Lemma 4.1.

Example 4.2. Let $(E, ||\cdot||_E)$ and $(F, ||\cdot||_F)$ be two normed vector spaces.

(i) If $f : E \to F$ is a continuous linear mapping, we have seen in Example 2.1 that is first-order derivative at any $x \in E$ reads

$$\forall h \in E, \quad \mathrm{d}f_x(h) = f(h),$$

that is, the mapping $\mathrm{d}f : E \to \mathcal{L}(E, F)$ is constant, equal to f. As a result, $\mathrm{d}^2 f_x = 0$.

(ii) Let $b : E \times E \to F$ be a continuous bilinear mapping, and let $f : E \to F$ be the quadratic function defined by $f(x) = b(x, x)$. We have seen in Example 2.1 that the differential of f at any point $x \in E$ reads:

$$\forall h \in E, \quad \mathrm{d}f_x(h) = b(h, x) + b(x, h).$$

A simple verification reveals that f is twice differentiable on E, with second-order derivative given by:

$$\forall h, k \in E, \quad \mathrm{d}^2 f_x(h)(k) = b(h, k) + b(k, h).$$

We now provide two interpretations that may shed some light on the quite abstract definition of second-order derivatives.

Lemma 4.2. *Let $(E, ||\cdot||_E)$ and $(F, ||\cdot||_F)$ be two normed vector spaces, and let $U \subset E$ be open. Let $f : U \to F$ be a twice differentiable function at some point $x \in U$, and let $k \in E$ be a fixed direction. Then the mapping $g : U \to F$ defined by $g(x) = \mathrm{d}f_x(k)$ is differentiable at x and its derivative reads*

$$\mathrm{d}g_x(h) = \mathrm{d}^2 f_x(h, k).$$

Proof. Let us simply consider, for all $h \in E$,

$$\begin{aligned}r_g(h) &:= g(x+h) - g(x) - \mathrm{d}^2 f_x(h, k) \\ &= \mathrm{d}f_{x+h}(k) - \mathrm{d}f_x(k) - \mathrm{d}^2 f_x(h)(k).\end{aligned}$$

By definition of the operator norm, we have:

$$||r_g(h)||_F \leq ||\mathrm{d}f_{x+h} - \mathrm{d}f_x - \mathrm{d}^2 f_x(h)||_{\mathcal{L}(E;F)}||k||_E.$$

Now, for any $\varepsilon > 0$, the definition of second-order differentiability states that there exists a number $\delta > 0$ such that

$$\forall h \in E,\ ||h||_E \leq \delta \ \Rightarrow\ ||\mathrm{d}f_{x+h} - \mathrm{d}f_x - \mathrm{d}^2 f_x(h)||_{\mathcal{L}(E;F)} \leq \varepsilon ||h||_E.$$

Hence,

$$\forall h \in E,\ ||h||_E \leq \delta, \quad \frac{||r_g(h)||_F}{||h||_E} \leq \varepsilon ||k||_E,$$

which means exactly that g is Fréchet differentiable at x, with derivative $h \mapsto \mathrm{d}^2 f_x(h,k)$. $\square$

Lemma 4.3. *Let $(E, ||\cdot||_E)$ and $(F, ||\cdot||_F)$ be two normed vector spaces, and let $U \subset E$ be open. Let $f : U \to F$ be a twice differentiable function at some point $x \in U$. The following formula holds:*

$$\mathrm{d}^2 f_x(h,k) = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\mathrm{d}}{\mathrm{d}s} f(x+th+sk)\Big|_{s=0} \right)\Big|_{t=0}.$$

Proof. This result is an immediate consequence of the previous Lemma 4.2 since

$$\frac{\mathrm{d}}{\mathrm{d}s} f(x+th+sk)\Big|_{s=0} = \mathrm{d}f_{x+th}(k), \text{ and } \frac{\mathrm{d}}{\mathrm{d}t}\mathrm{d}f_{x+th}(k)\Big|_{t=0} = \mathrm{d}^2 f_x(h,k).$$

$\square$

One remarkable property of the second-order derivative is its symmetry when it is seen as a bilinear mapping from $E \times E$ to F.

Theorem 4.2 (Schwarz theorem)**.** *Let $(E, ||\cdot||_E)$ and $(F, ||\cdot||_F)$ be two normed vector spaces, and let $U \subset E$ be open. Let $f : U \to F$ be a twice differentiable function; then the second-order derivative $\mathrm{d}^2 f_x \in \mathcal{L}(E, E; F)$ is a symmetric, continuous bilinear form.*

Proof(*). We start with a lemma, whose proof is postponed for clarity

Lemma 4.4. *Let $f : U \to F$ be a twice differentiable function at some point $x \in U$, and let $\delta > 0$ be such that $\overline{B(x,\delta)} \subset U$. Let the quantity $A : B(x,\delta) \times B(x,\delta) \to F$ be defined by*

$$A(h,k) := f(x+h+k) - f(x+h) - f(x+k) + f(x).$$

Then,

$$\lim_{(h,k)\to(0,0)} \frac{||A(h,k) - \mathrm{d}^2 f_x(h,k)||_F}{||h||_E^2 + ||k||_E^2} = 0.$$

In loose terms, Lemma 4.4 states that $\mathrm{d}^2 f_x(h,k)$ is close to the quantity $A(h,k)$, up to leading order in terms of h and k. Since the latter is symmetric in the variables h and k, the expected symmetry of $\mathrm{d}^2 f_x(h,k)$ follows easily. More precisely, let us write

$$\mathrm{d}^2 f_x(h,k) - \mathrm{d}^2 f_x(k,h) = \mathrm{d}^2 f_x(h,k) - A(h,k) + A(k,h) - \mathrm{d}^2 f_x(k,h).$$

Let $\varepsilon > 0$ be given; using Lemma 4.4 and the triangle inequality, there exists $\delta > 0$ such that

$$||h||_E \leq \delta \text{ and } ||k||_E \leq \delta \Rightarrow ||\mathrm{d}^2 f_x(h,k) - \mathrm{d}^2 f_x(k,h)||_F \leq \varepsilon\Big(||h||_E^2 + ||k||_E^2\Big).$$

Now, for any $h \neq 0$ and $k \neq 0$, let us define $\tilde{h} := \delta \frac{h}{||h||_E}$ and $\tilde{k} := \delta \frac{k}{||k||_E}$, which satisfy $||\tilde{h}||_E \leq \delta$ and $||\tilde{k}||_E \leq \delta$. Hence, we obtain

$$||\mathrm{d}^2 f_x(\tilde{h},\tilde{k}) - \mathrm{d}^2 f_x(\tilde{k},\tilde{h})||_F \leq \varepsilon\Big(||\tilde{h}||_E^2 + ||\tilde{k}||_E^2\Big).$$

Since all the quantities involved in the previous inequality are bilinear in terms of h and k, multiplying both sides by $\frac{1}{\delta}||h||_E||k||_E$ yields:

$$\forall h,k \in E, \quad ||\mathrm{d}^2 f_x(h,k) - \mathrm{d}^2 f_x(k,h)||_F \leq \varepsilon\Big(||h||_E||k||_E + ||k||_E||h||_E\Big) \leq \varepsilon\Big(||h||_E^2 + ||k||_E^2\Big);$$

since ε is arbitrary, this implies that $\mathrm{d}^2 f_x(h,k) = \mathrm{d}^2 f_x(k,h)$. $\square$

We now provide the proof of the missing link in the above argument.

Proof of Lemma 4.4. Let us consider the mapping $B : B(x,\delta) \times B(x,\delta) \to F$ given by:

$$\begin{array}{rcl} B(h,k) & = & A(h,k) - \mathrm{d}^2 f_x(h,k) \\ & = & f(x+h+k) - f(x+h) - f(x+k) + f(x) - \mathrm{d}^2 f_x(h,k). \end{array}$$

The function B is differentiable with respect to the second variable at any $(h,k) \in B(x,\delta) \times B(x,\delta)$, and its partial derivative reads

$$\frac{\partial B}{\partial k}(h,k)(\widehat{k}) = \mathrm{d}f_{x+h+k}(\widehat{k}) - \mathrm{d}f_{x+k}(\widehat{k}) - \mathrm{d}^2 f_x(h,\widehat{k}).$$

Let us rewrite this expression as:

$$\frac{\partial B}{\partial k}(h,k)(\widehat{k}) = \Big(\mathrm{d}f_{x+h+k}(\widehat{k}) - \mathrm{d}f_x(\widehat{k}) - \mathrm{d}^2 f_x(h+k,\widehat{k})\Big) - \Big(\mathrm{d}f_{x+k}(\widehat{k}) - \mathrm{d}f_x(\widehat{k}) - \mathrm{d}^2 f_x(k,\widehat{k})\Big).$$

Now, let $\varepsilon > 0$ be given; by the definitions of operator norm and second-order differentiability, there exists $\delta > 0$ such that for $||h||_E \leq \delta$ and $||k||_E \leq \delta$, the first term in the above right-hand side can be estimated as:

$$\begin{array}{rcl} \Big|\Big|\mathrm{d}f_{x+h+k}(\widehat{k}) - \mathrm{d}f_x(\widehat{k}) - \mathrm{d}^2 f_x(h+k,\widehat{k})\Big|\Big|_F & \leq & \Big|\Big|\mathrm{d}f_{x+h+k} - \mathrm{d}f_x - \mathrm{d}^2 f_x(h+k)\Big|\Big|_{\mathcal{L}(E;F)} ||\widehat{k}||_E \\ & \leq & \varepsilon ||h+k||_E ||\widehat{k}||_E. \end{array}$$

By the same token, up to decreasing the value of $\delta > 0$, one has:

$$\forall h,k \in E,\ \ ||h||_E \leq \delta,\ ||k||_E \leq \delta, \quad \Big|\Big|\mathrm{d}f_{x+k}(\widehat{k}) - \mathrm{d}f_x(\widehat{k}) - \mathrm{d}^2 f_x(k,\widehat{k})\Big|\Big|_F \leq ||k||_E ||\widehat{k}||_E.$$

Combining both inequalities, we see that

$$\forall h,k \in E,\ \ ||h||_E \leq \delta,\ ||k||_E \leq \delta \ \Rightarrow \ \left|\left|\frac{\partial B}{\partial k}(h,k)\right|\right|_{\mathcal{L}(E;F)} \leq \varepsilon\Big(||h||_E + ||k||_E\Big).$$

We now apply the generalized mean value Theorem 4.1 (and more precisely, Corollary 4.1) to the function $k \mapsto B(h,k)$: For all $||h||_E \leq \delta$ and $||k||_E \leq \delta$,

$$||B(h,k)||_{\mathcal{L}(E;F)} = ||B(h,k) - B(h,0)||_{\mathcal{L}(E;F)} \leq \varepsilon\Big(||h||_E + ||k||_E\Big)||k||_E.$$

This terminates the proof of the lemma. ☐

The above Definition 4.1 of the second-order derivative of a function $f : U \to F$ can be generalized to derivatives up to an arbitrary order. For brevity, we solely provide the main definitions and properties attached to this notion.

Definition 4.2. *Let $f : U \to F$ be a function and let $n \geq 1$. f is called n times differentiable at some point $x \in U$ if there exists an open neighborhood V of x such that f is differentiable on V, and if the mapping $V \ni x \mapsto \mathrm{d}f_x$, from V into $\mathcal{L}(E;F)$ is $(n-1)$ times differentiable at x.*

The function f is said to be of class $\mathcal{C}^n$ on U if it is differentiable on U, and if the derivative $x \mapsto \mathrm{d}f_x \in \mathcal{L}(E;F)$ is of class $\mathcal{C}^{n-1}$.

The following result defines the higher-order derivatives of a function in terms of symmetric, multilinear mappings.

Theorem-Definition 4.1. *Let $(E, ||\cdot||_E)$ and $(F, ||\cdot||_F)$ be two Banach spaces, and let $U \subset E$ be open. For any integer $n \geq 1$, one function $f : U \to F$ is n times differentiable at some point $x \in U$ if there exist*

- *A neighborhood $V \subset U$ of x;*
- *For all $p \leq n-1$, a mapping $L^p : V \to \mathcal{L}_s(E^p;F)$;*
- *An n-linear symmetric continuous mapping $\ell^n \in \mathcal{L}_s(E^n;F)$.*

such that:

- *L^1 coincides with the first-order differential $\mathrm{d}f : V \to \mathcal{L}(E;F)$ of f;*
- *For all $p \leq n-2$, L^p is Fréchet differentiable on V and for all $y \in V$:*

$$\forall h_1,\dots,h_{p+1} \in E, \quad L^{p+1}(y)(h_1,\dots,h_{p+1}) = \frac{\partial g^p}{\partial y}(h_1,\dots,h_p,h_{p+1}),$$

where we have defined $g^p : E^p \times V$ by $g^p(h_1,\dots,h_p,y) := L^p(y)(h_1,\dots,h_p)$.

- *L^{n-1} is Fréchet differentiable at x and*
$$\forall h_1, \dots, h_n \in E, \quad \ell^n(h_1, \dots, h_n) = \frac{\partial g^{n-1}}{\partial y}(h_1, \dots, h_{n-1}, h_n),$$
where we have defined $g^{n-1} : E^{n-1} \times V$ by $g^{n-1}(h_1, \dots, h_{n-1}, y) := L^{n-1}(y)(h_1, \dots, h_{n-1})$.

For $p = 1, \dots, n-1$, the symmetric p-linear continuous mapping $L^p(x) \in \mathcal{L}_s(E^p; F)$ is called the p^{th} order differential of f at x and the symmetric n-linear continuous mapping $\ell^n \in \mathcal{L}_s(E^n; F)$ is called the n^{th} order differential of f at x.

Let us now observe that the stability of the differentiable character of functions by composition expressed in the chain rule of Theorem 2.1 carries over to the setting of higher-order differentiability.

Theorem 4.3 (Chain rule for higher-order derivatives)**.** *Let $(E, ||\cdot||_E)$, $(F, ||\cdot||_F)$ and $(G, ||\cdot||_G)$ be three normed vector spaces, and let $U \subset E$ and $V \subset F$ be open subsets. Let $f : U \to F$ and $g : V \to G$ be two functions; we assume that $f(U) \subset V$ so that the composite mapping $g \circ f : U \to G$ is well-defined. For $n \geq 1$,*

(i) If $x \in U$ is a point such that f is n times differentiable at x and g is n times differentiable at $f(x)$, then, $g \circ f : U \to G$ is n times differentiable at x.
(ii) If f is of class $\mathcal{C}^n$ on U and g is of class $\mathcal{C}^n$ on V, then $g \circ f$ is of class $\mathcal{C}^n$ on U.

Proof. On several occurrences in the proof, we shall rely on the fact that the mapping
$$\begin{array}{ccc} \mathcal{L}(E; F) \times \mathcal{L}(F; G) & \longrightarrow & \mathcal{L}(E; G) \\ (u, v) & \longmapsto & v \circ u \end{array}$$
is bilinear and continuous, and thus of class $\mathcal{C}^\infty$, see Example 4.2. We now prove both statements of the theorem by induction on $n \geq 1$.

Case $n = 1$: If f and g are differentiable at x and $f(x)$, respectively, the chain rule of Theorem 2.1 shows that the composite mapping $g \circ f$ is differentiable at x, and that is derivative reads:
$$\mathrm{d}(g \circ f)_x = \mathrm{d}g_{f(x)} \circ \mathrm{d}f_x,$$
so that (i) holds for $n = 1$. If f and g are of class $\mathcal{C}^1$ on U and V respectively, both mappings
$$U \ni z \mapsto \mathrm{d}f_z \in \mathcal{L}(E; F) \text{ and } V \ni y \mapsto \mathrm{d}g_y \in \mathcal{L}(F; G)$$
are continuous. Since $U \ni z \mapsto f(z) \in V$ is also continuous, it follows that $U \ni z \mapsto \mathrm{d}g_{f(z)} \in \mathcal{L}(F; G)$ is also continuous. From our preliminary remark, it follows that
$$U \ni x \mapsto \mathrm{d}g_{f(x)} \circ \mathrm{d}f_x \in \mathcal{L}(E; G)$$
is continuous, so that $g \circ f$ is of class $\mathcal{C}^1$ on U, as desired.

