Introductory Functional Analysis

Juan Peypouquet

February 10, 2022

Contents

In these short lecture notes we recall the basic concepts and results from functional analysis and calculus that are useful for a general purpose. A first chapter is devoted to general normed spaces. We begin by establishing some of their main properties, with an emphasis on the linear functions between spaces. This leads us to bounded linear functionals and the topological dual. Second, we review the Hahn-Banach Separation Theorem, a very powerful tool with important consequences. Next, we discuss some relevant results concerning the weak topology, especially in terms of closedness and compactness. Finally, we include a subsection on differential calculus, which also provides an introduction to standard smooth optimization techniques. The second chapter deals with Hilbert spaces, and their very rich geometric structure, including the ideas of projection and orthogonality. We also revisit some of the general concepts from the first section (duality, reflexivity, weak convergence) in the light of this geometry.

For a comprehensive presentation, the reader is referred to [1] and [6]. In what follows, all vector spaces are defined over $\mathbf{R}$.

These notes are a partial reproduction of a preprint version of [5], exclusively for pedagogic purposes, in accordance with the copyright agreement between the author and Springer.

Chapter 1

Normed spaces

A norm on a real vector space X is a function $\|\cdot\| : X \to \mathbf{R}$ such that

i) $\|x\| > 0$ for all $x \neq 0$;

ii) $\|\alpha x\| = |\alpha|\|x\|$ for all $x \in X$ and $\alpha \in \mathbf{R}$;

iii) The triangle inequality $\|x+y\| \leq \|x\| + \|y\|$ holds for all $x, y \in X$.

A normed space is a vector space where a norm has been specified.

Example 1.1. The space $\mathbf{R}^N$ with the norms: $\|x\|_\infty = \max\limits_{i=1,\dots,N} |x_i|$, or $\|x\|_p = \left(\sum\limits_{i=1}^{N} |x_i|^p\right)^{1/p}$, for $p \geq 1$.

Example 1.2. The space $\mathscr{C}([a,b];\mathbf{R})$ of continuous real-valued functions on the interval $[a,b]$, with the norm $\|\cdot\|_\infty$ defined by $\|f\|_\infty = \max\limits_{t \in [a,b]} |f(t)|$.

Example 1.3. The space $L^1(a,b;\mathbf{R})$ of Lebesgue-integrable real-valued functions on the interval $[a,b]$, with the norm $\|\cdot\|_1$ defined by $\|f\|_1 = \int_a^b |f(t)|\,dt$.

Given $r > 0$ and $x \in X$, the open ball of radius r centered at x is the set

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

The closed ball is

$$\bar{B}_X(x,r) = \{y \in X : \|y-x\| \leq r\}.$$

We shall omit the reference to the space X whenever it is clear from the context. A set is bounded if it is contained in some ball.

In a normed space one can define a canonical topology as follows: a set V is a neighborhood of a point x if there is $r > 0$ such that $B(x,r) \subset V$. We call it the strong topology, in contrast with the weak topology to be defined later on.

We say that a sequence (x_n) in X converges (strongly) to $\bar{x} \in X$, and write $x_n \to \bar{x}$, as $n \to \infty$ if $\lim_{n\to\infty} \|x_n - \bar{x}\| = 0$. The point $\bar{x}$ is the limit of the sequence. On the other hand, (x_n) has the Cauchy property, or it is a Cauchy sequence, if $\lim_{n,m\to\infty} \|x_n - x_m\| = 0$. Every convergent sequence has the Cauchy property, and every Cauchy sequence is bounded. A Banach space is a normed space in which every Cauchy sequence is convergent, a property called completeness.

Example 1.4. The spaces in Examples 1.1, 1.2 and 1.3 are Banach spaces.

We have the following result:

Theorem 1.5 (Baire's Category Theorem). Let X be a Banach space and let (C_n) be a sequence of closed subsets of X. If each C_n has empty interior, then so does $\bigcup_{n\in\mathbf{N}} C_n$.

Proof. A set $C \subset X$ has empty interior if, and only if, every open ball intersects C^c. Let B be an open ball. Take another open ball B' whose closure is contained in B. Since C_0^c has empty interior, $B' \cap C_0^c \neq \emptyset$. Moreover, since both B' and C_0^c are open, there exist $x_1 \in X$ and $r_1 \in (0, \frac{1}{2})$ such that $B(x_1, r_1) \subset B' \cap C_0^c$. Analogously, there exist $x_2 \in X$ and $r \in (0, \frac{1}{4})$ such that $B(x_2, r_2) \subset B(x_1, r_1) \cap C_1^c \subset B' \cap C_0^c \cap C_1^c$. Inductively, one defines $(x_m, r_m) \in X \times \mathbf{R}$ such that $x_m \in B(x_n, r_n) \cap \left(\bigcap_{k=0}^{n} C_k^c\right)$ and $r_m \in (0, 2^{-m})$ for each $m > n \geq 1$. In particular, $\|x_m - x_n\| < 2^{-n}$ whenever $m > n \geq 1$. It follows that (x_n) is a Cauchy sequence and so, it must converge to some $\bar{x}$, which must belong to $\overline{B'} \cap \left(\bigcap_{k=0}^{\infty} C_k^c\right) \subset B \cap \left(\bigcup_{k=0}^{\infty} C_k\right)^c$, by construction. □

We shall find several important consequences of this result, especially the Banach-Steinhaus Uniform Boundedness Principle (Theorem 1.9) and the continuity of convex functions in the interior of their domain (see, for instance, [5]).

1.1 Bounded linear operators and functionals, topological dual

Bounded linear operators

Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be normed spaces. A linear operator $L : X \to Y$ is bounded if

$$\|L\|_{\mathscr{L}(X;Y)} := \sup_{\|x\|_X = 1} \|L(x)\|_Y < \infty.$$

The function $\|\cdot\|_{\mathscr{L}(X;Y)}$ is a norm on the space $\mathscr{L}(X;Y)$ of bounded linear operators from $(X, \|\cdot\|_X)$ to $(Y, \|\cdot\|_Y)$. For linear operators, boundedness and (uniform) continuity are closely related. This is shown in the following result:

Proposition 1.6. Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be normed spaces and let $L : X \to Y$ be a linear operator. The following are equivalent:

i) L is continuous in 0;

ii) L is bounded; and

iii) L is uniformly Lipschitz-continuous in X.

Proof. Let i) hold. For each $\varepsilon > 0$ there is $\delta > 0$ such that $\|L(h)\|_Y \le \varepsilon$ whenever $\|h\|_X \le \delta$. If $\|x\|_X = 1$, then $\|L(x)\|_Y = \delta^{-1}\|L(\delta x)\|_Y \le \delta^{-1}\varepsilon$ and so, $\sup_{\|x\|=1} \|L(x)\|_Y < \infty$. Next, if ii) holds, then

$$\|L(x) - L(z)\|_Y = \|x - z\|_X \left\| L\left(\frac{x-z}{\|x-z\|_X}\right)\right\| \le \|L\|_{\mathscr{L}(X;Y)} \|x - z\|_X$$

and L is uniformly Lipschitz-continuous. Clearly, iii) implies i). □

We have the following:

Proposition 1.7. If $(X, \|\cdot\|_X)$ is a normed space and $(Y, \|\cdot\|_Y)$ is a Banach space, then $(\mathscr{L}(X;Y), \|\cdot\|_{\mathscr{L}(X;Y)})$ is a Banach space.

Proof. Let (L_n) be a Cauchy sequence in $\mathscr{L}(X;Y)$. Then, for each $x \in X$, the sequence $(L_n(x))$ has the Cauchy property as well. Since Y is complete, there exists $L(x) = \lim_{n\to\infty} L_n(x)$. Clearly, the function $L : X \to Y$ is linear. Moreover, since (L_n) is a Cauchy sequence, it is bounded. Therefore, there exists $C > 0$ such that $\|L_n(x)\|_Y \le \|L_n\|_{\mathscr{L}(X;Y)} \|x\|_X \le C\|x\|_X$ for all $x \in X$. Passing to the limit, we deduce that $L \in \mathscr{L}(X;Y)$ and $\|L\|_{\mathscr{L}(X;Y)} \le C$. Finally, from the Cauchy property, we easily deduce that $\lim_{n\to\infty} \|L_n - L\|_{\mathscr{L}(X;Y)} = 0$. □

The kernel of $L \in \mathscr{L}(X;Y)$ is the set

$$\ker(L) = \{ x \in X \, : \, L(x) = 0 \} = L^{-1}(0),$$

which is a closed subspace of X. The range of L is

$$R(L) = L(X) = \{ L(x) \, : \, x \in X \}.$$

It is a subspace of Y, but it is not necessarily closed.

