Antoine Levitt

Mathematical tools for engineers

2021-2022

École des Ponts - 1st year

Preface

This is the lecture notes for the course "Outils mathématiques pour l'ingénieur", the first applied mathematics course at École des Ponts. Its purpose is to introduce basic tools in the analysis of functions, and to act as an introduction to further lectures in applied mathematics. The three-day lectures will cover the material in Chapters 1, 2 and 3. The appendices, for interested students, contain proofs of some statements that are assumed in the main text, as well as the Banach fixed-point theorem and its applications in differential equations, geometry and optimization.

Central formal mathematical definitions are in **bold**. Important or subtle points are highlighted in *italics*.

This is the first edition of these lecture notes, and as such very likely contain numerous typos and errors. Feedback and corrections are very welcome!

Contents

1

Completeness and Banach spaces

It is of great interest in practical problems to be able to manipulate functions as if they were vectors, and to apply to functions the powerful techniques of vector calculus and linear algebra. For instance, we want to be able to solve equations whose unknowns are functions, minimize over a set of functions, project a function onto a subspace of functions, etc. Spaces of functions are however infinite-dimensional (because functions have an infinite number of degrees of freedom), and a number of properties true in the finite-dimensional setting (completeness, compactness of closed bounded sets, boundedness of linear maps...) do not carry over. The goal of *functional analysis* is to identify those that do. Possibly the most important is the notion of completeness and Banach spaces.

1.1 Completeness

The real numbers are a central concept of elementary mathematics, and are usually manipulated without thought about their rigorous properties, or indeed their definition. The main conceptual problem is that the easily understood rational numbers are not sufficient to represent all interesting numbers: for instance, $\sqrt{2}$ is not a rational number. Consider the decimal expansion of $\sqrt{2}$ with n digits, which converges to $\sqrt{2}$ in $\mathbb{R}$. However, if we do not know what real numbers are and only know about rationals, all we see is a sequence of rationals that does not converge, although it "ought to" converge to something: $\mathbb{Q}$ has "holes". We formalize this with the notion of Cauchy sequence: a sequence of numbers x_n is Cauchy if and only if[†]

$$\lim_{N\to\infty} \sup_{n\geqslant N, m\geqslant N} |x_n - x_m| = 0.$$

A set X whose Cauchy sequences all converge is called *complete*: $\mathbb{Q}$ is not complete, $\mathbb{R}$ is[‡].

Completeness is a fundamental concept because it allows the *construction* of objects using limit processes; completeness is almost always used in some way in the theorems that guarantee that a certain object exists.

Exercise 1.1. *Using only the fact that* $\mathbb{R}$ *is complete, show*

- *That a bounded non-decreasing sequence converges*
- *The intermediate value theorem*
- *That a series that converges absolutely converges*

† When there is no potential for confusion, we will write "the sequence x_n" for "the sequence $(x_n)_{n\in\mathbb{N}}$".

‡ This last statement is strictly speaking non-mathematical because we have not defined the real numbers. However, all definitions are equivalent and this property is true no matter what definition is chosen. In fact, one standard way to define $\mathbb{R}$ is as the completion of $\mathbb{Q}$ (the equivalence class of the set of Cauchy sequences with values in $\mathbb{Q}$, quotiented by the relation that two Cauchy sequences are equivalent if the difference between them tends to zero).

In particular, using the intermediate value theorem one can *define* $\sqrt{2}$ as the only positive number such that $(\sqrt{2})^2 = 2$. This is impossible to do in $\mathbb{Q}$, which is not complete. In the same way, by identifying which infinite-dimensional spaces are complete, we will be able to construct by approximation solutions to various problems.

1.2 Banach spaces

In the following, the field $\mathbb{K}$ will generally be that of real or complex numbers; the exception will be when we discuss differentiation, where we will require $\mathbb{K} = \mathbb{R}$. Recall that a **normed $\mathbb{K}$-vector space** is a $\mathbb{K}$-vector space E together with a norm $\|\cdot\| : E \to \mathbb{R}_+$ satisfying the axioms of homogeneity ($\|\lambda u\| = |\lambda|\|u\|$), definiteness ($\|u\| = 0 \Rightarrow u = 0$) and triangle inequality ($\|u + v\| \leqslant \|u\| + \|v\|$). A subset U of E is said to be **open** if for all $x \in U$, there is $r > 0$ such that $B(x, r) = \{y \in E, \|y - x\| < r\} \subset U$. A subset Z of E is said to be **closed** if $E \setminus Z$ is open; equivalently, if every sequence of elements of Z that converges in E has its limit in Z. A **neighborhood** of a point x in a normed vector space in an open set containing x.

If N_1 and N_2 are two norms on E with a constant $C > 0$ such that, for all $u \in E$, $N_1(u) \leqslant CN_2(u)$, the norm N_2 is said to be **stronger** than N_1: a sequence converging for N_2 converges for N_1. Two norms N_1 and N_2 are said to be **equivalent** if N_1 is both stronger and weaker than N_2; explicitly, if there are two constants $c, C > 0$ such that, for all $u \in E$,

$$cN_1(u) \leqslant N_2(u) \leqslant CN_1(u).$$

If N_1 and N_2 are equivalent, sequences converging for N_1 converge for N_2, and vice-versa: equivalent norms yield the same *topology* (notion of convergence).

The definition of Cauchy sequences seen above extends naturally to the case of normed vector spaces:

Definition 1.2. *Let $(E, \|\cdot\|)$ be a normed vector space. A sequence $(u_n)_{n\in\mathbb{N}}$ in E is said to be a **Cauchy sequence** if*

$$\lim_{N\to\infty} \sup_{n\geqslant N, m\geqslant N} \|u_n - u_m\| = 0.$$

Explicitly, this means that for every $\varepsilon > 0$, there is $N \in \mathbb{N}$ such that, for all $n \geqslant N, m \geqslant N$,

$$\|u_n - u_m\| \leqslant \varepsilon.$$

*If all Cauchy sequences converge (in E), $(E, \|\cdot\|)$ is said to be **complete**. The vector space E is said to be a **Banach space** (for its norm).*

We emphasize that the Banach property refers to a pair $(E, \|\cdot\|)$ of a space and a norm; some spaces might be Banach for a norm but not for another.

Exercise 1.3. *Prove that*

- *A convergent sequence is Cauchy*
- *A Cauchy sequence is bounded*
- *If a Cauchy sequence has a converging subsequence, it converges*
- *A closed subspace of a Banach space is a Banach space*

The elements of a Cauchy sequence get arbitrarily close together as the sequence progresses. They are sequences that "ought to converge" to something, and indeed they usually do, at least in a generalized sense. However, their limit might not belong to E (for instance, when E is a space of functions or sequences with certain properties that are not satisfied by the limit); as we will see, not all infinite-dimensional normed spaces are Banach spaces. The notion of Banach spaces is an extremely useful one, because a sequence (for instance, the iterations of an algorithm) can often be shown to be Cauchy, without reference to the (unknown) limit. Once a space is shown to be a Banach space, this implies convergence of the sequence.

1.2.1 The finite-dimensional case

The basis example of a finite-dimensional space is $\mathbb{K}^d$. The three basic norms† on $\mathbb{K}^d$ are

$$\|u\|_1 = \sum_{i=1}^{d} |u_i|, \qquad \|u\|_2 = \sqrt{\sum_{i=1,\dots,d} |u_i|^2}, \qquad \|u\|_\infty = \max_{i=1,\dots,d} |u_i|$$

Exercise 1.4. *Prove that these are norms.*

In finite dimensions, all norms are equivalent (see Appendix 5 for a proof). Therefore, one can simply talk about convergent, Cauchy sequences, complete vector spaces, etc. without having to specify the norm.

Note that in this chapter we will discuss the properties of spaces of vectors (or sequences); these notions require handling sequences of vectors. If $(u_n)_{n\in\mathbb{N}}$ is a sequence in $\mathbb{R}^d$, then we denote by $u_{n,i} = (u_n)_i$ the i-th component of u_n.