Case $n \geq 2$: Let us assume that f and g are n times differentiable at x and $f(x)$, respectively. By definition, there exist open neighborhoods $\widetilde{U} \subset U$ and $\widetilde{V} \subset V$ of x and $f(x)$, respectively, such that the mappings
$$\widetilde{U} \ni z \mapsto \mathrm{d}f_z \in \mathcal{L}(E; F) \text{ and } \widetilde{V} \ni y \mapsto \mathrm{d}g_y \in \mathcal{L}(F; G)$$
are differentiable.Thus, up to decreasing the size of $\widetilde{U}$ so that $f(\widetilde{U}) \subset \widetilde{V}$, $g \circ f$ is differentiable on $\widetilde{U}$, with differential
$$\forall z \in \widetilde{U}, \quad \mathrm{d}(g \circ f)_z = \mathrm{d}g_{f(z)} \circ \mathrm{d}f_z.$$
Here, by assumption, $z \mapsto \mathrm{d}f_z \in \mathcal{L}(E; F)$ is $(n-1)$ times differentiable at x, $y \mapsto \mathrm{d}g_y \in \mathcal{L}(F; G)$ is $(n-1)$ times differentiable at $f(x)$ and $z \mapsto f(z)$ is $(n-1)$ times differentiable at x; by the induction hypothesis, $\widetilde{U} \ni z \mapsto \mathrm{d}g_{f(z)} \in \mathcal{L}(F; G)$ is $(n-1)$ times differentiable at x, and again, by our preliminary remark,
$$\widetilde{U} \ni x \mapsto \mathrm{d}g_{f(x)} \circ \mathrm{d}f_x \in \mathcal{L}(E; G)$$
$(n-1)$ times differentiable at x. This shows (i). The proof of (ii) is conducted in a similar way. □

Example 4.3 (Example 2.2 continued)**.** Let $(E, ||\cdot||_E)$ and $(F, ||\cdot||_F)$ be two normed vector spaces, and let $U \subset E$ be open. For some $n \geq 1$, let $f : U \to F$ be a function which is n times differentiable on U. Let also $h \in E$ be such that the whole segment $\{x + th,\ t \in [0, 1]\}$ is contained in U, so that the composite function $g : [0, 1] \to F$ defined by

$$\forall t \in [0, 1], \quad g(t) = f(x + th)$$

is well-defined; the latter reads $g = f \circ \ell$, where $\ell : [0, 1] \to U$ is given by $\ell(t) = x + th$. The chain rule Theorem 4.3 reveals that g is n times differentiable on $(0, 1)$. We have seen in Example 2.2 that is first-order derivative reads:

$$\forall t \in (0, 1), \quad g'(t) = \mathrm{d}f_{x+th}(h).$$

Lemma 4.2 shows that, if $n \geq 2$, the second-order derivative of g equals:

$$\forall t \in (0, 1), \quad g''(t) = \mathrm{d}^2 f_{x+th}(h, h).$$

By induction, using Theorem-Definition 4.1 about the nature of the n^{th}-order differential of f, we obtain that the n^{th}-order derivative of g is:

$$\forall t \in (0, 1), \quad g^{(n)}(t) = \mathrm{d}^n f_{x+th}(\underbrace{h, \dots, h}_{n \text{ times}}).$$

We conclude this section by adapting the previous concepts to the particular case where U is an open subset of the Euclidean space $E = \mathbb{R}^d$, and $(F, ||\cdot||_F)$ is an arbitrary normed vector space.

As we have seen in Section 2.5, if $f : U \to F$ is differentiable at some point $x \in U$, it admits partial derivatives $\frac{\partial f}{\partial x_i}(x)$ at x, which are defined by:

$$\frac{\partial f}{\partial x_i}(x) = \lim_{t \to 0} \frac{f(x + te_i) - f(x)}{t}, \quad i = 1, \dots, d.$$

The differential $\mathrm{d}f_x \in \mathcal{L}(\mathbb{R}^d; F)$ has then the following expression:

$$\forall h = (h_1, \dots, h_d) \in E, \quad \mathrm{d}f_x(h) = \sum_{i=1}^{d} \frac{\partial f}{\partial x_i}(x) h_i.$$

If $f : U \to F$ is now assumed to be twice differentiable at x, there exists a neighborhood V of x in U where f is differentiable, so that the partial derivatives $z \mapsto \frac{\partial f}{\partial x_i}(z)$ are well-defined on V. In addition, f admits second-order partial derivatives $\frac{\partial^2 f}{\partial x_i \partial x_j}(x)$ at x, defined by:

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \lim_{t \to 0} \frac{\frac{\partial f}{\partial x_j}(x + te_i) - \frac{\partial f}{\partial x_j} f(x)}{t}, \quad i, j = 1, \dots, d.$$

The second-order derivative of f at x then reads:

$$\forall h = (h_1, \dots, h_d),\ k = (k_1, \dots, k_d) \in \mathbb{R}^d, \quad \mathrm{d}^2 f_x(h, k) = \sum_{i=1}^{d} \frac{\partial^2 f}{\partial x_i \partial x_j}(x) h_i k_j,$$

and the Schwarz Theorem 4.2 states that the second-order partial derivatives of f at x are symmetric, i.e.:

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \frac{\partial^2 f}{\partial x_j \partial x_i}(x), \quad i, j = 1, \dots, d.$$

It is easy to carry over this discussion to define higher-order partial derivatives for a sufficiently differentiable function $f : U \to F$, and to express is higher-order differential in terms of those.

Let us eventually introduce two useful differential operators in this finite-dimensional context.

Definition 4.3. *Let U be an open subset of $\mathbb{R}^d$ and let $f : U \to \mathbb{R}$ be a twice differentiable function at some point $x \in U$.*

- *The Hessian of f at x is the $d \times d$ matrix $\nabla^2 f(x)$ gathering its second-order partial derivatives:*

$$\forall i, j = 1, \dots, d, \quad (\nabla^2 f(x))_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}(x).$$

- *The Laplacian $\Delta f(x)$ of f at x is:*

$$\Delta f(x) = \text{div}(\nabla f)(x) = \text{tr}\Big(\nabla^2 f(x)\Big) = \sum_{i=1}^{d} \frac{\partial^2 f}{\partial x_i^2}(x).$$

4.3. Taylor-Young's formula

In this section, we initiate our discussion about Taylor's formulas. As we have hinted at, the differentiability of a function $f : U \subset E \to F$ between two normed vector spaces at some point $x \in U$ is associated to the following approximation of f in the vicinity of x:

$$\text{For } h \in E \text{ "small enough"}, \quad f(x+h) = f(x) + \mathrm{d}f_x(h) + \mathrm{o}(||h||_E);$$

in other words, $f(x+h)$ is approximated by the affine function $h \mapsto f(x) + \mathrm{d}f_x(h)$ near x, up to a remainder $\mathrm{o}(||h||_E)$ which is "small" at first-order.

The purpose of Taylor's formulas is to provide more precise approximations of $f(x+h)$, by n^{th}-order polynomial functions, involving the n^{th}-order derivatives of f at x, up to a "smaller" remainder $\mathrm{o}(||h||_E^n)$.

Several versions of Taylor's formula are available. The one that we discuss in the present section is the most general one: it applies to functions defined between arbitrary normed vector spaces (in particular, it does not require any completeness assumption about E or F), and it assumes minimal regularity about the considered function f. The conclusion is, however, fairly weak insofar as pretty much nothing is known about the remainder term $\mathrm{o}(||h||_E^n)$.

Theorem 4.4. *Let $(E, ||\cdot||_E)$ and $(F, ||\cdot||_F)$ be two normed vector spaces, and let $U \subset E$ be an open subset. Let $f : U \to F$ be a function which is n times differentiable at some point $x \in U$. Then, it holds:*

$$f(x+h) = \sum_{p=0}^{n} \frac{1}{p!} \mathrm{d}^p f_x(h, \dots, h) + \mathrm{o}(||h||_E^n). \tag{4.7}$$

Remark 4.1. *Let us recall the precise meaning of the formulation (4.7): for any $\varepsilon > 0$, there exists $\delta > 0$ such that*

$$\forall h \in E, \ ||h||_E \leq \delta \ \Rightarrow \ \left|\left| f(x+h) - \sum_{p=0}^{n} \frac{1}{p!} \mathrm{d}^p f_x(h, \dots, h) \right|\right|_F \leq \varepsilon ||h||_E^n. \tag{4.8}$$

*Proof**. The proof goes by induction on the order n of the formula.

For $n = 1$: Formula (4.7) with $n = 1$ is exactly the definition of the differentiability of f at x, see Proposition-Definition 2.1.

Let now $n \geq 1$ be fixed; we assume that the statement of the theorem holds up to the order n, for any function which is n times differentiable between any normed vector spaces.

Let $f : U \to F$ be a function which is $(n+1)$ times differentiable at $x \in U$, and let $\delta > 0$ be so small that $\overline{B(x, \delta)} \subset U$. We define the function $r : B(x_0, \delta) \to F$ by

$$r(h) = f(x+h) - \sum_{p=0}^{n+1} \frac{1}{p!} \mathrm{d}_x^p f(h, \dots, h).$$

Observing that the mapping $h \mapsto \mathrm{d}f_x^p(h, \dots, h)$ is a symmetric, p-linear continuous mapping, it follows that r is differentiable in the neighborhood of 0 in E with differential

$$\begin{aligned} \forall k \in E, \quad \mathrm{d}r_h(k) &= \mathrm{d}f_{x+h}(k) - \sum_{p=1}^{n+1} \tfrac{p}{p!} \mathrm{d}f_x^p(\underbrace{h, \dots, h}_{(p-1) \text{ times}}, k) \\ &= \mathrm{d}f_{x+h}(k) - \sum_{p=1}^{n} \tfrac{1}{p!} \mathrm{d}^p(\mathrm{d}f_x(k))(\underbrace{h, \dots, h}_{p \text{ times}}), \end{aligned}$$

see Example 2.1 (v). When passing from the first line to the second one, we have used the definition of the $(p+1)^{\text{th}}$-order derivative of f at x as the p^{th}-order derivative of the differential $\mathrm{d}f$ at x, see Theorem-Definition 4.1; the notation $\mathrm{d}^p(\mathrm{d}f_x(k))$ stands for the p^{th} order differential of the mapping $z \mapsto \mathrm{d}f_z(k)$ at x for fixed k.

Applying the induction hypothesis to the mapping $\mathrm{d}f : U \to \mathcal{L}(E; F)$, we see that, for all $\varepsilon > 0$, up to decreasing the value of δ, it holds

$$\forall h \in E, \ \ ||h||_E \leq \delta \Rightarrow \left|\left|\mathrm{d}f_{x+h} - \sum_{p=0}^{n} \frac{1}{p!} \mathrm{d}^p(\mathrm{d}f_{x+\cdot})(h, \dots, h)\right|\right|_{\mathcal{L}(E;F)} \leq \varepsilon ||h||_E^n,$$

which means exactly that

$$\forall h \in E, \ \ ||h||_E \leq \delta \Rightarrow \forall k \in E, \ \ \left|\left|\mathrm{d}f_{x+h}(k) - \sum_{p=0}^{n} \frac{1}{p!} \mathrm{d}^{p+1} f_x(h, \dots, h, k)\right|\right|_F \leq \varepsilon ||h||_E^n ||k||_E.$$

In turns, this implies

$$\forall h \in E, \ \ ||h||_E \leq \delta \Rightarrow ||\mathrm{d}r_h||_{\mathcal{L}(E;F)} \leq \varepsilon ||h||_E^n.$$

Using the Mean Value inequality of Corollary 4.2, we now see that

$$\forall h \in E, \ \ ||h||_E \leq \delta \ \Rightarrow \ ||r(h) - r(0)||_F \leq \varepsilon ||h||_E^n ||h||_E,$$

which is the desired result (4.8). □

4.4. **The Taylor-Lagrange formula**

The second version of Taylor's formula that we now present demands a little more regularity of the considered function $f : U \to F$ than the Taylor Young's formula of Theorem 4.4, but it offers a better understanding of the remainder term.

Theorem 4.5. *Let $(E, ||\cdot||_E)$ and $(F, ||\cdot||_F)$ be two normed vector spaces, and let $U \subset E$ be an open subset. Let $f : U \to F$ be a function which is $(n+1)$ times differentiable on U. Let $x \in U$ and $h \in E$ be so that the whole segment $\{x + th, \ \ t \in [0, 1]\}$ is contained in U. Then, it holds:*

$$\left|\left|f(x + h) - \sum_{p=0}^{n} \frac{1}{p!} \mathrm{d}^p f_x(h, \dots, h)\right|\right|_F \leq \frac{M}{(n+1)!} ||h||_E^{n+1}, \tag{4.9}$$

where $M > 0$ is any bound such that

$$\sup_{t \in [0,1]} ||\mathrm{d}^{n+1} f_{x+th}||_{\mathcal{L}(E^{n+1};F)} \leq M.$$

Proof. Let us define the functions $g, m : [0, 1] \to F$ by

$$\forall t \in [0, 1], \quad g(t) = f(x + th), \text{ and } m(t) = \sum_{p=0}^{n} \frac{(1-t)^p}{p!} g^{(p)}(t).$$

Since f is $(n+1)$ times differentiable on U, m is differentiable on $(0, 1)$, and a simple calculation yields:

$$\begin{aligned} \forall t \in (0, 1), \quad m'(t) &= -\sum_{p=1}^{n} \frac{(1-t)^{p-1}}{(p-1)!} g^{(p)}(t) + \sum_{p=0}^{n} \frac{(1-t)^p}{p!} g^{(p+1)}(t) \\ &= -\sum_{p=0}^{n-1} \frac{(1-t)^p}{p!} g^{(p+1)}(t) + \sum_{p=0}^{n} \frac{(1-t)^p}{p!} g^{(p+1)}(t) \\ &= \frac{(1-t)^n}{n!} g^{(n+1)}(t). \end{aligned}$$

The calculation conducted in Example 4.3 allows to estimate the $(n+1)^{\text{th}}$ order derivative of g as:

$$\forall t \in (0, 1), \quad ||g^{(n+1)}(t)||_F \leq \left(\sup_{t \in [0,1]} ||\mathrm{d}^{n+1} f_{x+th}||_{\mathcal{L}(E^{n+1};F)} \right) ||h||_E^{n+1} \leq M ||h||_E^{n+1},$$

and so

$$\forall t \in (0, 1), \quad ||m'(t)||_F \leq \varphi'(t), \text{ where } \varphi(t) = \frac{(1-t)^{n+1}}{(n+1)!} M ||h||_E^{n+1}.$$

Now applying the Mean Value Theorem 4.1, we obtain that:

$$||m(1) - m(0)||_F \leq \frac{M}{(n+1)!}||h||_E^{n+1},$$

which is exactly the desired estimate (4.9). $\square$

Remark 4.2. *In the setting of Theorem 4.5, when the arrival space F is $\mathbb{R}$, a slightly more precise result holds about the remainder term. Indeed, applying the Mean Value Theorem 1.1 for real-valued functions in the above proof instead of the generalized version Theorem 4.1 leads to the following conclusion: there exists a real value $s \in (0,1)$ such that:*

$$f(x+h) = \sum_{p=0}^{n} \frac{1}{p!} f^{(p)}(x) h^p + \frac{1}{(n+1)!} f^{(n+1)}(x+sh) h^{n+1}.$$

4.5. The integral version of Taylor's formula

We eventually discuss yet another avatar of Taylor's formula, which requires a little more regularity than the previous ones, but where the remainder term is explicit: it takes the form of an integral involving the derivatives of the considered function f. For this reason, the result only holds when the arrival space F is a Banach space, so that the construction of Section 3.5 of the integral of F-valued functions be available.

Theorem 4.6 (Taylor's formula with integral rest)**.** *Let $(E, ||\cdot||_E)$ be a normed vector space, $(F, ||\cdot||_F)$ be a Banach space and let $U \subset E$ be open. Let $f : U \to F$ be a function of class $\mathcal{C}^{n+1}$ for some $n \geq 0$. Then, for any point $x \in U$ and vector $h \in E$ such that the whole segment $\{x + th,\ \ t \in [0,1]\}$ is contained in U, the following expansion holds:*

$$f(x+h) = \sum_{p=0}^{n} \frac{1}{p!} \mathrm{d}^p f_x(h, \dots, h) + \int_0^1 \frac{(1-t)^n}{n!} \mathrm{d}^{n+1} f_{x+th}(h, \dots, h)\, \mathrm{d}t. \tag{4.10}$$

Proof. The proof is simple and useful, insofar as it allows to retrieve the correct formula without any risk to get mixed up with indices. Let us define the function $g : [0,1] \to F$ by $g(t) = f(x+th)$; we have seen in Example 4.3 that g is of class $\mathcal{C}^{n+1}$ on $(0,1)$ and that for any $1 \leq p \leq n+1$, it holds

$$\forall t \in (0,1), \quad g^{(p)}(t) = \mathrm{d}^p f_{x+th}(h, \dots, h).$$

Hence, to prove the theorem, it is enough to verify that for any function $g : [0,1] \to F$ of class $\mathcal{C}^{n+1}$ on $(0,1)$, one has:

$$g(1) = \sum_{p=0}^{n} \frac{1}{p!} g^{(p)}(0) + \int_0^1 \frac{(1-t)^n}{n!} g^{(n+1)}(t)\, \mathrm{d}t, \tag{4.11}$$

which we prove by induction on $n \geq 0$.

- *Case where $n = 0$:* If $g : [0,1] \to \mathbb{R}$ is a function of class $\mathcal{C}^1$ on $(0,1)$, the properties of the integral directly imply that

$$g(1) = g(0) + \int_0^1 g'(t)\, \mathrm{d}t,$$

  see Proposition-Definition 3.2 (iii). This equality is exactly the desired formula (4.11) in the case where $n = 0$.
- Let $n \geq 0$ be given, and let $g : [0,1] \to F$ be a function of class $\mathcal{C}^{n+2}$ on $(0,1)$. Since g is in particular of class $\mathcal{C}^{n+1}$ on $(0,1)$, the induction hypothesis implies that:

$$g(1) = \sum_{p=0}^{n} \frac{1}{p!} g^{(p)}(0) + \int_0^1 \frac{(1-t)^n}{n!} g^{(n+1)}(t)\, \mathrm{d}t, \tag{4.12}$$

Since $g^{(n+1)}$ is a function of class $\mathcal{C}^1$ on $(0,1)$, an integration by parts in the last integral in the above right-hand side yields

$$\begin{aligned}\int_0^1 \frac{(1-t)^n}{n!} g^{(n+1)}(t)\,\mathrm{d}t &= \left[-\frac{(1-t)^{n+1}}{(n+1)!} g^{(n+1)}(t)\right]_0^1 - \int_0^1 -\frac{(1-t)^{n+1}}{(n+1)!} g^{(n+2)}(t)\,\mathrm{d}t \\ &= \frac{1}{(n+1)!} g^{(n+1)}(0) + \int_0^1 \frac{(1-t)^{n+1}}{(n+1)!} g^{(n+2)}(t)\,\mathrm{d}t.\end{aligned}$$

Injecting this expression into (4.12), we obtain the desired formula (4.10) at the order $(n+1)$.