An operator $L \in \mathscr{L}(X;Y)$ is invertible if there exists an operator in $\mathscr{L}(Y;X)$, called the inverse of L, and denoted by L^{-1}, such that $L^{-1} \circ L(x) = x$ for all $x \in X$ and $L \circ L^{-1}(y) = y$ for all $y \in Y$. The set of invertible operators in $\mathscr{L}(X;Y)$ is denoted by $\mathrm{Inv}(X;Y)$. We have the following:

Proposition 1.8. Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be Banach spaces. The set $\mathrm{Inv}(X;Y)$ is open in $\mathscr{L}(X;Y)$ and the function $\Phi : \mathrm{Inv}(X;Y) \to \mathrm{Inv}(Y;X)$, defined by $\Phi(L) = L^{-1}$, is continuous.

Proof. Let $L_0 \in \mathrm{Inv}(X;Y)$ and let $L \in B(L_0, \|L_0^{-1}\|^{-1})$. Let I_X be the identity operator in X and write $M = I_X - L_0^{-1} \circ L = L_0^{-1} \circ (L_0 - L)$. Denote by M^k the composition of M with itself k times and define $M_n = \sum_{k=0}^{n} M^k$. Since $\|M\| < 1$,

(M_n) is a Cauchy sequence in $\mathscr{L}(X;X)$ and must converge to some $\bar{M}$. But $M \circ M_n = M_n \circ M = M_{n+1} - I_X$ implies $\bar{M} \circ (I_X - M) = (I_X - M) \circ \bar{M} = I_X$, which in turn gives

$$(\bar{M} \circ L_0^{-1}) \circ L = L \circ (\bar{M} \circ L_0^{-1}) = I_X.$$

It ensues that $L \in \mathrm{Inv}(X;Y)$ and $L^{-1} = \bar{M} \circ L_0^{-1}$. For the continuity, since

$$\|L^{-1} - L_0^{-1}\| \leq \|L^{-1} \circ L_0 - I_X\| \cdot \|L_0^{-1}\| = \|L_0^{-1}\| \cdot \|\bar{M}^{-1} - I_X\|$$

and

$$\|M_{n+1} - I_X\| = \|M \circ M_n\| \leq \|M\| \cdot (1 - \|M\|)^{-1},$$

we deduce that

$$\|L^{-1} - L_0^{-1}\| \leq \frac{\|L_0^{-1}\|^2}{1 - \|M\|} \|L - L_0\|.$$

Observe that Φ is actually Lipschitz-continuous in every closed ball $\bar{B}(L_0, R)$ with $R < \|L_0^{-1}\|^{-1}$. □

This and other useful calculus tools for normed spaces can be found in [2].

A remarkable consequence of linearity and completeness is that pointwise boundedness implies boundedness in the operator norm $\|\cdot\|_{\mathscr{L}(X;Y)}$. More precisely, we have the following consequence of Baire's Category Theorem 1.5:

Theorem 1.9 (Banach-Steinhaus Uniform Boundedness Principle). Let $(L_\lambda)_{\lambda \in \Lambda}$ be a family of bounded linear operators from a Banach space $(X, \|\cdot\|_X)$ to a normed space $(Y, \|\cdot\|_Y)$. If

$$\sup_{\lambda \in \Lambda} \|L_\lambda(x)\|_Y < \infty$$

for each $x \in X$, then

$$\sup_{\lambda \in \Lambda} \|L_\lambda\|_{\mathscr{L}(X;Y)} < \infty.$$

Proof. For each $n \in \mathbf{N}$, the set

$$C_n := \{x \in X : \sup_{\lambda \in \Lambda} \|L_\lambda(x)\|_Y \leq n\}$$

is closed, as intersection of closed sets. Since $\cup_{n \in \mathbf{N}} C_n = X$ has nonempty interior, Baire's Category Theorem 1.5 shows the existence of $N \in \mathbf{N}$, $\hat{x} \in X$ and $\hat{r} > 0$ such that $B(\hat{x}, \hat{r}) \subset C_N$. This implies

$$r\|L_\lambda(h)\| \leq \|L_\lambda(\hat{x} + rh)\|_Y + \|L_\lambda(\hat{x})\|_Y \leq 2N$$

for each $r < \hat{r}$ and $\lambda \in \Lambda$. It follows that $\sup_{\lambda \in \Lambda} \|L_\lambda\|_{\mathscr{L}(X;Y)} < \infty$. □

The topological dual and the bidual

The topological dual of a normed space $(X,\|\cdot\|)$ is the normed space $(X^*,\|\cdot\|_*)$, where $X^*=\mathscr{L}(X;\mathbf{R})$ and $\|\cdot\|_*=\|\cdot\|_{\mathscr{L}(X;\mathbf{R})}$. It is actually a Banach space, by Proposition 1.7. Elements of $X*$ are bounded linear functionals. The bilinear function $\langle\cdot,\cdot\rangle_{X^*,X}:X^*\times X\to\mathbf{R}$, defined by

$$\langle L,x\rangle_{X^*,X}=L(x),$$

is the duality product between X and X^*. If the space can be easily guessed from the context, we shall write $\langle L,x\rangle$ instead of $\langle L,x\rangle_{X^*,X}$ to simplify the notation.

The orthogonal space or annihilator of a subspace V of X is

$$V^\perp=\{L\in X^*:\langle L,x\rangle=0\quad\text{for all}\quad x\in V\},$$

which is a closed subspace of X^*, even if V is not closed.

The topological bidual of $(X,\|\cdot\|)$ is the topological dual of $(X^*,\|\cdot\|_*)$, which we denote by $(X^{**},\|\cdot\|_{**})$. Each $x\in X$ defines a linear function $\mu:X\to\mathbf{R}$ by

$$\mu_x(L)=\langle L,x\rangle_{X*,X}$$

for each $L\in X^*$. Since $\langle L,x\rangle\le\|L\|_*\|x\|$ for each $x\in X$ and $L\in X^*$, we have $\|\mu_x\|_{**}\le\|x\|$, so actually $\mu_x\in X^{**}$. The function $\mathscr{J}:X\to X^{**}$, defined by $\mathscr{J}(x)=\mu_x$, is the canonical embedding of X into X^{**}. Clearly, the function $\mathscr{J}$ is linear and continuous. We shall see later (Proposition 1.17) that $\mathscr{J}$ is an isometry. This fact will imply, in particular, that the canonical embedding $\mathscr{J}$ is injective. The space $(X,\|\cdot\|)$ is reflexive if $\mathscr{J}$ is also surjective. In other words, if every element μ of X^{**} is of the form $\mu=\mu_x$ for some $x\in X$. Necessarily, $(X,\|\cdot\|)$ must be a Banach space since it is homeomorphic to $(X^{**},\|\cdot\|_{**})$.

The adjoint operator

Let $(X,\|\cdot\|_X)$ and $(Y,\|\cdot\|_Y)$ be normed spaces and let $L\in\mathscr{L}(X;Y)$. Given $y^*\in Y^*$, the function $\zeta_{y^*}:X\to\mathbf{R}$ defined by

$$\zeta_{y^*}(x)=\langle y^*,Lx\rangle_{Y^*,Y}$$

is linear and continuous. In other words, $\zeta_{y^*}\in X^*$. The adjoint of L is the operator $L^*:Y^*\to X^*$ defined by

$$L^*(y^*)=\zeta_{y^*}.$$

In other words, L and L^* are linked by the identity

$$\langle L^*y^*,x\rangle_{X^*,X}=\langle y^*,Lx\rangle_{Y^*,Y}$$

We immediately see that $L^* \in \mathscr{L}(Y^*;X^*)$ and

$$\|L^*\|_{\mathscr{L}(Y^*;X^*)} = \sup_{\|y^*\|_{Y^*}=1} \left[\sup_{\|x\|_X=1} \langle y^*, Lx \rangle_{Y^*,Y} \right] \leq \|L\|_{\mathscr{L}(X;Y)}.$$

We shall verify later (Corollary 1.18) that the two norms actually coincide.