Proposition 1.5. *If E is finite-dimensional, it is complete.*

Proof. Let us first treat the case $\mathbb{K} = \mathbb{R}$. Using a basis and the equivalence of norms, we can assume that $E = \mathbb{R}^d$ equipped with the infinity norm. Let u_n be a Cauchy sequence in $\mathbb{R}^d$. For any $i = 1, \dots, d, n \in \mathbb{N}, m \in \mathbb{N}$,

$$|u_{n,i} - u_{m,i}| \leqslant \|u_n - u_m\|_\infty$$

and therefore the sequence $(u_{n,i})_{n\in\mathbb{N}}$ is Cauchy in $\mathbb{R}$. By completeness of $\mathbb{R}$, it converges to some $u_{*,i}$. Now,

$$\|u_n - u_*\|_\infty = \max_{i=1,\dots,d} |u_{n,i} - u_{*,i}| \to 0$$

and so u_n converges to u_*.

The same proof together with the identification of $\mathbb{C}$ to $\mathbb{R}^2$ shows that $\mathbb{C}$ is complete, and the proof of the result follows in the case $\mathbb{K} = \mathbb{C}$. □

This statement is *not* true in infinite-dimensional spaces.

1.3 Examples of Banach and non-Banach spaces

1.3.1 Spaces of sequences

The simplest infinite-dimensional spaces are the spaces of sequences $\mathbb{K}^\mathbb{N}$: $(u_i)_{i\in\mathbb{N}}$. On $\mathbb{K}^\mathbb{N}$, one can try to define the three usual norms in the same way as before:

$$\|u\|_1 = \sum_{i\in\mathbb{N}} |u_i|, \qquad \|u\|_2 = \sqrt{\sum_{i\in\mathbb{N}} |u_i|^2}, \qquad \|u\|_\infty = \sup_{i\in\mathbb{N}} |u_i|$$

The difference now is that these quantities (always well-defined as positive or infinite numbers, as sums or sups of non-negative numbers) can be infinite, and therefore are not norms on $\mathbb{K}^\mathbb{N}$. We must restrict the space:

$$\ell^p = \{u \in \mathbb{K}^\mathbb{N}, \|u\|_p < \infty\}.$$

† These are by no means the only possible norms. Other commonly-used norms are the p-norm $\|u\|_p = \left(\sum_{i=1}^d |u_i|^p\right)^{1/p}$, and the change-of-variable norms $\|Av\|$ with A an invertible matrix.

Clearly, we have

$$\|u\|_\infty \leqslant \|u\|_2 \leqslant \|u\|_1$$

but there are no reverse inequalities: the norms are not equivalent. The norm $\|\cdot\|_1$ is the *strongest* of the three: we have $\ell^1 \subset \ell^2 \subset \ell^\infty$, and if a sequence of sequences converges in the $\|\cdot\|_1$ norm, it converges also in the $\|\cdot\|_2$ and $\|\cdot\|_\infty$ norm.

Proposition 1.6. *The space $(\ell^p, \|\cdot\|_q)$ is either not a normed space, a non-Banach normed vector space or a Banach space, according to the following table:*

Space p/Norm q	1	2	∞
ℓ^1	Banach	Non-Banach normed	Non-Banach normed
ℓ^2	Not normed	Banach	Non-Banach normed
ℓ^∞	Not normed	Not normed	Banach

Proof.
- The diagonal: $p = q$. It is clear that $\|\cdot\|_p$ defines a norm on ℓ^p, just as in finite dimension.
  Case $p = \infty$. Let u_n be a Cauchy sequence in $(\ell^\infty, \|\cdot\|_\infty)$. For all $i \in \mathbb{N}$,

  $$|u_{n,i} - u_{m,i}| \leqslant \|u_n - u_m\|_\infty,$$

  and so $(u_{n,i})_{n\in\mathbb{N}}$ is Cauchy in $\mathbb{K}$. Let $u_{*,i}$ be its limit.
  Let $\varepsilon > 0$. There is $N \geqslant 0$ such that, for all $i \in \mathbb{N}$, $n \geqslant N, m \geqslant N$,

  $$|u_{n,i} - u_{m,i}| \leqslant \varepsilon.$$

  We pass to the limit $m \to \infty$ and obtain

  $$|u_{n,i} - u_{*,i}| \leqslant \varepsilon.$$

  This proves that $u \in \ell^\infty$, and that $u_n \to u$ in $(\ell^\infty, \|\cdot\|_\infty)$.
  Case $p = 1, 2$. We prove the case $p = 1$; the case $p = 2$ is similar. We construct u_i as in the case $p = \infty$.
  Let $\varepsilon > 0$. There is $N \in \mathbb{N}$ such that, for all $m \geqslant N$, for all $I \in \mathbb{N}$,

  $$\sum_{i=1}^{I} |u_{N,i} - u_{m,i}| \leqslant \varepsilon.$$

  We pass to the limit $m \to \infty$ and obtain

  $$\sum_{i=1}^{I} |u_{N,i} - u_{*,i}| \leqslant \varepsilon.$$

  We can then pass to the limit $I \to \infty$ and obtain both that $u \in \ell^1$ and that $\|u_N - u\|_1 \leqslant \varepsilon$, which proves the result.
- Lower triangle: ($p > q$, norm too strong). The sequence $u_i = 1$ is in ℓ^∞ but has infinite 1- and 2-norm. The sequence $u_i = 1/i$ is in ℓ^2 but has infinite 1-norm.
- Upper triangle: ($p < q$, norm too weak). We show that $(\ell^1, \|\cdot\|_\infty)$ is not complete; the other cases are similar. Let $u_{n,i} = 1/i$ if $i \leqslant n$, 0 otherwise. We have $\|u_n - u_m\|_\infty \leqslant 1/\min(n, m)$, so that the sequence u_n of elements of ℓ^1 is Cauchy for the $\|\cdot\|_\infty$ norm. If it converged in $(\ell^1, \|\cdot\|_\infty)$, it would necessarily converge to the sequence $1/i$, which is impossible since this sequence is not in ℓ^1.

□

This example provides some intuition for Banach spaces: they are spaces whose norm is precisely adapted to the space. If the norm is too strong, it is not globally defined on the space (it is infinite on some vectors), and does not define a normed vector space. On the other hand, if the norm is too weak, it allows sequences that converge to some element outside the space (and therefore, Cauchy sequences that are non-convergent inside the space).

1.3.2 Spaces of functions on a bounded interval

Consider functions on a *bounded* closed interval I, with values in $\mathbb{K}$.

Recall that a sequence of functions u_n can converge to a function in a variety of senses: *pointwise* (for all $x \in I$, $u_n(x) \to u(x)$), *uniformly* ($\sup_{x\in I} |u_n(x) - u(x)| \to 0$), *in* L^1 (or *in mean*, $\int |u_n(x) - u(x)|dx \to 0$), or *in* L^2 (or *in square mean*, $\int_I |u_n(x) - u(x)|^2 dx \to 0)^\dagger$. Uniform convergence is stronger than (implies) pointwise convergence. On a bounded interval I, uniform is stronger than L^2, which is stronger than L^1. All other implications are false.

The norms that correspond to these modes of convergence (L^1, L^2, uniform) are

$$\|u\|_1 = \int_I |u(x)|dx, \qquad \|u\|_2 = \sqrt{\int_I |u(x)|^2 dx}\,, \qquad \|u\|_\infty = \sup_{x\in I} |u\,(x)\,|.$$

When defined, we can easily show that

$$\|u\|_1 \leqslant \sqrt{|I|}\,\|u\|_2 \leqslant |I|\,\|u\|_\infty$$

and therefore the ∞ norm is the strongest of the three. Note that this hierarchy is the reverse as for sequences. This is because the larger p is, the more emphasis is put where u is large, and therefore on local spikes; the lower it is, the more it is put where u is small, and therefore on long tails at infinity. Sequences cannot have arbitrarly narrow local spikes, and therefore the 1 norm is the strongest; on the other hand, functions on a bounded interval cannot have tails at infinity, and therefore the ∞ norm is the strongest. Functions on unbounded intervals combine both difficulties (they can have both local spikes and tails at infinity), and there is no hierarchy there.