□

4.6. **Exercises**

Exercise 4.1 (Proof of Point *(ii)* of Proposition 2.2). Let $(E_1, ||\cdot||_{E_1})$, $(E_2, ||\cdot||_{E_2})$ and $(F, ||\cdot||_F)$ be three normed vector spaces, and let $U \subset E_1$, $V \subset E_2$ be open subsets. Let $f : U \times V \to F$ be a function and let $(x_0, y_0) \in U \times V$. We assume that there exist open neighborhoods $U' \subset U$ and $V' \subset V$ of x_0 and y_0 respectively such that the partial derivatives

$$U' \times V' \ni (x,y) \mapsto \frac{\partial f}{\partial x}(x,y) \in \mathcal{L}(E_1; F) \text{ and } U' \times V' \ni (x,y) \mapsto \frac{\partial f}{\partial y}(x,y) \in \mathcal{L}(E_2; F)$$

e Show that f is Fréchet differentiable at (x_0, y_0) and that its derivative reads:

$$\forall (h,k) \in E_1 \times E_2, \quad \mathrm{d}f_{(x_0,y_0)}(h,k) = \frac{\partial f}{\partial x}(x_0,y_0)(h) + \frac{\partial f}{\partial y}(x_0,y_0)(k).$$

[Hint: For $h \in E_1$ and $k \in E_2$ small enough, write:

$$\begin{aligned}&f(x_0+h, y_0+k) - f(x_0,y_0) - \frac{\partial f}{\partial x}(x_0,y_0)(h) - \frac{\partial f}{\partial y}(x_0,y_0)(k) = \\ &\quad \left(f(x_0+h, y_0+k) - f(x_0, y_0+k) - \frac{\partial f}{\partial x}(x_0,y_0)(h)\right) + \left(f(x_0, y_0+k) - f(x_0,y_0) - \frac{\partial f}{\partial y}(x_0,y_0)(k)\right).\end{aligned}$$

The second parenthesis in the above right-hand side can be treated directly thanks to the definition of the partial differentiability of f with respect to the second variable. As for the first parenthesis, consider the function g, defined on a small enough ball $B(0,\delta) \subset E_1$ into F by

$$g(h) = f(x_0+h, y_0) - \frac{\partial f}{\partial x}(x_0,y_0)(h)$$

and apply Corollary 4.1 with the continuity of the partial derivative $(x,y) \mapsto \frac{\partial f}{\partial x}(x,y) \in \mathcal{L}(E_1; F)$.]

Exercise 4.2 (Trapeze integration formula). Let $(E, ||\cdot||_E)$ be a Banach space and let $f : [a,b] \to E$ be a function of class $\mathcal{C}^2$. Show that there exists a constant $C > 0$ which does not depend on f such that

$$\left|\left|\int_a^b f(t)\,\mathrm{d}t - (b-a)\frac{f(a)+f(b)}{2}\right|\right|_E \le C \sup_{s\in[a,b]} ||f''(s)||_E (b-a)^3.$$

Exercise 4.3. Let $(E, ||\cdot||_E)$ be a Banach space, and let $U \subset E$ be open. Let $f : U \to \mathbb{R}$ be differentiable, and let $C \subset U$ be a convex subset.

(i) Show that f is convex on C if and only if

$$\forall x, y \in C, \quad f(y) \ge f(x) + \mathrm{d}f_x(y-x).$$

(ii) Show that, if f is convex on C, one has, for all x, $y \in C$ and for all $0 < s < t < 1$,

$$\frac{f(x+s(y-x)) - f(x)}{s} \le \frac{f(x+t(y-x)) - f(x)}{t}.$$

(iii) Show that f is strictly convex on C if and only if

$$\forall x, y \in C, \quad f(y) > f(x) + \mathrm{d}f_x(y-x).$$

(iv) We now assume that f is twice differentiable on U. Show that f is convex on C if and only if

$$\forall x, y \in C, \quad \mathrm{d}^2 f_x(y-x, y-x) \ge 0.$$

(v) Show that f is strictly convex on C if and only if

$$\forall x, y \in C, \quad \mathrm{d}^2 f_x(y-x, y-x) > 0.$$

Exercise 4.4 (Jacobi's identity). Let $(E, ||\cdot||_E)$ be a normed vector space, and let $U \subset E$ be open. Let u, $v : U \to E$ be two functions of class $\mathcal{C}^2$.

(i) Show that the mapping $g : U \to E$ defined by

$$g(x) = \mathrm{d}v_x(u(x)) - \mathrm{d}u_x(v(x))$$

is of class $\mathcal{C}^1$ on U. This function is denote by $g := [u, v]$ in the following.

(ii) For any functions u, v, $w : U \to E$ of class $\mathcal{C}^2$, we define

$$\Phi(u, v, w)(x) = \mathrm{d}^2 u_x(v(x), w(x)) - \mathrm{d}v_x(\mathrm{d}u_x(w(x))) - \mathrm{d}w_x(\mathrm{d}u_x(v(x))).$$

Show that Φ is trilinear and symmetric in its arguments u, v and w.

(iii) Calculate $[[u, v], w]$ in terms of $\Phi(u, v, w)$ and $\Phi(v, u, w)$.

(iv) Deduce from the result that

$$[[u, v], w] + [[v, w], u] + [[w, u], v] = 0.$$

Exercise 4.5. We recall that the length of a curve γ is defined by

$$\ell(\gamma) = \int_0^1 |\gamma'(t)| \,\mathrm{d}t.$$

We define

$$F(u) = \ell(u \circ \gamma).$$

Show that F is Fréchet differentiable at $u = 0$, and that its Fréchet derivative reads:

$$\mathrm{d}F_0(h) = \int_0^1 \left\langle e(h(\gamma(t)))\frac{\gamma'(t)}{|\gamma\prime(t)|}, \frac{\gamma'(t)}{|\gamma\prime(t)|} \right\rangle |\gamma\prime(t)| \,\mathrm{d}t$$

Exercise 4.6. Let $U \subset \mathbb{R}^3$ be a connected open set, and let $u : U \to \mathbb{R}^3$ be a vector field of class $\mathcal{C}^2$ with null strain, i.e. $e(u) = \frac{1}{2}\left(\nabla u + \nabla u^T\right) = 0$ on U.

(i) Show that all the second-order derivatives of u vanish on U, i.e.

$$\forall i, j, k = 1, \dots, 3, \quad \frac{\partial^2 u_k}{\partial x_i \partial x_j} = 0.$$

(ii) Infer from the previous question and one of the Taylor's formulas that there exist an antisymmetric matrix $M \in M_3(\mathbb{R})$ and a vector $b \in \mathbb{R}^3$ such that

$$\forall x \in \mathbb{R}^3, \quad u(x) = Mx + b.$$

5. The fixed point theorem and some applications

This section is perhaps a little more conceptual than the previous, calculus-oriented ones. It is devoted to the so-called Banach fixed point theorem, sometimes referred to as the Contraction Mapping Principle, which is a central result in the theory of complete spaces. We present this theorem and some of its variants in Section 5.1, before turning to two (among many) interesting illustrations, in the analysis of the Newton-Raphson method, and the proof of the Cauchy-Lipschitz theorem, in Sections 5.2 and 5.3 respectively.

5.1. Statement and proof of the fixed point theorem

Let us start with a definition.

Definition 5.1. *Let X be a closed subset of a Banach space $(E, ||\cdot||_E)$. A function $T : X \to X$ is called a contraction if it is Lipschitz continuous with ratio $k < 1$, that is*

$$\forall x, y \in X, \quad ||T(x) - T(y)||_E \leq k||x - y||_E. \tag{5.1}$$

We now state and prove the Banach fixed point theorem in the context of a closed subset of a Banach space, however the result holds true in the more general context of a complete metric space.

Theorem 5.1. *Let X be a closed subset of a Banach space $(E, ||\cdot||_E)$, and let $T : X \to X$ be a contraction. Then T admits a unique fixed point $x^* \in X$, i.e. a point $x^* \in X$ such that*

$$T(x^*) = x^*.$$

The latter is the limit of any sequence of the form $\{T^n(x)\}_{n\geq 0}$, where $x \in X$ is arbitrary.

Proof. Let us first observe that T has at most one fixed point. Indeed, if $x, y \in X$ are two points such that $T(x) = x$ and $T(y) = y$, it follows from the contraction property (5.1) that

$$||x - y||_E = ||T(x) - T(y)||_E \leq k||x - y||_E;$$

since $k < 1$, this imposes that $x = y$.

Let us now turn to the existence of a fixed point for T. As suggested by the statement, taking any $x \in X$, we shall prove that the sequence of iterates $T^n(x)$ is a Cauchy sequence in X, and thus converges to some $x^* \in X$. We shall then see that this limit x^* is one fixed point of T.

Let us then consider a point $x \in X$; a simple calculation shows that, for all n, $p \geq 0$,

$$\begin{array}{rcl} ||T^{n+p}(x) - T^n(x)||_E & \leq & k^n||T^p(x) - z||_E \\ & \leq & k^n \sum\limits_{i=0}^{p-1} ||T^{i+1}(x) - T^i(x)||_E \\ & \leq & k^n \sum\limits_{i=0}^{p-1} k^i||T(x) - (x)||_E \\ & \leq & k^n \sum\limits_{i=0}^{\infty} k^i||T(x) - (x)||_E \\ & = & \frac{k^n}{1-k}||T(x) - (x)||_E. \end{array}$$

Since $k < 1$, it follows from this estimate that the sequence $T^n(x)$ is a Cauchy sequence of elements of X; as E is complete and X is closed, it thus converges to an element $x^* \in X$, as desired.

We are left to show that x^* is one fixed point of T. Since T is a continuous function, we see that

$$T(x^*) = T\left(\lim_{n\to\infty} T^n(x)\right) = \lim_{n\to\infty} T^n(T(x)) = \lim_{n\to\infty} T^{n+1}(x) = x^*,$$

which terminates the proof. □

Remark 5.1. *In a practical application where T is differentiable on the Banach space E, the contraction property (5.1) required by the Banach fixed point Theorem 5.1 is often established by estimating the differential of the function T, and applying the mean value Theorem 4.1.*

Remark 5.2. *The fixed point Theorem 5.1 lends itself to an interesting algorithmic implementation, which is ubiquitous in numerical analysis. Under its assumptions, the unique fixed point $x^* \in X$ of T can be calculated as the limit of the following iterative procedure:*

- *Initialization: Let $x_0 = x$ be any point in X;*
- *For $n = 0, \ldots$: The next iterate x_{n+1} is obtained by applying T to x_n: $x_{n+1} = T(x_n)$.*

It is even possible to quantify the convergence rate of this method:

$$\forall n \geq 0, \quad ||x_n - x^*||_E \leq \frac{k^n}{1-k}||T(x) - x||_E,$$

see Exercise 5.1.

The Banach fixed point Theorem 5.1 is often applied via the following strengthened version, which is a mere corollary of the above statement.

Corollary 5.1. *Let X be a closed subset of a Banach space $(E, ||\cdot||_E)$, and let $T : X \to X$ be a function. Assume that there exists $n \geq 1$ such that T^n is a contraction; then T has a unique fixed point in X.*

Proof. The function T^n being a contraction, it has a unique fixed point $x^* \in X$ by virtue of Theorem 5.1. Hence, if $x \in X$ is a fixed point of T, it holds $T(x) = x$, and so $T^n(x) = x$. This implies that $x = x^*$, which reveals that T has at most one fixed point in X.

On the other hand, the fixed point x^* of T^n satisfies $T^n(x^*) = x^*$, and so

$$T(T^n(x^*)) = T^n(T(x^*)) = T(x^*).$$

Hence, $T(x^*)$ is another fixed point of T^n. Since there exists exactly one such fixed point, one has necessarily $T(x^*) = x^*$. We have proved that x^* is also a fixed point of T, which completes the proof. $\square$

We now state without proof a stronger version of the above fixed point Theorem 5.1, where the contraction operator now reads $T : \Lambda \times X \to X$: T additionally depends on a parameter $\lambda \in \Lambda$, and under suitable assumptions, the mapping $T(\lambda, \cdot) : X \to X$ has a unique fixed point $x^*(\lambda) \in X$ for any given value λ. The main conclusion of the next statement is that the mapping $\lambda \mapsto x^*(\lambda)$ is "as regular" as the mapping T itself.

Theorem 5.2. *Let X be a closed subset of a Banach space $(E, ||\cdot||_E)$, and let Λ be an open subset of another Banach space $(F, ||\cdot||_F)$. We consider the fixed point equation*

$$T(\lambda, x) = x, \tag{5.2}$$

where the function $T : \Lambda \times X \to X$ is of class $\mathcal{C}^p$ for some $p \geq 0$. We assume that there exists a real number $k < 1$ such that for all $\lambda \in \Lambda$, $x \mapsto T(\lambda, x)$ is a contraction with ratio k. Then for all $\lambda \in \Lambda$, (5.2) has a unique solution $x^*(\lambda) \in X$, and the mapping

$$\Lambda \ni \lambda \mapsto x^*(\lambda) \in X$$

is of class $\mathcal{C}^p$.

5.2. Application I: convergence of the Newton-Raphson algorithm

In this section, we briefly discuss the Newton-Raphson method. The latter is a fundamental idea for finding zeroes of a function f, defined from a Banach space E into itself. An elementary convergence proof of this method can be achieved by means of the fixed point Theorem 5.1. Although the conclusion is not optimal, the material of this section is a good application example of this key idea.

Let U be an open subset of a Banach space $(E, ||\cdot||_E)$, and let $f : U \to E$ be a function. We search for zeroes of f, that is, points $x^* \in E$ such that

$$f(x^*) = 0. \tag{5.3}$$

Let us first provide a formal, non rigorous sketch of the Newton-Raphson method. We assume that f has a zero x^* that we aim to find. The Newton-Raphson method is a local, iterative procedure to achieve this goal: starting from a point $x_0 \in U$ which is "close enough" from x^*, it produces a sequence x_n, $n = 0, \dots$ of iterates which are hopefully closer and closer to x^*. The first iterate x_1 is sought under the form $x_1 = x_0 + h$, where h is a "small perturbation"; the latter is characterized by the requirement that it should solve the linearized version of the equation $f(x_0 + h) = 0$ at x_0, that is:

$$f(x_0) + \mathrm{d}f_{x_0}(h) = 0,$$

where we recall that the differential $\mathrm{d}f_{x_0} : E \to E$ of f at x_0 is a continuous linear mapping. Assuming that $\mathrm{d}f_{x_0}$ is invertible, the above equation has a unique solution $h = -\mathrm{d}f_{x_0}^{-1}(f(x_0))$, so that x_1 equals:

$$x_1 = x_0 - \mathrm{d}f_{x_0}^{-1}(f(x_0)).$$

Iterating this procedure we get the sketch described in Algorithm 1.

Algorithm 1 The Newton-Raphson method.

---

initialization: Point $x_0 \in U$ "close enough" from x^*.
for $n = 0, ...,$ until convergence: **do**
  $x_{n+1} = T(x_n)$, where $T(x) = x - \mathrm{d}f_x^{-1}(f(x))$.
end for
return x_n

Obviously, the well-posedness of this strategy hinges on a number of assumptions. In particular, f has to be differentiable at all the iterates x_n, and the differential $\mathrm{d}f_{x_n} : E \to E$ has to be an invertible mapping. Precisely, the convergence of the method is guaranteed by the following theorem.

Theorem 5.3. *Let U be an open subset of a Banach space $(E, ||\cdot||_E)$, and let $f : U \to E$ be a function of class $\mathcal{C}^2$. Assume that there exists a point $x^* \in U$ such that*

$$f(x^*) = 0, \text{ and the differential } \mathrm{d}f_{x^*} \in \mathcal{L}(E;E) \text{ is an isomorphism.}$$

Then there exists $\delta > 0$ such that for all $x_0 \in B(x^*, \delta)$, the sequence x_n of iterates produced by the Newton-Raphson Algorithm 1 is well-defined, and converges to x^*.

Proof. Our first task is to verify that the mapping T featured in the definition of the Newton-Raphson Algorithm 1 is well-defined. To achieve this, we use the fact that f is, in particular, a function of class $\mathcal{C}^1$ on U, so that the mapping

$$U \ni x \mapsto \mathrm{d}f_x \in \mathcal{L}(E;E)$$

is continuous. Since $\mathrm{d}f_{x^*} \in \mathcal{L}(E;E)$ is invertible and since the subset of invertible mappings of $\mathcal{L}(E;E)$ is open (a fact which is proved thanks to Neumann series, see Exercise 1.8), there exists $\delta > 0$ such that for all x in the closed set $X := \overline{B(x^*, \delta)}$, the mapping $\mathrm{d}f_x$ is also invertible. We may then define the mapping $T : X \to E$ by the formula

$$\forall x \in X, \quad T(x) = x - \mathrm{d}f_x^{-1}(f(x)),$$

which is of class $\mathcal{C}^1$ on a neighborhood of X because f is of class $\mathcal{C}^2$. With this definition, the sought zero x^* of f obviously satisfies the fixed point equation $x^* = T(x^*)$. We shall then verify that, up to decreasing the value of δ (and updating accordingly the definition of the set X), T is a contraction mapping from X into itself. The fixed point Theorem 5.1 will then yield the conclusion of the present theorem.