1.2 The Hahn-Banach Separation Theorem

The Hahn-Banach Separation Theorem is a cornerstone in Functional and Convex Analysis. As we shall see in next chapter, it has several important consequences.

Let X be a real vector space. A set $C \subseteq X$ is convex if for each pair of points of C, the segment joining them also belongs to C. In other words, if the point $\lambda x + (1-\lambda)y$ belongs to C whenever $x, y \in C$ and $\lambda \in (0,1)$.

Theorem 1.10 (Hahn-Banach Separation Theorem). Let A and B be nonempty, disjoint convex subsets of a normed space $(X, \|\cdot\|)$.

i) If A is open, there exists $L \in X^* \setminus \{0\}$ such that $\langle L, x \rangle < \langle L, y \rangle$ for each $x \in A$ and $y \in B$.

ii) If A is compact and B is closed, there exist $L \in X^* \setminus \{0\}$ and $\varepsilon > 0$ such that $\langle L, x \rangle + \varepsilon \leq \langle L, y \rangle$ for each $x \in A$ and $y \in B$.

Remarks

Before proceeding with the proof of the Hahn-Banach Separation Theorem 1.10, some remarks are in order:

First, Theorem 1.10 is equivalent to

Theorem 1.11. Let C be a nonempty, convex subset of a normed space $(X, \|\cdot\|)$ not containing the origin.

i) If C is open, there exists $L \in X^* \setminus \{0\}$ such that $\langle L, x \rangle < 0$ for each $x \in C$.

ii) If C is closed, there exist $L \in X^* \setminus \{0\}$ and $\varepsilon > 0$ such that $\langle L, x \rangle + \varepsilon \leq 0$ for each $x \in C$.

Clearly, Theorem 1.11 is a special case of Theorem 1.10. To verify that they are actually equivalent, simply write $C = A - B$ and observe that C is open if A is, while C is closed if A is compact and B is closed.

Second, part ii) of Theorem 1.11 can be easily deduced from part i) of Theorem 1.10 by considering asufficiently small open ball A around the origin (not intersecting C), and writing $B = C$.

Finally, in finite-dimensional spaces, part i) of Theorem 1.11 can be obtained without any topological assumptions on the sets involved. More precisely, we have the following:

Proposition 1.12. Given $N \geq 1$, let C be a nonempty and convex subset of $\mathbf{R}^N$ not containing the origin. Then, there exists $v \in \mathbf{R}^N \setminus \{0\}$ such that $v \cdot x \leq 0$ for each $x \in C$. In particular, if $N \geq 2$ and C is open, then

$$V = \{ x \in \mathbf{R}^N : v \cdot x = 0 \}$$

is a nontrivial subspace of $\mathbf{R}^N$ that does not intersect C.

Proof. Let $(x_n) \in C$ such that the set $\{x_n : n \geq 1\}$ is dense in C. Let C_n be the convex hull of the set $\{x_k : k = 1, \ldots, n\}$ and let p_n be the least-norm element of C_n. By convexity, for each $x \in C_n$ and $t \in (0,1)$, we have

$$\|p_n\|^2 \leq \|p_n + t(x - p_n)\|^2 = \|p_n\|^2 + 2t\, p_n \cdot (x - p_n) + t^2 \|x - p_n\|^2.$$

Therefore,

$$0 \leq 2\|p_n\|^2 \leq 2\, p_n \cdot x + t\|x - p_n\|^2.$$

Letting $t \to 0$, we deduce that $p_n \cdot x \geq 0$ for all $x \in C_n$. Now write $v_n = -p_n/\|p_n\|$. The sequence (v_n) lies in the unit sphere, which is compact. We may extract a subsequence that converges to some $v \in \mathbf{R}^N$ with $\|v\| = 1$ (thus $v \neq 0$) and $v \cdot x \leq 0$ for all $x \in C$. □

Proof of Theorem 1.11

Many standard Functional Analysis textbooks begin by presenting a general form of the Hahn-Banach Extension Theorem (see Theorem 1.13 below) and used to prove Theorem 1.11. We preferred to take the opposite path here, which has a more convex-analytic flavor. The same approach can be found, for instance, in [7].

Step 1: If the dimension of X is at least 2, there is a nontrivial subspace of X not intersecting C.

Take any two-dimensional subspace Y of X. If $Y \cap C = \emptyset$ there is nothing to prove. Otherwise, identify Y with $\mathbf{R}^2$ and use Proposition 1.12 to obtain a subspace of Y disjoint from $Y \cap C$, which clearly gives a subspace of X not intersecting C.

Step 2: There is a closed subspace M of X such that $M \cap C = \emptyset$ and the quotient space X/M has dimension 1.

Let $\mathscr{M}$ be the collection of all subspaces of X not intersecting C, ordered by inclusion. Step 1 shows that $\mathscr{M} \neq \emptyset$. According to Zorn's Lemma (see, for instance, [1, Lemma 1.1]), $\mathscr{M}$ has a maximal element M, which must be a closed subspace of X not intersecting C. The dimension of the quotient space X/M is at least 1 because $M \neq X$. The canonical homomorphism $\pi : X \to X/M$ is continuous and open. If the dimension of X/M is greater than 1, we use Step 1 again with $\tilde{X} = X/M$ and $\tilde{C} = \pi(C)$ to find a nontrivial subspace $\tilde{M}$ of $\tilde{X}$ that does not intersect $\tilde{C}$. Then $\pi^{-1}(\tilde{M})$ is a subspace of X that does not intersect C and is strictly greater than M, contradicting the maximality of the latter.

Step 3: Conclusion.

Take any (necessarily continuous) isomorphism $\phi : X/M \to \mathbf{R}$ and set $L = \phi \circ \pi$. Then, either $\langle L, x\rangle < 0$ for all $x \in C$, or $\langle -L, x\rangle < 0$ for all $x \in C$.

A few direct but important consequences

The following is known as the Hahn-Banach Extension Theorem:

Theorem 1.13. Let M be a subspace of X and let $\ell : M \to \mathbf{R}$ be a linear function such that $\langle \ell, x\rangle \le \alpha \|x\|$ for some $\alpha > 0$ and all $x \in M$. Then, there exists $L \in X^*$ such that $\langle L, x\rangle = \langle \ell, x\rangle$ for all $x \in M$ and $\|L\|_* \le \alpha$.

Proof. Define

$$\begin{aligned} A &= \{ (x,\mu) \in X \times \mathbf{R} : \mu > \alpha\|x\|, x \in X\} \\ B &= \{ (y,\nu) \in X \times \mathbf{R} : \nu = \langle \ell, y\rangle, y \in M\}. \end{aligned}$$

By the Hahn-Banach Separation Theorem 1.10, there is $(\tilde{L}, s) \in X^* \times \mathbf{R} \setminus \{(0,0)\}$ such that

$$\langle \tilde{L}, x\rangle + s\mu \le \langle \tilde{L}, y\rangle + s\nu$$

for all $(x,\mu) \in A$ and $(y,\nu) \in B$. Taking $x = y = 0$, $\mu = 1$ and $\nu = 0$, we deduce that $s \le 0$. If $s = 0$, then $\langle \tilde{L}, x - y\rangle \le 0$ for all $x \in X$ and so $\tilde{L} = 0$, which contradicts the fact that $(\tilde{L}, s) \neq (0,0)$. We conclude that $s > 0$. Writing $L = -\tilde{L}/s$, we obtain

$$\langle L, x\rangle - \mu \le \langle L - \ell, y\rangle$$

for all $(x,\mu) \in A$ and $y \in M$. Passing to the limit as $\mu \to \alpha\|x\|$, we get

$$\langle L, x\rangle \le \langle L - \ell, y\rangle + \alpha\|x\|$$

for all $x \in X$ and all $y \in M$. It follows that $L = \ell$ on M and $\|L\|_* \le \alpha$. □

Another consequence of the Hahn-Banach Separation Theorem 1.10 is the following:

Corollary 1.14. For each $x \in X$, there exists $\ell_x \in X^*$ such that $\|\ell_x\|_* = 1$ and $\langle \ell_x, x\rangle = \|x\|$.