The definition of the 1 and 2 norms using the Riemann integral requires (piecewise) continuous functions. Unfortunately, these norms are too weak for the space of continuous functions: a sequence of continuous functions converging in L^1 does not necessarily converge to a continuous function.

Exercise 1.7. *Give an example of continuous functions converging for the* $\|\cdot\|_1$ *norm towards a discontinuous function. Show that* $(C^0(I), \|\cdot\|_1)$ *is a normed vector space, but not a Banach space.*

Finding a Banach space adapted to the 1- and 2-norms requires the construction of the Lebesgue integral and the Lebesgue L^p spaces, which will be the topic of future courses. On the other hand, since uniformly convergent continuous functions converge towards continuous functions, there is hope that $(C^0(I), \|\cdot\|_\infty)$ is a Banach space, and indeed this is the case:

Proposition 1.8. $(C^0(I), \|\cdot\|_\infty)$ *is a Banach space.*

Proof. It is immediate to see that this is a normed vector space. For completeness, we will proceed in much the same way as for ℓ^∞, by constructing the limit object pointwise. Let u_n be a Cauchy sequence in $(C^0(I), \|\cdot\|_\infty)$. Let $x \in I$. Then, for all $n, m \in \mathbb{N}$,

$$|u_n(x) - u_m(x)| \leqslant \|u_n - u_m\|_\infty.$$

The sequence $(u_n(x))_{n\in\mathbb{N}}$ is Cauchy in $\mathbb{K}$ and therefore converges. Let $u(x)$ be its limit.

For all $\varepsilon > 0$, there is $N \in \mathbb{N}$ such that, for all $x \in I$, $m \geqslant N$, we have

$$|u_N(x) - u_m(x)| \leqslant \varepsilon$$

Passing to the limit $m \to \infty$ and then taking the supremum over x, we have $\|u_N - u\|_\infty \leqslant \varepsilon$ and therefore $(u_N)_{N\in\mathbb{N}}$ converges to u uniformly.

We now have to prove that the uniform limit u of the continuous functions u_n is continuous. Let $x \in I$, and $\varepsilon > 0$. There is $N \in \mathbb{N}$ such that

† Note that these notions are distinct from the various modes of convergence usually studied in probability theory, where the object of interest is a measure rather than a function.

$$\sup_{z\in I}|u_N(z)-u(z)|\leqslant\varepsilon.$$

It follows that, for all $y\in I$,

$$|u(x)-u(y)|\leqslant|u(x)-u_N(x)|+|u_N(x)-u_N(y)|+|u_N(y)-u(y)|\leqslant 2\varepsilon+|u_N(x)-u_N(y)|.$$

We can now choose δ such that $|u_N(x)-u_N(y)|\leqslant\varepsilon$ whenever $|x-y|\leqslant\delta$, which completes the proof. □

Proposition 1.9. *For any function $u\in C^1(I)$, let*

$$\|u\|_{C^1}=\|u\|_\infty+\|u'\|_\infty.$$

Show that $(C^1(I),\|\cdot\|_{C^1(I)})$ is a Banach space.

This immediately generalizes to the space $C^k(I)$ of k times continuously differentiable functions.

Proof. Both u_n and u_n' are Cauchy sequences in $(C^0(I),\|\cdot\|_\infty)$, and therefore converge uniformly to two continuous functions u and v. We now have to prove the differentiation under the limit sign: that if u_n converges uniformly to u and u_n' converges uniformly to v, then u is C^1 and $u'=v$.

For this, let x_0 be an arbitrary point in I, and $\widetilde{u}(x)=u(x_0)+\int_{x_0}^x v(y)dy$. Clearly $\widetilde{u}$ is C^1 with $\widetilde{u}'=v$, so we only have to show that $u=\widetilde{u}$. Let $x\in I$ and $\varepsilon>0$. There is $N\in\mathbb{N}$ such that

$$\sup_{z\in I}|u_N(z)-u(z)|+\sup_{z\in I}|u_N'(z)-v(z)|\leqslant\varepsilon.$$

For any $x\in I$, we have $u_N(x)=u_N(x_0)+\int_{x_0}^x u_N'$ and so

$$|u(x)-\widetilde{u}(x)|\leqslant\varepsilon+\left|u_N(x_0)-u(x_0)+\int_{x_0}^x(u_N'-v)\right|\leqslant 2\varepsilon+|x-x_0|\varepsilon$$

and the result follows. □

Note that the vocabulary of Banach spaces summarizes the "X under the Y sign" theorems in a concise way. For instance, consider the differentiation under the limit sign theorem, which says that if $u_n\to u$ uniformly and $u_n'\to v$ uniformly, then u is C^1 and $u'=v$. Since u_n and u_n' converge, they are Cauchy in $C^0(I)$, and therefore u_n is Cauchy in $C^1(I)$. Since $C^1(I)$ is a Banach space, there is $w\in C^1(I)$ such that $u_n\to w$ and $u_n'\to w'$ uniformly; by uniqueness of the limit, $u=w$ and $v=w'$, which shows that $u'=v$.

The examples above generalize to the case where the sequences or functions have values in a general Banach space F instead of $\mathbb{K}$.

In the case where I is unbounded, the continous functions do not necessarily have finite infinity norm, and therefore the correct Banach space is the space of *bounded* continuous functions $C_B^0(I)$. Bounded functions are not necessarily integrable, and so $(C_B^0(I),\|\cdot\|_1)$ is not a normed vector space.

1.4 Normal convergence

Definition 1.10. *A series $\sum_{n\in\mathbb{N}}u_n$ in a normed vector space is said to* ***converge normally*** *if $\sum_{n\in\mathbb{N}}\|u_n\|<\infty$.*

When the underlying Banach space is $\mathbb{R}$ or $\mathbb{C}$, norms are simply absolute values, and normal convergence is accordingly called **absolute convergence**†.

† Note that in some literature the terminology of normally convergent series is used in the more restrictive context of the Banach space $(C^0(I),\|\cdot\|_\infty)$.

Proposition 1.11. *Normally convergent series in a normed vector space E are convergent if and only if E is a Banach space.*

Proof. • If E is a Banach space, normally convergent series are convergent. Let

$$S_n = \sum_{k=0}^{n} u_n.$$

For $N \leqslant n < m$, we have

$$\|S_n - S_m\| \leqslant \sum_{k=n+1}^{m} \|u_n\| \leqslant \sum_{k=N}^{\infty} \|u_n\|$$

which tends to zero as $N \to \infty$.

• If normally convergent series are convergent, E is a Banach space. Let u_n be a Cauchy sequence in E. We extract from u_n a "rapidly Cauchy" subsequence in the following way. For all $\varepsilon > 0$, let $N(\varepsilon)$ be an integer such that, for all $p \geqslant N(\varepsilon), q \geqslant N(\varepsilon)$, $\|u_p - u_q\| \leqslant \varepsilon$. The sequence $(v_n)_{n\in\mathbb{N}}$ defined by

$$v_n = u_{N\left(\frac{1}{2^n}\right)}$$

is a subsequence of u_n such that

$$\|v_{n+1} - v_n\| \leqslant \frac{1}{2^n}$$

The sequence v_n converges normally, and therefore converges. This implies that u_n converges. □

When applicable, the theory of normal convergence allows particularly simple answers to the question of the convergence and interchange of differentiation and sum signs for series of functions:

Exercise 1.12. *Let*

$$S_N(x) = \sum_{n=-N}^{N} \frac{\cos(n)}{1 + |n|^{42}} e^{inx}.$$

Show that S_N converges pointwise to a continuously differentiable function $S(x)$ and that, for all $x \in \mathbb{R}$,

$$S'(x) = \lim_{N\to\infty} S'_N(x).$$

What can we say about higher-order derivatives?