Let us calculate, for $h \in E$, $||h||_E \leq \delta$:

$$\begin{array}{rcl} T(x^*+h) - T(x^*) & = & h - \mathrm{d}f^{-1}_{x^*+h}(f(x^*+h)) \\ & = & h - \mathrm{d}f^{-1}_{x^*+h}(f(x^*+h) - f(x^*)) \\ & = & h - \mathrm{d}f^{-1}_{x^*+h}(\mathrm{d}f_{x^*}(h)) - \mathrm{d}f_{x^*+h}(r_f(h)) \\ & = & h - \mathrm{d}f^{-1}_{x^*+h}(\mathrm{d}f_{x^*+h}(h)) + \mathrm{d}f^{-1}_{x^*+h}(\mathrm{d}f_{x^*+h}(h) - \mathrm{d}f_{x^*}(h)) - \mathrm{d}f_{x^*+h}(r_f(h)) \end{array} \tag{5.4}$$

where we have introduced the first-order expansion of f at x^* to pass from the second to the third line of the above series of equalities:

$$f(x^*+h) = f(x^*) + \mathrm{d}f_{x^*}(h) + r_f(h), \text{ where } r_f(h) \to 0 \text{ as } h \to 0.$$

Using the fact that f is twice differentiable at x^* together with Lemma 4.2, we also obtain the following estimate:

$$\mathrm{d}f_{x^*+h}(h) - \mathrm{d}f_{x^*}(h) = \mathrm{o}(||h||_E),$$

and since the mapping $X \in x \mapsto \mathrm{d}f_x^{-1} \in \mathcal{L}(E;E)$ is continuous (again, see Exercise 1.8, we obtain:

$$\mathrm{d}f^{-1}_{x^*+h}(\mathrm{d}f_{x^*+h}(h) - \mathrm{d}f_{x^*}(h)) = \mathrm{o}(||h||_E).$$

Eventually, (5.4) rewrites:

$$\begin{array}{rcl} T(x^*+h) - T(x^*) & = & h - \mathrm{d}f^{-1}_{x^*+h}(\mathrm{d}f_{x^*+h}(h)) + \mathrm{o}(||h||_E) \\ & = & h - h + \mathrm{o}(||h||_E) \\ & = & \mathrm{o}(||h||_E). \end{array}$$

As a result, we see that the differential of T at x^* reads $\mathrm{d}T_{x^*} = 0$. Since the mapping $X \ni x \mapsto \mathrm{d}T_x \in \mathcal{L}(E;E)$ is continuous, up to decreasing the value of δ, we have:

$$\forall x \in X, \quad ||\mathrm{d}T_x||_{\mathcal{L}(E;E)} \leq \frac{1}{2}.$$

We now apply Corollary 4.1 of the mean value theorem, which reveals that:

$$\forall x, y \in X, \quad ||T(x) - T(y)||_E \leq \frac{1}{2}||x-y||_E. \tag{5.5}$$

The first conclusion of this estimate, obtained by letting $y = x^* = T(x^*)$ in the above estimate, is that:

$$\forall x \in \overline{B(x^*, \delta)}, \quad ||T(x) - x^*||_E \leq \frac{1}{2}||x - x^*||_E \leq \delta,$$

and so T maps X into itself. In addition, (5.5) expresses the fact that T is a contraction of X into itself, with ratio $\frac{1}{2}$. It follows from the fixed point Theorem 5.1 that T has a unique fixed point in X, which is

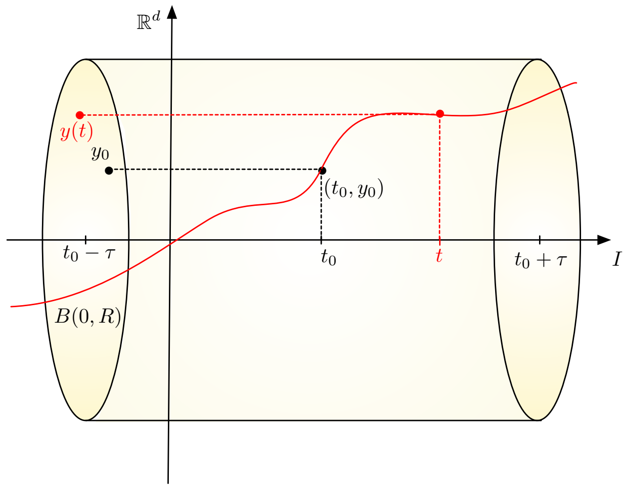

FIGURE 11. *Any solution $t \mapsto y(t)$ to the ordinary differential equation (5.6) (red line) stays in a security cylinder.*

necessarily x^*. The latter is the limit of any sequence of the form $T^n(x)$, for $x \in X$, which is the desired result. □

Remark 5.3. *As we have hinted at, the analysis of this section is not optimal: the convergence rate of x_n to x^* supplied by the analysis, see Remark 5.2 and Exercise 5.1, is actually "weaker" than that provided by a more careful analysis, see §6.1.2 in [2] about this matter.*

5.3. Application II: the Cauchy-Lipschitz theorem

In this section, we make a short foray into the field of differential equations; our purpose is to illustrate another use of the fixed point Theorem 5.1, as the pivotal ingredient of the proof of the famous Cauchy-Lipschitz theorem. We refer to the monograph [6] for a much more in-depth treatment of ordinary differential equations.

The situation of interest in this section is the following: let I be an open interval of $\mathbb{R}$, $d \geq 1$ and let $U \subset \mathbb{R}^d$ be an open set. Let $f : I \times U \to \mathbb{R}^d$ be a continuous function; for a given couple $(t_0, y_0) \in I \times U$ of initial data, we consider the ordinary differential equation:

$$\left\{ \begin{array}{l} y'(t) = f(t, y(t)), \\ y(t_0) = y_0. \end{array} \right. \tag{5.6}$$

Let us first make precise what we intend by solving this equation.

Definition 5.2. *A solution to the ordinary differential equation (5.6) on an open interval $J \subset I$ containing t_0 is a function $y : J \to U$ of class $\mathcal{C}^1$ such that:*

$$y(t_0) = y_0, \text{ and } \forall t \in J, \; y'(t) = f(t, y(t)).$$

We now introduce the notion of *security cylinder* attached to the differential equation (5.6), which roughly speaking accounts for a combination of a subinterval of I (in time) and a subregion of U (in space) such that any solution to (5.6) over the given time region stays inside the space region, see Fig. 11.

Definition 5.3. *Let $(t_0, y_0) \in I \times U$ be given; a security cylinder around (t_0, y_0) for the equation (5.6) is a set of the form $C = [t_0 - \tau, t_0 + \tau] \times \overline{B(y_0, R)}$, for some $\tau > 0$ and $R > 0$, which enjoys the following property: any solution to (5.6) on an open interval $J \subset [t_0 - \tau, t_0 + \tau]$ satisfies:*

$$\forall t \in J, \quad y(t) \in \overline{B(y_0, R)}.$$

This definition is not empty owing to the following lemma.

Lemma 5.1. *For any $(t_0, y_0) \in I \times U$, there exist security cylinders around (t_0, y_0) attached to the differential equation (5.6).*

Proof. Let $R > 0$ be such that $\overline{B(y_0, R)} \subset U$. Let $J \subset I$ be a time interval and let $y : J \to U$ be any solution to the equation (5.6) on J. Then,

$$\forall t \in J, \quad y(t) - y_0 = \int_{t_0}^{t} y'(s)\,\mathrm{d}s = \int_{t_0}^{t} f(s, y(s))\,\mathrm{d}s,$$

and so

$$\forall t \in J, \;\; |y(t) - y_0| \le |t - t_0| M \le \tau M, \;\text{ where we have defined } M := \sup_{(t,y) \in C} |f(t, y)|.$$

Hence, choosing τ so small that $\tau \le \frac{R}{M}$, the claim follows. $\square$

The Cauchy-Lipschitz theorem is our next statement.

Theorem 5.4 (The Cauchy-Lipschitz theorem). *Let I be an open interval of $\mathbb{R}$ and let $U \subset \mathbb{R}^d$ be an open set. Assume additionally that the continuous function $f : I \times U \to \mathbb{R}^d$ is locally Lipschitz continuous in the second variable: for all $(t_0, y_0) \in I \times U$, there exist open neighborhoods $J \subset I$ and $V \subset U$ of t_0 and y_0 respectively, and a constant $k > 0$ such that*

$$\forall t \in J, \;\forall y_1, y_2 \in V, \quad |f(t, y_1) - f(t, y_2)| \le k|y_1 - y_2|.$$

Let $(t_0, y_0) \in I \times U$ be arbitrary, and let $C = [t_0 - \tau, t_0 + \tau] \times \overline{B(y_0, R)}$ be an associated security cylinder. Then, the differential equation (5.6) has a unique solution y on $(t_0 - \tau, t_0 + \tau)$.

Remark 5.4. *By a compactness argument based on the Borel-Lebesgue property of Theorem 1.3, the locally Lipschitz character of f is equivalent to the following fact: for any compact subsets $A \subset I$ and $K \subset U$, there exists a constant $k > 0$ such that:*

$$\forall t \in A, \;\forall y_1, y_2 \in K, \quad |f(t, y_1) - f(t, y_2)| \le k|y_1 - y_2|.$$

Proof. At first, let us observe with Remark 5.4 that there exists a constant $k > 0$ such that the following estimate holds:

$$\forall t \in [t_0 - \tau, t_0 + \tau], \;\forall y_1, y_2 \in \overline{B(y_0, R)}, \quad |f(t, y_1) - f(t, y_2)| \le k|y_1 - y_2|. \tag{5.7}$$

The first step in the proof consists in giving the ordinary differential equation (5.6) the equivalent form of an integral equation. It is immediate to verify that a function y is solution to (5.6) on $(t_0 - \tau, t_0 + \tau)$ if and only if the following three conditions hold:

(i) The mapping $t \mapsto y(t)$ is continuous on $(t_0 - \tau, t_0 + \tau)$;
(ii) For all $t \in (t_0 - \tau, t_0 + \tau)$, $y(t) \in U$;
(iii) y satisfies the following equation:

$$\forall t \in (t_0 - \tau, t_0 + \tau), \quad y(t) = y_0 + \int_{t_0}^{t} f(s, y(s))\,\mathrm{d}s. \tag{5.8}$$

We now solve (5.8) by applying a fixed point strategy to an adequate mapping, defined on a suitable set. More precisely, let E be the space of continuous functions $y : [t_0 - \tau, t_0 + \tau] \to \mathbb{R}^d$ equipped with the norm:

$$||y||_E := \sup_{t \in [t_0 - \tau; t_0 + \tau]} |y(t)|.$$

We have already proved that E is a Banach space, see Exercise 1.7. Let us also define the closed subset of E

$$X = \{y \in E, \;\forall t \in [t_0 - \tau, t_0 + \tau], \;|y(t)| \le R\}\,.$$

Introducing the mapping $T : X \to E$ given by

$$\forall y \in X, \;\forall t \in [t_0 - \tau, t_0 + \tau], \quad T(y)(t) = y_0 + \int_{t_0}^{t} f(s, y(s))\,\mathrm{d}s;$$

the fact that $C = [t_0 - \tau, t_0 + \tau] \times \overline{B(y_0, R)}$ is a security cylinder for the equation (5.6) and the data (t_0, y_0) shows that T actually maps the set X into itself.

In order to complete the proof, we eventually show that the composition T^n is a contraction mapping for a certain integer $n \geq 1$, which will allow us to apply Corollary 5.1. To achieve this, let $y_1, y_2 \in X$ be two functions; we estimate:

$$\forall t \in [t_0 - \tau, t_0 + \tau], \quad T(y_1)(t) - T(y_2)(t) = \int_{t_0}^{t} \Big(f(s, y_1(s)) - f(s, y_2(s)) \Big) \, \mathrm{d}s,$$

and so

$$\forall t \in [t_0 - \tau, t_0 + \tau], \quad |T(y_1)(t) - T(y_2)(t)| \leq \int_{t_0}^{t} |f(s, y_1(s)) - f(s, y_1(s))| \, \mathrm{d}s \leq |t - t_0| k ||y_1 - y_2||_E,$$

where we have used the Lipschitz estimate (5.7). Iterating this calculation, we obtain

$$\begin{aligned} \forall t \in [t_0 - \tau, t_0 + \tau], \quad |T^2(y_1)(t) - T^2(y_2)(t)| &\leq \int_{t_0}^{t} |f(s, T(y_1)(s)) - f(s, T(y_2)(s))| \, \mathrm{d}s \\ &\leq \int_{t_0}^{t} k^2 |s - t_0| ||y_1 - y_2||_E \, \mathrm{d}s \\ &= \frac{k^2 ||y_1 - y_2||_E |t - t_0|^2}{2}. \end{aligned}$$

By an easy induction argument, we obtain that, for all $n \geq 1$, it holds

$$\forall t \in [t_0 - \tau, t_0 + \tau], \quad |T^n(y_1)(t) - T^n(y_2)(t)| \leq \frac{k^n |t - t_0|^n}{n!} ||y_1 - y_2||_E,$$

and so

$$||T^n(y_1) - T^n(y_2)||_E \leq \frac{k^n \tau^n}{n!} ||y_1 - y_2||_E.$$

Since $\frac{k\tau^n}{n!} < 1$ for n sufficiently large, Corollary 5.1 allows to conclude the proof of the theorem. □

5.4. **Exercises**

Exercise 5.1. Let X be a closed subset of a Banach space $(E, ||\cdot||_E)$ and let $T : X \to X$ be a contraction with ratio $k < 1$. Let $x \in X$ be arbitrary, and define the sequence $x_n \in X$ by:

$$x_0 = x, \text{ and } \forall n \geq 0, \; x_{n+1} = T(x_n).$$

Show that the following estimate holds:

$$\forall n \geq 0, \quad ||x_n - x^*||_E \leq \frac{k^n}{1-k} ||T(x) - x||_E,$$

where $x^* \in X$ is the unique fixed point of T.

[Hint: Adapt the proof of Theorem 5.1.]

Exercise 5.2. Let $(E, ||\cdot||_E)$ be a Banach space and let $f, g : E \to E$ be two contraction mappings from E into itself. Show that there exist two points x_0, $y_0 \in E$ such that:

$$y_0 = f(x_0) \text{ and } x_0 = g(y_0).$$

6. The implicit function theorem and some of its applications

This section is devoted to the implicit function theorem and its close avatar, the local inverse theorem. These key results in the setting of Banach spaces have countless applications in fields so diverse as differential geometry, ordinary and partial differential equations, etc.

After the short Section 6.1 dedicated to an intuitive and non rigorous introduction to the main ideas, we state the implicit function theorem in Section 6.2 and provide illustrative applications of the latter. The local inverse theorem is then tackled in Section 6.3.

6.1. **Intuitive presentation**

In a nutshell, the implicit function theorem is a key tool for proving the existence, uniqueness and regularity of solutions to non linear equations depending on parameters. The typical application context is as follows: we consider an equation of the form

$$\text{Search for } u \text{ s.t. } \mathcal{F}(\theta, u) = 0, \tag{6.1}$$

where

- The sought variable u belongs to a subset U of a Banach space E;
- The parameter θ belongs to a subset O of another Banach space Θ;
- The mapping $\mathcal{F} : O \times U \to F$ is non linear and takes values in a Banach space F.

Let us glean some intuition about what we should expect regarding the solvability of the equation (6.1) by considering the particular case where all spaces are finite-dimensional, and the mapping $\mathcal{F}$ is linear. The parameter θ belongs to $\mathbb{R}^m$ and the variable u is sought in $\mathbb{R}^n$: for each value $\theta \in \mathbb{R}^m$, we search for the solution $u = u(\theta) \in \mathbb{R}^n$ to the equation

$$\mathcal{F}(\theta, u) = 0.$$

Since $\mathcal{F} : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^n$ is assumed to be linear, it can be written under the form

$$\mathcal{F}(\theta, u) = A\theta + Bu, \tag{6.2}$$

where A is a matrix with size $n \times m$ and B is a matrix with size $n \times n$. In this setting, the resolution of (6.1) is quite simple: assuming that B is an invertible matrix, (6.1) has a unique solution

$$u(\theta) = -B^{-1}A\theta.$$

The implicit function theorem is meant to generalize this pattern to the general context where θ and u belong to infinite-dimensional Banach spaces, and $\mathcal{F}$ is a smooth, possibly non linear mapping. Similar results to those of our model situation hold true, up to the following adaptations:

- The result applies locally, around one particular solution (θ_0, u_0) to the equation (6.1): for θ "close enough" to the reference value θ_0, (6.1) has a unique solution u which is "close enough" to u_0;
- The linearized version of $\mathcal{F}$ with respect to the variable u at the point (θ_0, u_0) has to be a linear isomorphism (this generalizes the above requirement that the block B be invertible in (6.2));

6.2. **Statement of the implicit function theorem**

Let us now state rigorously the implicit function theorem.