Proof. Set $A = B(0, \|x\|)$ and $B = \{x\}$. By Theorem 1.10, there exists $L_x \in X \setminus \{0\}$ such that $\langle L_x, y\rangle \le \langle L_x, x\rangle$ for all $y \in A$. This implies $\langle L_x, x\rangle = \|L_x\|_* \|x\|$. The functional $\ell_x = L_x/\|L_x\|_*$ has the desired properties. □

The functional ℓ_x, given by Corollary 1.14, is a support functional for x. The normalized duality mapping is the set-valued function $\mathscr{F}: X \to \mathscr{P}(X^*)$ given by

$$\mathscr{F}(x) = \{\, x^* \in X^* : \|x^*\|_* = 1 \text{ and } \langle x^*, x\rangle = \|x\| \,\}.$$

The set $\mathscr{F}(x)$ is always convex and it need not be a singleton, as shown in the following example:

Example 1.15. Consider $X = \mathbf{R}^2$ with $\|(x_1,x_2)\| = |x_1| + |x_2|$ for $(x_1,x_2) \in X$. Then, $X^* = \mathbf{R}^2$ with $\langle (x_1^*,x_2^*),(x_1,x_2)\rangle = x_1^* x_1 + x_2^* x_2$ and $\|(x_1^*,x_2^*)\|_* = \max\{|x_1^*|,|x_2^*|\}$ for $(x_1^*,x_2^*) \in X^*$. Moreover, $\mathscr{F}(1,0) = \{\,(1,b) \in \mathbf{R}^2 : b \in [-1,1]\}$.

From Corollary 1.14 we deduce the following:

Corollary 1.16. For every $x \in X$, $\|x\| = \max\limits_{\|L\|_*=1} \langle L, x\rangle$.

Recall from Subsection 1.1 that the canonical embedding $\mathscr{J}$ of X into X^{**} is defined by $\mathscr{J}(x) = \mu_x$, where μ_x satisfies

$$\langle \mu_x, L\rangle_{X^{**},X^*} = \langle L, x\rangle_{X^*,X}.$$

Recall also that $\mathscr{J}$ is linear, and $\|\mathscr{J}(x)\|_{**} \leq \|x\|$ for all $x \in X$. We have the following:

Proposition 1.17. The canonical embedding $\mathscr{J}: X \to X^{**}$ is a linear isometry.

Proof. It remains to prove that $\|x\| \leq \|\mu_x\|_{**}$. To this end, simply notice that

$$\|\mu_x\|_{**} = \sup_{\|L\|_*=1} \mu_x(L) \geq \mu_x(\ell_x) = \langle \ell_x, x\rangle_{X^*,X} = \|x\|,$$

where ℓ_x is the functional given by Corollary 1.14. □

Another consequence of Corollary 1.16 concerns the adjoint of a bounded linear operator, defined in Subsection 1.1:

Corollary 1.18. Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be normed spaces. Given $L \in \mathscr{L}(X;Y)$, let $L^* \in \mathscr{L}(Y^*;X^*)$ be its adjoint. Then $\|L^*\|_{\mathscr{L}(Y^*;X^*)} = \|L\|_{\mathscr{L}(X;Y)}$.

Proof. We already proved that $\|L^*\|_{\mathscr{L}(Y^*;X^*)} \leq \|L\|_{\mathscr{L}(X;Y)}$. For the reverse inequality, use Corollary 1.16 to deduce that

$$\|L\|_{\mathscr{L}(X;Y)} = \sup_{\|x\|_X=1} \left[\max_{\|y^*\|_{Y^*}=1} \langle L^*(y^*), x\rangle_{X^*,X} \right] \leq \|L^*\|_{\mathscr{L}(Y^*;X^*)},$$

which gives the result. □

1.3 The weak topology

By definition, each element of X^* is continuous as a function from $(X, \|\cdot\|)$ to $(\mathbf{R}, |\cdot|)$. However, there are other topologies on X for which every element of X^* is continuous.[1] The coarsest of such topologies (the one with the fewest open sets) is called the weak topology and will be denoted by $\sigma(X)$, or simply σ, if there is no possible confusion.

Given $x_0 \in X$, $L \in X^*$ and $\varepsilon > 0$, every set of the form

$$\mathscr{V}_L^{\varepsilon}(x_0) = \{x \in X \ : \ \langle L, x - x_0 \rangle < \varepsilon\} = L^{-1}\Big((-\infty, L(x_0) + \varepsilon)\Big)$$

is open for the weak topology and contains x_0. Moreover, the collection of all such sets generates a base of neighborhoods of x_0 for the weak topology in the sense that if V_0 is a neighborhood of x_0, then there exist $L_1, \ldots, L_N \in X^*$ and $\varepsilon > 0$ such that

$$x_0 \in \bigcap_{k=1}^{N} \mathscr{V}_{L_k}^{\varepsilon}(x_0) \subset V_0.$$

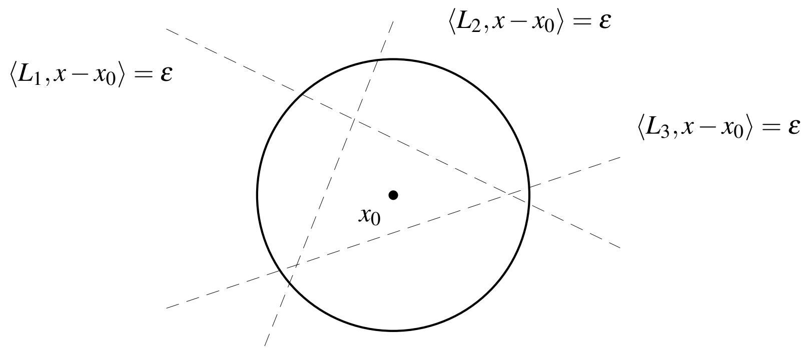
Recall that a Hausdorff space is a topological space in which every two distinct points admit disjoint neighborhoods.

Proposition 1.19. (X, σ) is a Hausdorff space.

Proof. Let $x_1 \neq x_2$. Part ii) of the Hahn-Banach Separation Theorem 1.10 shows the existence of $L \in X^*$ such that $\langle L, x_1 \rangle + \varepsilon \leq \langle L, x_2 \rangle$. Then $\mathscr{V}_L^{\frac{\varepsilon}{2}}(x_1)$ and $\mathscr{V}_{-L}^{\frac{\varepsilon}{2}}(x_2)$ are disjoint weakly open sets containing x_1 and x_2, respectively. □

It follows from the definition that every weakly open set is (strongly) open. If X is finite-dimensional, the weak topology coincides with the strong topology. Roughly speaking, the main idea is that, inside any open ball, one can find a weak neighborhood of its center that can be expressed as a finite intersection of open half-spaces.

[1] A trivial example is the discrete topology (for which every set is open), but it is not very useful for our purposes.



On the other hand, if X is infinite-dimensional, the weak topology is strictly coarser than the strong topology, as shown in the following example:

Example 1.20. If X is infinite-dimensional, the ball $B(0,1)$ is open for the strong topology but not for the weak topology. Roughly speaking, the reason is that no finite intersection of half-spaces can be bounded in all directions.

In other words, there are open sets which are not weakly open. Of course, the same is true for closed sets. However, closed convex sets are weakly closed.

Proposition 1.21. A convex subset of a normed space is closed for the strong topology if, and only if, it is closed for the weak topology.

Proof. By definition, every weakly closed set must be strongly closed. Conversely, let $C \subset X$ be convex and strongly closed. Given $x_0 \notin C$, we may apply part ii) of the Hahn-Banach Separation Theorem 1.10 with $A = \{x_0\}$ and $B = C$ to deduce the existence of $L \in X^* \setminus \{0\}$ and $\varepsilon > 0$ such that $\langle L, x_0 \rangle + \varepsilon \leq \langle L, y \rangle$ for each $y \in C$. The set $\mathscr{V}_L^{\varepsilon}(x_0)$ is a weak neighborhood of x_0 that does not intersect C. It follows that C^c is weakly open. □

Weakly convergent sequences

We say that a sequence (x_n) in X converges weakly to $\bar{x}$, and write $x_n \rightharpoonup \bar{x}$, as $n \to \infty$ if $\lim_{n\to\infty} \langle L, x_n - \bar{x} \rangle = 0$ for each $L \in X^*$. This is equivalent to saying that for each weakly open neighborhood $\mathscr{V}$ of $\bar{x}$ there is $N \in \mathbf{N}$ such that $x_n \in \mathscr{V}$ for all $n \geq N$. The point $\bar{x}$ is the weak limit of the sequence.