1.5 Summary

Infinite-dimensional vector spaces are important tools for the analysis of functions. Unless the vector spaces are chosen carefully, there is the risk that the space has "holes": Cauchy sequences that "ought to converge" but do not (often because they converge to an element outside the space). Working in Banach spaces prevents such issues, and allows many tools of finite-dimensional vector calculus to be used. The vocabulary of Banach function spaces unify the maze of "X under the Y sign" type theorems, where X is limit or differentiation, and Y is sum or limit. Banach spaces for integral norms (Lebesgue L^p spaces) can be constructed, but this requires the tool of the Lebesgue integral, which will be introduced in later lectures. Anticipating slightly on this and with an abuse of notation in the case $p = \infty$, the following table summarizes the important Banach spaces.

Nature	Space	Norm	Comments
Vectors	$(\mathbb{K}^d, \\|\cdot\\|_p)$	$\\|u\\|_p = (\sum_{i=1}^d \|u_i\|^p)^{1/p}$	Finite dimensional
Sequences	ℓ^p	$\\|u\\|_p = (\sum_{i=1}^\infty \|u_i\|^p)^{1/p}$	$\ell^1 \subset \ell^2 \subset \ell^\infty$
Continuous functions	$C^0(I)$	$\\|u\\|_\infty = \sup_{x\in I} \|u(x)\|$	Summarizes continuity of limits/sums
C^k functions	$C^k(I)$	$\\|u\\|_{C^k} = \sum_{l=0}^k \\|u^{(l)}\\|_\infty$	Summarizes differentiability of limits/sums
Measurable functions	$L^p(I)$	$\\|u\\|_p = (\int_I \|u\|^p)^{1/p}$	I bounded $\Rightarrow L^\infty \subset L^2 \subset L^1$

2

Maps between Banach spaces

2.1 Linear maps

A **linear map** (or **linear operator**) between two vector spaces E and F is a mapping $A : E \to F$ such that $A(\lambda u + \mu v) = \lambda A(u) + \mu A(v)$ for all $u, v \in E, \lambda, \mu \in \mathbb{K}$.

Definition 2.1. *A linear map $A : E \to F$ is said to be **bounded** if*

$$\|\|A\|\| := \sup_{u \in E, u \neq 0} \frac{\|Au\|}{\|u\|} < \infty$$

We write $L(E, F)$ the set of bounded linear maps from E to F. Note that this sense of bounded is specific to linear maps, and distinct from that of bounded functions (linear maps that are not identically zero can never define bounded functions).

Exercise 2.2. *Prove that:*

- *A linear map is bounded if and only if it is continuous*
- *$\|\|\cdot\|\|$ is a norm on $L(E, F)$*
- *If $A \in L(E, F)$ and $B \in L(F, G)$, then $AB \in L(E, G)$ and $\|\|AB\|\| \leqslant \|\|A\|\| \|\|B\|\|$*

Just as for matrices, a bounded linear map $A : E \to F$ is said to be **invertible** if there is another bounded linear map B such that $AB = BA = \mathrm{Id}$: in this case we write $A^{-1} = B$.

In finite dimension, linear maps can be represented by matrices: if $(e_j)_{j=1,\dots,d_E}$ is a basis for E, $(f_i)_{i=1,\dots,d_F}$ a basis for F, and A is linear from E to F, then

$$A \sum_{j=1}^{d_E} x_j e_j = \sum_{i=1}^{d_F} \left(\sum_{j=1}^{d_E} A_{ij} x_j \right) f_i$$

for some coefficients A_{ij}. The matrix $(A_{ij})_{i=1,\dots,d_F, j=1,\dots,d_E}$ is the representation of A in the chosen bases. From the triangular inequality it is easy to see that all finite-dimensional linear maps are bounded.

Exercise 2.3. *Let $A_{ij} \in \mathbb{K}$ be an infinite matrix, acting on sequences by*

$$(Au)_i = \sum_{j \in \mathbb{N}} A_{ij} u_j.$$

Show that if $\sup_{i \in \mathbb{N}} \sum_{j \in \mathbb{N}} |A_{ij}| < \infty$, A defines a bounded linear map from $\ell^\infty(\mathbb{N})$ to itself. Give an example of infinite matrix with bounded coefficients that does not define a bounded operator from $\ell^\infty(\mathbb{N})$ to itself.

Exercise 2.4. *Let I be a bounded and closed interval, and $K : I \times I \to \mathbb{R}$ be continuous. Show that the linear map K acting on $C^0(I)$ as*

$$(Ku)(x) = \int_I K(x,y)u(y)dy$$

is a bounded linear map.

Exercise 2.5. *Show that the linear map A defined by $(Au)(x) = u(-x)$ is bounded on $C^0([-1,1])$. Show that the linear map $u \to u'$ is bounded from $C^1([-1,1])$ to $C^0([-1,1])$.*

Unlike in the finite-dimensional case, not all linear maps are bounded†:

Exercise 2.6. *Show that the map $u \mapsto u(0)$ is not bounded from $(C^0([-1,1]), \|\cdot\|_1)$ to $\mathbb{R}$. Show that the map $u \mapsto u$ is not bounded from $(C^0([-1,1]), \|\cdot\|_1)$ to $(C^0([-1,1]), \|\cdot\|_\infty)$.*

In practice, many linear operations of interest (for instance, differentiation $u \to u'$) can not be defined as bounded linear maps on a single Banach space (for instance, from $C^0(I)$ to itself), but only on a subset of the Banach space (for instance, from $C^1(I)$ to $C^0(I)$). This is an important complication in the theory of differential equations.

2.2 Functional calculus

Proposition 2.7. *When E and F are Banach spaces, the set $L(E,F)$ of linear maps from E to F together with the norm $|||\cdot|||$ is a Banach space.*

Proof. It is easily seen that this space is a normed vector space. Let A_n be a Cauchy sequence in $L(E,F)$. As usual, our first task is to construct the limit "pointwise". For all $u \in E$,

$$\|A_n u - A_m u\| \leqslant |||A_n - A_m||| \|u\|$$

and so $(A_n u)_{n\in\mathbb{N}}$ is Cauchy and therefore converges. We call Au its limit. The map $u \mapsto Au$ is clearly linear. For all $\varepsilon > 0$, there is $N \geqslant 0$ such that, for all $m \geqslant N$, $u \in E$,

$$\|(A_N - A_m)u\| \leqslant \varepsilon\|u\|.$$

Using the triangular inequality and passing to the limit $m \to \infty$, we see that $\|A_N u - Au\| \leqslant \varepsilon\|u\|$, which shows both that A is bounded and that that $A_N \to A$ in $L(E,F)$. □

In particular, the set $L(E) = L(E,E)$ of bounded operators on a Banach space to itself is a Banach *algebra* (an algebra which is also a Banach space, with $|||AB||| \leqslant |||A||| |||B|||$). Several techniques of real calculus, and in particular power series, generalize to this setting. Recall that to a formal power series $f(z) = \sum a_n z^n$ we can associate a *radius of convergence* $r \in \mathbb{R}+$. When $|z| < r$, the series converges absolutely, i.e. $\sum_{n\in\mathbb{N}} |a_n||z|^n < \infty$. If now A is a linear map between Banach spaces such that $\|A\| < r$, then the series

$$\sum_{n\in\mathbb{N}} a_n A^n$$

converges normally (and therefore in norm) to a bounded operator which we call $f(A)$.

Exercise 2.8. *Let $A \in L(E)$ with E a Banach space.*

† Note that these counter-examples rely on the fact that $(C^0([-1,1]), \|\cdot\|_1)$ is not a Banach space. Linear maps between Banach spaces may also be unbounded, but there are no simple examples of such maps because their existence relies on subtle arguments involving the axiom of choice. See `https://en.wikipedia.org/wiki/Discontinuous_linear_map` or [1] for more details.