Theorem 6.1 (Implicit function theorem)**.** *Let $(\Theta, ||\cdot||_\Theta)$, $(E, ||\cdot||_E)$ and $(F, ||\cdot||_F)$ be three Banach spaces, and let $O \subset \Theta$ and $U \subset E$ be open subsets. Let*

$$\mathcal{F} : O \times U \to F$$

be a mapping of class $\mathcal{C}^k$ for some $k \geq 1$. Let $(\theta_0, u_0) \in O \times U$ satisfy $\mathcal{F}(\theta_0, u_0) = 0$; we assume that the partial Fréchet derivative

$$\frac{\partial \mathcal{F}}{\partial u}(\theta_0, u_0) : E \to F$$

is a linear isomorphism. Then, there exist open subsets $O_0 \subset O$ and $U_0 \subset U$ as well as a mapping $g : O_0 \to U_0$ of class $\mathcal{C}^k$ such that $u_0 = g(\theta_0)$ and

$$\forall \theta \in O_0,\ \ u \in U_0, \quad \mathcal{F}(\theta, u) = 0 \text{ if and only if } u = g(\theta).$$

This result is illustrated on Fig. 12.

Remark 6.1. *The derivative of the mapping $g : O_0 \to U_0$ supplied by the implicit function Theorem 6.1 can be calculated owing to the following observation. We know that both mappings $\mathcal{F} : O \times U \to F$ and $g : O_0 \to U_0$ are of class $\mathcal{C}^k$, and that the following relation holds:*

$$\forall \theta \in O_0, \quad \mathcal{F}(\theta, g(\theta)) = 0.$$

Taking derivatives in this equation and using the chain rule (see Theorem 2.1), we obtain

$$\forall h \in \Theta, \quad \frac{\partial \mathcal{F}}{\partial \theta}(\theta, g(\theta))(h) + \frac{\partial \mathcal{F}}{\partial u}(\theta, g(\theta))\mathrm{d}_\theta g(h) = 0.$$

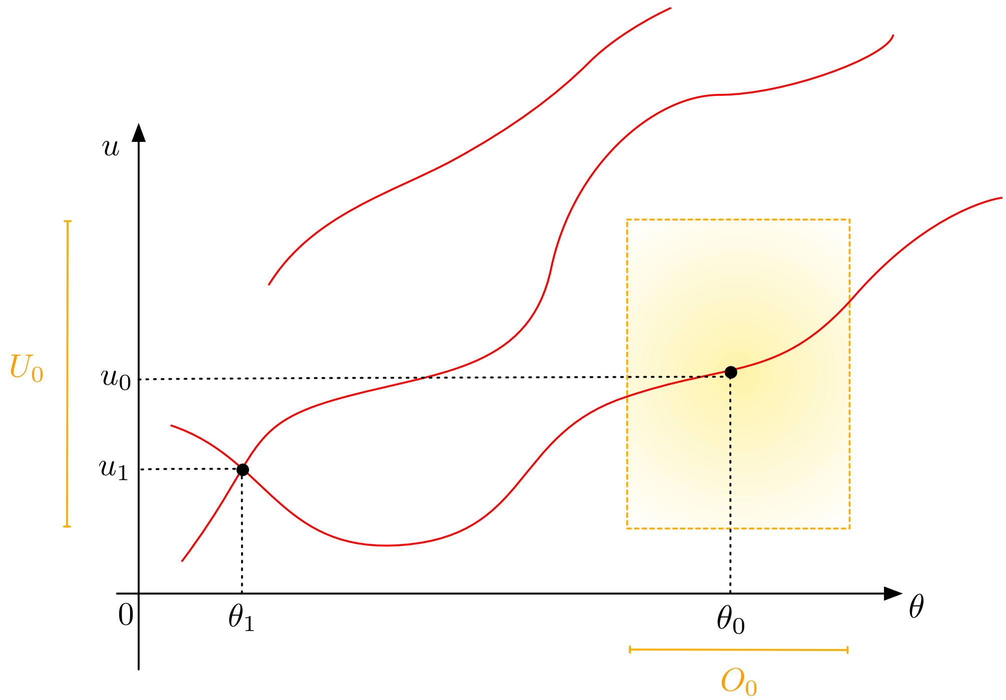

Figure 12. *The implicit function theorem: the set of solutions (θ, u) to (6.1) consists of the red curves in the (θ, u) plane; at (θ_0, u_0), the partial derivative $\frac{\partial \mathcal{F}}{\partial u}(\theta_0, u_0)$ is invertible and for θ close to θ_0, the equation $\mathcal{F}(\theta, u) = 0$ has a unique solution u close enough to u_0; at (θ_1, u_1), $\frac{\partial \mathcal{F}}{\partial u}(\theta_0, u_0)$ is not invertible, and several branches of solutions may meet.*

By the assumption of Theorem 6.1, the partial derivative $\frac{\partial \mathcal{F}}{\partial u}(\theta_0, g(\theta_0)) : E \to F$ is invertible. Since the subset of $\mathcal{L}(E; F)$ consisting of invertible mappings is open (a fact which follows from the consideration of Neumann series, see Exercise 1.8), it follows that, up to decreasing the open set O_0, the mapping $\frac{\partial \mathcal{F}}{\partial u}(\theta, g(\theta)) \in \mathcal{L}(E; F)$ is actually invertible for all $\theta \in O_0$. We thus obtain:

$$\forall h \in \Theta, \quad \mathrm{d}_\theta g(h) = -\left[\frac{\partial \mathcal{F}}{\partial u}(\theta, g(\theta))\right]^{-1} \left(\frac{\partial \mathcal{F}}{\partial \theta}(\theta, g(\theta))(h)\right).$$

We now provide a few typical applications of the implicit function Theorem 6.1.

Example 6.1 (Differentiability of the roots of a polynomial around a simple root)**.** This first example aims to show that the roots of a polynomial depend in a smooth way on its coefficients, as long as these roots are simple. More precisely, let $n \geq 1$; we consider the polynomial of degree n:

$$P(x) = a_0 + a_1 x + \ldots + a_n x^n,$$

which we assume to have simple roots $z_1, \ldots, z_n$. Let us introduce the function $\mathcal{F} : \mathbb{R}^{n+1} \times \mathbb{R}$ defined by

$$\forall (b_0, \ldots, b_n) \in \mathbb{R}^{n+1},\ x \in \mathbb{R}, \quad \mathcal{F}((b_0, \ldots, b_n), x) = b_0 + b_1 x + \ldots b_n x^n.$$

By assumption, the non linear equation

$$\text{Search for } x \in \mathbb{R} \text{ s.t. } \mathcal{F}((a_0, \ldots, a_n), x) = 0$$

has exactly n distinct solutions $z_1, \ldots, z_n$. For any of these roots z_i, the partial derivative of $\mathcal{F}$ with respect to the last variable at $((a_0, \ldots, a_n), z_i)$ reads:

$$\frac{\partial \mathcal{F}}{\partial x}((a_0, \ldots, a_n), z_i) = a_1 + 2a_2 z_i + \ldots + n a_n z_i^{n-1},$$

which is a non zero real number as z_i is a simple root of P.

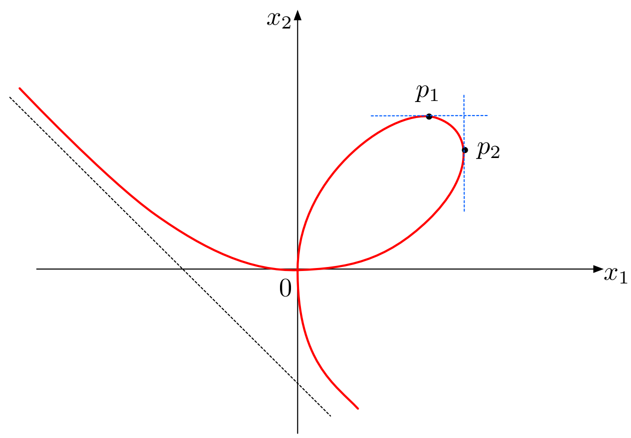

FIGURE 13. *The folium S studied in Example 6.2.*

By the implicit function Theorem 6.1, there exists a neighborhood U of $(a_0, \ldots, a_n)$ in $\mathbb{R}^{n+1}$, a neighborhood V of z_i in $\mathbb{R}$ and a function $g : U \to V$ of class $\mathcal{C}^\infty$ such that for all $(b_0, \ldots, b_n) \in U$, the polynomial $q(x) = b_0 + b_1 x + \ldots + b_n x^n$ has a unique root $g(b_0, \ldots, b_n)$ in V (which thus depends in a smooth way on $(b_0, \ldots, b_n)$).

Example 6.2 (Regularity of a curve in 2d)**.** In this example, we aim to study the set $S \subset \mathbb{R}^2$ defined by

$$S = \left\{x = (x_1, x_2) \in \mathbb{R}^2,\ \mathcal{F}(x_1, x_2) = 0\right\}, \text{ where } \mathcal{F}(x_1, x_2) = x_1^3 + x_2^3 - 3x_1 x_2.$$

This set, called the Descartes folium, is depicted on Fig. 13.

The function $\mathcal{F}$ is differentiable on $\mathbb{R}^2$ (actually, it is of class $\mathcal{C}^\infty$), and its partial derivatives read:

$$\forall x = (x_1, x_2) \in \mathbb{R}^2, \quad \frac{\partial \mathcal{F}}{\partial x_1}(x_1, x_2) = 3(x_1^2 - x_2), \text{ and } \frac{\partial \mathcal{F}}{\partial x_2}(x_1, x_2) = 3(x_2^2 - x_1).$$

A simple calculation shows that the only points $x = (x_1, x_2) \in S$ where the partial derivative $\frac{\partial \mathcal{F}}{\partial x_1}(x_1, x_2)$ vanishes are:

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

Hence, for any point $x = (x_1, x_2) \in S \setminus \left\{0, p^1\right\}$, the partial Fréchet derivative

$$\mathbb{R} \ni h \mapsto \frac{\partial \mathcal{F}}{\partial x_1}(x_1, x_2) h \in \mathbb{R}$$

is invertible; the implicit function Theorem 6.1 guarantees that there exist open neighborhoods U of x_1 and V of x_2 and a function $f : V \to U$ of class $\mathcal{C}^\infty$ such that

$$\forall z_1 \in U,\ z_2 \in V,\ \ (z_1, z_2) \in S \ \Leftrightarrow z_1 = f(z_2).$$

Likewise, the only points in S where $\frac{\partial \mathcal{F}}{\partial x_2}(x_1, x_2)$ vanishes are:

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

A similar argument as above shows that for any point $x = (x_1, x_2) \in S \setminus \left\{0, p^2\right\}$, there exist open neighborhoods U of x_1 and V of x_2 and a function $g : U \to V$ of class $\mathcal{C}^\infty$ such that

$$\forall z_1 \in U,\ z_2 \in V,\ \ (z_1, z_2) \in S \ \Leftrightarrow z_2 = g(z_1).$$

The set S has a horizontal tangent line at p_1, which is why the partial derivative $\frac{\partial \mathcal{F}}{\partial x_1}$ vanishes at p^1, but $\frac{\partial \mathcal{F}}{\partial x_2}$ does not: roughly speaking, near p^1, the x_2 coordinate can be expressed in terms of the x_1 coordinate in a smooth way, but not the other way around. Likewise, S has a vertical tangent line at p^2. The point 0 is more particular: both partial derivatives $\frac{\partial \mathcal{F}}{\partial x_1}$ and $\frac{\partial \mathcal{F}}{\partial x_2}$ vanish at that point, which reveals the singular character of S at this point.

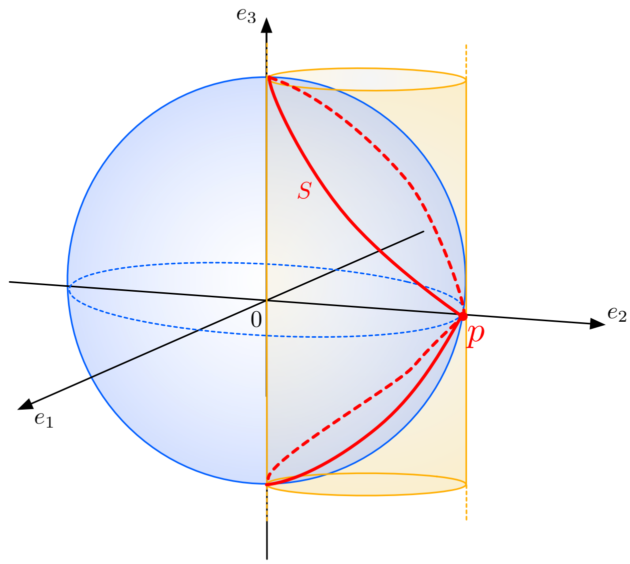

FIGURE 14. *Illustration of the set S considered in Example 6.3.*

Example 6.3 (Regularity of a surface in 3d)**.** Let $\mathcal{F} : \mathbb{R}^3 \to \mathbb{R}^2$ be the mapping of class $\mathcal{C}^\infty$ defined by

$$\forall x = (x_1, x_2, x_3) \in \mathbb{R}^3, \quad \mathcal{F}(x_1, x_2, x_3) = \begin{pmatrix} x_1^2 + x_2^2 + x_3^2 - 1 \\ x_1^2 + x_2^2 - x_2 \end{pmatrix}.$$

We are interested in the study of the set $S = \big\{x = (x_1, x_2, x_3) \in \mathbb{R}^3,\ \mathcal{F}(x_1, x_2, x_3) = 0\big\}$. Visually, S is the intersection between the unit sphere of $\mathbb{R}^3$ and the cylinder with axis e_3 and radius $\frac{1}{2}$, passing through the point $(0, \frac{1}{2}, 0)$, see Fig. 14.

A simple calculation yields the Jacobian matrix of $\mathcal{F}$:

$$\forall x = (x_1, x_2, x_3) \in \mathbb{R}^3, \quad \nabla\mathcal{F}(x_1, x_2, x_3) = \begin{pmatrix} 2x_1 & 2x_2 & 2x_3 \\ 2x_1 & 2x_2 - 1 & 0 \end{pmatrix}.$$

Let $p = (0, 1, 0) \in S$. We observe that for any point $x \in S \setminus \{p\}$, the Jacobian matrix $\nabla\mathcal{F}(x)$ has full rank (its two rows are linearly independent). In particular, the mapping

$$\mathbb{R}^2 \ni (h_1, h_2) \mapsto \frac{\partial \mathcal{F}}{\partial x_1}(x_1, x_2, x_3)h_1 + \frac{\partial \mathcal{F}}{\partial x_3}(x_1, x_2, x_3)h_2 = \begin{pmatrix} 2x_1h_1 + 2x_3h_2 \\ 2x_1h_1 \end{pmatrix} \in \mathbb{R}^2$$

is a linear isomorphism. It follows from the implicit function Theorem 6.1 that there exists a neighborhood U of (x_1, x_3) in $\mathbb{R}^2$ and V of x_2 in $\mathbb{R}$, as well as a mapping $g : U \to V$ of class $\mathcal{C}^\infty$ such that:

$$\forall (z_1, z_3) \in \mathbb{R}^2,\ z_2 \in \mathbb{R},\ \ (z_1, z_2, z_3) \in S \ \Leftrightarrow z_2 = g(z_1, z_3).$$

The degeneracy of the Jacobian matrix $\nabla\mathcal{F}$ at p accounts for the singularity of the set S at this point, where several branches cross.

6.3. The local inverse theorem

The local inverse theorem gives a simple sufficient condition for a function to be invertible in the vicinity of a point x_0. Let us start with a definition.

Definition 6.1. *Let $(E, ||\cdot||_E)$ and $(F, ||\cdot||_F)$ be two normed vector spaces, and let $U \subset E$ and $V \subset F$ be two open sets. One mapping $f : U \to V$ is called a diffeomorphism of class $\mathcal{C}^k$ between U and V for some integer $k \geq 1$ if f is a mapping of class $\mathcal{C}^k$, if it realizes a bijection between U and V and if the inverse mapping $f^{-1} : V \to U$ is also of class $\mathcal{C}^k$.*

The main result of interest in this section is the following.

Theorem 6.2. *Let $(E, ||\cdot||_E)$ and $(F, ||\cdot||_F)$ be two Banach spaces, let $U \subset E$ be an open subset, and let $f : U \to F$ be a function of class $\mathcal{C}^k$ for some $k \geq 1$. Let $x_0 \in U$ be a point such that the differential $\mathrm{d}f_{x_0} : E \to F$ of f at x_0 is a linear isomorphism. Then there exist two open neighborhoods $U_0 \subset U$ of x_0 in E and V_0 of $f(x_0)$ in F such that $f : U_0 \to V_0$ is a diffeomorphism of class $\mathcal{C}^k$.*

6.4. **Exercises**

Exercise 6.1. For $t \in \mathbb{R}$, let us consider the system of equations

$$\begin{cases} x + y + z + t = 0 \\ x^2 + y^2 + z^2 + t = 2 \\ x^3 + y^3 + z^3 + t^2 = 0 \end{cases}$$

(i) Show that $(x, y, z) = (0, -1, 1)$ is one solution to this system for $t = 0$.
(ii) Show that there exist neighborhoods U of 0 in $\mathbb{R}$ and V of $(0, -1, 1)$ in $\mathbb{R}^3$, as well as a smooth function $f : U \to V$ such that for all $t \in U$, this system has a unique solution in V given by $(x, y, z) = f(t)$.
(iii) Calculate the derivative of the function f.