Since $|\langle L, x_n - \bar{x} \rangle| \leq \|L\|_* \, \|x_n - \bar{x}\|$, strongly convergent sequences are weakly convergent and the limits coincide.

Proposition 1.22. Let (x_n) converge weakly to $\bar{x}$ as $n \to \infty$. Then:

i) (x_n) is bounded.

ii) $\|\bar{x}\| \leq \liminf_{n\to\infty} \|x_n\|$.

iii) If (L_n) is a sequence in X^* that converges strongly to $\bar{L}$, then $\lim_{n\to\infty}\langle L_n, x_n\rangle = \langle \bar{L}, \bar{x}\rangle$.

Proof. For i), write $\mu_n = \mathscr{J}(x_n)$, where $\mathscr{J}$ is the canonical injection of X into X^{**}, which is a linear isometry, by Proposition 1.17. Since $\lim_{n\to\infty}\mu_n(L) = \langle L, \bar{x}\rangle$ for all $L \in X^*$, we have $\sup_{n\in\mathbf{N}} \mu_n(L) < +\infty$. The Banach-Steinhaus Uniform Boundedness Principle (Theorem 1.9) implies $\sup_{n\in\mathbf{N}} \|x_n\| = \sup_{n\in\mathbf{N}} \|\mu_n\|_{**} \leq C$ for some $C > 0$. For ii), use Corollary 1.14 to deduce that

$$\|x\| = \langle \ell_{\bar{x}}, x - x_n\rangle + \langle \ell_{\bar{x}}, x_n\rangle \leq \langle \ell_{\bar{x}}, x - x_n\rangle + \|x_n\|,$$

and let $n \to \infty$. Finally, by part i), we have

$$\begin{aligned} |\langle L_n, x_n\rangle - \langle \bar{L}, \bar{x}\rangle| &\leq |\langle L_n - \bar{L}, x_n\rangle| + |\langle \bar{L}, x_n - \bar{x}\rangle| \\ &\leq C\|L_n - \bar{L}\|_* + |\langle \bar{L}, x_n - \bar{x}\rangle|. \end{aligned}$$

As $n \to \infty$, we obtain iii). □

More on closed sets

Since we have defined two topologies on X, there exist (strongly) closed sets and weakly closed sets. It is possible and very useful to define some sequential notions as well. A set $C \subset X$ is sequentially closed if every convergent sequence of points in C has its limit in C. Analogously, C is weakly sequentially closed if every weakly convergent sequence in C has its weak limit in C. The relationship between the various notions of closedness is summarized in the following result:

Proposition 1.23. Consider the following statements concerning a nonempty set $C \subset X$:

i) C is weakly closed.

ii) C is weakly sequentially closed.

iii) C is sequentially closed.

iv) C is closed.

Then $i) \Rightarrow ii) \Rightarrow iii) \Leftrightarrow iv) \Leftarrow i)$. If C is convex, the four statements are equivalent.

Proof. It is easy to show that $i) \Rightarrow ii)$ and $iii) \Leftrightarrow iv)$. We also know that $i) \Rightarrow iv)$ and $ii) \Rightarrow iii)$ because the weak topology is coarser than the strong topology. Finally, if C is convex, Proposition 1.21 states precisely that $i) \Leftrightarrow iv)$, which closes the loop. □

The topological dual revisited: the weak* topology

The topological dual X^* of a normed space $(X, \|\cdot\|)$ is a Banach space with the norm $\|\cdot\|_*$. As in Subsection 1.3, we can define the weak topology $\sigma(X^*)$ in X^*. Recall that a base of neighborhoods for some point $L \in X^*$ is generated by the sets of the form

$$\{\ell \in X^* : \langle \mu, \ell - L\rangle_{X^{**},X^*} < \varepsilon\}, \tag{1.1}$$

with $\mu \in X^{**}$ and $\varepsilon > 0$.

Nevertheless, since X^* is, by definition, a space of functions, a third topology can be defined on X^* in a very natural way. It is the topology of pointwise convergence, which is usually referred to as the weak* topology in this context. We shall denote it by $\sigma^*(X^*)$, or simply σ^* if the space is clear from the context. For this topology, a base of neighborhoods for a point $L \in X^*$ is generated by the sets of the form

$$\{\ell \in X^* : \langle \ell - L, x\rangle_{X^*,X} < \varepsilon\}, \tag{1.2}$$

with $x \in X$ and $\varepsilon > 0$. Notice the similarity and difference with (1.1). Now, since every $x \in X$ determines an element $\mu_x \in X^{**}$ by the relation

$$\langle \mu_x, \ell\rangle_{X^{**},X^*} = \langle \ell, x\rangle_{X^*,X},$$

it is clear that every set that is open for the weak* topology must be open for the weak topology as well. In other words, $\sigma^* \subset \sigma$.

Reflexivity and weak compactness

In infinite-dimensional normed spaces, compact sets are rather scarce. For instance, in such spaces the closed balls are not compact (see [1, Theorem 6.5]). One of the most important properties of the weak* topology is that, according to the Banach-Alaoglu Theorem (see, for instance, [1, Theorem 3.16]), the closed unit ball in X^* is compact for the weak* topology. Recall, from Subsection 1.1, that X is reflexive if the canonical embedding $\mathscr{J}$ of X into X^{**} is surjective. This implies that the spaces (X, σ) and (X^{**}, σ^*) are homeomorphic, and so, the closed unit ball of X is compact for the weak topology. Further, we have the following:

Theorem 1.24. Let $(X, \|\cdot\|)$ be a Banach space. The following are equivalent:

i) X is reflexive.

ii) The closed unit ball $\bar{B}(0,1)$ is compact for the weak topology.

iii) Every bounded sequence in X has a weakly convergent subsequence.

We shall not include the proof here for the sake of brevity. The interested reader may consult [3, Chapters II and V] for full detail, or [1, Chapter 3] for abridged commentaries.

An important consequence is the following:

Corollary 1.25. Let (y_n) be a bounded sequence in a reflexive space. If every weakly convergent subsequence has the same weak limit $\hat{y}$, then (y_n) must converge weakly to $\hat{y}$ as $n \to \infty$.

Proof. Suppose (y_n) does not converge weakly to $\hat{y}$. Then, there exist a weakly open neighborhood $\mathscr{V}$ of $\hat{y}$, and a subsequence (y_{k_n}) of (y_n) such that $y_{k_n} \notin \mathscr{V}$ for all $n \in \mathbf{N}$. Since (y_{k_n}) is bounded, it has a subsequence $(y_{j_{k_n}})$ that converges weakly as $n \to \infty$ to some $\check{y}$ which cannot be in $\mathscr{V}$ and so $\check{y} \neq \hat{y}$. This contradicts the uniqueness of $\hat{y}$. □

1.4 Differential calculus

Consider a nonempty open set $A \subset X$ and function $f : A \to \mathbf{R}$. The directional derivative of f at $x \in A$ in the direction $d \in X$ is

$$f'(x;d) = \lim_{t \to 0^+} \frac{f(x+td) - f(x)}{t},$$

whenever this limit exists. The function f is differentiable (or simply Gâteux-differentiable) at x if $f'(x;d)$ exists for all $d \in X$ and the function $d \mapsto f'(x;d)$ is linear and continuous. In this situation, the Gâteaux derivative (or gradient) of f at x is $\nabla f(x) = f'(x;\cdot)$, which is an element of X^*. On the other hand, f is differentiable in the sense of Fréchet (or Fréchet-differentiable) at x if there exists $L \in X^*$ such that

$$\lim_{\|h\| \to 0} \frac{|f(x+h) - f(x) - \langle L, h \rangle|}{\|h\|} = 0.$$

If it is the case, the Fréchet derivative of f at x is $Df(x) = L$. An immediate consequence of these definitions is

Proposition 1.26. If f is Fréchet-differentiable at x, then it is continuous and Gâteaux-differentiable there, with $\nabla f(x) = Df(x)$.

