Topology and differential calculus

Pierre-Antoine Guihéneuf

26th January 2018

Contents

I Topology 2

1 Sets and maps 4

2 Normed vector spaces and their topology 7
- 2.1 Norms . . . 7
- 2.2 Open and closed sets . . . 9
- 2.3 Continuity . . . 10
- 2.4 Uniformly continuous and Lipschitz maps . . . 10
- 2.5 Denseness . . . 11
- 2.6 Sequential characterisations . . . 12
- 2.7 limsup, liminf . . . 13

3 Compactness 14
- 3.1 Definition and first properties . . . 14
- 3.2 Some applications . . . 16

4 Completeness 19
- 4.1 Definitions and first properties . . . 19
- 4.2 Applications . . . 20

II Differential calculus 24

5 Differentials 26
- 5.1 Differentiable maps . . . 26
- 5.2 C^1 maps . . . 29
- 5.3 Extrema: order 1 conditions . . . 30
- 5.4 Implicit function theorem . . . 32

6 Order two differentials 35
- 6.1 C^2 maps . . . 35
- 6.2 Local extrema: order two conditions . . . 37

Part I

Topology

From an etymological viewpoint, the topology is the study of the place. It studies mathematical objects up to continuous deformations (more precisely, up to homeomorphisms), it is sometimes called “mathematics of rubber.

Among others, some examples of use of topology:

- Classification of surfaces, and more generally classification of manifolds *via* algebraic topology. This can be used in AI to classify geometrical shapes, for example by using persistent homology.
- Proof of mathematical theorems by using the denseness of simple objects among big sets of more complicated objects. This is the fundamental idea of functional analysis; in particular it is used to define the Fourier transform on the space $L^2(\mathbb{R})$.
- Optimisation problems: the theory of compact spaces gives criteria of existence of optima.
- Definition of numbers and functions by the mean of series. The main example of this is the theory of power series, which is used for example to define the functions exp, sin, cos (and thus the number π). The matrix exponential is also defined from a series, it is the main ingredient for the resolution of linear differential equations.
- Fixed point theorems, and in particular that of Picard. From this theorem one gets proofs of theorems like Cauchy-Lipschitz, local inversion, etc. This fixed point theorem can also be applied to see if Newton method[1] converges or not.
- Knot theory, which has applications to dynamical systems, or to the study of the DNA’s structure.
- Number theory: one way to define p-adic fields $\mathbb{Q}_p$ is to endow $\mathbb{Q}$ with an appropriate metric and to take its completion with respect to this metric.

Two additional references:

- La première, en français : Mémo de topologie – Frédéric Le Roux `https://webusers.imj-prg.fr/~frederic.le-roux/enseignement.html`
- The second one, in English: Topology without tears – Sidney A. Morris `http://www.topologywithouttears.net/topbook.pdf`

[1]This consists in an algorithm which finds approximate solutions for equations $f(x) = 0$.

Chapter 1

Sets and maps

The goal of this small chapter is to revise some concepts of basic set theory.

Let us start with a very simple example of map.

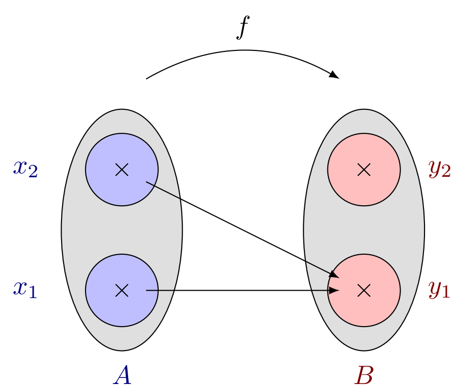


Figure 1.1: The map f

On this figure, A and B are two sets, each one having two elements: $A = \{x_1, x_2\}$ and $B = \{y_1, y_2\}$. The map f takes its arguments in the set A and gives outputs in the set B; it is defined by

$$\begin{array}{rcl} f : A & \to & B \\ x & \mapsto & y_1. \end{array}$$

We also denote $A_1 = \{x_1\}$, $A_2 = \{x_2\}$ (which are subsets of A), $B_1 = \{y_1\}$ and $B_2 = \{y_2\}$ (which are subsets of B). With this example we will play with image and inverse image sets.

Exercise 1. 1. Give the list of the subsets of A and the list of the subsets of B.

2. Determine all possible image sets for f. Determine all possible inverse image sets for f.
3. Determine if the following equalities hold:
   (a) $f\big(A_1 \cup A_2\big) = f\big(A_1\big) \cup f\big(A_2\big)$;
   (b) $f\big(A_1 \cap A_2\big) = f\big(A_1\big) \cap f\big(A_2\big)$;
   (c) $f^{-1}\big(B_1 \cap B_2\big) = f^{-1}\big(B_1\big) \cap f^{-1}\big(B_2\big)$;
   (d) $f^{-1}\big(B_1 \cup B_2\big) = f^{-1}\big(B_1\big) \cup f^{-1}\big(B_2\big)$;
   (e) $f^{-1}(f(A_2)) = A_2$;
   (f) $f(f^{-1}(B)) = B$.

It turns out that what happens on this very simple example is true in general, as explained by the following propositions. The first one treats the case of image sets.

Proposition 1.1. *Let A and B be two sets, and f be a map from A to B. Then for any subsets A_1, A_2 of A, one has:*

$$f\big(A_1 \cup A_2\big) = f\big(A_1\big) \cup f\big(A_2\big)$$

and

$$f\big(A_1 \cap A_2\big) \subseteq f\big(A_1\big) \cap f\big(A_2\big).$$

The second proposition treats the case of inverse image sets.

Proposition 1.2. *Let A and B be two sets, and f be a map from A to B. For any subsets B_1, B_2 of B, one has:*

$$f^{-1}\big(B_1 \cap B_2\big) = f^{-1}\big(B_1\big) \cap f^{-1}\big(B_2\big)$$

and

$$f^{-1}\big(B_1 \cup B_2\big) = f^{-1}\big(B_1\big) \cup f^{-1}\big(B_2\big).$$

This third and last proposition shows what happens when we play with both images and inverse images.

Proposition 1.3. *Let A and B be two sets, and f be a map from A to B. Let $A_1 \subset A$, then*

$$f^{-1}(f(A_1)) \supseteq A_1.$$

Moreover, if f is injective, then the equality holds: $f^{-1}(f(A_1)) = A_1$.

Let $B_1 \subset B$, then

$$f(f^{-1}(B_1)) \subseteq B_1.$$

Moreover, if f is surjective, then the equality holds: $f(f^{-1}(B_1)) = B_1$.

Supplementary exercises

Exercise 2. Let $f : \mathbb{R} \to \mathbb{R}$ the map defined by $f(0) = 0$ and $f(x) = \sin(1/x)$ if $x \neq 0$. Determine $f(]1/2, 2[)$ and $f^{-1}(]1/2, 2[)$.

Exercise 3. Prove that if f is a map from A to B and $A_1 \subset A$, $B_1 \subset B$, then

$$f^{-1}(B_1^{\complement}) = f^{-1}(B_1)^{\complement}.$$

Prove that if f is injective, then $f(A_1^{\complement}) \subseteq f(A_1)^{\complement}$, and if f is surjective, then $f(A_1^{\complement}) \supseteq f(A_1)^{\complement}$.

Chapter 2

Normed vector spaces and their topology

In all this chapter, all the considered vector spaces are real (i.e. $\mathbb{R}$*-vector spaces).*

2.1 Norms

On the line $\mathbb{R}$, the plane $\mathbb{R}^2$ or the space $\mathbb{R}^3$, there is a natural way to define a distance between two vectors by the mean of the Euclidean norm. For example, for $v = (x, y, z) \in \mathbb{R}^3$, this norm is defined as

$$\|v\|_2 = \sqrt{x^2 + y^2 + z^2}.$$

The *Euclidean distance* between two vectors u and v is then given by the norm of their difference: $\|u - v\|_2$. More generally, the vector space $\mathbb{R}^n$ is endowed with the Euclidean norm: for $v = (v_1, \dots, v_n) \in \mathbb{R}^n$, one has

$$\|v\|_2 = \sqrt{\sum_{i=1}^{n} v_i^2}$$

and this norm allows to measure distances between vectors.

However, in some circumstances, one would like to measure the distance between vectors in a different manner. For example in New-York, the good way to measure distances is to consider the “1-norm”, defined by

$$\|v\|_1 = |x| + |y|$$

for $v = (x, y)$. This distance is much more practical than the Euclidean distance for a New-Yorker to measure traveling time in the city (see Figure 2.1).

More generally, there is no “best” choice of a distance on a vector space. Rather, for each specific problem there can be more practical choices of distances than others. To be easily manipulated, these distances have to satisfy some good properties; classically one will use distances arising from a norm:

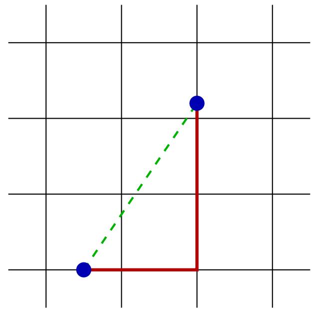


Figure 2.1: City map: in New-York, the good way to measure distance between two points is to measure the length of the smallest street path between these two points (red); this distance is usually greater then the Euclidean distance (greed, dotted).

Definition 2.1. Let V be a vector space. A *norm* on V is a map $N : V \to \mathbb{R}_+$ satisfying the following properties:

1. $\forall x \in V$, $N(x) = 0 \implies x = 0$ (separation);
2. $\forall x \in V$, $\lambda \in \mathbb{R}$, $N(\lambda x) = |\lambda| N(x)$ (absolute homogeneity);
3. $\forall x, y \in V$, $N(x + y) \leq N(x) + N(y)$ (triangle inequality).