Exercise 6.2. Let $(E, ||\cdot||_E)$ and $(F, ||\cdot||_F)$ be two Banach spaces, and let $U \subset E$ be open. Let $f : U \to F$ be a differentiable mapping whose differential is invertible at all points $x \in U$. Show that $f(U)$ is an open subset of F.

[A mapping such that the image of any open set is open is called an open mapping.]

Exercise 6.3 (The method of perturbations of identity). Let E be the vector space of vector fields $\theta : \mathbb{R}^d \to \mathbb{R}^d$ of class $\mathcal{C}^1$, which are bounded as well as their derivatives. A simple variation of Exercise 1.7 shows that E is a Banach space when equipped with the norm

$$||\theta||_E := \max\Big(\sup_{x\in\mathbb{R}^d} |\theta(x)|, \sup_{x\in\mathbb{R}^d} |\nabla\theta(x)| \Big),$$

where we have introduced the matrix norm

$$\forall M \in M_d(\mathbb{R}), \quad ||M|| := \sup_{|x|=1} |Mx|, \text{ where } |x| := (|x_1|^2 + \ldots + |x_d|^2)^{\frac{1}{2}} \text{ is the Euclidean norm.}$$

(i) Show that, when $||\theta||_E < 1$, the mapping $(\mathrm{Id} + \theta) : \mathbb{R}^d \to \mathbb{R}^d$ is bijective.
[Hint: Apply the fixed point Theorem 5.1 with the mapping $T_y : \mathbb{R}^d \ni x \mapsto y - \theta(x) \in \mathbb{R}^d$, for an arbitrary, fixed $y \in \mathbb{R}^d$.]
(ii) Show that, still under the assumption $||\theta||_E < 1$, $(\mathrm{Id} + \theta)$ is a $\mathcal{C}^1$ diffeomorphism of $\mathbb{R}^d$.
[Hint: Use the local inverse Theorem 6.2.]

7. A first encounter with optimality conditions for optimization problems

In this section, we use the previous concepts of differential calculus to analyze optimization problems – a (very) wide subject that we only broach for the sake of illustrating our developments. More precisely, we formulate first- and second-order conditions for a point to be one local minimum of a function f defined on a subset of a normed vector space in terms of the derivatives of f.

7.1. **Local minimizers of unconstrained optimization problems in normed vector spaces**

In this first section, we deal with an unconstrained minimization problem of a function f, that is, a problem of the form

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

where X is a subset of a normed vector space and $f : X \to \mathbb{R}$ is a function. Let us start with some definitions.

Definition 7.1. *Let $(E, ||\cdot||_E)$ be a normed vector space, $X \subset E$ be a subset of E, and let $f : X \to \mathbb{R}$ be a function. Then,*

- *f is said to have a local minimum at some point $x_0 \in X$ if there exists an open subset U in E containing x_0 such that:*
$$\forall x \in U \cap X, \quad f(x_0) \leq f(x).$$
This local minimum is strict if the above inequality is strict for $x \neq x_0$.

- *f has a global minimum at $x_0 \in X$ if*

$$\forall x \in X, \quad f(x_0) \leq f(x).$$

This global minimum is strict if the above inequality is strict for $x \neq x_0$.

Without further assumption on the considered function f and the subset X, there is no guarantee that a minimizer (even local) of f on X should exist. In practice, one is often satisfied with finding local minimizers of f (if any); as we shall see, the task of characterizing those is easier than global minimizers, since one can rely on the derivatives of f to this end. See nevertheless Exercises 7.1 to 7.3 for situations where a function f can be proved to have global minimizers.

Theorem 7.1. *Let U be an open subset of a normed vector space $(E, ||\cdot||_E)$, and let $f : U \to \mathbb{R}$ be a function. Assume that f has a local minimizer $x \in U$ and that f is differentiable at x; then $\mathrm{d}f_x = 0$ in E^*.*

Proof. Since x is a local minimizer of f on U, there exists an open neighborhood $V \subset U$ of x such that:

$$\forall y \in V, \quad f(x) \leq f(y).$$

Let us fix a direction $h \in E$; there exists $\delta > 0$ such that $x + th$ belongs to V for $0 < t < \delta$, and so

$$\forall 0 < t \leq t_0, \quad f(x) \leq f(x + th). \tag{7.1}$$

Subtracting $f(x)$ from both sides and dividing by $t > 0$ we obtain:

$$\frac{f(x + th) - f(x)}{t} \geq 0;$$

now letting t tend to 0 yields:

$$f'(x; h) = 0,$$

where we recall that $f'(x; h)$ stands for the one-sided directional derivative of f at x in the direction $h \in E$, see Definition 2.4. Since f is differentiable at x and $h \in E$ is arbitrary, the claim follows. □

In the above statement, the requirement that the considered local minimizer x be interior to the domain of definition of f is crucial, it makes it possible to test the minimal character of x with respect to perturbations $x + th$ of x in arbitrary directions $h \in E$, see (7.1). The terminology "unconstrained minimization" for the above setting comes from this feature. Actually, the exact same argument allows to prove the following more general statement, which is illustrated on Fig. 15.

Theorem-Definition 7.1. *Let K be a subset of a normed vector space $(E, ||\cdot||_E)$, and let $f : K \to \mathbb{R}$ be a function. One direction $h \in E$ is admissible at some point $x \in K$ if there exists $\delta > 0$ such that*

$$\forall\, 0 \leq t < \delta, \quad x + th \in K.$$

If x is a local minimum of f over K, and if f is Gateaux-differentiable at x, it holds

$$f'(x; h) \geq 0 \text{ for any admissible direction } h \in E. \tag{7.2}$$

Let us provide two examples where the above statement takes on a more explicit form:

- If K is a convex subset of the normed vector space $(E, ||\cdot||_E)$, then for any point $y \in K$ the direction $(y - x)$ is admissible at x since the whole segment $\{(1-t)x + ty,\ t \in [0,1]\}$ is contained in K. Hence, if x is a local minimizer of the Gateaux-differentiable function f over K, (7.2) rewrites:

$$\forall y \in K, \quad f'(x; y - x) \geq 0.$$

- If K is an affine subspace of E, i.e. of the form

$$K = \{x_0 + z,\ z \in F\} \text{ for some point } x_0 \in E \text{ and a vector subspace} F \subset E,$$

then (7.2) rewrites:

$$\forall h \in F, \quad f'(x; h) \geq 0.$$

We now turn to additional necessary conditions for a point x to be a local minimizer of a function, involving the second-order derivatives of f. Under some appropriate assumptions, these may besides turn out to be sufficient.

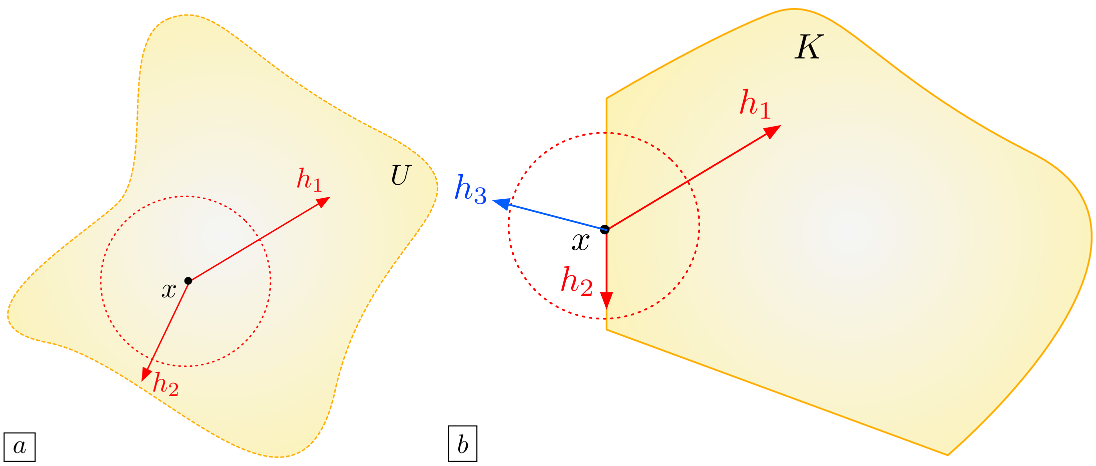

FIGURE 15. *(a) When the considered local minimum x is interior to the domain of definition U of f, all directions $h \in E$ are admissible: for $t > 0$ small enough, $x + th$ belongs to U; (b) when the domain of definition U is not open, x may lie on its boundary, in which case all directions (such as that h_3) are not admissible.*

Theorem 7.2. *Let U be an open subset of a normed vector space $(E, ||\cdot||_E)$ and let $f : U \to \mathbb{R}$ be a function. We assume that f has a local minimum at some point $x \in U$ and that it is twice differentiable at x. Then, the following conditions hold:*

$$\forall h \in E, \quad \mathrm{d}f_x(h) = 0 \text{ and } \mathrm{d}^2 f_x(h,h) \geq 0. \tag{7.3}$$

Conversely, if f is twice differentiable at some point $x \in U$ and if

$$\forall h \in E, \ \mathrm{d}f_x(h) = 0 \text{ and } \exists c > 0 \text{ s.t. } \forall h \in E, \quad \mathrm{d}^2 f_x(h,h) \geq c||h||_E^2, \tag{7.4}$$

then f has a strict local minimum at x.

The conditions (7.3) (resp. (7.4)) are often referred to as the *necessary* (resp. *sufficient*) *second-order conditions for local optimality.*

Proof. Let us first show that the conditions (7.3) hold, under the assumption that f has a local minimum at $x \in U$. We have already seen with Theorem 7.1 that $\mathrm{d}f_x = 0$ in this case. Besides, since f is twice differentiable at x, the Taylor Young's formula from Theorem 4.4 at order 2 yields the existence of a real number $\delta > 0$ and a function $r_f : E \to \mathbb{R}$ such that:

$$r_f(h) \to 0 \text{ as } h \to 0 \text{ and } \forall h \in E, \ ||h||_E \leq \delta, \quad f(x+h) = f(x) + \frac{1}{2}\mathrm{d}f_x^2(h,h) + ||h||_E^2 r_f(h).$$

Let now $h \in E$ be a fixed direction. Since by assumption x is a local minimum for f, up to decreasing the value of δ, we have:

$$\forall 0 < t < \delta, \ f(x+th) \geq f(x), \text{ and so } \frac{t^2}{2}\mathrm{d}f_x^2(h,h) + t^2||h||_E^2 r_f(th) \geq 0.$$

Dividing both sides of the above inequality by t^2 and letting t tend to 0, we obtain the desired relation (7.3).

Conversely, let us assume that the conditions (7.4) hold at some point $x \in U$. Again, since f is twice differentiable at x, the Taylor Young's formula at order 2 yields the existence of a number $\delta > 0$ and a function $r_f : E \to \mathbb{R}$ such that $r_f(h) \to 0$ as $h \to 0$, and

$$\begin{aligned} \forall h \in E, \ ||h||_E \leq \delta, \ f(x+h) &= f(x) + \mathrm{d}f_x(h) + \tfrac{1}{2}\mathrm{d}^2 f_x(h,h) + ||h||_E^2 r_f(h) \\ &\geq f(x) + c||h||_E^2 + ||h||_E^2 r_f(h). \end{aligned} \tag{7.5}$$

Now, let $\varepsilon < c$ be fixed; since $r_f(h) \to 0$ as $h \to 0$, up to decreasing the value of δ, one has $|r_f(h)| \leq \varepsilon$ as soon as $||h||_E \leq \delta$. Then, the estimate (7.5) becomes:

$$\forall h \in E, \;\; ||h||_E \leq \delta, \quad f(x+h) \geq f(x) + (c-\varepsilon)||h||_E^2.$$

In particular, for $||h||_E \leq \delta$, $h \neq 0$, we see that $f(x+h) > f(x)$, which means that f has a strict local minimum at x. □

7.2. Local minimizers of constrained optimization problems

The previous section has mainly focused on deriving necessary and sufficient local optimality conditions at points x which are interior to the domain of definition of the considered function f. As we have seen, for these results to hold, it is necessary that x can be perturbed in all directions $h \in E$.

Situations where this is not possible are usually much more delicate to handle; they fall into the general framework of *constrained optimization*. We shall see in this section that, when the set where the optimized variable x is sought has a particular structure, appropriate necessary conditions can be derived. More precisely, we shall investigate optimization problems of a function $f : U \to \mathbb{R}$ defined on an open set $U \subset E$, where the set C for the optimization variable x arises as the set of zeroes of certain constraint functions $g_1, \ldots, g_p : U \to \mathbb{R}$:

$$\min_{x \in C} f(x), \text{ where } C := \{x \in U, \;\; g_1(x) = \ldots = g_p(x) = 0\}.$$

Remark 7.1. *It is possible to deal with more general constraint sets, including inequality constraints of the form:*

$$h_1(x) \leq 0, \ldots, h_q(x) \leq 0,$$

for some suitable functions $h_1, \ldots, h_q : U \to \mathbb{R}$. Since the treatment of such constraints is more technical, we prefer to stick to equality constraints in this course.

Definition 7.2. *Let U be an open subset of a normed vector space $(E, ||\cdot||_E)$, let $f, g_1, \ldots, g_p : U \to \mathbb{R}$ be functions, and let the set $C \subset U$ be defined by $C = \{x \in U, \;\; g_1(x) = \ldots = g_p(x)\}$. One says that f has a local minimum at some point $x_0 \in U$ under the constraints $g_1(x) = \ldots = g_p(x) = 0$ if $x \in C$ and there exists an open neighborhood $V \subset U$ of x_0 such that*

$$\forall x \in V \cap C, \quad f(x) \geq f(x_0).$$

The following result, that we shall admit, yields a necessary conditions for optimality.

Theorem 7.3. *Let U be an open subset of a Banach space $(E, ||\cdot||_E)$, and let $f, g_1, \ldots, g_p : U \to \mathbb{R}$ be functions of class $\mathcal{C}^1$. Let us assume that f has a local minimum at $x_0 \in C$ under the constraints $g_1(x_0) = \ldots = g_p(x_0) = 0$, and that the constraints are qualified at x_0 in the sense that:*

$$\{\mathrm{d}g_{1,x_0}, \ldots, \mathrm{d}g_{p,x_0}\} \text{ is a linearly independent family in } E^*.$$

Then, then there exist $\lambda_1, \ldots, \lambda_p \in \mathbb{R}$ such that

$$\mathrm{d}f_{x_0} = \lambda_1 \mathrm{d}g_{1,x_0} + \ldots + \lambda_p \mathrm{d}g_{p,x_0}. \tag{7.6}$$

Remark 7.2. *When the ambient vector space E has finite dimension, say $E = \mathbb{R}^d$, the necessary condition (7.6) can be rewritten in terms of the gradients of f, $g_1, \ldots, g_p$ at x_0: there exist $\lambda_1, \ldots, \lambda_p \in \mathbb{R}$ such that:*

$$\nabla f(x_0) = \lambda_1 \nabla g_1(x_0) + \ldots + \lambda_p \nabla g_p(x_0).$$

The fact that the relation (7.6) encodes the local optimality of f at x_0 under the constraints $g_1(x) = \ldots = g_p(x) = 0$ can be understood in multiple ways. On the one hand, let us assume that the relation (7.6) holds at some point $x_0 \in C$ for some $\lambda_1, \ldots, \lambda_p \in \mathbb{R}$. Obviously, if we are to modify (decrease) the value of f at first order, a direction h has to be chosen such that $\mathrm{d}f_{x_0}(h) < 0$. But then, by (7.6), one of the quantities $\mathrm{d}g_{i,x_0}(h)$ has to be different from 0, meaning that the point $x_0 + h$ will leave the set of constraints C. Hence, the only means to improve the value of f from x_0 is to get out of the set of constraints, which agrees with the intuition of optimality. Another, geometric interpretation of (7.6) is exemplified on Fig. 16.

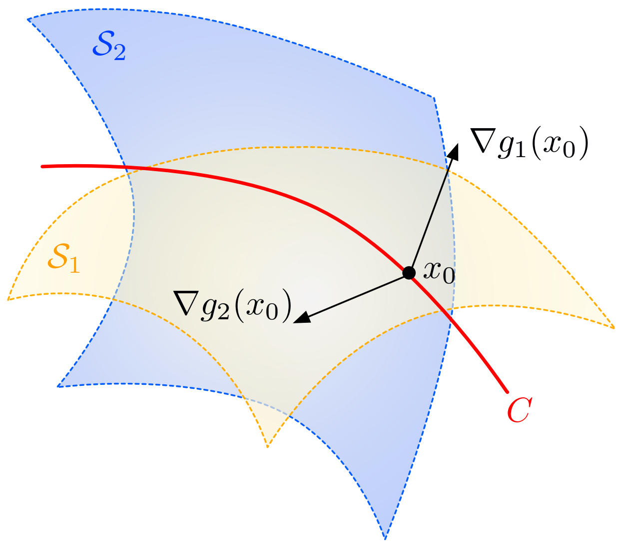

FIGURE 16. *Optimization of a function $f : \mathbb{R}^3 \to \mathbb{R}$ under two constraints induced by functions $g_1, g_2 : \mathbb{R}^3 \to \mathbb{R}$. Each constraint corresponds to a hypersurface in $\mathbb{R}^3$ (in yellow and blue), with normal vectors oriented along ∇g_1 and ∇g_2; the constraint set C is the intersection of these hypersurfaces (red line). A point $x_0 \in C$ is a local minimum for f if the value of f cannot be decreased by moving tangentially to C: the only way to decrease the value of f is to move "outside" the set of constraints, i.e. along $\nabla g_1(x_0)$ or $\nabla g_2(x_0)$.*

7.3. Towards the calculus of variations

In this section, we propose an interesting calculation, which is fundamental in an area of mathematics called the calculus of variations, and is meanwhile a good opportunity to handle the material of Sections 2 to 4.