As usual, f is differentiable (in the sense of Gâteaux or Fréchet) on A if it is so at every point of A.

Example 1.27. Let $B : X \times X \to \mathbf{R}$ be a bilinear function:

$$B(x+\alpha y, z) = B(x,z) + \alpha B(y,z) \quad \text{and} \quad B(x, y+\alpha z) = B(x,y) + \alpha B(x,z)$$

for all $x,y,z \in X$ and $\alpha \in \mathbf{R}$. Suppose also that B is continuous: $|B(x,y)| \leq \beta \|x\| \|y\|$ for some $\beta \geq 0$ and all $x, y \in X$. The function $f : X \to \mathbf{R}$, defined by $f(x) = B(x,x)$, is Fréchet-differentiable and $Df(x)h = B(x,h) + B(h,x)$. Of course, if B is symmetric: $B(x,y) = B(y,x)$ for all $x, y \in X$, then $Df(x)h = 2B(x,h)$.

Example 1.28. Let X be the space of continuously differentiable functions defined on $[0,T]$ with values in $\mathbf{R}^N$, equipped with the norm

$$\|x\|_X = \max_{t\in[0,T]} \|x(t)\| + \max_{t\in[0,T]} \|\dot{x}(t)\|.$$

Given a continuously differentiable function $\ell : \mathbf{R}\times\mathbf{R}^N\times\mathbf{R}^N \to \mathbf{R}$, define $J : X \to \mathbf{R}$ by

$$J[u] = \int_0^T \ell(t,x(t),\dot{x}(t))\,dt.$$

Then J is Fréchet-differentiable and

$$DJ(x)h = \int_0^T \Big[\nabla_2\ell(t,x(t),\dot{x}(t))\cdot h(t) + \nabla_3\ell(t,x(t),\dot{x}(t))\cdot \dot{h}(t)\Big]\,dt,$$

where we use ∇_i to denote the gradient with respect to the i-th set of variables.

It is to note that the Gâteaux-differentiability does not imply continuity. In particular, it is weaker than Fréchet-differentiability.

Example 1.29. Define $f : \mathbf{R}^2 \to \mathbf{R}$ by

$$f(x,y) = \begin{cases} \dfrac{2x^4y}{x^6+y^3} & \text{if} \quad (x,y) \neq (0,0) \\ 0 & \text{if} \quad (x,y) = (0,0). \end{cases}$$

A simple computation shows that $\nabla f(0,0) = (0,0)$. However, $\lim_{z\to 0} f(z,z^2) = 1 \neq f(0,0)$.

If the gradient of f is Lipschitz-continuous, we can obtain a more precise first-order estimation for the values of the function:

Lemma 1.30 (Descent Lemma). If $f : X \to \mathbf{R}$ is Gâteaux-differentiable and ∇f is Lipschitz-continuous with constant L, then

$$f(y) \leq f(x) + \langle \nabla f(x), y-x\rangle + \frac{L}{2}\|y-x\|^2$$

for each $x,y \in X$. In particular, f is continuous.

Proof. Write $h = y-x$ and define $g : [0,1] \to \mathbf{R}$ by $g(t) = f(x+th)$. Then $\dot{g}(t) = \langle \nabla f(x+th), h\rangle$for each $t \in (0,1)$, and so

$$\int_0^1 \langle \nabla f(x+th), h\rangle\,dt = \int_0^1 \dot{g}(t)\,dt = g(1)-g(0) = f(y)-f(x).$$

Therefore,

$$\begin{aligned} f(y)-f(x) &= \int_0^1 \langle \nabla f(x),h\rangle\,dt + \int_0^1 \langle \nabla f(x+th)-\nabla f(x),h\rangle\,dt \\ &\leq \langle \nabla f(x),h\rangle + \int_0^1 \|\nabla f(x+th)-\nabla f(x)\|\,\|h\|\,dt \\ &\leq \langle \nabla f(x),h\rangle + L\|h\|^2\int_0^1 t\,dt \\ &= \langle \nabla f(x),y-x\rangle + \frac{L}{2}\|y-x\|^2, \end{aligned}$$

as claimed. □

Second derivatives

If $f : A \to \mathbf{R}$ is Gâteaux-differentiable in A, a valid question is whether the function $\nabla f : A \to X^*$ is differentiable. As before, one can define a directional derivative

$$(\nabla f)'(x;d) = \lim_{t\to 0^+} \frac{\nabla f(x+td)-\nabla f(x)}{t},$$

whenever this limit exists (with respect to the strong topology of X^*). The function f is twice differentiable in the sense of Gâteaux (or simply twice Gâteaux-differentiable) in x if f is Gâteaux-differentiable in a neighborhood of x, $(\nabla f)'(x;d)$ exists for all $d \in X$, and the function $d \mapsto (\nabla f)'(x;d)$ is linear and continuous. In this situation, the second Gâteaux derivative (or Hessian) of f at x is $\nabla^2 f(x) = (\nabla f)'(x;\cdot)$, which is an element of $\mathscr{L}(X;X^*)$. Similarly, f is twice differentiable in the sense of Fréchet (or twice Fréchet-differentiable) at x if there exists $M \in \mathscr{L}(X;X^*)$ such that

$$\lim_{\|h\|\to 0} \frac{\|Df(x+h)-Df(x)-M(h)\|_*}{\|h\|} = 0.$$

The second Fréchet derivative of f at x is $D^2 f(x) = M$.

We have the following:

Proposition 1.31 (Second-order Taylor Approximation). Let A be an open subset of X and let $x \in A$. Assume $f : A \to \mathbf{R}$ is twice Gâteaux-differentiable in x. Then, for each $d \in X$, we have

$$\lim_{t\to 0} \frac{1}{t^2}\left| f(x+td)-f(x)-t\langle \nabla f(x),d\rangle - \frac{t^2}{2}\langle \nabla^2 f(x)d,d\rangle \right| = 0.$$

Proof. Define $\phi : I \subset R \to \mathbf{R}$ by $\phi(t) = f(x+td)$, where I is a sufficiently small open interval around 0 such that $\phi(t)$ exists for all $t \in I$. It is easy to see that $\phi'(t) = \langle \nabla f(x+td), d\rangle$ and $\phi''(0) = \langle \nabla^2 f(x)d, d\rangle$. The second-order Taylor expansion for ϕ in $\mathbf{R}$ yields

$$\lim_{t\to 0} \frac{1}{t^2}\left|\phi(t) - \phi(0) - t\phi'(0) - \frac{t^2}{2}\phi''(0)\right| = 0,$$

which gives the result. □

Of course, it is possible to define derivatives of higher order, and obtain the corresponding Taylor approximations.

Optimality conditions for differentiable optimization problems

The following is the keynote necessary condition for a point $\hat{x}$ to minimize a Gâteaux-differentiable function f over a convex set C.

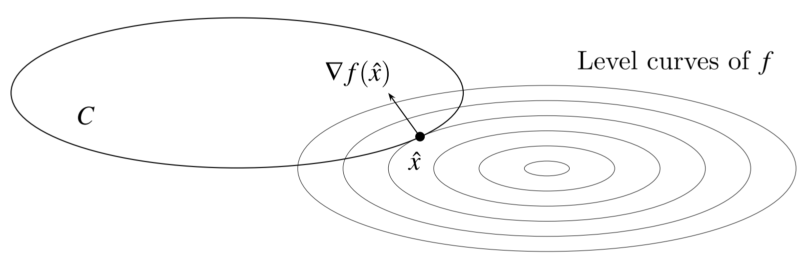
Theorem 1.32 (Fermat's Rule). Let C be a convex subset of a normed space $(X, \|\cdot\|)$ and let $f : X \to \mathbf{R} \cup \{+\infty\}$. If $f(\hat{x}) \leq f(y)$ for all $y \in C$ and if f is Gâteaux-differentiable at $\hat{x}$, then

$$\langle \nabla f(\hat{x}), y - \hat{x}\rangle \geq 0$$

for all $y \in C$. If moreover $\hat{x} \in \operatorname{int}(C)$, then $\nabla f(\hat{x}) = 0$.