The *distance* associated to a norm N on the vector space V is then the map

$$\begin{aligned} d_N : V \times V &\longrightarrow \mathbb{R}^+ \\ (x, y) &\longmapsto N(x - y). \end{aligned}$$

Exercise 4. Prove that the absolute value is a norm on $\mathbb{R}$. Reciprocally, prove that for any norm N on $\mathbb{R}$, there is a number $\mu > 0$ such that $N(x) = \mu|x|$ for any $x \in \mathbb{R}$.

Exercise 5. Prove that the maps $x \mapsto x^2$ and $x \mapsto 2x$ are not norms on $\mathbb{R}$:

Example 2.2. For any $n \in \mathbb{N}^*$ and any $p \in \mathbb{R}$, $p \geq 1$, one can define on the vector space $\mathbb{R}^n$ the "p-norm", by

$$\|v\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$

for $v = (x_1, \ldots, x_n)$.

Exercise 6. Prove that the map

$$\begin{array}{rccc} \| \cdot \|_\infty : & \mathbb{R}^n & \longrightarrow & \mathbb{R}_+ \\ & (x_1, \ldots, x_n) & \longmapsto & \max\big(|x_1|, \ldots, |x_n|\big) \end{array}$$

defines a norm on $\mathbb{R}^n$.

Exercise 7. Prove that for any vectors $x, y \in V$, one has:

$$\big|N(x) - N(y)\big| \leq N(x-y).$$

This is called the *inverse triangle inequality.*

2.2 Open and closed sets

For now on, V is a vector space endowed with a norm $\|\cdot\|$.

Definition 2.3. For $x \in V$ and $r > 0$, the *ball* of center x and radius r is the set of points whose distance to x is smaller than r:

$$B(x,r) = \big\{y \in V \mid \|x-y\| < r\big\}.$$

Exercise 8. In $\mathbb{R}^2$, draw the unit open balls (i.e. the balls $B(O,1)$, where O is the origin) for the norms $\|\cdot\|_1$, $\|\cdot\|_2$ and $\|\cdot\|_\infty$ (as defined in Example 2.2 and Exercise 6).

Definition 2.4. A subset O of V is *open* if any point of O is the center of a ball included in O:

$$\forall x \in O, \exists r > 0 : \; B(x,r) \subset O.$$

Let X be a subset of V. A subset O of X is *open in X* if

$$\forall x \in O, \exists r > 0 : \; B(x,r) \cap X \subset O.$$

Exercise 9. The line $\mathbb{R}$ is endowed with the distance given by the absolute value. Show that for any real numbers $a < b$ the interval $]a,b[$ is open. Show that the interval $]0,1]$ is not open but is open in $]-\infty,1]$.

Example 2.5. Any open ball is open.

Proposition 2.6. *Any finite intersection of open sets is open. Any union of open sets is open.*

Exercise 10. Show that the intersection

$$\bigcap_{n=1}^{\infty} \left]-\frac{1}{n}, \frac{1}{n}\right[$$

is not open in $\mathbb{R}$. Thus, an infinite intersection of open sets can be not open.

Definition 2.7. A subset F of V is called *closed* if its complementary is open.

Exercise 11. Prove that the segment $[0,1]$ is closed.

Be careful, some sets can be both closed *and* open (for example the whole vector space V), and many sets are neither closed nor open (for example $[0,1[$ in $\mathbb{R}$).

Exercise 12. Are the following subsets of $\mathbb{R}$ closed or open: $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R} \setminus \mathbb{Q}$?

2.3 Continuity

In this section, V and W are two vector spaces endowed respectively with the norms $\|\cdot\|_V$ and $\|\cdot\|_W$, and X is a subset of V.

Definition 2.8. A map $f : X \to W$ is said to be *continuous at* $x \in X$ if:

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

The map f is said to be *continuous* if it is continuous at every point $x \in X$, i.e.

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

Exercise 13. Write with quantifiers the fact that a map f is not continuous.

Prove that the map $\chi_{\mathbb{R}_+}$, defined on $\mathbb{R}$ by $\chi_{\mathbb{R}_+}(x) = 0$ if $x < 0$ and $\chi_{\mathbb{R}_+}(x) = 1$ if $x \geq 0$, is not continuous.

Reminder: the sum of to continuous functions with values in the same vectorial space is continuous; the product of two real continuous functions is continuous.

There is a practical characterisation of continuity in terms of open sets.

Proposition 2.9. *A map $f : X \to W$ is continuous if and only if the inverse image of any open subset of W is open in X.*

Equivalently, f is continuous if and only if the inverse image of any closed subset of W is closed in X.

This proposition can be used in two different manners. Of course, one can prove that a map is continuous by proving that the inverse image of every open set is open. One can also start from a map which we know to be continuous, and use it to prove that a subset O of X is open by showing that it is the inverse image of an open subset of W.

From this proposition we can deduce easily the following:

Proposition 2.10. *The composition of two continuous maps is continuous.*

In practical, when we have a map defined with a formula and want to prove it to be continuous, we combine this proposition with the usual properties stating that the product, sum, inverse, etc. of continuous maps is continuous.

Exercise 14. Prove that the set $GL_n(\mathbb{R})$ of invertible matrices is open in the set $M_n(\mathbb{R})$ of $n \times n$ matrices. Prove that the set $SL_n(\mathbb{R})$ of matrices with determinant 1 is closed in $M_n(\mathbb{R})$.

2.4 Uniformly continuous and Lipschitz maps

Here we define two notions that say that a map is "better than continuous": uniform continuity and Lipschitz continuity.

Definition 2.11. A map $f : X \to W$ is said to be *uniformly continuous* if:

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

This concept of uniform continuity in central in functional analysis. It can be used to prove some results of density among continuous maps; for example the fact that for any interval I, piecewise affine and continuous maps are dense among continuous maps from I to $\mathbb{R}$. It is also a key notion for defining Fourier transform of L^2 maps.

Exercise 15. Write with quantifiers the fact that a map f is not uniformly continuous.

Prove that the map

$$\begin{aligned} f : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto x^2 \end{aligned}$$

is continuous but not uniformly continuous.

Definition 2.12. Let $k > 0$, a map $f : X \to W$ is said to be *k-Lipschitz* if:

$$\forall x, y \in X, \ \|f(x) - f(y)\|_W \leq k\|x - y\|_V.$$

A particular class of Lipschitz maps – contractions, which are k-Lipschitz maps with $k < 1$ – will play a key role in the theory of complete spaces, where they will be proved to possess a fixed point. Also, we will see in the differential calculus part that a C^1 map is Lipschitz (via the mean value theorem); this fact is very important in analysis where it is used to bound norms of functions.

Exercise 16. Write with quantifiers the fact that a map f is not Lipschitz.

Prove that the map

$$\begin{aligned} f : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto \sqrt{x} \end{aligned}$$

is not Lipschitz. Prove that it is uniformly continuous.

Proposition 2.13. *Lipschitz maps are uniformly continuous. Uniformly continuous maps are continuous.*

2.5 Denseness

Definition 2.14. A set $D \subset X$ is said to be *dense in* X if any ball of X meets D:

$$\forall x \in X, \forall \varepsilon > 0, \ B(x, \varepsilon) \cap D \neq \emptyset.$$

Example 2.15. The set $\mathbb{Q}$ of rational numbers is dense in the set $\mathbb{R}$ of real numbers.

Example 2.16. The segment $[0, 1]$ is not dense in the segment $[-1, 2]$.

2.6 Sequential characterisations

In this section we give some equivalent characterisations of the previous notions in terms of sequences. As in the previous sections, V and W are two vector spaces endowed respectively with the norms $\|\cdot\|_V$ and $\|\cdot\|_W$, and X is a subset of V. We first recall the definition of a limit.

Definition 2.17. A sequence $(u_n)_{n\in\mathbb{N}} \in V^{\mathbb{N}}$ admits $\ell \in V$ as a limit if

$$\forall \varepsilon > 0,\ \exists N \in \mathbb{N} :\ \forall n \geq N,\ \|u_n - \ell\|_V < \varepsilon.$$

Exercise 17. Prove that if a sequence $(u_n)_{n\in\mathbb{N}} \in V^{\mathbb{N}}$ converges to ℓ_1 and ℓ_2, then $\ell_1 = \ell_2$.

Proposition 2.18. *A subset F of X is closed if and only if for any sequence $(u_n)_{n\in\mathbb{N}} \in F^{\mathbb{N}}$ of points of F which converges to $\ell \in X$, the limit ℓ is in F.*

Example 2.19. The set $\mathbb{Q}$ of rational numbers is not closed in $\mathbb{R}$, as there exists a sequence of rational numbers converging to $\sqrt{2} \notin \mathbb{Q}$ (for example the sequence of decimal approximations of $\sqrt{2}$).

Proposition 2.20. *A map $f : X \to F$ is continuous if and only if for any sequence $(u_n)_{n\in\mathbb{N}} \in X^{\mathbb{N}}$ converging to $\ell \in X$, the sequence $\big(f(u_n)\big)_{n\in\mathbb{N}}$ converges to $f(\ell)$.*

Proposition 2.21. *A subset D of X is dense in X if and only if for any point $x \in X$, there exists a sequence $(u_n)_{n\in\mathbb{N}} \in D^{\mathbb{N}}$ of points of D which converges to x.*

Example 2.22. The set $GL_n(\mathbb{R})$ of invertible matrices is dense in the set $M_n(\mathbb{R})$ of $n \times n$ matrices.