- *Show that if* $\|A\| < 1$*, then* $\mathrm{Id} - A$ *is invertible, and*

$$(\mathrm{Id} - A)^{-1} = \sum_{n\in\mathbb{N}} A^n$$

(the Neumann series). Using this result, show that small perturbations of invertible bounded linear operators are invertible.
- *Show that the map* $\mathbb{R} \to L(E)$ *given by* $e(t) = e^{tA} = \sum_{n\geqslant 0} \frac{(tA)^n}{n!}$ *is continuous.*
- *Show that*

$$\lim_{n\to\infty} \left(\mathrm{Id} + \frac{A}{n}\right)^n = e^A.$$

Exercise 2.9. *Let* I *be a bounded interval, and let* $f : I \to \mathbb{R}$ *and* $K : I \times I \to \mathbb{R}$ *be continuous. Show that the equation*

$$u(x) = f(x) + \varepsilon \int_I K(x,y)u(y)dy$$

has a unique solution for ε *small enough. Link this result with the "expansion in powers of* ε*" method of writing* $u(x) = u_0(x) + \varepsilon u_1(x) + \varepsilon^2 u_2(x) + \ldots$*, replacing in the equation and identifying powers of* ε*.*

2.3 Nonlinear maps and differentials

Definition 2.10. *Let* $f : E \to F$ *be a nonlinear map, with* E *and* F *real normed vector spaces. We say that* f *is (Fréchet-)**differentiable** at* x *if and only if there is a bounded linear map* $A_x : E \to F$ *such that*

$$\|f(x+h) - (f(x) + A_x h)\| = o(\|h\|).$$

The (unique) linear map A_x *is called the **differential** of* f *at* x*, noted* $df(x)$*.*

f is continuously differentiable (C^1) at $x \in E$ if f is differentiable in a neighborhood of x and if df is continuous from a neighborhood of x to $L(E,F)$.

The usual tools of vector calculus apply; for instance, the chain rule is

$$d(f \circ g)(x) = df(g(x))dg(x).$$

Consider now the case where both E and F are finite-dimensional. The matrix representation of the differential of a nonlinear map is called the **Jacobian**, a $d_F \times d_E$ matrix, often also denoted $df(x)$. If $f : \mathbb{R}^{d_E} \to \mathbb{R}^{d_F}$ is C^1 and has components f_i for $i = 1, \ldots, d_F$, then we have the Taylor expansion

$$f_i(x_1 + h_1, \ldots, x_{d_E} + h_{d_E}) = \sum_{j=1}^{d_E} \frac{\partial f_i}{\partial x_j}(x_1, \ldots, x_{d_E})h_j + o(\|h\|).$$

from where it follows that the Jacobian J of f at x is the matrix of **partial derivatives** $J_{ij} = \frac{\partial f_i}{\partial x_j}(x)$.

When $d_E = d_F = 1$, the Jacobian is a single number, the **derivative**. When $d_E \geqslant 1$ but $d_F = 1$, f is called a **functional**. Its differential is a linear map from E to $\mathbb{R}$, a **linear form**. Its jacobian is a $1 \times d_E$ matrix (row vector); the transpose of this vector (an element of E) is the **gradient**†: $\nabla f(x) = df(x)^T$.

† Note that the definition of the gradient depends on the choice of basis. More generally, in infinite dimensions, the gradient can be defined using an inner product structure, in Hilbert spaces.

When it exists, the differential of the differential $d^2 f(x)$ is a linear map from E to $L(E, \mathbb{R})$. This can also be seen as a bilinear form: $B(h_1, h_2) = (d^2 f(x) \cdot h_1) \cdot h_2$, which is symmetric by the Schwarz theorem. The matrix representation of this bilinear form, i.e. the matrix H defined by

$$B(h_1, h_2) = h_1^T H h_2,$$

is called the **Hessian**; it is the matrix of second derivatives.

Exercise 2.11. *Compute the differential, gradient and Hessian of the following nonlinear map from $\mathbb{R}^d$ to $\mathbb{R}$:*

$$F(x) = \frac{1}{2} x^T A x - b^T x$$

where $A \in \mathbb{R}^{d \times d}$ and $b \in \mathbb{R}^d$.

Exercise 2.12. *Let E be a Banach space, and $A \in L(E)$. Show that the map $\mathbb{R} \to L(E)$ given by $e(t) = e^{tA}$ is continuously differentiable, with $e'(t) = Ae(t)$.*

Exercise 2.13. *Show that the following maps are differentiable from $C^0([-1, 1])$ to itself, and compute their differentials*

- $f_1(u)(x) = u(0) \cos(x)$
- $f_2(u)(x) \to \left(\int_0^1 u\right)^2 e^x$
- $f_3(u)(x) = \int_0^x e^{u(y)} dy$

Exercise 2.14. *For all $\alpha \in \mathbb{R}$, let T_α be the map*

$$(T_\alpha u)(x) = u(x - \alpha)$$

defined from $(C^2(\mathbb{R}), \|\cdot\|_\infty + \|\cdot'\|_\infty + \|\cdot''\|_\infty)$ to $(C^0(\mathbb{R}), \|\cdot\|_\infty)$. Show that $T : \mathbb{R} \to L(C^2(\mathbb{R}), C^0(\mathbb{R}))$ is differentiable, and compute its differential.

2.4 Summary

Bounded linear maps between Banach spaces extend the notion of matrices to infinite-dimensional spaces. They provide a generalization of the notion of derivative to Banach spaces, in the form of differentials. Bounded operators of interest include infinite matrices of the spaces of sequences, and integral operators on $C^0(I)$. They share many of the properties of matrices, and one can talk for instance of the exponential of an operator, or solve linear equations with operators. Many interesting operators, such as differential operators, are *not* bounded from a Banach space to itself, and will require special tools to be developed in later lectures.

3

Ordinary differential equations

3.1 Classification of differential equations

Differential equations are generally separated in two major types: **initial value problems** (IVP) such as Newton's equations or the heat equation where the initial state of a system and a set of evolution laws are specified, and **boundary value problems** (BVP) such as the Laplace equation where a differential equation is specified in the interior of a region of space, together with a set of boundary conditions. Equations in a single dimension (either time or space) are called **ordinary differential equations** (ODE); equations in more dimensions are called **partial differential equations** (PDE). An equation is **linear** if linear combinations of solutions are still solutions. An IVP is called **autonomous** if the evolution law does not depend explicitly on time.

3.2 Linear ordinary differential equations

We will focus here on linear autonomous IVP ODEs, of the form

$$x' = Ax, \quad x(0) = x_0, \tag{3.1}$$

where $x_0 \in E$ and $A \in L(E)$, with E a Banach space. The theory of such equations is very similar whether E is finite- or infinite-dimensional; however, it is important that $A \in L(E)$, which excludes for instance the case of the heat equation.

Theorem 3.1. *The function*

$$x(t) = e^{tA} x_0$$

is the unique solution in $C^1(\mathbb{R}, E)$ of (3.1).

Proof. From Exercice 2.12, it is immediate to see that $x(t)$ is a solution. Now let $y(t)$ be another solution, and set $y(t) = e^{tA}\widetilde{y}(t)$; $\widetilde{y}(t)$ satisfies the equation $\widetilde{y}' = 0$, and therefore $\widetilde{y}(t) = x_0$, hence the result. □

Assume now that $E = \mathbb{R}^d$, and A is diagonalizable : $A = PDP^{-1}$ with D diagonal. Then we can change variables into the eigenbasis: $x(t) = Py(t)$. $y(t)$ satisfies the equation

$$y'(t) = Dy, \quad y(0) = P^{-1}x_0$$

These equations are now *decoupled*: for all $i = 1, \dots, d$,

$$y_i' = \lambda_i y_i, \quad y_i(0) = y_{i,0}$$

where the eigenvalue λ_i is the $i-$th element of the diagonal of D. The solution is

$$y_i(t) = e^{\lambda_i t} y_{i,0}$$

(this result can also be obtained by showing that $e^{tA} = Pe^{tD}P^{-1}$).