Throughout this section, E is the vector space of functions of class $\mathcal{C}^1$ on the real interval $[0,1]$, equipped with the norm

$$||u||_E := \sup(||u||_\infty, ||u'||_\infty);$$

let us recall from Exercise 1.7 that $(E, ||\cdot||_E)$ is a Banach space.

We consider the function $\mathcal{L} : E \to \mathbb{R}$ defined by

$$\mathcal{L}(u) = \int_0^T j(t, u(t), u'(t))\,\mathrm{d}t, \tag{7.7}$$

where $j : \mathbb{R}_t \times \mathbb{R}_u \times \mathbb{R}_p \to \mathbb{R}$ is a function of class $\mathcal{C}^1$. The following result is devoted to the calculation of the differential of $\mathcal{L}$.

Proposition 7.1. *Let $u \in E$ be given; then the function $\mathcal{L}$ defined in (7.7) is Fréchet differentiable at u, and its derivative reads:*

$$\forall h \in E, \quad \mathrm{d}\mathcal{L}_u(h) = \int_0^T \left(\frac{\partial j}{\partial u}(t, u(t), u'(t))h(t) + \frac{\partial j}{\partial p}(t, u(t), u'(t))h'(t) \right) \mathrm{d}t. \tag{7.8}$$

If, in addition, u and j are of class $\mathcal{C}^2$ on their respective domains of definition, this implies, in particular, that for any direction $h \in E$ such that $h(0) = h(T) = 0$, it holds:

$$\mathrm{d}\mathcal{L}_u(h) = \int_0^T \left(\frac{\partial j}{\partial u}(t, u(t), u'(t)) - \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial j}{\partial p}(t, u(t), u'(t)) \right) \right) h(t)\,\mathrm{d}t. \tag{7.9}$$

Proof. Let $u \in E$ be fixed, and let $h \in E$ be an arbitrary direction. For any time $t \in [0,T]$, the chain rule of Theorem 2.1 reveals that the mapping

$$[0,1] \ni s \mapsto j(t, u(t) + sh(t), u'(t) + sh'(t)) \in \mathbb{R}$$

is of class $\mathcal{C}^1$, so that an application of the Taylor formula with integral remainder from Theorem 4.6 yields:

$$j(t,u(t)+h(t),u'(t)+h'(t)) - j(t,u(t),u'(t)) =$$
$$\int_0^1 \left(\frac{\partial j}{\partial u}(t,u(t)+sh(t),u'(t)+sh'(t))h(t) + \frac{\partial j}{\partial p}(t,u(t)+sh(t),u'(t)+sh'(t))h'(t)\right) \mathrm{d}s.$$

Hence, the difference $\mathcal{L}(u+h)-\mathcal{L}(u)$ equals:

(7.10)
$$\begin{aligned}\mathcal{L}(u+h)-\mathcal{L}(u) &= \int_0^T \int_0^1 \left(\frac{\partial j}{\partial u}(t,u(t)+sh(t),u'(t)+sh'(t))h(t) + \frac{\partial j}{\partial p}(t,u(t)+sh(t),u'(t)+sh'(t))h'(t)\right) \mathrm{d}s \\ &= \int_0^T \left(\frac{\partial j}{\partial u}(t,u(t),u'(t))h(t) + \frac{\partial j}{\partial p}(t,u(t),u'(t))h'(t)\right) \mathrm{d}t + \int_0^T A(t)\,\mathrm{d}t,\end{aligned}$$

where the first integral in the above right-hand side is exactly the quantity featured in (7.8); it defines a linear and continuous mapping of the variable $h \in E$. We have also defined the remainder

$$\begin{aligned}A(t) \;&:=\; \left(\int_0^1 \frac{\partial j}{\partial u}(t,u(t)+sh(t),u'(t)+sh'(t))\,\mathrm{d}s - \frac{\partial j}{\partial u}(t,u(t),u'(t))\right) h(t) \\ &\quad + \left(\int_0^1 \frac{\partial j}{\partial p}(t,u(t)+sh(t),u'(t)+sh'(t))\,\mathrm{d}s - \frac{\partial j}{\partial p}(t,u(t),u'(t))\right) h'(t) \\ &= \int_0^1 \Bigg(\left(\frac{\partial j}{\partial u}(t,u(t)+sh(t),u'(t)+sh'(t)) - \frac{\partial j}{\partial u}(t,u(t),u'(t))\right) h(t) \\ &\quad + \left(\frac{\partial j}{\partial p}(t,u(t)+sh(t),u'(t)+sh'(t)) - \frac{\partial j}{\partial p}(t,u(t),u'(t))\right) h'(t)\Bigg)\,\mathrm{d}s.\end{aligned}$$

We now proceed to estimate this remainder; to this end, using the fact that j is of class $\mathcal{C}^1$ together with Corollary 4.1, we see that for all $\varepsilon > 0$, there exists $\delta > 0$ such that

(7.11) $\forall t \in [0,T],\ u_1, u_2 \in \mathbb{R}$ and $p_1, p_2 \in \mathbb{R}$,
$$\Big(|u_1-u_2| \leq \delta \text{ and } |p_1-p_2| \leq \delta\Big) \Rightarrow \left|\frac{\partial j}{\partial u}(t,u_2,p_2) - \frac{\partial j}{\partial u}(t,u_1,p_1)\right| + \left|\frac{\partial j}{\partial p}(t,u_2,p_2) - \frac{\partial j}{\partial p}(t,u_1,p_1)\right| \leq \varepsilon.$$

As a result, if the direction $h \in E$ satisfies $||h||_E \leq \delta$, so that for all $t \in [0,T]$, $|h(t)| \leq \delta$ and $|h'(t)| \leq \delta$, we obtain:
$$\forall h \in E,\ \ ||h||_E \leq \delta, \quad \forall t \in [0,T],\ \ |A(t)| \leq \varepsilon ||h||_E.$$

Insertion of this estimate into (7.10) yields the desired expression (7.8). The second formula (7.9) follows easily from an integration by parts. □

Remark 7.3. *Functionals of the form (7.7) often show up in physics, where the bear the name of Lagrangian. They represent the so-called action associated to the motion $t \mapsto u(t)$ of an objective in time. The least action principle states that the motion adopted by a particle is that u which minimizes the action among all admissible motions.*

Using Theorem 7.1, we see that any local minimizer $u \in E$ of $\mathcal{L}$ satisfies:
$$\forall u \in \mathrm{d}\mathcal{L}_u(h).$$

Assuming that j and u are of class $\mathcal{C}^2$, this implies that:
$$\forall h \in E, \quad \int_0^T \left(\frac{\partial j}{\partial u}(t,u(t),u'(t)) - \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial j}{\partial p}(t,u(t),u'(t))\right)\right) h(t)\,\mathrm{d}t = 0,$$

and using the fact that a continuous function which vanishes when integrated against any continuous function is necessarily null, it follows:
$$\forall t \in [0,T], \quad \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial j}{\partial p}(t,u(t),u'(t))\right) = \frac{\partial j}{\partial u}(t,u(t),u'(t)).$$

This equation, which is sometimes sufficient to characterize the motion $u(t)$ is called the Euler-Lagrange equation of the system.

7.4. **Exercises**

Exercise 7.1 (Existence of global minimizers (I)). Let X be a compact subset of a normed vector space $(E, ||\cdot||_E)$ and let $f : X \to \mathbb{R}$ be a continuous function. We consider the minimization problem

$$\inf_{x \in X} f(x).$$

(i) Show that there exists a minimizing sequence for this problem, that is, a sequence $x_n \in X$ such that

$$\lim_{n\to\infty} f(x_n) = \inf_{x\in X} f(x).$$

(ii) Show that one may extract a subsequence x_{n_k} from x_n which converges to some point $x^* \in X$.
(iii) Show that x^* is a global minimizer of f on X.

[Hint: This technique for proving the existence of a global minimizer for a function f can be adapted and generalized to various more challenging contexts; it is often referred to as the "direct method of the calculus of variations" for proving existence of minimizers.]

Exercise 7.2 (Existence of global minimizers (II)). Let $d \geq 1$ and let $f : \mathbb{R}^d \to \mathbb{R}$ be a continuous function which is "infinite at infinity", that is:

$$\forall M > 0,\ \exists R > 0 \text{ s.t. } \forall x \in \mathbb{R}^d,\ |x| \geq R \ \Rightarrow\ f(x) \geq M.$$

Show that f has a global minimizer on $\mathbb{R}^d$.

Exercise 7.3. Let $(E, ||\cdot||_E)$ be a normed vector space and let $C \subset E$ be a convex subset. Let $f : C \to \mathbb{R}$ be a convex function which has a local minimum $x^* \in C$. Show that x^* is a global minimizer of f on C.

Exercise 7.4. Let $a, b \in \mathbb{R}$, and let the function $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by

$$\forall x = (x_1, x_2) \in \mathbb{R}^2, \quad f(x) = (x_1 - a)^2 (x_2 - b)^2.$$

(i) Calculate the first- and second-order derivatives of f.
(ii) Search for all the critical points of f, and search for its global and local minima.

Exercise 7.5. Let $\gamma : [0, T] \to \mathbb{R}^d$ be a curve; we define its length by

$$L(\gamma) = \int_0^T \sqrt{1 + |\gamma'(t)|^2}\, \mathrm{d}t.$$

- Show that this definition is independent by change of parametrization.
- Show that the curve with minimum length between two points is a line segment.

8. Differential calculus in regular domains of $\mathbb{R}^d$

This last part of the course is intended as a rough introduction to differential calculus in regular domains of the Euclidean space $\mathbb{R}^d$. This topic is ubiquitous in many fields of mathematics, such as the study of partial differential equations, differential geometry, the calculus of variations, etc. It is admittedly quite technical, and we have preferred to focus on the presentation of the main concepts and results, even it means omitting intricated proofs.

8.1. **Hypersurface in $\mathbb{R}^d$ and subdomains of $\mathbb{R}^d$**

The central notion in this section is that of hypersurface of class $\mathcal{C}^k$ in $\mathbb{R}^d$, which can be defined in three equivalent ways, as depicted on Fig. 17.

Definition 8.1. *Let $k \geq 1$; a subset $\mathcal{S}$ of $\mathbb{R}^d$ is called a hypersurface of class $\mathcal{C}^k$ if one of the following three equivalent situations applies:*

(i) (Graph representation): For any point $x_0 \in \mathcal{S}$, there exists an open neighborhood V of x_0 in $\mathbb{R}^d$, an orthonormal basis $(e_1, \dots, e_d)$ of $\mathbb{R}^d$ and a function $\psi : \mathbb{R}^{d-1} \to \mathbb{R}$ of class $\mathcal{C}^k$ such that:

$$\mathcal{S} \cap V = \left\{ x = \sum_{i=1}^d x_i e_i \in V \ s.t.\ x_d = \psi(x_1, \dots, x_{d-1}) \right\}.$$

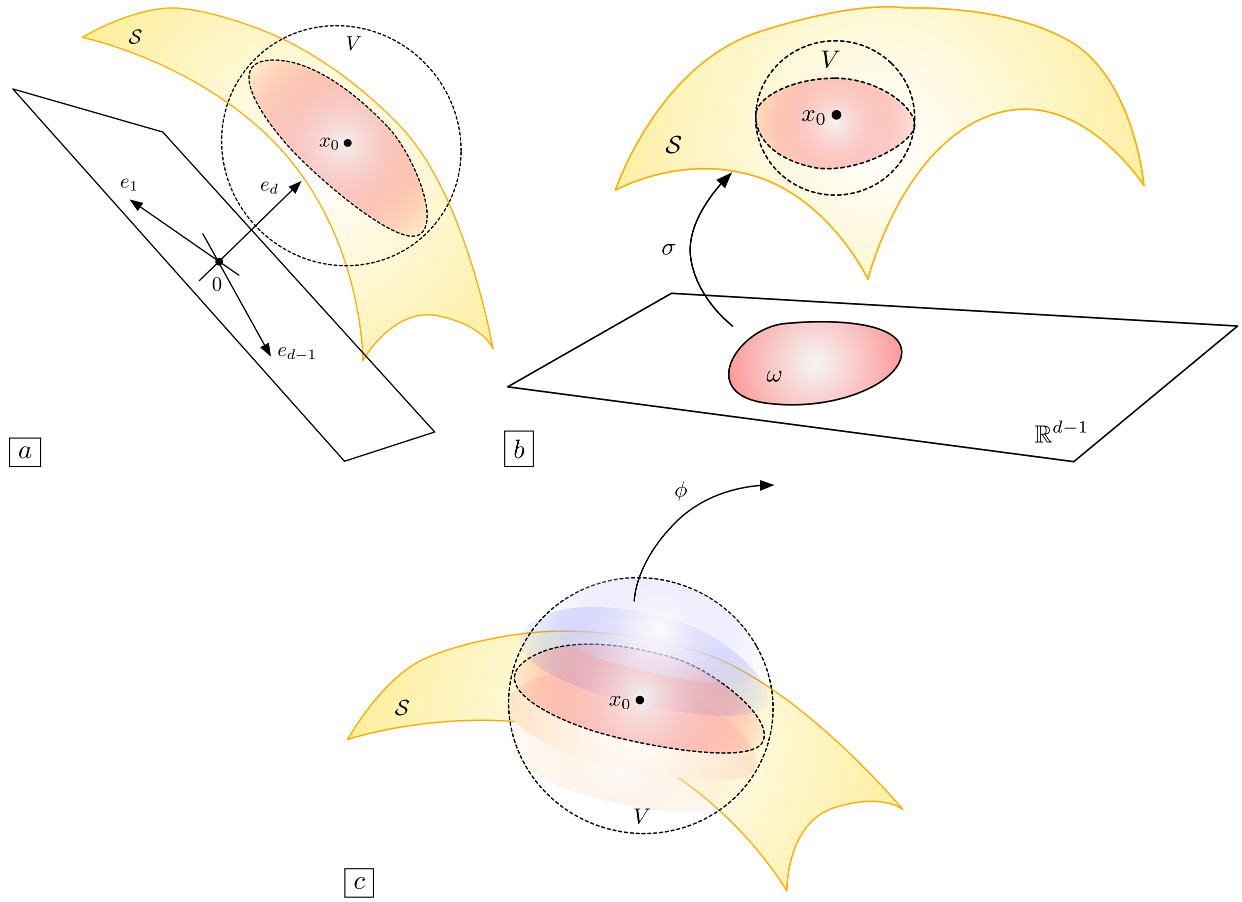

FIGURE 17. *Three different representations for a hypersurface $\mathcal{S} \subset \mathbb{R}^d$: (a) Near the point $x_0 \in \mathcal{S}$, the hypersurface $\mathcal{S}$ is the graph of a function ψ; (b) $\mathcal{S} \cap V$ is the image $\sigma(\omega)$ of an open subset $\omega \subset \mathbb{R}^{d-1}$ by a suitable parametrization mapping $\sigma : \omega \to \mathbb{R}^d$; (c) near x_0, $\mathcal{S}$ can be represented as the 0 isosurface of a function $\phi : \mathbb{R}^d \to \mathbb{R}$ (isosurfaces associated to values different from 0 are also depicted).*

(ii) (Parametrization): For any point $x_0 \in \mathcal{S}$, there exists an open neighborhood V of x_0 in $\mathbb{R}^d$, an open neighborhood ω of 0 in $\mathbb{R}^{d-1}$ and a mapping $\sigma : \omega \to \mathbb{R}^d$ of class $\mathcal{C}^k$ such that $\mathcal{S} \cap V = \sigma(\omega)$;
(iii) (Definition by an implicit function): For any point $x_0 \in \mathcal{S}$, there exists an open neighborhood V of x_0 in $\mathbb{R}^d$ and a function $\phi : V \to \mathbb{R}$ of class $\mathcal{C}^k$ such that:

$$\forall x \in V, \quad \nabla\phi(x) \neq 0, \text{ and } \mathcal{S} \cap V = \{x \in V, \ \phi(x) = 0\}\,. \tag{8.1}$$

Remark 8.1.

- *The above Definition 8.1 precludes the possibility for a hypersurface $\mathcal{S}$ to have a boundary: visually, $\mathcal{S}$ must be a "closed" surface, as exemplified on Fig. 18. It is possible to define a corresponding notion of "open hypersurface" in $\mathbb{R}^d$, but we shall not enter into these details.*
- *The above notion of hypersurface of class $\mathcal{C}^k$ falls into the general theory of submanifolds of $\mathbb{R}^d$. More precisely, a hypersurface $\mathcal{S} \subset \mathbb{R}^d$ is a codimension 1 submanifold of $\mathbb{R}^d$, but it is possible to define submanifolds with higher codimension; for instance, a curve in $\mathbb{R}^3$ is a codimension 2 submanifold.*

We now turn to the notion of orientation for a hypersurface of class $\mathcal{C}^k$.