To see this, one can use the canonical form of matrices under equivalence: by a Gaussian elimination, one can prove that for any matrix $M \in M_n(\mathbb{R})$, there exists two invertible matrices $P, Q \in GL_n(\mathbb{R})$ such that

$$M = P \begin{pmatrix} 1 & & & & & \\ & \ddots & & & (0) & \\ & & 1 & & & \\ & & & 0 & & \\ & (0) & & & \ddots & \\ & & & & & 0 \end{pmatrix} Q.$$

Then, the sequence $(M_k)_{k\geq 1}$ defined by

$$M_k = P \begin{pmatrix} 1 & & & & & \\ & \ddots & & & (0) & \\ & & 1 & & & \\ & & & 1/k & & \\ & (0) & & & \ddots & \\ & & & & & 1/k \end{pmatrix} Q$$

is a sequence of invertible matrices which converges to M.

2.7 limsup, liminf

Let $(u_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ be a real sequence. There are a lot of examples for which such a sequence admit no limit in $+\infty$: take for example the sequence defined by $u_n = (-1)^n$. To bypass this problem, it is possible to define two numbers – called supremum and infimum limits – which bound away from above and below all the asymptotics of the sequence.

Definition 2.23. Let $(u_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ be a *real* sequence. The supremum limit of this sequence is the number (possibly equal to $+\infty$ or $-\infty$)

$$\limsup_{n\to+\infty} u_n = \lim_{n\to+\infty} \sup\{u_m \mid m \geq n\}.$$

Similarly, the infimum limit of this sequence is the number (possibly equal to $+\infty$ or $-\infty$)

$$\liminf_{n\to+\infty} u_n = \lim_{n\to+\infty} \inf\{u_m \mid m \geq n\}.$$

The crucial fact of this definition is that infimum and supremum limits are well defined. Indeed, the sequence $(\sup\{u_m \mid m \geq n\})_{n\in\mathbb{N}}$ is decreasing[1] and a decreasing sequence always has a well defined limit. Remark that the supremum limit can be equal to $+\infty$ (in the case where at each time, the supremum is equal to $+\infty$) or to $-\infty$ (in the case where the suprema form a sequence of numbers tending to $-\infty$).

Exercise 18. Let $(u_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ be defined as $u_n = (-1)^n$. Compute $\limsup_{n\to+\infty} u_n$ and $\liminf_{n\to+\infty} u_n$.

Proposition 2.24. *A real sequence $(u_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ converges to $\ell \in \mathbb{R} \cup \{\pm\infty\}$ if and only if $\limsup_{n\to+\infty} u_n = \liminf_{n\to+\infty} u_n = \ell$.*

Supplementary exercises

Exercise 19. Let V be a vector space, and $x \in V$, $R \geq 0$. Prove that any closed ball, defined by

$$\overline{B}(x, R) = \{y \in V \mid \|x - y\| \leq R\}$$

is closed. *Indication: use the sequential characterization of Proposition 2.18.*

Exercise 20. Show that if a sequence converges, then it is *bounded*, i.e. there exists $R > 0$ such that $u_n \in B(0, R)$ for any $n \in \mathbb{N}$.

[1] To see this, observe that $\{u_m \mid m \geq n+1\} \subset \{u_m \mid m \geq n\}$.

Chapter 3

Compactness

First, a warning: I will cheat a lot during this chapter, as we will not see any of the two classical definitions of compacity...

Throughout this chapter, V is a finite dimensional vector space equipped with a norm $\|\cdot\|$.

3.1 Definition and first properties

Definition 3.1. A set $K \subset V$ is called *compact* if any real continuous function $f : K \to \mathbb{R}$ is bounded and attains its bounds, i.e.:

$$\exists a, b \in K : \ \forall x \in K, f(a) \leq f(x) \leq f(b).$$

Exercise 21. Prove that any finite set is compact.

Example 3.2. The set $\mathbb{R}$ is not compact, as the map $\mathrm{Id}_{\mathbb{R}}$ is continuous but not bounded.

In fact, this definition is about a property we want to be satisfied by some sets, but is quite inconvenient to use. So we need some equivalent characterisations of compact sets.

Definition 3.3. A set $B \subset V$ is *bounded* if it is included in some ball centered at the origin:

$$\exists R > 0 : \ B \subseteq B(0, R).$$

Theorem 3.4. *Let V be a finite dimensional vector space. A set $K \subset V$ is compact if and only if it is closed and bounded.*

It is easy to see that a set which is not closed is not compact. It is a bit more complicated to prove that a set which is not closed is not compact, and even more complicated to prove that in finite dimension, a set which is closed and bounded is compact. One of the proofs of this statement uses supremum limits. We outline a proof of one of the implications of this theorem in the case where $V = \mathbb{R}$.

Proof. Let $K \subset V$ be a closed and bounded set; our goal is to prove that K is compact. Let $f : K \to \mathbb{R}$ be a continuous map. We know that there exists a sequence $(x_n)_{n\in\mathbb{N}} \in K^n$ of elements of K realizing the supremum of f: $\lim_{n\to+\infty} f(x_n) = \sup_{x\in K} f(x)$. We denote $\ell = \limsup_{n\to+\infty} x_n$. Then, it is possible to prove that $\lim_{n\to+\infty} f(x_n) = f(\limsup_{n\to+\infty} x_n) = f(\ell)$. Moreover, as K is closed and bounded, it is also possible to prove that $\ell \in \mathbb{R}$ and then that $\ell \in K$. This proves that $\sup_{x\in K} f(x) = f(\ell)$, with $\ell \in K$; thus for any $x \in K$ one has $f(x) \leq f(\ell)$. A similar argument with $\liminf$ instead of $\limsup$ shows that there exists $\ell' \in K$ such that for any $x \in K$, $f(x) \geq f(\ell')$. This proves that K is compact. □

Proposition 3.5. *Let F be a nonempty closed subset of $\mathbb{R}^n$ which is not bounded, and $f : F \to \mathbb{R}$ a continuous map such that*

$$\lim_{x\in F, \|x\|\to+\infty} f(x) = +\infty.$$

Then there exists $y \in F$ satisfying $f(y) = \inf_{x\in F} f(x)$.

In other words, f is bounded from below and attains this bound.

Proof. By the hypothesis on the limit of f, we know that

$$\forall A \in \mathbb{R}, \exists R > 0 : \ \forall x \in F, \|x\| \geq R \implies f(x) \geq A.$$

As F is nonempty, we can pick $x_0 \in F$ and choose $A = f(x_0) + 1$. This gives us $R_0 > 0$ such that if $x \in F$ satisfies $\|x\| \geq R_0$, then $f(x) \geq f(x_0) + 1$.

Now, the closed ball

$$B_f(0, R_0) = \{x \in \mathbb{R}^n \mid \|x\| \geq R_0\}$$

is closed (as the inverse image of $[R_0, +\infty[$ by the continuous map $\|\cdot\|$) and bounded (by the very definition of a bounded set). So the set $K = F \cap B_f(0, R_0)$ is also closed (as the intersection of two closed sets) and bounded, in other words it is compact.

Hence, the map $f_{|K}$ is bounded and attains its bounds, in particular it attains its infimum at at least one point $y \in K$. But $x_0 \in K$ (because every point $x \in F \setminus K$ satisfies $f(x) \geq f(x_0) + 1$), so this bound is less than or equal to $f(x_0)$. This means that $\inf_{x\in F} f(x) = \inf_{x\in K} f(x) = f(y)$ □

Exercise 22.
1. Application: Show that if ABC is a triangle of the plane, then there exists a point M of the plane minimizing the sum of the distances to A, B and C.
2. Application 2: Show d'Alembert-Gauss theorem: any nonconstant polynom $P \in \mathbb{C}[X]$ has a root in $\mathbb{C}$. *Indication: if $|P(z_0)| = \inf_{z\in\mathbb{C}} |P(z)| > 0$, build a complex number $z \in \mathbb{C}$ (close to z_0) such that $|P(z)| < |P(z_0)|$.*

Proposition 3.6. *The image of a compact set by a continuous map is compact.*

3.2 Some applications

Theorem 3.7 (Heine)**.** *Let W be a normed vector space, K a compact subset of V and $f : K \to W$ a continuous map. Then f is uniformly continuous.*

As an example of application of this theorem, the is the fact that the set of continuous and affine maps is dense in the set of continuous maps from $[0, 1]$ to $\mathbb{R}$.

Proof. Let $\varepsilon > 0$, and define the set

$$E = \big\{(x, y) \in K \times K \mid \|f(x) - f(y)\| \geq \varepsilon\big\}.$$

Let us first prove that this set is compact by proving that it is closed and bounded. It is closed, as it is the inverse image of the closed interval $[\varepsilon, +\infty[$ by the continuous map $(x, y) \mapsto \|f(x) - f(y)\|$. It is bounded as a subset of the bounded set $K \times K$ (each K is bounded because it is compact).

Now, consider the map $\varphi : E \to \mathbb{R},\ (x, y) \mapsto \|x-y\|$. This map is continuous, and E is compact, so it attains its bounds: there exists $(x_0, y_0) \in E$ such that

$$\varphi(x, y) \geq \varphi(x_0, y_0) \quad \text{for any} \quad (x, y) \in E. \tag{3.1}$$

But as $(x_0, y_0) \in E$, one has $\|f(x_0) - f(y_0)\| \geq \varepsilon$, so $f(x_0) \neq f(y_0)$, so $x_0 \neq y_0$ and hence $\varphi(x_0, y_0) = \|x_0 - y_0\| > 0$. Setting $\delta = \varphi(x_0, y_0) > 0$, (3.1) can be rewritten as $\|x - y\| = \varphi(x, y) \geq \delta$ for any $(x, y) \in E$.