It is then apparent from $e^{\lambda_i t} = e^{\mathrm{Re}(\lambda_i)t} e^{\mathrm{Im}(\lambda_i)t}$ that the eigenvalues, and in particular the sign of their real parts, determines the nature of the dynamics. If all eigenvalues have a strictly negative real part, all trajectories converge to zero exponentially, possibly in a spiral if the eigenvalues have a nonzero imaginary part. If all eigenvalues have a strictly positive real part, all trajectories diverge to infinity, again possibly in a spiral. If at least one eigenvalue λ_i has a strictly positive real part, almost all trajectories diverge, but some (those for which $y_{i,0} = 0$) can remain bounded. If all eigenvalues have a zero real part, all trajectories are oscillatory.

In the non-diagonalizable case, the solutions are of a similar form, but can acquire an additional polynomial prefactor:

Exercise 3.2. *Solve the equation* (3.1) *in the case*

$$A = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix},$$

either by solving the differential equation directly or by computing the exponential.

3.3 Introduction to nonlinear systems

The existence and uniqueness of nonlinear systems is a much more subtle question. The major existence result is the Cauchy-Lipschitz theorem†:

Theorem 3.3. *Suppose that $f \in C^1(\mathbb{R}^d, \mathbb{R}^d)$, and let $x_0 \in \mathbb{R}^d$. There is $T > 0$ such that the equation*

$$x' = f(x), \quad x(0) = x_0 \tag{3.2}$$

has a unique solution in the interval $[0, T]$.

When f is not C^1, uniqueness might fail; when f grows too fast with x, the solution might only exist up to a finite time $T < \infty$. This is illustrated by the following two examples:

Exercise 3.4. *Solve the equations*

$$x' = \sqrt{x}, \quad x(0) = 0$$
$$x' = x^2, \quad x(0) = 1$$

The global behavior of nonlinear dynamics can be much more complicated than for linear systems. For instance, nonlinear dynamics can be *chaotic* (extremely sensitive to initial conditions). The *local* behavior on the other hand can be studied using linear theory. Zeros of f are stationary points of the dynamics. Near a stationary point x_*, f can be *linearized*:

$$x' = df(x_*)(x - x_*) + O\left(\|x - x_*\|^2\right).$$

The study of the dynamics of the linear ODE $(x - x_*)' = df(x_*)(x - x_*)$ then yields useful information on the nonlinear dynamics close to the stationary point; a representative result is

Theorem 3.5. *Let f be C^1 in a neighborhood U of $x_* \in \mathbb{R}^d$, with $f(x_*) = 0$, and assume that the eigenvalues of $df(x_*)$ all have a negative real part. Then, for all x_0 in a neighborhood $V \subset U$ of x_*, the differential equation $x' = f(x), x(0) = x_0$ has a unique solution in $\mathbb{R}^+$, and* $\lim_{t\to\infty} x = x_*$.

† This theorem further generalizes to the case where f is not C^1 but only Lipschitz in a neighborhood of x_0, and depends on time in a continuous way, but this version will suffice for our purposes.

Similarly one can show that when an eigenvalue of $df(x_*)$ has a positive real part, almost all trajectories starting close to x_* go away from it. In the other cases (involving eigenvalues of $df(x_*)$ zero real part), the behavior of the system is not determined by the linearization and depends on the higher-order terms.

Exercise 3.6. *Consider the spread of an epidemic in a population. The simplest model is the SIR model, where the total population $N(t)$ is split between susceptibles $S(t)$, infected $I(t)$ and recovered $R(t)$. Susceptible people get infected with some probability when they come in contact with infected people; infected people recover at a constant rate.*

- *Show that this can be modeled by the equations*

$$\frac{dS}{dt} = -\frac{\beta IS}{N}$$
$$\frac{dI}{dt} = \frac{\beta IS}{N} - \gamma I$$
$$\frac{dR}{dt} = \gamma I$$

- *What are the fixed-points and the eigenvalues of their jacobians? What does this imply?*
- *If $S(0) > 0$, $I(0) \geqslant 0$, $R(0) \geqslant 0$, assuming the existence of a solution, show that S, I and R stay non-negative and bounded for all positive times.*
- *Under the above condition on $S(0), I(0), R(0)$, show that the equation has a unique solution for all positive times. Hint: generalize the proof of the Cauchy-Lipschitz theorem to show that T can be taken to be uniformly bounded away from zero.*

3.4 Numerical integration

How to compute the solution of ordinary differential equations like (3.2) in practice? In the case of linear, autonomous systems we can simply expand e^{tA} in series, but this does not generalize to more complicated (non-autonomous, non-linear) systems. A much more general algorithm is the **explicit Euler method**, where we subdivide a time interval $[0, T]$ into discrete times $t_n = \frac{nT}{N}$ for $n = 0, \ldots, N$, and use a finite difference approximation for $x'(t_n) \approx \frac{x_{n+1} - x_n}{\Delta t}$, which leads to

$$x_{n+1} = x_n + \Delta t f(x_n)$$

with $\Delta t = \frac{T}{N}$[†].

Exercise 3.7. *Assume that $f(x) = Ax$ on $\mathbb{R}^d$. Show that, for all $T \in \mathbb{R}$, there is a constant C_T such that, for all $N \in \mathbb{N}$,*

$$\|x_N - e^{TA} x_0\| \leqslant C_T \Delta t$$

Note here that the *total* error at time T is of order Δt, while the *local* error $e^{\Delta t A} - (\mathrm{Id} + \Delta t A)$ (the one introduced by the finite difference approximation at each step) is of order $(\Delta t)^2$. The fact that the total error is bounded by a multiple of N times the local error reflects the *stability* (non-catastrophic amplification) of errors along the dynamics; this is easy to prove in our context, but non-trivial for complicated non-linear dynamics.

[†] Note that this method is presented here for pedagogical reasons, but should almost never be used directly: much more sophisticated (higher-order, and with automatic selection of the time-step) are implemented in well-established software, and should be used in practice.

3.5 Summary

Linear autonomous equations with bounded operators can be solved using exponentials, and allow a complete existence and uniqueness theory. In finite dimension, the eigenvalues determine the type of dynamics (divergent, convergent, oscillatory). The fixed-points or nonlinear differential equations can be studied by linearization. Differential equations can be solved numerically using the explicit Euler method.

4

French-English differences in mathematical notation

Most terms are transparent. Some non-obvious translations or *faux amis*:

French	English
application	map
borne	bound
calcul différentiel	calculus
démonstration	proof
dénombrable	countable
dérivée	derivative
dériver	differentiate
ensemble	set
inversible	invertible
Jacobienne	Jacobian
Jacobien	Jacobian determinant
suite	sequence
suite extraite	subsequence
vectoriel	vector

Notations are mostly the same, with a few notable exceptions:

- Open intervals use open brackets in French $[a, b[$, parentheses in English $[a, b)$.
- Transposes of matrices are still sometimes denoted ${}^tA$ in French (mostly by CPGE teachers...), but always A^T in English.
- Inequalities in English are strict by default:

Math	French	English
$x \geqslant 0$	x est positif	x is non-negative
$x > 0$	x est strictement positif	x is positive
$x \geqslant y$	x est plus grand que y	x is greater or equal to y
$x > y$	x est strictement plus grand que y	x is greater than y
$\forall n > m, x_n \geqslant x_m$	x_n est croissante	x_n is non-decreasing
$\forall n > m, x_n > x_m$	x_n est strictement croissante	x_n is increasing

5

Appendix A: Compactness

5.1 Compactness

A sequence $(y_m)_{m\in\mathbb{N}}$ is said to be a **subsequence** of $(x_n)_{n\in\mathbb{N}}$ if there is a (strictly) increasing sequence of integers $(n_m)_{m\in\mathbb{N}}$ such that $y_m = x_{n_m}$ for all $m \in \mathbb{N}$.