Proof. Take $y \in C$. Since C is convex, for each $\lambda \in (0,1)$, the point $y_\lambda = \lambda y + (1-\lambda)\hat{x}$ belongs to C. The inequality $f(\hat{x}) \leq f(y_\lambda)$ is equivalent to $f(\hat{x} + \lambda(y-\hat{x})) - f(\hat{x}) \geq 0$. It suffices to divide by λ and let $\lambda \to 0$ to deduce that $f'(\hat{x}; y-\hat{x}) \geq 0$ for all $y \in C$. □

To fix the ideas, consider a differentiable function on $X = \mathbf{R}^2$. Theorem 1.32 states that the gradient of f at $\hat{x}$ must point inwards, with respect to C. In other words, f can only decrease by leaving the set C. This situation is depicted below:



As we shall see in the next chapter, the condition given by Fermat's Rule (Theorem 1.32) is not only necessary, but also sufficient, for convex functions. In the general case, one can provide second-order necessary and sufficient conditions for optimality. We state this result in the unconstrained case ($C = X$)

for simplicity.

An operator $M \in \mathscr{L}(X;X^*)$ is positive semidefinite if $\langle Md, d\rangle \geq 0$ for all $d \in X$; positive definite if $\langle Md, d\rangle > 0$ for all $d \neq 0$; and uniformly elliptic with constant $\alpha > 0$ if $\langle Md, d\rangle \geq \frac{\alpha}{2}\|d\|^2$ for all $d \in X$.

Theorem 1.33. Let A be an open subset of a normed space $(X, \|\cdot\|)$, let $\hat{x} \in A$, and let $f : A \to \mathbf{R}$.

i) If $f(\hat{x}) \leq f(y)$ for all y in a neighborhood of $\hat{x}$, and f is twice Gâteaux-differentiable at $\hat{x}$, then $\nabla f(\hat{x}) = 0$ and $\nabla^2 f(\hat{x})$ is positive semidefinite.

ii) If $\nabla f(\hat{x}) = 0$ and $\nabla^2 f(\hat{x})$ is uniformly elliptic, then $f(\hat{x}) < f(y)$ for all y in a neighborhood of $\hat{x}$.

Proof. For i), we already know by Theorem 1.32 that $\nabla f(\hat{x}) = 0$. Now, if $d \in X$ and $\varepsilon > 0$, by Proposition 1.31, there is $t_0 > 0$ such that

$$\frac{t^2}{2}\langle \nabla^2 f(\hat{x})d, d\rangle > f(\hat{x}+td) - f(\hat{x}) - \varepsilon t^2 \geq -\varepsilon t^2$$

for all $t \in [0, t_0]$. It follows that $\langle \nabla^2 f(\hat{x})d, d\rangle \geq 0$.

For ii), assume $\nabla^2 f(\hat{x})$ is uniformly elliptic with constant $\alpha > 0$ and take $d \in X$. Set $\varepsilon = \frac{\alpha}{4}\|d\|^2$. By Proposition 1.31, there is $t_1 > 0$ such that

$$f(\hat{x}+td) > f(\hat{x}) + \frac{t^2}{2}\langle \nabla^2 f(\hat{x})d, d\rangle - \varepsilon t^2 \geq f(\hat{x})$$

for all $t \in [0, t_1]$. □

Chapter 2

Hilbert spaces

Hilbert spaces are an important class of Banach spaces with rich geometric properties.

2.1 Basic concepts, properties and examples

An inner product in a real vector space H is a function $\langle \cdot, \cdot \rangle : H \times H \to \mathbf{R}$ such that:

i) $\langle x, x \rangle > 0$ for every $x \neq 0$;

ii) $\langle x, y \rangle = \langle y, x \rangle$ for each $x, y \in H$;

iii) $\langle \alpha x + y, z \rangle = \alpha \langle x, z \rangle + \langle y, z \rangle$ for each $\alpha \in \mathbf{R}$ and $x, y, z \in H$.

The function $\| \cdot \| : H \to \mathbf{R}$, defined by $\|x\| = \sqrt{\langle x, x \rangle}$, is a norm on H. Indeed, it is clear that $\|x\| > 0$ for every $x \neq 0$. Moreover, for each $\alpha \in \mathbf{R}$ and $x \in H$, we have $\|\alpha x\| = |\alpha| \|x\|$. It only remains to verify the triangle inequality. We have the following:

Proposition 2.1. For each $x, y \in H$ we have

i) The Cauchy-Schwarz inequality: $|\langle x, y \rangle| \leq \|x\| \, \|y\|$.

ii) The triangle inequality: $\|x + y\| \leq \|x\| + \|y\|$.

Proof. The Cauchy-Schwarz inequality is trivially satisfied if $y = 0$. If $y \neq 0$ and $\alpha > 0$, then

$$0 \leq \|x \pm \alpha y\|^2 = \langle x \pm \alpha y, x \pm \alpha y \rangle = \|x\|^2 \pm 2\alpha \langle x, y \rangle + \alpha^2 \|y\|^2.$$

Therefore,

$$|\langle x, y \rangle| \leq \frac{1}{2\alpha} \|x\|^2 + \frac{\alpha}{2} \|y\|^2$$

for each $\alpha > 0$. In particular, taking $\alpha = \|x\|/\|y\|$, we deduce i). Next, we use i) to deduce that

$$\|x+y\|^2 = \|x\|^2 + 2\langle x,y\rangle + \|y\|^2 \leq \|x\|^2 + 2\|x\|\,\|y\| + \|y\|^2 = (\|x\| + \|y\|)^2,$$

whence ii) holds. □

If $\|x\| = \sqrt{\langle x,x\rangle}$ for all $x \in X$, we say that the norm $\|\cdot\|$ is associated to the inner product $\langle\cdot,\cdot\rangle$. A Hilbert space is a Banach space, whose norm is associated to an inner product.

Example 2.2. The following are Hilbert spaces:

i) The Euclidean space $\mathbf{R}^N$ is a Hilbert space with the inner product given by the dot product: $\langle x,y\rangle = x\cdot y$.

ii) The space $\ell^2(\mathbf{N};\mathbf{R})$ of real sequences $\mathbf{x} = (x_n)$ such that

$$\sum_{n\in\mathbf{N}} x_n^2 < +\infty,$$

equipped with the inner product defined by $\langle \mathbf{x},\mathbf{y}\rangle = \sum_{n\in\mathbf{N}} x_n y_n$.

iii) Let Ω be a bounded open subset of $\mathbf{R}^N$. The space $L^2(\Omega;\mathbf{R}^M)$ of (classes of) measurable vector fields $\phi : \Omega \to \mathbf{R}^M$ such that

$$\int_\Omega \phi_m(\zeta)^2\, d\zeta < +\infty, \qquad \text{for} \quad m = 1,2,\ldots,M,$$

with the inner product $\langle \phi,\psi\rangle = \sum_{m=1}^M \int_\Omega \phi_m(\zeta)\,\psi_m(\zeta)\,d\zeta$.

By analogy with $\mathbf{R}^N$, it seems reasonable to define the angle γ between two nonzero vectors x and y by the relation

$$\cos(\gamma) = \frac{\langle x,y\rangle}{\|x\|\,\|y\|}, \qquad \gamma \in [0,\pi].$$

We shall say that x and y are orthogonal, and write $x \perp y$, if $\cos(\gamma) = 0$. In a similar fashion, we say x and y are parallel, and write $x \,\|\, y$, if $|\cos(\gamma)| = 1$. With this notation, we have

i) Pythagoras Theorem: $x \perp y$ if, and only if, $\|x+y\|^2 = \|x\|^2 + \|y\|^2$;

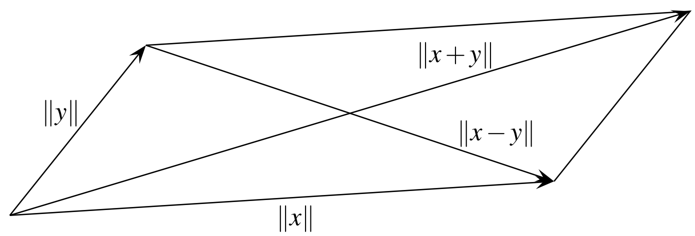
ii) The colinearity condition: $x \,\|\, y$ if, and only if, $x = \lambda y$ with $\lambda \in \mathbf{R}$.