This implies that if $\|x-y\| < \delta$, then $(x, y) \notin E$ and thus $\|f(x)-f(y)\| < \varepsilon$. As ε is arbitrary, one gets that:

$$\forall \varepsilon > 0, \exists \delta > 0 :\ \forall x, y \in K, \|x - y\| < \delta \implies \|f(x) - f(y)\| < \varepsilon,$$

which is the very definition of a uniformly continuous map. □

Definition 3.8. Let N and N' be two norms on V. We say that they are *equivalent* if there exists two positive numbers $0 < c < C$ such that for any $v \in V$, one has:

$$cN(v) \leq N'(v) \leq CN(v).$$

This usefulness of this definition comes from the fact that if two norms are equivalent, they will define the same open and closed sets, and a map which is continuous for one norm will be continuous for the other one.

Theorem 3.9. *If V is a finite dimensional vector space, then all the norms on V are equivalent.*

Thus, in a finite dimensional space, it is not necessary to precise the norm we use when we talk about open or closed sets, continuous maps

Proof. Let N be a norm on V. To prove the theorem it is sufficient to prove that N is equivalent to the infinite norm $\|\cdot\|_\infty$.

Let $(e_1, \dots, e_n)$ be a basis of V. Then, any vector $x \in V$ can be written as $v = \sum_{i=1^n} x_i e_i$, where the x_i's are real numbers. Hence, by triangle inequality, one gets

$$N(v) = N\left(\sum_{i=1^n} x_i e_i\right) \leq \sum_{i=1}^{n} N(x_i e_i) = \sum_{i=1}^{n} |x_i| N(e_i).$$

By definition of the infinite norm, each $|x_i|$ is smaller than $\|v\|_\infty$, thus, by denoting $C = \sum_{i=1}^n N(e_i)$, one gets

$$N(v) \leq \|v\|_\infty \sum_{i=1}^{n} N(e_i) = C\|v\|_\infty.$$

This proves two things. First, this gives one of the two inequalities we need to get the equivalence of norms. But second, this implies that the map $v \mapsto N(v)$ is continuous for the norm $\|\cdot\|_\infty$ (because it is C-Lipschitz).

As this map N is continuous, it is bounded on the unit sphere of the norm $\|\cdot\|_\infty$ (as this sphere is compact): there exists $a \in V$ with $\|a\|_\infty = 1$ such that for any $v \in V$ satisfying $\|v\|_\infty = 1$, one has $N(a) \leq N(v)$. As $\|a\|_\infty = 1$, one has $a \neq 0$ and so $c = N(a) > 0$. Now, take any $v \in V \setminus \{0\}$. The vector $w = \frac{1}{\|v\|_\infty} v$ satisfies $\|w\|_\infty = 1$, so $N(w) \geq c$, hence

$$N(v) = \|v\|_\infty N\left(\frac{1}{\|v\|_\infty} v) = \|v\|_\infty N(w)\right) \geq c\|v\|_\infty,$$

which gives us the missing inequality to prove the equivalence of norms N and $\|\cdot\|_\infty$. $\square$

Proposition 3.10. *Let V and W be two finite dimensional vector spaces of respective dimensions n and m, and with respective norms $\|\cdot\|_V$ and $\|\cdot\|_V$. Let also $L : V \to W$ be a linear map. Then L is Lipschitz (and thus continuous). Moreover, the following defines a norm on the $m \times n$-dimensional vector space made of linear maps*[1] *from V to W:*

$$|||L||| = \sup_{\|x\|_V = 1} \|L(x)\|_W.$$

Proof. To prove the continuity of L, we use equivalence of norms in finite dimension: it is sufficient to get that L is Lipschitz with respect to the infinite norm, and this fact is immediate when one writes the map L in coordinates.

This implies in particular that $|||L|||$ is finite for any linear map L. The proof of the fact that $|||\cdot|||$ is a norm is a simple verification. $\square$

[1] This set can be identified with $M_{m,n}(\mathbb{R})$.

Proposition 3.11. *This operator (or triple) norm is sub-multiplicative:*

$$|||MN||| \leq |||M|||\;|||N|||.$$

This property is crucial in the definition of the matrix exponential.

Proof. Let $M : W \to W'$ and $N : V \to W$, where V, W and W' are endowed respectively with norms $\|\cdot\|_V$, $\|\cdot\|_W$ and $\|\cdot\|_{W'}$. We compute:

$$|||MN||| = \sup_{\|x\|_V=1} \|MN(x)\|_{W'} = \sup_{\substack{\|x\|_V=1 \\ \|N(x)\|_W \neq 0}} \left(\left\| M\left(\frac{N(x)}{\|N(x)\|_W}\right) \right\|_{W'} \|N(x)\|_W \right).$$

The last equation is obtained by multiplying and dividing by $\|N(x)\|_W$. One can restrict the supremum to vectors x satisfying $\|N(x)\|_W \neq 0$ because vectors such that $\|N(x)\|_W = 0$ do not contribute to the supremum (they satisfy $MN(x) = 0$). As the supremum of a product is smaller than the product of the supremums (for nonnegative numbers), one gets

$$|||MN||| \leq \sup_{\substack{\|x\|_V=1 \\ \|N(x)\|_W \neq 0}} \left\| M\left(\frac{N(x)}{\|N(x)\|_W}\right) \right\|_{W'} \sup_{\substack{\|x\|_V=1 \\ \|N(x)\|_W \neq 0}} \|N(x)\|_W.$$

The second supremum is equal to $|||N|||$. For the first supremum, one notices that the vector $N(x)/\|N(x)\|_W$ is of norm 1, so by definition of the triple norm of M,

$$\left\| \|M\left(\frac{N(x)}{\|N(x)\|_W}\right) \right\|_{W'} \leq |||M|||.$$

So finally,

$$|||MN||| \leq |||M|||\;|||N|||.$$

□

Exercise 23. Consider the map

$$L \;:\; \mathbb{R}^2 \to \mathbb{R}^2,\; (x, y) \mapsto (x,\; 2y - x).$$

Compute $|||L|||$ for the norms $\|\cdot\|_1$ and $\|\cdot\|_\infty$.

Chapter 4

Completeness

Let V be a real vector space, endowed with a norm $\|\cdot\|$.

4.1 Definitions and first properties

Definition 4.1. A *Cauchy* sequence of V is a sequence $(u_n)_{n\in\mathbb{N}} \in V^{\mathbb{N}}$ satisfying:

$$\forall \varepsilon > 0,\ \exists N > 0 : \forall p, q \geq N,\ \|u_p - u_q\| \leq \varepsilon.$$

In some sense, a Cauchy sequence is a sequence whose terms are eventually very close one to each other.

Exercise 24. Prove that the sequence defined by $u_n = 1/n$ (for any $n \geq 1$) is Cauchy.

Exercise 25. Prove that any Cauchy sequence is bounded.

Proposition 4.2. *Any convergent sequence is Cauchy.*

The converse is not true in general. The sets where it is true are called complete.

Definition 4.3. A set $X \subset V$ is called *complete* if any Cauchy sequence of X converges in X. If the vector space V is itself complete, it is called a *Banach space.*

Example 4.4. The set $X =]0,1[$ is not complete. Indeed, take the sequence $u_n = 1/n$ for $n \geq 1$. Then this sequence converges to 0 in $\mathbb{R}$, in particular it is Cauchy. So it is Cauchy in $]0,1[$. However, it does not converge in $]0,1[$ (as it converges to $0 \notin]0,1[$).

The good point with this notion of completeness is that it encompasses all finite dimensional vector spaces.

Theorem 4.5. *Any closed subset of a finite dimensional vector space on $\mathbb{R}$ is complete.*

In particular, the line $\mathbb{R}$ is complete.

4.2 Applications

As a first application of the notion of completeness, we address the problem of convergence of series.

Definition 4.6. Let $(u_n)_{n\in\mathbb{N}} \in V^n$ be a sequence. The series $\sum_{n\geq 0} u_n$ is said *normally convergent* if the series of positive real numbers $\sum_{n\geq 0} \|u_n\|$ converges in $\mathbb{R}$ (that is, is $< +\infty$).

Proposition 4.7. *If V is a Banach space, then any normally convergent series is convergent.*

Proof. As V is complete, to prove that the series is convergent, it is sufficient to prove that the sequence of partial sums

$$U_k = \sum_{n=0}^{k} u_n \in V$$

is Cauchy, that is

$$\forall \varepsilon > 0,\ \exists N \in \mathbb{N} :\ \forall p, q \geq N,\ \|U_p - U_q\| \leq \varepsilon.$$

But for $p > q$, one has

$$U_p - U_q = \sum_{n=0}^{p} u_n - \sum_{n=0}^{q} u_n = \sum_{n=q+1}^{p} u_n,$$

thus

$$\|U_p - U_q\| = \left\| \sum_{n=q+1}^{p} u_n \right\| \leq \sum_{n=q+1}^{p} \|u_n\|. \tag{4.1}$$

But by hypothesis, the series $\sum_{n\geq 0} \|u_n\|$ converges in $\mathbb{R}$, so in particular it is Cauchy:

$$\forall \varepsilon > 0,\ \exists N \in \mathbb{N} :\ \forall p > q \geq N,\ \sum_{n=0}^{p} \|u_n\| - \sum_{n=0}^{q} \|u_n\| = \sum_{n=q+1}^{p} \|u_n\| \leq \varepsilon.$$

Combing this with (4.1), one gets that the sequence $(U_k)_{k\in\mathbb{N}}$ of partial sums is Cauchy, and thus converges. □

Exercise 26. By using an operator norm, prove that for any matrix $M \in M_n(\mathbb{R})$, its *exponential* is well defined:

$$\exp(M) = \sum_{n=0}^{\infty} \frac{1}{n!} M^n.$$

Exercise 27 (a simple criterion of convergence of a power series). Let $k > 0$ and $(u_n)_{n\in\mathbb{N}}$ a sequence of real numbers such that for any $n \geq 1$, one has $|u_{n+1}| \leq k|u_n|$.