Definition 5.1. *A subset X of a normed vector space E is* ***compact***† *if and only if all sequences in X have a convergent subsequence with a limit in X.*

Compact subsets are bounded (because divergent sequences do not have convergent subsequences) and closed (because if a sequence is converging, all its subsequences converge to the same limit).

The motivation for compactness is

Theorem 5.2 (Bolzano-Weierstrass). *In a finite-dimensional vector space, closed and bounded sets are compact.*

Proof. We first prove the result for subsets of $\mathbb{R}$. Let $X \subset I_0 = [a_0, b_0]$ be a closed and bounded set in $\mathbb{R}$, and x_n a sequence in X. We use a dichotomy process. At least one of $[a_0, (a_0 + b_0)/2]$ or $[(a_0 + b_0)/2, b_0]$ must necessarily contain an infinite number of terms of x_n. We set I_1 to be one such interval, and $(x^1_n)_{n\in\mathbb{N}}$ be the subsequence of terms of x_n that belong to I_1. Repeating the process, we obtain a sequence of nested intervals $I_k = [a_k, b_k]$ and of nested subsequences $(x^k_n)_{n\in\mathbb{N}}$ of x_n in I_k such that $b_k - a_k = \frac{b_0 - a_0}{2^k}$. The sequence a_k is non-decreasing and bounded and therefore converges; so does b_k, to the same limit x_*. The subsequence $(x^k_0)_{k\in\mathbb{N}}$ converges to x_*, which by closedness belongs to X.

Now for a subset X of $\mathbb{R}^d$, we extract a subsequence y_n of x_n such that $(y_{n,1})_{n\in\mathbb{N}}$ converges. From this subsequence, we extract another subsequence z_n such that $(z_{n,2})_{n\in\mathbb{N}}$ converges, and repeat this process d times. □

Importantly, this is *not* true in infinite dimension, even in Banach spaces:

Exercise 5.3. *Find a bounded sequence in ℓ^∞ with no convergent subsequence.*

5.2 Applications: optimization and equivalence of norms

A classical application of compactness is in optimization

Proposition 5.4. *A real-valued continuous function on a compact set is bounded and attains its bounds.*

† Strictly speaking, our definition is that of *sequential compactness*. There are alternative definitions of compactness, equivalent in the case of normed vector spaces. This definition will suffice for our purposes.

Proof. Let X be compact, and $f : X \to \mathbb{R}$ a continuous function. Assume that f is not bounded on X: this means that there exists a sequence x_n of elements of X such that $f(x_n) \geqslant n$. By compactness, we can extract a subsequence y_n which converges to some $x_* \in X$. By continuity, $\lim_{n\to\infty} f(y_n) = f(x_*)$, which is impossible.

Now let x_n be a *minimizing sequence*, i.e. a sequence of elements of X such that $f(x_n) \to \inf_{x\in X} f(x)$ (such a sequence exists by definition of the infinimum). By compactness, we can extract a converging subsequence; by continuity, f at that limit is equal to $\inf_{x\in X} f(x)$, hence f attains its minimum. We repeat the same argument for the maximum. □

This gives a particularly simple proof of the important

Proposition 5.5. *If E is finite-dimensional, all norms are equivalent.*

Proof. Using a basis, we can assume that $E = \mathbb{K}^d$. Let N be a norm. We will show that N is equivalent to the infinity norm, which will by transitivity show that all norms are equivalent. We have for all $u = \sum_{i=1}^d u_i e_i \in E$

$$N(u) = N\left(\sum_{i=1}^d u_i e_i\right) \leqslant \sum_{i=1}^d |u_i| N(e_i) \leqslant \left(\sum_{i=1}^d N(e_i)\right) \|u\|_\infty.$$

By the triangle inequality, N is continuous from $(E, \|\cdot\|_\infty)$ to $\mathbb{R}$. By the Bolzano-Weierstrass theorem, the unit sphere $S = \{u \in E, \|u\|_\infty = 1\}$ is compact (for the infinity norm). It follows that N is bounded on S and attains its bounds. Assume that $\inf_{u\in S} N(u) = 0$: this would imply the existence of a nonzero $u \in S$ with $N(u) = 0$, which is impossible. It follows that there are $c, C > 0$ such that, for all $u \in S, c\|u\|_\infty \leqslant N(u) \leqslant C\|u\|_\infty$. The result follows by homogeneity. □

6

Appendix B: The Banach fixed-point theorem and applications

We prove the fundamental Banach fixed-point theorem, and apply it to two striking results of large practical importance: the Cauchy-Lipschitz theorem for ODEs, and the method of Lagrange multipliers for constrained optimization.

6.1 The Banach fixed-point theorem

In the following, E is a Banach space.

Theorem 6.1. *Let U a closed set in E, and let $F : E \to E$ be contractive in U, in the sense that there is $\alpha < 1$ such that, for all $x, y \in U$,*

$$\|F(x) - F(y)\| \leqslant \alpha \|x - y\|.$$

There is a unique fixed-point x_* of F in U. For all $x_0 \in U$, the iteration

$$x_{n+1} = F(x_n)$$

converges to x_*.

Proof. Let $x_0 \in U$, and define x_n as above. Then we have

$$\|x_{n+1} - x_n\| \leqslant \alpha \|x_n - x_{n-1}\|$$

from which it follows that the telescopic series $\sum_{n \in \mathbb{N}} x_{n+1} - x_n$ converges normally, and therefore that x_n converges to some x_*. Contractivity implies continuity; passing to the limit in $x_{n+1} = F(x_n)$ shows that x_* is a fixed-point of F. Uniqueness follows from contractivity. □

Contractivity can usually be proven from the following criterion:

Proposition 6.2. *Let U be a convex closed set, and F a C^1 map from U to U with $\alpha < 1$ such that, for all $x \in U$, $|||df(x)||| \leqslant \alpha$. Then F is contractive.*

Proof. Let $x, y \in U$. By the fundamental theorem of calculus applied to the C^1 function $\phi(t) = F(x + t(y - x))$, we have

$$F(y) - F(x) = \int_0^1 df(x + t(y - x))(y - x)dt$$
$$\|F(y) - F(x)\| \leqslant \alpha \|y - x\|.$$

□

Exercise 6.3. *Let $A \in L(E)$ with $|||A||| < 1$. Prove that* $\mathrm{Id} - A$ *is invertible, and link the iterates of the fixed-point iteration with the Neumann series.*

It is hard to overstate the importance of the Banach fixed-point theorem: it is one of the few ways to prove the existence of solutions to nonlinear equations. Furthermore, it is constructive, and provides an iterative algorithm that converges to the solution.

6.2 The Cauchy-Lipschitz theorem

Theorem 6.4 (Cauchy-Lipschitz). *Suppose that $f \in C^1(\mathbb{R}^d, \mathbb{R}^d)$, and let $x_0 \in \mathbb{R}^d$. There is $T > 0$ such that the equation*

$$x' = f(x), \quad x(0) = x_0$$

has a unique solution in the interval $[0, T)$.