Proposition-Definition 8.1. *Let $\mathcal{S}$ be hypersurface of class $\mathcal{C}^k$ of $\mathbb{R}^d$ for some $k \geq 1$; according to the implicit representation of $\mathcal{S}$, for each point $x_0 \in \mathcal{S}$, there exists a neighborhood V of x_0 in $\mathbb{R}^d$ and a function*

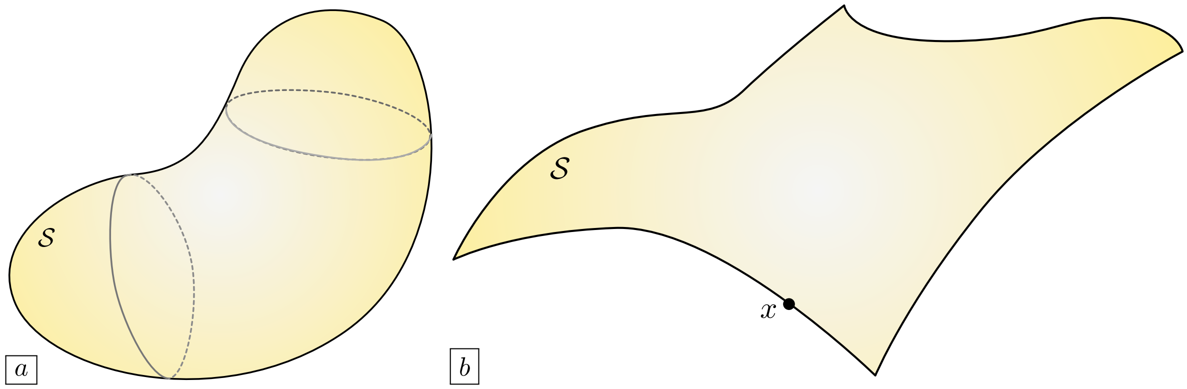

FIGURE 18. *(a) $\mathcal{S}$ is a hypersurface (without boundary) in the sense of Definition 8.1; (b) $\mathcal{S}$ is not a hypersurface in the sense of Definition 8.1 since it presents a boundary; in particular, near the point x, $\mathcal{S}$ cannot be mapping diffeomorphically to an open subset of $\mathbb{R}^{d-1}$ (but rather to a half-ball).*

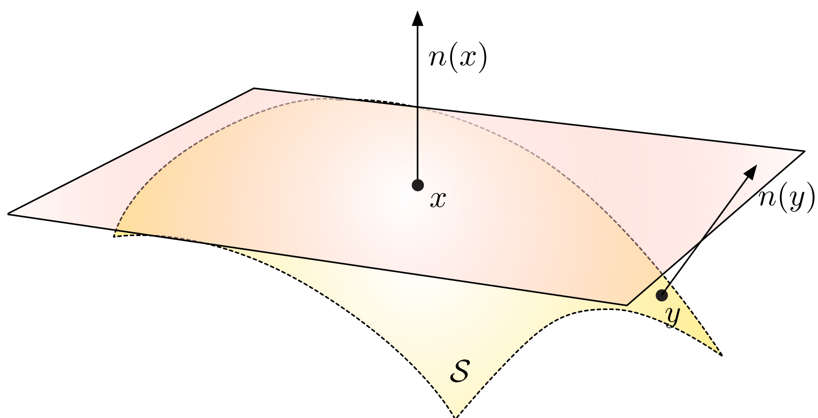

FIGURE 19. *At every point x of a hypersurface of class $\mathcal{C}^k$, the normal vector $n(x)$ is "orthogonal" to $\mathcal{S}$; the tangent plane (in red) to $\mathcal{S}$ at x is the affine plane with normal direction $n(x)$ passing through x.*

$\phi : V \to \mathbb{R}$ such that (8.1) holds. Then, the unit vector field $\frac{\nabla\phi(x)}{|\nabla\phi(x)|}$ is uniquely defined (i.e. independently of the function ϕ chosen to represent $\mathcal{S}$ in a neighborhood of x_0 as long as it satisfies (8.1)), up to a sign. One says that $\mathcal{S}$ is oriented if there exists a selection of these vectors which is continuous on $\mathcal{S}$. The resulting vector field is denoted by n and is called the unit normal vector field to $\mathcal{S}$.

Visually, for all $x \in \mathcal{S}$, the normal vector $n(x)$ is "perpendicular" to the surface $\mathcal{S}$ at x; it allows to generalize the notion of tangent line to a curve to the case of a hypersurface, see Fig. 19.

Definition 8.2. *Let $\mathcal{S}$ be an oriented hypersurface of class $\mathcal{C}^k$ in $\mathbb{R}^d$ for some $k \geq 1$; the affine plane of $\mathbb{R}^d$ passing through x and normal to $n(x)$ is called the tangent plane to $\mathcal{S}$ at x.*

We now turn to the notion of domain of $\mathbb{R}^d$ of class $\mathcal{C}^k$, see Fig. 20.

Definition 8.3. *Let Ω be an open subset of $\mathbb{R}^d$; Ω is called a domain of class $\mathcal{C}^k$ for some $k \geq 1$ if $\partial\Omega = \partial(\mathbb{R} \setminus \overline{\Omega})$, and if its boundary $\partial\Omega$ is an oriented hypersurface of class $\mathcal{C}^k$. Then, the unit normal vector $n(x)$ can always be chosen to point outward Ω.*

In many applications, it is of utmost interest to be able to integrate on an oriented hypersurface, and notably on the boundary $\partial\Omega$ of a regular domain. As we have seen in Section 3, this requires the datum of

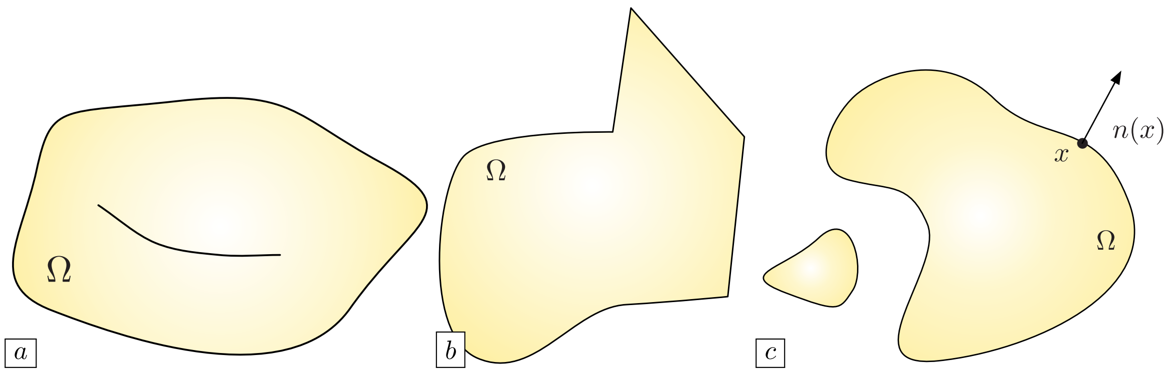

FIGURE 20. *(a) Ω is not a regular domain, since it is not locally on one side of its boundary; (b) Ω is not a regular domain, since it shows some sharp angles (corners): it is a Lipschitz domain; (c) Ω is a (non connected) domain of class $\mathcal{C}^1$.*

a suitable σ-algebra in $\mathcal{S}$, and of a (positive) measure on the induced measure space. The adequate notion to this end is that of *surface measure*, whose technical construction is omitted.

Definition 8.4. *Let $\mathcal{S}$ be an oriented hypersurface of class $\mathcal{C}^k$ in $\mathbb{R}^d$, for some $k \geq 1$. There exists a positive measure $\mathrm{d}s$ on the σ-algebra of $\mathcal{S}$ generated by its open subsets, which is called the surface measure on $\mathcal{S}$, and generalizes the notion of the length of a curve in $\mathbb{R}^2$ and surface in $\mathbb{R}^3$. In particular,*

- *The length $\ell(\gamma)$ of a curve $\gamma \subset \mathbb{R}^2$ equals*
$$\ell(\gamma) = \int_\gamma \mathrm{d}s.$$
- *The area $\mathcal{A}(\mathcal{S})$ of a piece of surface $\mathcal{S} \subset \mathbb{R}^3$ equals*
$$\mathcal{A}(\mathcal{S}) = \int_{\mathcal{S}} \mathrm{d}s.$$

Remark 8.2. *The above definitions may change slightly from one textbook to the other; in particular, domains are sometimes required to be connected subsets.*

We finally arrive at the main theoretical result of this section, which is the celebrated Green's formula.

Theorem 8.1 (Green's formula)**.** *Let $\Omega \subset \mathbb{R}^d$ be a bounded domain of class $\mathcal{C}^1$, and let u be a function of class $\mathcal{C}^1$ on $\overline{\Omega}$ (that is, u is the restriction to $\overline{\Omega}$ of a function of class $\mathcal{C}^1$ on an open neighborhood of $\overline{\Omega}$). For every index $i = 1, \ldots, d$, one has:*
$$\int_\Omega \frac{\partial u}{\partial x_i}\,\mathrm{d}x = \int_{\partial\Omega} u n_i \,\mathrm{d}s,$$
where $n = (n_1, \ldots, n_d)$ is the unit normal vector field to $\partial\Omega$, pointing outward Ω.

Let us provide a few useful alternative version of this formula:

Corollary 8.1. *Let $\Omega \subset \mathbb{R}^d$ be a bounded domain of class $\mathcal{C}^1$.*

- *Let $f : \overline{\Omega} \to \mathbb{R}^d$ be a vector field of class $\mathcal{C}^1$; it holds:*
$$\int_\Omega \mathrm{div}(f)\,\mathrm{d}x = \int_{\partial\Omega} \langle f, n\rangle \,\mathrm{d}s, \tag{8.2}$$
where the divergence of f is defined by $\mathrm{div}(f)(x) = \sum_{i=1}^d \frac{\partial f_i}{\partial x_i}(x)$, see Definition 2.6.
- *Let $f, g : \overline{\Omega} \to \mathbb{R}$ be two functions of class $\mathcal{C}^1$; then, for any index $i = 1, \ldots, d$, one has:*
$$\int_\Omega \frac{\partial f}{\partial x_i} g \,\mathrm{d}x = \int_{\partial\Omega} f g n_i \,\mathrm{d}s - \int_\Omega f \frac{\partial g}{\partial x_i}\,\mathrm{d}x. \tag{8.3}$$

Remark 8.3. *The Green's formula is the natural generalization of the idea of integration by parts to the case of functions of more than one real variable. Indeed, let $\Omega = (a, b)$ be an open interval of the real line, and let $f, g : [a, b] \to \mathbb{R}$ be two functions of class $\mathcal{C}^1$. The classical integration by parts formula*

$$\int_a^b f'(x)g(x)\,\mathrm{d}x = \Big(f(b)g(b) - f(a)g(a)\Big) - \int_a^b f(x)g'(x)\,\mathrm{d}x$$

can be rewritten in a form very close to (8.3), if we introduce the "normal vector" $n(b) = 1$, $n(a) = 1$ to $\partial\Omega$, pointing outward Ω, and if we agree that integrating on the "boundary" of (a, b) amounts to summing over its endpoints, we obtain:

$$\int_\Omega f'(x)g(x)\,\mathrm{d}x = \int_{\partial\Omega} f(x)g(x)n(x)\,\mathrm{d}s - \int_\Omega f(x)g'(x)\,\mathrm{d}x,$$

which is exactly the one-dimensional counterpart to (8.3).

8.2. Application: establishing the thermal conductivity equation

In this section, we exemplify how the notions introduced in the previous section are naturally involved in the mathematical modeling and treatment of physical phenomena. Here, we aim to derive the boundary-value problem governing the equilibrium state of the temperature inside a cavity from the physical conservation laws.

Let Ω be a bounded domain of class $\mathcal{C}^1$ of the Euclidean space $\mathbb{R}^d$; physically, Ω accounts for a cavity insulated from the outside, which is filled by a material with thermal conductivity $\gamma > 0$. Let $u : \Omega \to \mathbb{R}$ denote the temperature field within Ω; we aim to characterize it by means of a boundary-value problem, i.e. the combination of a partial differential equation inside Ω and a boundary condition. We proceed to this end under the assumption that u is of class $\mathcal{C}^2$ on $\overline{\Omega}$, so that all the forthcoming developments are legitimate.

The key ingredient in the modeling of thermal phenomena is the heat flux $j : \Omega \to \mathbb{R}^d$ within Ω; this vector field (that we also assume to be "regular enough") allows to calculate the amount of energy crossing any surface $\mathcal{S} \subset \Omega$, oriented by the normal vector n, via the formula:

$$\int_{\mathcal{S}} \langle j(x), n(x)\rangle\,\mathrm{d}s.$$

Note that, in particular, the quantity vanishes when the heat flux is tangential to the surface $\mathcal{S}$ (i.e. when $\langle j(x), n(x)\rangle = 0$ on $\mathcal{S}$).

The heat flux is related to the temperature u within Ω and the conductivity γ of the constituent material via the famous *Fourier's law*:

$$j(x) = -\gamma\nabla u(x), \;\; x \in \Omega.$$

Intuitively, this law encodes the fact that the heat flux is oriented from the regions of Ω with large temperature (the "hot" regions of Ω) to those with low temperature (the "cold" regions).

Let now $f : \Omega \to \mathbb{R}$ be the heat source at play in the medium. By definition, the amount of heat produced in an arbitrary subdomain $\omega \subset \Omega$ equals:

$$\int_\omega f(x)\,\mathrm{d}x.$$

We now suppose that the system is at equilibrium. This means that, for any subdomain $\omega \subset \Omega$, the amount of heat produced inside ω must be equal to that leaving ω; this corresponds to the equality:

$$\int_\omega f(x)\,\mathrm{d}x = \int_{\partial\omega} \langle j(x), n(x)\rangle\,\mathrm{d}s = -\gamma\int_{\partial\omega} \frac{\partial u}{\partial n}(x)\,\mathrm{d}s,$$

where we have introduced the *normal derivative* $\frac{\partial u}{\partial n} = \langle\nabla u, n\rangle$ of u.

Now using the version (8.2) of Green's formula, we obtain:

$$\int_\omega f(x)\,\mathrm{d}x = -\gamma\int_\omega \Delta u(x)\,\mathrm{d}x, \tag{8.4}$$

where we recall that the Laplace operator Δ is defined by $\Delta u = \mathrm{div}(\nabla u)$, see Definition 4.3. Since the relation (8.4) holds for all subsets $\omega \subset \Omega$, it follows that

$$-\gamma\Delta u = f \quad \text{in} \quad \Omega.$$

This is the expected partial differential equation governing u inside Ω.

We now seek to characterize the behavior of u on the boundary $\partial\Omega$. To this end, we rely on the assumption that Ω is insulated from the outer medium, which implies that the amount of energy $-\gamma \int_V \frac{\partial u}{\partial n}\,\mathrm{d}s$ crossing every subregion $V \subset \partial\Omega$ should vanish. Hence,

$$\gamma \frac{\partial u}{\partial n} = 0 \quad \text{on} \quad \partial\Omega.$$

Summarizing, we have shown that u satisfies the Laplace equation with homogeneous Neumann boundary conditions:

$$\begin{cases} -\gamma \Delta u = f & \text{in } \Omega, \\ \gamma \frac{\partial u}{\partial n} = 0 & \text{on } \partial\Omega. \end{cases}$$

8.3. **Exercises**

Exercise 8.1. Let $\Omega \subset \mathbb{R}^d$ be a bounded domain of class $\mathcal{C}^1$.

(i) Show that the volume of Ω equals:

$$|\Omega| = \int_\Omega \mathrm{d}x = \int_{\partial\Omega} x_i n_i \,\mathrm{d}s.$$

(ii) Show that the center of mass of Ω equals:

$$\frac{1}{|\Omega|} \int_\Omega x \,\mathrm{d}x = \frac{1}{|\Omega|} \int_{\partial\Omega} \sum_{i=1}^d x_i^2 n_i \,\mathrm{d}s.$$

[Remark: These formulas are handful in practice, since they allow to calculate volume quantities from the sole knowledge of the boundary $\partial\Omega$.]

References

[1] S. Benzoni-Gavage, *Calcul différentiel et équations différentielles-2e éd.: Cours et exercices corrigés*, Dunod, 2014.
[2] F. Bonnans, *Optimisation continue.*
[3] J.-M. Bony, *Cours d'analyse: théorie des distributions et analyse de Fourier*, Editions Ecole Polytechnique, 2001.
[4] H. Cartan, *Differential calculus*, Hermann, 1983.
[5] A. J. Chorin, J. E. Marsden, and J. E. Marsden, *A mathematical introduction to fluid mechanics*, vol. 3, Springer, 1990.
[6] J.-P. Demailly, *Analyse numérique et équations différentielles-4eme Ed*, EDP sciences, 2016.
[7] L. C. Evans, *Partial differential equations*, vol. 19, American mathematical society, 2022.
[8] D. J. Garling, *A Course in Mathematical Analysis*, vol. 3, Cambridge University Press, 2014.
[9] S. G. Krantz, *The integral: a crux for analysis*, Synthesis Lectures on Mathematics and Statistics, 4 (2011), pp. 1–105.