1. Prove that for any $n \in \mathbb{N}$, one has $|u_n| \leq k^n |u_0|$.
2. Prove that the power series $\sum_{n \geq 0} u_n x^n$ converges for any $x \in \mathbb{R}$ satisfying $|x| < 1/k$.

As a second application of complete spaces, we now state the Picard's fixed point theorem.

Definition 4.8. A *contraction* is a map which is k-Lipschitz for some $k < 1$. A *fixed point* of a map $f : V \to V$ is a point $y \in V$ such that $f(y) = y$.

Theorem 4.9 (Picard, Banach)**.** *Let V be a Banach space and $f : V \to V$ be a contraction. Then f admits a unique fixed point. Moreover, for any $x_0 \in V$, the sequence defined by the recurrence relation $x_{n+1} = f(x_n)$ converges towards this fixed point.*

This theorem has numerous applications, among them Cauchy-Lipschitz, local inversion and implicit function theorems.

Proof. We first prove that f hat at least one fixed point.

Let $x_0 \in V$, and define the sequence $(x_n)_{n\in\mathbb{N}} \in V^n$ by the recurrence relation $x_{n+1} = f(x_n)$. Then, for any $n > 1$, as f is k-Lipschitz,

$$\|x_{n+1} - x_n\| = \|f(x_n) - f(x_{n-1})\| \leq k\|x_n - x_{n-1}\|.$$

Iterating this procedure, a simple recurrence shows that

$$\|x_{n+1} - x_n\| \leq k^n \|x_1 - x_0\|.$$

This shows that the series $\sum_{n\geq 0}(u_{n+1} - u_n)$ is normally convergent (as $k < 1$), thus (by Proposition 4.7) is convergent. But

$$\sum_{n=0}^{N}(u_{n+1} - u_n) = u_{N+1} - u_0,$$

so the sequence $(u_N)_{N\in\mathbb{N}}$ converges, to a point that we denote by $y \in V$.

Passing to the limit in the equality $u_{n+1} = f(u_n)$, one gets that $y = f(y)$: y is a fixed point for f.

Finally, we prove that f admits a unique fixed point. Suppose that y and z are fixed points of f. Then

$$\|y - z\| = \|f(y) - f(z)\| \leq k\|y - z\|.$$

This implies that $(1 - k)\|y - z\| \leq 0$ and so (as $k < 1$), $\|y - z\| \leq 0$. As $\|y - z\| \geq 0$, one gets $\|y - z\| = 0$ and the separation axiom then implies that $y = z$. $\square$

A small remark: by the proof, one can see that

$$\|u_n - y\| = \lim_{N\to+\infty} \left\| \sum_{k=n}^{N} u_{n+1} - u_n \right\| \leq \sum_{k=n}^{\infty} k^n \|x_1 - x_0\| = \frac{k^n}{1-k}\|x_1 - x_0\|,$$

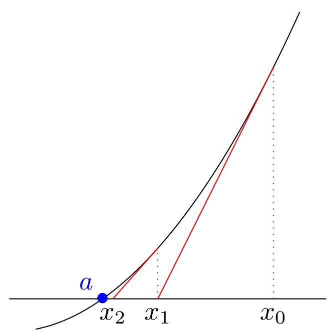
so the sequence $(x_n)_{n\in\mathbb{N}}$ converges exponentially fast towards y.

Exercise 28. Let $f : I \to \mathbb{R}$ a map of class C^1 on an open interval I, and $a \in I$ a fixed point of f. We suppose that $|f'(a)| < 1$. Show that there exists a closed interval $J \subset I$, stable by f (i.e. $f(J) \subset J$) and containing a, such that for any $x_0 \in J$, the sequence defined by the recurrence relation $x_{n+1} = f(x_n)$ converges to a.

Exercise 29 (Newton's method). Let $c < d \in \mathbb{R}$, and $f : [c, d] \to \mathbb{R}$ be a C^2 map. We suppose that $f(c) < 0 < f(d)$ and $f'(x) > 0$ for every $x \in [c, d]$.

1. Prove that f has a unique zero in the interval $[c, d]$, i.e. a point $a \in [c, d]$ such that $f(a) = 0$.

The goal of this exercise is to state a numerical algorithm that finds quickly approximations of the zero a. To do this, we define a sequence $(x_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ by recurrence in a geometric manner: x_{n+1} is the intersection between the tangent to the curve of f at the point $(x_n, f(x_n))$ and the axis $y = 0$.



2. Prove that
$$x_{n+1} = F(x_n) \quad \text{with} \quad F(x) = x - \frac{f(x)}{f'(x)}.$$

At his point, a small computation leads to $F'(a) = 0$, so it is possible to apply Exercise 28 (and so Picard's theorem) and deduce that for any x_0 sufficiently close to a, the sequence $(x_n)_{n\in\mathbb{N}}$ converges to a. In the following questions we specify the speed of this convergence.

3. By applying Taylor formula to f between a and x, show that for any $x \in [c, d]$, one has[1]
$$|F(x) - a| \leq \frac{1}{2} \frac{\sup_{z\in[(a,x)]} |f''(z)|}{f'(x)} (x - a)^2.$$

[1]Where $[(a, x)] = [a, x]$ if $a \leq x$ and $[(a, x)] = [x, a]$ if $x \leq a$.

Deduce that there exists $C > 0$ such that for any $x \in [c, d]$, one has

$$\big|F(x) - a\big| \le C|x - a|^2.$$

4. Show that there exists $\alpha > 0$ such that the segment $[a - \alpha, a + \alpha]$ is stable by F (i.e. $F([a - \alpha, a + \alpha]) \subsetneq [a - \alpha, a + \alpha]$).
5. Finally, show that for any $x_0 \in [a - \alpha, a + \alpha]$, one has

$$C|x_n - a| \le (C\alpha)^{2^n}.$$

We say that the convergence of x_n towards a is *of order 2* (or *quadratic*).

This algorithm is implemented in most of programing languages as the basic one to approximate solutions of equations $f(x) = 0$. It can be generalised to higher dimensional cases.

Part II

Differential calculus

Pour cette partie, deux références en français : la première est la suite de la référence pour la topologie : Mémo de calcul différentiel – Frédéric Le Roux `https://webusers.imj-prg.fr/~frederic.le-roux/enseignement.html`
La seconde est un excellent livre contenant beaucoup d'exercices : Petit guide de calcul différentiel à l'usage de la licence et de l'agrégation – François Rouvière.

And one reference in English: Differential Calculus – Pierre Schapira `https://webusers.imj-prg.fr/~pierre.schapira/lectnotes/CalD.pdf`

Chapter 5

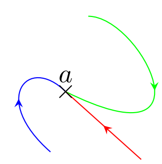
Differentials

5.1 Differentiable maps

One would like to generalize to higher dimensional cases the notion of derivative for a map $f : \mathbb{R} \to \mathbb{R}^p$:

$$\lim_{h \to 0} \frac{f(a+h) - f(a)}{h}.$$

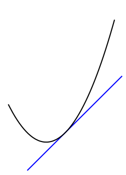
If we try to apply this formula to a map $f : \mathbb{R}^n \to \mathbb{R}^p$, we are lead to consider small *vectors* $h \in \mathbb{R}^n$. The problem is that in $\mathbb{R}^n$, there are many ways to approach a. For example on the figure, if the above limit exists for h going to a along each path, these three limits could be different (and they are in most of the cases).



As a first attempt to try generalizing the derivatives to higher dimensional cases, it is possible to define partial derivatives:

Definition 5.1. Let $(e_1, \dots, e_n)$ be a basis of $\mathbb{R}^n$ and $f : U \to \mathbb{R}^p$, where U is an open subset of $\mathbb{R}^n$. We say that f admits a *partial derivative* at $a \in U$ with respect to the i-th variable if the map $\mathbb{R} \to \mathbb{R}^p$, $t \mapsto f(a + te_i)$ is differentiable at 0. In this case, its derivative is denoted by $\frac{\partial f}{\partial x_i}(a)$.

The problem with this definition is that it does not take into account all the possible manners for h to tend to 0 (as in the previous picture).



The solution to this problem is to consider derivatives from a different viewpoint: for a map $f : \mathbb{R} \to \mathbb{R}$, the derivative $f'(a)$ represents the slope of the line that fits the best the curve of f at the point $(a, f(a))$. In other words, close to a, the map f is close to the affine map $x \mapsto f(a) + (x - a)f'(a)$, and one writes a *Taylor expansion of order 1*:

$$f(a+h) = f(a) + hf'(a) + o(h),$$

where $o(h)$ is a function which is negligible in comparison with $\|h\|$ when h tends to 0.

This is the great idea of differential calculus: approach a map by an affine one. This allows us to reduce the study of the initial map (in general difficult) to the one of an affine map (which is in general very simple). In higher dimensions, this gives the following definition:

Definition 5.2. Let U be an open set of $\mathbb{R}^n$, and $f : U \to \mathbb{R}^p$. We say that f is *differentiable* at $a \in U$ if there exists a linear map $L : \mathbb{R}^n \to \mathbb{R}^p$ such that

$$f(a+h) = f(a) + L(h) + o(h).$$

The map L is unique; it is called the *differential* of f at the point a, and is denoted by $Df(a)$. So we have:

$$f(a+h) = f(a) + Df(a).h + o(h).$$

If a map is differentiable at every point $a \in U$, we say that it is *differentiable in* U.