Proof. We first reformulate the ODE as an integral equation by integrating:

$$x(t) = x_0 + \int_0^t f(x(t'))dt'$$

We show that this equation has a unique solution in $X = C^0([0, T), \mathbb{R}^d)$ for T small enough. Let $F : X \to X$ be the application defined by

$$(F(y))(t) = x_0 + \int_0^t f(y(t'))dt'.$$

We try to find fixed points of this equation. Fix an arbitrary $r \in \mathbb{R}$. Note first that, if $\|y - x_0\|_X \leqslant r$ then, for all $t \in [0, T)$, $|y(t) - x_0| \leqslant r$ and so

$$\|F(y) - x_0\|_X \leqslant T \sup_{x \in B(x_0, r)} |f(x)|.$$

It follows that, for T small enough, F maps $U = B(x_0, r)$ to itself. Now we compute the differential of F:

$$(F(y+h))(t) = x_0 + \int_0^t f(y(t') + h(t'))dt' = F(y)(t) + \int_0^t \big(f'(y(t'))h(t') + O(h(t')^2)\big)dt'$$
$$F(y+h) = F(y) + F'(y)h + O(\|h\|_X^2)$$

with

$$(F'(y)h)(t') = \int_0^t f'(y(t'))h(t')dt'$$

There is a constant $C > 0$ such that, for all $y \in U$, we have $\|F'(y)\| \leqslant TC$. Choosing T small enough, F is a contraction on U and the result follows from the Banach fixed-point theorem. □

6.3 The implicit function theorem

Theorem 6.5 (Implicit function theorem). *Let E be a Banach space, and $G : E \times \mathbb{R} \to E$ be C^1. Assume that there is x_0, α_0 such that $G(x_0, \alpha_0) = 0$, and that $\partial_x G'(x_0, \alpha_0)$ is invertible. Then there is a neighborhood I of α_0 and a neighborhood U of x_0 with a C^1 curve $x(\alpha)$ from I to U such that*

$$G(x(\alpha), \alpha) = 0$$

for all $\alpha \in I$. $x(\alpha)$ is the unique solution of the equation $G(x, \alpha) = 0$ in U.

Proof. We give the proof in the case where G is C^2, where the connection to the Newton method is clearer. The proof in the C^1 case is a slight modification of the arguments below.

When G is C^2, the exercise 2.8 on perturbations of invertible operators shows that the mapping $x \to \partial_x G(x, \alpha)^{-1}$ is C^1 on $U \times I$ for some neighborhoods U and I. Inspired by the Newton method

$$x_{n+1} = x_n - f'(x_n)^{-1} f(x_n)$$

for finding zeros of a function $f : E \to E$, we define for all $\alpha \in I$ a mapping

$$F_\alpha(x) = x - \partial_x G(x, \alpha)^{-1} G(x, \alpha)$$

This mapping is C^1 from U to E, and $F_{\alpha_0}(x_0) = x_0$ Furthermore,

$$F'_\alpha(x) = \mathrm{Id} - JG(x, \alpha) - \mathrm{Id},$$

where J is the differential at x of the mapping $x \mapsto \partial_x G(x, \alpha)^{-1}$. It follows that $F'_{\alpha_0}(x_0) = 0$, and therefore that $\|F_\alpha(x) - x_0\| = O(|\alpha - \alpha_0|(1 + \|x - x_0\|) + \|x - x_0\|^2)$. By restricting the diameters of U and I appropriately, we can ensure that for all $\alpha \in I$, F_α is a contraction from U to itself. The result follows by the Banach fixed-point theorem. □

Note that this shows as a byproduct the convergence of the Newton method. This theorem generalizes easily to the case where α is a vector. As an easy consequence, taking $G(x, y) = f(x) - y$ we obtain the inverse function: a function f with an invertible differentiable is locally invertible. Another typical application of the implicit function theorem is to weakly nonlinear equations:

Exercise 6.6. *Let I be a bounded interval, f continuous on I and K continuous on $I \times I$. Show that the equation*

$$u(x) = f(x) + \varepsilon u(x)^2 + \delta \int_I K(x, y) u(y) dy$$

has a unique solution for ε, δ small enough.

6.4 Local parametrizations and constrained optimization

When $E = \mathbb{R}^N$, the implicit function theorem states that, locally, d non-degenerate equations in d unknowns specify a unique solution. We can generalize to the case of $k < d$ equations. For this we need the following deep but elementary linear algebra result, which gives conditions on b for $Ax = b$ to have solutions when A is non-square or non-invertible.

Lemma 6.7 ("Fundamental theorem of linear algebra"). *If A is a (not necessarily square) matrix, then* $\mathrm{Im}(A) = \mathrm{Ker}(A^T)^\perp$.

Proof. Note that

$$z^T(Ax) = (A^T z)^T x.$$

It follows that if $z \in \mathrm{Ker}(A^T)$, then $z \perp \mathrm{Im}(A)$. Vice-versa, if $z \perp \mathrm{Im}(A)$, $A^T z$ is orthogonal to every vector, and so $z \in \mathrm{Ker}(A^T)$. □

Proposition 6.8 (Local parametrization). *Let $x_0 \in \mathbb{R}^d$ and $g : \mathbb{R}^d \to \mathbb{R}^k$, with $k \leqslant d$ be C^2 in a neighborhood of x_0, and such that $g(x_0) = 0$ and $g'(x_0) : \mathbb{R}^d \to \mathbb{R}^k$ is surjective. Then there exists a mapping x from a neighborhood U of 0 in $\mathbb{R}^{d-k}$ to a neighborhood E of x_0 such that $g(x(t)) = 0$ for all $t \in U$, and the set $x(U)$ contains all the solutions of $g(x) = 0$ in E.*

Proof. Split the space $\mathbb{R}^d$ as the orthogonal sum $E_1 + E_2$, where $E_1 = \mathrm{Im}(g'(x_0)^T)$ and $E_2 = \mathrm{Im}(g'(x_0)^T)^\perp = \mathrm{Ker}(g'(x_0))$. The columns of the matrix $U_1 = g'(x_0)^T$ form a basis of E_1. Let U_2 be an orthogonal basis of E_2. In a linear approximation, $g'(x_0)(x - x_0) = 0$: the constraint $g(x)$ forces $x - x_0$ to be of the form $U_2 t$ for some $t = U_2^T(x - x_0) \in \mathbb{R}^{d-k}$, and the variable t enumerates all solutions.

In the non-linear regime, we set out to solve the system of $d = k + (d - k)$ equations in the d unknowns x

$$
\begin{aligned}
g(x) &= 0 \\
U_2^T(x - x_0) &= t
\end{aligned}
$$

for all values of $t \in \mathbb{R}^{d-k}$ small enough. The jacobian at x_0 of this system of equations with respect to x is the $d \times d$ matrix

$$
\begin{pmatrix} g'(x_0) \\ U_2^T \end{pmatrix}
$$

This is the transpose of the matrix (U_1, U_2), whose columns form a basis of $\mathbb{R}^d$. Therefore, this matrix is invertible, and the result follows from the implicit function theorem. $\square$

Theorem 6.9 (Lagrange multipliers). *Let $x_* \in \mathbb{R}^d$ be a local minimum of the optimization problem*

$$
\begin{aligned}
&\min f(x) \\
&g(x) = 0
\end{aligned}
$$

where $f : \mathbb{R}^d \to \mathbb{R}$ is C^1, $g : \mathbb{R}^d \to \mathbb{R}$ is C^1, and $g'(x_*) \neq 0$. Then there is $\lambda \in \mathbb{R}$ such that $\nabla f(x_*) = \lambda \nabla g(x_*)$.

Proof. By the local parametrization theorem, there is a parametrization $x(t)$ of the solutions of $g(x) = 0$ for t in a neighborhood of 0 in $\mathbb{R}^{d-1}$. $t = 0$ is an unconstrained local minimum of $f(x(t))$, and so $x'(0)^T \nabla f(x_*) = 0$. From the considerations in the proof of the local parametrization theorem, the columns of $x'(0)$ are a basis for $\mathrm{Ker}(g'(x_*)) = \nabla g(x_*)^{\perp}$. It follows that $\nabla f(x_*) \in (\nabla g(x_*)^{\perp})^{\perp}$, hence the result. $\square$

This easily generalizes to several constraints, as long as their gradients are linearly independent.

References

[1] H. Brézis, *Analyse fonctionnelle : théorie et applications* (Dunod, 1999).
Complete and concise course in functional analysis, in French. For those who don't like interruptions in their definition-theorem-proofs.

[2] E. Kreyszig, *Introductory functional analysis with applications* (Wiley New York, 1978).
Elementary, pedagogical and verbose treatment of functional analysis.

[3] M. Reed and B. Simon, *Functional analysis* (Elsevier, 2012).
Somewhat of a compromise between the previous two extremes.