We recall that $o(h)$ is a notation for a map satisfying:

$$\lim_{h\to 0, h\neq 0} \frac{\|o(h)\|}{\|h\|} = \lim_{h\to 0, h\neq 0} \frac{\|f(a+h) - f(a) - L(h)\|}{\|h\|} = 0.$$

Exercise 30. Show that the differential of a map, if it exists, is unique.

Remark 5.3. A priori, the notion of differentiability could depend on the norms chosen for the spaces $\mathbb{R}^n$ and $\mathbb{R}^p$. However, as all norms are equivalent in finite dimension, this is not the case: a map differentiable for a choice of norms will also be differentiable for another choice of norms. This property is no longer true in the infinite dimensional case.

Proposition 5.4. *A map which is differentiable at $a \in U$ is continuous at a.*

Example 5.5. Let $L : \mathbb{R}^n \to \mathbb{R}^p$ be an linear map, $b \in \mathbb{R}^p$ and $f(x) = b + L(x)$ be the corresponding affine map. Then f is differentiable and its differential is everywhere equal to L.

Exercise 31. Let $f : \mathbb{R} \to \mathbb{R}$ be a differentiable function. Express the differential of f in terms of the derivative of f.

Exercise 32. The formula $f(x,y) = xe^{3y}$ defines a map from $\mathbb{R}^2$ to $\mathbb{R}$. By using a Taylor expansion, prove that f is differentiable at the point $(2,1)$.

Exercise 33. Let $|||\cdot|||$ be a matrix norm on $M_n(\mathbb{R})$ (endowed with an operator norm). Prove that if $H \in M_n(\mathbb{R})$ satisfies $|||H||| < 1$, then one has

$$(I_n + H)^{-1} = \sum_{k=0}^{\infty} (-H)^k.$$

Deduce that the differential of the map $f : M \mapsto M^{-1}$ at I_n is $-\mathrm{Id}$.

We recall that we have seen in Exercise 14 that $\mathrm{GL}_n(\mathbb{R})$ is an open subset of $M_n(\mathbb{R})$. With the formula $(A+H)^{-1} = A^{-1}(I_n + HA^{-1})^{-1}$, deduce the differentiability and the differential of f at any point $A \in \mathrm{GL}_n(\mathbb{R})$.

A map f which is differentiable at $a \in U$ possesses a directional derivative in every direction at a, that is, for any $v \in \mathbb{R}^n$:

$$\lim_{t\to 0, t\neq 0} \frac{f(a+tv)-f(a)}{t} = Df(a).v.$$

In particular, when $(e_1, \ldots, e_n)$ is a basis of $\mathbb{R}^n$, the i-th partial derivative is the directional derivative in the direction e_i, and one has:

$$\frac{\partial f}{\partial x_i}(a) = Df(a).e_i.$$

So when $v = \sum_{i=1}^n v_i e_i$ is a vector of $\mathbb{R}^n$, as $Df(a)$ is linear, one has

$$Df(a).v = \sum_{i=1}^n v_i Df(a).e_i = \sum_{i=1}^n v_i \frac{\partial f}{\partial x_i}(a).$$

If $(e'_1, \ldots, e'_p)$ is a basis of $\mathbb{R}^p$, the *Jacobian matrix* of f at a, denoted by $Jf(a)$, is the matrix representation of $Df(a)$ in the bases (e_i) and (e'_j). What we have just seen implies that

$$Jf(a) = \begin{pmatrix} \frac{\partial f}{\partial x_1} & \cdots & \frac{\partial f}{\partial x_n} \end{pmatrix}.$$

In practical, if we know that a map is differentiable, this allows us to compute in practical the differential. However, there exist examples of maps admitting partial derivatives but not differentiable.

Exercise 34. Let f be the map:

$$\begin{aligned} f \quad : \mathbb{R}^2 &\longrightarrow \mathbb{R} \\ (x,y) &\longmapsto \begin{cases} 0 & \text{if } (x,y) = (0,0) \\ \frac{xy^2}{x^2+y^2} & \text{if } (x,y) \neq (0,0). \end{cases} \end{aligned}$$

Prove that f is continuous at $(0,0)$, that it admits directional derivatives for any direction, but that it is not differentiable at $(0,0)$.

So we need to get some tools to prove that a map is differentiable or not.

Proposition 5.6. *If $f_1, f_2 : U \to \mathbb{R}^p$ are differentiable at $a \in U$, then:*

- *f_1+f_2 is differentiable at a and one has $D(f_1+f_2)(a) = Df_1(a)+Df_2(a)$;*
- *if $W = \mathbb{R}$, then $f_1 f_2$ is differentiable at a and one has $D(f_1 f_2)(a) = f_1(a).Df_2(a) + f_2(a).Df_1(a)$.*

If $f : U \to W$ can be decomposed into two components: $f = (f_1, f_2)$, and if f_1 and f_2 are differentiable at $a \in U$, then $Df(a) = (Df_1(a), Df_2(a)$.

Proposition 5.7. *Let $f : U \to \mathbb{R}^p$ and $g : U' \to \mathbb{R}^{p'}$, where U' is an open subset of $\mathbb{R}^p$. If f is differentiable at $a \in U$ and g is differentiable at $f(a) \in U'$, then $g \circ f$ is differentiable at a and one has:*

$$D(g \circ f)(a) = Dg(f(a)).Df(a).$$

In other words the differential at a of the composition of g and f is the compositions of the differentials of g at $f(a)$ and of f at a.

Exercise 35. Let $f : \mathbb{R}^2 \to \mathbb{R}^3$ defined by $f(x, y) = (\sin(x + y), 2xy^2, y)$.

1. (a) Compute the partial derivatives $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}$ at a point (x_0, y_0) and write the Jacobian matrix.

   (b) Recall the link between partial derivatives and differential, and give the value of $Df(x_0, y_0)(\vec{h})$ for any vector $\vec{h} = (h_x, h_y)$.

   (c) Write the approximation $f(x_0 + \vec{h})$ given by the differential.

2. Same questions for the map defined by $g(a, b, c) = (2a + b^2c, ae^b)$.

3. How can we get the matrix of the linear map $D(f \circ g)$ from those of Df et Dg ?

Exercise 36. Let I be an open interval of $\mathbb{R}$ and $f : I^2 \to \mathbb{R}$ a differentiable map. Show that the map $g : I \to \mathbb{R}$ defined by $g(x) = f(x, x)$ is differentiable, and compute its derivative in terms of the partial derivatives of f.

5.2 C^1 maps

Theorem 5.8 (Mean value theorem)**.** *Let $U \subset \mathbb{R}^n$ be an open set and $f : U \to \mathbb{R}^p$ be a differentiable map. Suppose that $|||Df(x)||| \leq M$ for any $x \in U$. Then for any $a, b \in U$ such that $[a, b] \subset U$, we have*

$$\|f(b) - f(a)\| \leq M\|b - a\|.$$

In particular, if U is convex, then f is k-Lipschitz, with $k = \sup_{x \in U} |||Df(x)|||$.

Exercise 37. Show that the following system has a unique solution:

$$\begin{cases} x = \frac{1}{3}\cos(x + y) \\ y = \frac{1}{3}\sin(x - y) \end{cases}$$

Indication: use mean value theorem, and apply Picard's fixed point theorem.

Definition 5.9. A map $f : U \to \mathbb{R}^p$ is called C^1 if is is differentiable on U, and moreover the map $X \mapsto Df(x)$ is continuous (from the open set U to the set of linear maps from V to W).

It is straightforward that the sum, the product and the composition of C^1 maps is itself C^1.

Theorem 5.10. *The map f is of class C^1 if and only if*

- *f has partial derivatives at every point $x \in U$, and*
- *these partial derivatives are continuous.*

5.3 Extrema: order 1 conditions

If $f : U \to \mathbb{R}$ is differentiable at $a \in U$, it is possible to define the *gradient vector* of f at a as:

$$\nabla f(a) = \left(\frac{\partial f}{\partial x_1}, \cdots, \frac{\partial f}{\partial x_n} \right).$$

So in this case we have

$$Df(a).h = \langle \nabla f(a), h \rangle.$$

To summarize, $Df(a)$, $Jf(a)$ and $\nabla f(a)$ represent the same object, but the first one is a linear form, the second is a matrix and the last one is a vector.

The gradient vector $\nabla f(a)$ represents the direction of "greatest slope" at a. It is orthogonal to level sets of the function f.

Definition 5.11. A *level set* of a map $F : U \to \mathbb{R}$ is a set of the form $f^{-1}(c)$ for some $c \in \mathbb{R}$.

In general, level sets of a map from $\mathbb{R}^2$ to $\mathbb{R}$ are paths, while level sets of a map from $\mathbb{R}^3$ to $\mathbb{R}$ are surfaces. This is rigorously expressed by the *implicit function theorem*.

Exercise 38. Represent the lines of level c of the following maps (when they exist!):

1. $f(x, y) = \ln(x + y)$ for $c = -1, 0, 1, 2$;
2. $f(x, y) = x^2 + y^2$ for $c = -1, 0, 1, 2$.

On each picture, draw one gradient vector on one point of each level line.

Definition 5.12. Let $f : U \to \mathbb{R}$ be a map defined on an open set, and $a \in U$ a point where f is differentiable. The point a is called a *critical point* if $Df(a) = 0$.

Theorem 5.13. *Let $f : U \to \mathbb{R}$ be a map defined on an open set, and $a \in U$ a point where f is differentiable. If f admits a local extremum at a, then a is a critical point.*

Exercise 39. Consider the map f from $\mathbb{R}^2$ to $\mathbb{R}$ defined by

$$f(x, y) = x^5 - x^2y + y.$$

Show that f is of class C^1 on $\mathbb{R}^2$ and compute its Jacobian matrix and its gradient. Determine the critical points of f.

Exercise 40. Show that the map $f : \mathbb{R}^2 \to \mathbb{R}$, defined by

$$f(x, y) = x^2 + y^4 - e^{-y^2} + e^{-x^2},$$

admits a global minimum, and determine it.

Same question for the map $f : \mathbb{R} \times \mathbb{R}_+^* \to \mathbb{R}$ defined by

$$f(x, y) = y(x^2 + (\ln y)^2).$$

Same question for the map $f : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f(x, y) = \sin x \sin y \sin(x + y).$$

Exercise 41 (Fermat Point, continuation of Exercise 22). Let ABC be a non flat plane triangle. We want to find the minimum in the plane of the map

$$f(M) = MA + MB + MC.$$

1. Prove that the map $x \mapsto \|x\|_2$ is differentiable on $\mathbb{R}^2 \setminus \{0\}$. Deduce that f is differentiable on $\mathbb{R}^2 \setminus \{A, B, C\}$.
2. Prove that if $P \notin \{A, B, C\}$ is a local extremum of f, then

$$\frac{\vec{PA}}{PA} + \frac{\vec{PB}}{PB} + \frac{\vec{PC}}{PC} = 0$$

3. Deduce that $\widehat{APB} = \widehat{BPC} = \widehat{CPA} = 2\pi/3$.
4. Using the fact that for any angle α, the set of points M of the plane such that $\widehat{AMB} = \alpha$ is an arc of circle passing through A and B, prove that f has at most one local extremum different from A, B and C.
5. (difficult) Prove that if all angles of ABC are smaller than $2\pi/3$, then A, B and C are not local minima of f, and thus that f has a unique minimum, lying in the interior of ABC.

We now come to the problem of optimization *under constraints*.

Let $U \subset \mathbb{R}^n$ be an open set and $\varphi_1, \cdots, \varphi_k$ be maps from U to $\mathbb{R}$. We consider the set

$$S = \big\{x \in U \mid \varphi_1(x) = \cdots = \varphi_k(x) = 0\big\}.$$

We also consider a map $f : U \to \mathbb{R}$, and our aim is to find the extrema of $f_{|S}$, in other words we want to maximize or minimize the quantity $f(x)$ for $x \in S$.

Definition 5.14. A point $x \in S$ is called *regular* if the gradients $\nabla\varphi_1(x), \cdots, \nabla\varphi_k(x)$ form a free family of $\mathbb{R}^n$.

This definition is motivated by the implicit function theorem: around a regular point, the set S is in fact "locally as $\mathbb{R}^{n-k}$". The dimension $n - k$ comes from the heuristic: "we are in $\mathbb{R}^n$ so we have n variables, and there are k equations $\varphi_i(x) = 0$, thus the set of solutions is of dimension $n - k$.

For example, around a regular point,

- if $n = 2$ and $k = 1$ or if $n = 3$ and $k = 2$, the set S is locally a curve;
- if $n = 3$ and $k = 1$, the set S is locally a surface;

The Lagrange multipliers theorem[1] gives a necessary condition for a regular point of S to be a extremum of the map f.

[1] Appelé "théorème des extrema liés" en français.

Theorem 5.15 (Lagrange)**.** *Suppose that $a \in S$ is a regular point of S at which f is differentiable. If a is a local extremum of $f_{|S}$, then there exists numbers $\lambda_1, \cdots, \lambda_k \in \mathbb{R}$ such that*

$$\nabla f(a) = \lambda_1 \nabla \varphi_1(a) + \cdots + \lambda_k \nabla \varphi_k(a).$$

In other words, the gradient $\nabla f(a)$ belongs to the subspace spanned by the gradients $\nabla \varphi_1(a), \cdots, \nabla \varphi_k(a)$. This condition which can seem a bit obscure is the algebraic reformulation of a geometric property: at a local extremum a of $f_{|S}$, the level set $\{x \in U \mid f(x) = f(a)\}$ is tangent to the surface S; in other words the gradient of f is orthogonal to S.

In practical problems, we usually use both the condition given by Lagrange multipliers and the conditions $\varphi_i(x) = 0$.

Exercise 42. Let α and β be two real numbers. Using Lagrange multipliers, determine the maximum and the minimum on the unit circle $\mathbf{S}^1$ of $\mathbb{R}^2$ of the map $f(x, y) = \alpha x + \beta y$.

Exercise 43. Determine the extrema of the map $f : \mathbb{R}^3 \to \mathbb{R}$ defined by

$$f(x, y, z) = \frac{13}{2} x^2 + 3y - 4z$$

on the unit sphere of $\mathbb{R}^3$.

Exercise 44. Let $f : \mathbb{R}^n \to \mathbb{R}$ be defined by $f(x) = x_1 x_2 \dots x_n$, and X be the set

$$X = \{(x_1 \cdots, x_n) \in \mathbb{R}^n_+ \mid x_1 + \cdots + x_n = n\}.$$

1. Determine the maximum of f on X.
2. Deduce from it the inequality between arithmetic and geometric means: for any $(x_1, \dots, x_n) \in \mathbb{R}^n_+$,

$$\sqrt[n]{x_1 \dots x_n} \leq \frac{x_1 + \cdots + x_n}{n}.$$

5.4 Implicit function theorem

In this section, we give a flavor of the important applications of differential calculus to geometry.

In lot of problems, the parameter space is not an open subset of $\mathbb{R}^n$ but rather a subset of $\mathbb{R}^n$ which locally "look like" $\mathbb{R}^d$, with $d \leq n$. For example, the sphere – which is a subspace of $\mathbb{R}^3$ – look locally like the plane $\mathbb{R}^2$ (so much that for a long time, a lot of people have thought that the Earth was flat). Also, the parameter space of a system depending on two angles is a 2-dimensional torus, which again locally look like $\mathbb{R}^2$.

The problem is that for now, we can only use differentials with maps having an *open* domain. The definitions of *submanifold* and then *manifold* are done to solve it: a submanifold of $\mathbb{R}^n$ is a set which locally look like $\mathbb{R}^d$, for $d \leq n$; it is possible to locally rectify these sets to make them correspond to an open subset of $\mathbb{R}^n$.

One of the grounding results of this theory is the implicit function theorem; it says that under some natural conditions, a subset of $\mathbb{R}^{n+p}$ defined by p equations can be seen as the graph of a map from an open subset of $\mathbb{R}^n$ to $\mathbb{R}^p$.

Theorem 5.16 (Implicit function)**.** *Let U be an open subset of $\mathbb{R}^n \times \mathbb{R}^p$ and $f : U \to \mathbb{R}^p$ a C^1 map. Suppose that there exists $(a,b) \in U$ such that*

- *$f(a,b) = 0$;*
- *$D_y f(a,b)$ is invertible (in other words, the matrix formed by the partial derivatives $\frac{\partial f}{\partial x_{n+1}}(a,b), \dots, \frac{\partial f}{\partial x_{n+p}}(a,b)$ has nonzero determinant).*

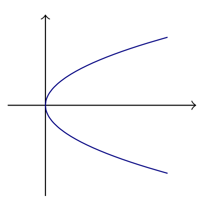
Then the equation $f(x,y) = 0$ can be solved locally with respect to the y variable: there exists a neighbourhood V of a in $\mathbb{R}^n$ and a neighbourhood W of b in $\mathbb{R}^p$, with $V \times W \subset U$, and a unique map $\varphi : V \to W$ such that, for any $(x,y) \in V \times W$,

$$f(x,y) = 0 \qquad \Longleftrightarrow \qquad y = \varphi(x).$$

Moreover, φ is of class C^1 on V.

For example, if $f : \mathbb{R}^2 \to \mathbb{R}$, and (a,b) is such that $f(a,b) = 0$ and $\partial f/\partial y(a,b) \neq 0$, then there exists a map φ such that for (x,y) close to (a,b), one has $f(x,y) = 0$ if and only if $y = \varphi(x)$.

This theorem can be visualized geometrically. The set defined by the cancellation of f is $S = \{(x,y) \in U \mid f(x,y) = 0\}$ (remark that it is the same set as for the Lagrange multipliers theorem!). The implicit function theorem says that under some condition, this set is the graph of a map φ.



Notice that the condition of the theorem is very natural. Take for example the subset of $\mathbb{R}^2$ defined by the equation $f(x,y) = x^2 - y = 0$. This curve is locally a graph $y = \varphi(x)$ around each of its points *except from* the point $(0,0)$ (if it was then some points should have two images by φ). This point is "bad" because the tangent to the curve is vertical. Implicit function theorem says that if this tangent is not vertical, then indeed the curve is locally the graph of some map.

In the case where the hypothesis of invertibility of the matrix is not satisfied, it is possible to permute the coordinates of $\mathbb{R}^{n+p}$ to try to apply the theorem. For instance, in the previous example, the curve cannot be defined as $y = \varphi(x)$ around $(0,0)$, but it can be defined as $x = \phi(y)$.

Exercise 45. Let $F(x,y) = x^2 + y^4 - 3xy + x$.

1. Compute and represent the gradient vector at the point $(2,1)$. What can be deduced for the level line
$$L_1 = \{(x,y) | F(x,y) = 1\} \ ?$$
2. Show that the equation $x^2 + y^4 - 3xy + x = 1$ defines implicitly y as a map of x in a neighbourhood of $(2,1)$.

3. Differentiate the equation $F(x, \varphi(x)) = 0$. Deduce the derivative $\varphi'(2)$.

This is a general procedure to find the differential of the map φ; it gives:

Proposition 5.17. *Under the hypothesis of implicit function theorem, one has*

$$D\varphi(a) = -\big(D_y f(a, b)\big)^{-1} \circ D_x f(a, b).$$

Exercise 46. Same exercise as Exercise 45, but with the equation $x^5 + 3xy - y^6 = 1$ and the point $(1, 0)$.

Exercise 47. Show that the equation $xy + yz + xz + 2x + 2y - z = 0$ defines implicitely a map $(x, y) \longmapsto z = \varphi(x, y)$ in a neighbourhood of $(0, 0, 0)$. Compute the differential of this map at the point $(0, 0)$.

Chapter 6

Order two differentials

6.1 C^2 maps

Let U be an open subset of $\mathbb{R}^n$ and $f : U \to \mathbb{R}^p$. Our goal in this chapter is to define second order differential and to use it to state order two conditions for extrema: we will get necessary and sufficient conditions for critical points to be local minima or maxima.

Definition 6.1. We say that f is of class C^2 if it is differentiable and if its differential $x \mapsto Df(x)$ is a C^1 map.

This definition is a bit more complicated that it seems. For every $x \in U$, the differential $Df(x)$ is a linear map from $\mathbb{R}^n$ to $\mathbb{R}^p$, i.e. $Df(x) \in L(\mathbb{R}^n, \mathbb{R}^p)$, with $L(\mathbb{R}^n, \mathbb{R}^p)$ a vector space of dimension np. Hence, $D(Df)(x)$ is a linear map from $\mathbb{R}^n$ to $L(\mathbb{R}^n, \mathbb{R}^p)$, i.e. $D(Df(x)) \in L(\mathbb{R}^n, L(\mathbb{R}^n, \mathbb{R}^p))$. In practical, this map

$$\begin{array}{rccc} D(Df)(x): & \mathbb{R}^n \times \mathbb{R}^n & \longrightarrow & \mathbb{R}^p \\ & (h,k) & \longmapsto & D(Df)(x).(h,k) \end{array}$$

is a bilinear map, that is, is linear in h and in k.

In practical, it is easier to see $D(Df) = D^2 f$ in coordinates. Indeed, we have the same result as in C^1 regularity, which links the fact of being C^2 with the order two partial derivatives:

$$\frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}\right) = \frac{\partial^2 f}{\partial x_i \partial x_j}$$

Theorem 6.2. *The map f is of class C^2 if and only if*

- *f has order two partial derivatives at every point $x \in U$, and*
- *these partial derivatives $x \mapsto \partial^2 f/\partial x_i \partial x_j$ are continuous.*

If the map f is C^2, then the map D^2f can be easily expressed in coordinates: if $h = (h_1, \dots, h_n)$ and $k = (k_1, \dots, k_n)$, then

$$D^2f(x).(h,k) = \sum_{1 \le i,j \le n} h_i k_j \frac{\partial^2 f}{\partial x_i \partial x_j}(x). \tag{6.1}$$

In the particular case where $p = 1$ (when f takes real values), the map D^2f can be represented as a matrix called *Hessian matrix*:

$$H(f)(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(x) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1}(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n}(x) \end{pmatrix},$$

and the formula (6.1) can be rewritten as

$$D^2f(x).(h,k) = \begin{pmatrix} h_1 & \cdots & h_n \end{pmatrix} H(f)(x) \begin{pmatrix} k_1 \\ \vdots \\ k_n \end{pmatrix}$$

Exercise 48. Let f be the function from $\mathbb{R}^2$ to $\mathbb{R}$ defined by $f(x,y) = x^4 + y^4 - 2xy$. Compute the order two partial derivatives of f.

You might have remarked that in this example, the crossed partial derivatives are equal: $\partial^2 f/\partial x \partial y = \partial^2 f/\partial y \partial x$. This is in fact true in general.

Theorem 6.3 (Schwarz)**.** *Suppose that f is a C^2 map, and let $x \in U$. Then for any $1 \le i, j \le n$,*

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \frac{\partial^2 f}{\partial x_j \partial x_i}(x).$$

This theorem can be rephrased in terms of differential: the map D^2f is symmetric, that is

$$D^2f(x).(h,k) = D^2f(x).(k,h).$$

A map which is bilinear, symmetric and with values in $\mathbb{R}$ is called a *quadratic form*. Schwarz theorem implies that if $f : U \to \mathbb{R}$ is of class C^2, then D^2f is a quadratic form.

Before stating the order two conditions of extremum, we have to give Taylor formula for order two:

Theorem 6.4 (Taylor)**.** *Let f be C^2 in an open set containing $x \in U$. Then*

$$f(x+h) = f(x) + Df(x).h + \frac{1}{2}D^2f(x).(h,h) + o^2(h),$$

where $o^2(h)$ is a notation for a map such that:

$$\lim_{h \to 0, h \neq 0} \frac{\|o^2(h)\|}{\|h\|^2} = \lim_{h \to 0, h \neq 0} \frac{\|f(x+h) - f(x) - Df(x).h - \frac{1}{2}D^2f(x).(h,h)\|}{\|h\|^2} = 0.$$

Exercise 49. Let $u : \mathbb{R}^n \to \mathbb{R}^p$ be a linear map, and $a \in \mathbb{R}^p$. We define $f : \mathbb{R}^n \to \mathbb{R}$ by $f(x) = \|u(x) - a\|_2^2$, where $\|\cdot\|_2$ is the Euclidean norm. Show that f is of class C^2 and compute $Df(x)$ and $D^2f(x)$ for any $x \in \mathbb{R}^n$.

6.2 Local extrema: order two conditions

Order two differentials can tell us whether a critical point is a local maximum, minimum or not a local extremum.

Theorem 6.5. *Suppose that $f : U \to \mathbb{R}$ is of class C^2. Then*

- *If x is a critical point for f and if for any $h \neq 0$, $D^2 f(x).(h,h) > 0$, then f admits a local strict minimum at x, i.e.*

$$\exists \varepsilon > 0 : \forall y \in B(x,\varepsilon) \setminus \{x\}, f(y) > f(x).$$

- *If x is a critical point for f and if there exists $h \neq 0$ and $k \neq 0$ such that $D^2 f(x).(h,h) > 0$ and $D^2 f(x).(k,k) < 0$, then f admits no local extremum at x.*
- *If f admits a local minimum at x, then x is a critical point for f and for any $h \neq 0$, $D^2 f(x).(h,h) \geq 0$.*

Of course, the same holds for maxima by reversing the inequalities.

Be careful with strict and large inequalities in this theorem: if for example $D^2 f(x).(h,h) = 0$ for some $h \neq 0$, then it is impossible to conclude.

This theorem can be easily understood applying Taylor theorem: close to x, the map f "looks like" its Taylor expansion at order two. If x is a critical point, then $f(x+h) \simeq f(x) + D^2 f(x).(h,h)$. In other words, f is almost equal to a constant plus a quadratic form, so morally, x is an extremal point for f if and only if it is an extremal point for this quadratic form[1].

Let us study quadratic forms in $\mathbb{R}^2$. We write

$$H(f)(a) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(x) & \frac{\partial^2 f}{\partial y \partial x}(x) \\ \frac{\partial^2 f}{\partial x \partial y}(x) & \frac{\partial^2 f}{\partial y \partial y}(x) \end{pmatrix} = \begin{pmatrix} r & s \\ s & t \end{pmatrix}.$$

Algebra tells us that such a matrix can be diagonalized, in other words in a good basis or $\mathbb{R}^2$,

$$D^2 f(a)(h,h) = \begin{pmatrix} h_1 & h_2 \end{pmatrix} \begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = \lambda h_1^2 + \mu h_2^2.$$

Note that $D^2 f(a)(0,0) = 0$, so the fact that $(0,0)$ is a local extremum of $D^2 f$ or not depends on the sign of the eigenvalues λ and μ. Moreover, we know that $\lambda\mu = \det H(a)$ and $\lambda + \mu = \operatorname{tr} H(a)$. So to summarize:

Proposition 6.6. *Let $f : U \to \mathbb{R}$ be of class C^2, where U is an open subset of $\mathbb{R}^2$, and $a \in U$ be a critical point of f. We write*

$$H(f)(a) = \begin{pmatrix} r & s \\ s & t \end{pmatrix}.$$

[1] however this heuristic is no longer true in the degenerate cases where there exists $h \neq 0$ such that $D^2 f(x).(h,h) = 0$.

- *If* $rt - s^2 < 0$, *then* a *is not a local extremum of* f.
- *If* $rt - s^2 > 0$ *and* $r + t > 0$, *then* a *is a local maximum of* f.
- *If* $rt - s^2 > 0$ *and* $r + t < 0$, *then* a *is a local minimum of* f.

The same kind of results holds in higher dimensions. To reduce quadratic forms there is a quick algorithm due to Gauss.

Exercise 50. Determine local extrema of the map $f : (\mathbb{R}_+^*)^2 \to \mathbb{R}$ defined by

$$f(x, y) = \frac{xy}{(1 + x)(1 + y)(x + y)}.$$

Same question for $f : \mathbb{R}^2 \to \mathbb{R}$

$$f(x, y) = x^3 + y^3 - 3xy.$$

Exercise 51. Let

$$f(x, y) = x^2 - y^2 + \frac{y^4}{4}.$$

Determine the extrema of f, and sketch its level lines.