Another important geometric property of the norm in a Hilbert space is the Parallelogram Identity, which states that

$$\|x+y\|^2 + \|x-y\|^2 = 2\left(\|x\|^2 + \|y\|^2\right) \tag{2.1}$$

for each $x,y \in H$. It shows the relationship between the length of the sides and the lengths of the diagonals in a parallelogram, and is easily proved by adding the following identities

$$\|x+y\|^2 = \|x\|^2 + \|y\|^2 + 2\langle x,y\rangle \quad \text{and} \quad \|x-y\|^2 = \|x\|^2 + \|y\|^2 - 2\langle x,y\rangle.$$



Example 2.3. The space $X = \mathscr{C}([0,1];\mathbf{R})$ with the norm $\|x\| = \max_{t\in[0,1]} |x(t)|$ is not a Hilbert space. Consider the functions $x, y \in X$, defined by $x(t) = 1$ and $y(t) = t$ for $t \in [0,1]$. We have $\|x\| = 1$, $\|y\| = 1$, $\|x+y\| = 2$ and $\|x-y\| = 1$. The parallelogram identity (2.1) does not hold.

2.2 Projection and orthogonality

An important property of Hilbert spaces is that given a nonempty, closed and convex subset K of H and a point $x \notin K$, there is a unique point in K which is the closest to x. More precisely, we have the following:

Proposition 2.4. Let K be a nonempty, closed and convex subset of H and let $x \in H$. Then, there exists a unique point $y^* \in K$ such that

$$\|x - y^*\| = \min_{y\in K} \|x - y\|. \tag{2.2}$$

Moreover, it is the only element of K such that

$$\langle x - y^*, y - y^* \rangle \leq 0 \quad \text{for all} \quad y \in K. \tag{2.3}$$

Proof. We shall prove Proposition 2.4 in three steps: first, we verify that (2.2) has a solution; next, we establish the equivalence between (2.2) and (2.3); and finally, we check that (2.3) cannot have more than one solution.

First, set $d = \inf_{y\in K} \|x - y\|$ and consider a sequence (y_n) in K such that $\lim_{n\to\infty} \|y_n - x\| = d$. We have

$$\begin{aligned} \|y_n - y_m\|^2 &= \|(y_n - x) + (x - y_m)\|^2 \\ &= 2\Big(\|y_n - x\|^2 + \|y_m - x\|^2 \Big) - \|(y_n + y_m) - 2x\|^2, \end{aligned}$$

by virtue of the parallelogram identity (2.1). Since K is convex, the midpoint between y_n and y_m belongs to K. Therefore,

$$\|(y_n + y_m) - 2x\|^2 = 4 \left\| \frac{y_n + y_m}{2} - x \right\|^2 \geq 4d^2,$$

according to the definition of d. We deduce that

$$\|y_n - y_m\|^2 \leq 2\left(\|y_n - x\|^2 + \|y_m - x\|^2 - 2d^2 \right).$$

Whence, (y_n) is a Cauchy sequence, and must converge to some y^*, which must lie in K by closedness. The continuity of the norm implies $d = \lim_{n\to\infty} \|y_n - x\| = \|y^* - x\|$.

Next, assume (2.2) holds and let $y \in K$. Since K is convex, for each $\lambda \in (0,1)$ the point $\lambda y + (1-\lambda)y^*$ also belongs to K. Therefore,

$$\begin{aligned} \|x - y^*\|^2 &\leq \|x - \lambda y - (1-\lambda)y^*\|^2 \\ &= \|x - y^*\|^2 + 2\lambda(1-\lambda)\langle x - y^*, y^* - y\rangle + \lambda^2\|y^* - y\|^2. \end{aligned}$$

This implies

$$\langle x - y^*, y - y^*\rangle \leq \frac{\lambda}{2(1-\lambda)}\|y^* - y\|^2.$$

Letting $\lambda \to 0$ we obtain (2.3). Conversely, if (2.3) holds, then

$$\|x - y\|^2 = \|x - y^*\|^2 + 2\langle x - y^*, y^* - y\rangle + \|y^* - y\|^2 \geq \|x - y^*\|^2$$

for each $y \in K$ and (2.2) holds.

Finally, if $y_1^*, y_2^* \in K$ satisfy (2.3), then

$$\langle x - y_1^*, y_2^* - y_1^*\rangle \leq 0 \quad \text{and} \quad \langle x - y_2^*, y_1^* - y_2^*\rangle \leq 0.$$

Adding the two inequalities we deduce that $y_1^* = y_2^*$. □

The point y^* given by Proposition 2.4 is the projection of x onto K and will be denoted by $P_K(x)$. The characterization of $P_K(x)$ given by (2.3) says that for each $x \notin K$, the set K lies in the closed half-space

$$S = \{\ y \in H : \langle x - P_K(x), y - P_K(x)\rangle \leq 0\ \}.$$

Corollary 2.5. Let K be a nonempty, closed and convex subset of H. Then $K = \cap_{x\notin K}\{y \in H : \langle x - P_K(x), y - P_K(x)\rangle\}$.

Conversely, the intersection of closed convex half-spaces is closed and convex.

For subspaces we recover the idea of orthogonal projection:

Proposition 2.6. Let M be a closed subspace of H. Then,

$$\langle x - P_M(x), u\rangle = 0$$

for each $x \in H$ and $u \in M$. In other words, $x - P_M(x) \perp M$.

Proof. Let $u \in M$ and write $v_\pm = P_M(x) \pm u$. Then $v_\pm \in M$ and so

$$\pm\langle x - P_M(x), u\rangle \leq 0.$$

It follows that $\langle x - P_M(x), u\rangle = 0$. □

Another property of the projection, bearing important topological consequences, is the following:

Proposition 2.7. Let K be a nonempty, closed and convex subset of H. The function $x \mapsto P_K(x)$ is nonexpansive.

Proof. Let $x_1, x_2 \in H$. Then $\langle x_1 - P_K(x_1), P_K(x_2) - P_K(x_1)\rangle \leq 0$ and $\langle x_2 - P_K(x_2), P_K(x_1) - P_K(x_2)\rangle \leq 0$. Summing these two inequalities we obtain

$$\|P_K(x_1) - P_K(x_2)\|^2 \leq \langle x_1 - x_2, P_K(x_1) - P_K(x_2)\rangle.$$

We conclude using the Cauchy-Schwarz inequality. □

2.3 Duality, reflexivity and weak convergence

The topological dual of a real Hilbert space can be easily characterized. Given $y \in H$, the function $L_y : H \to \mathbf{R}$, defined by $L_y(h) = \langle y, h\rangle$, is linear and continuous by the Cauchy-Schwarz inequality. Moreover, $\|L_y\|_* = \|y\|$. Conversely, we have the following:

Theorem 2.8 (Riesz Representation Theorem). Let $L : H \to \mathbf{R}$ be a continuous linear function on H. Then, there exists a unique $y_L \in H$ such that

$$L(h) = \langle y_L, h\rangle$$

for each $h \in H$. Moreover, the function $L \mapsto y_L$ is a linear isometry.

Proof. Let $M = \ker(L)$, which is a closed subspace of H because L is linear and continuous. If $M = H$, then $L(h) = 0$ for all $h \in H$ and it suffices to take $y_L = 0$. If $M \neq H$, let us pick any $x \notin M$ and define

$$\hat{x} = x - P_M(x).$$

Notice that $\hat{x} \neq 0$ and $\hat{x} \notin M$. Given any $h \in H$, set $u_h = L(\hat{x})h - L(h)\hat{x}$, so that $u_h \in M$. By Proposition 2.6, we have $\langle \hat{x}, u_h\rangle = 0$. In other words,

$$0 = \langle \hat{x}, u_h\rangle = \langle \hat{x}, L(\hat{x})h - L(h)\hat{x}\rangle = L(\hat{x})\langle \hat{x}, h\rangle - L(h)\|\hat{x}\|^2.$$

The vector

$$y_L = \frac{L(\hat{x})}{\|\hat{x}\|^2}\,\hat{x}$$

has the desired property and it is straightforward to verify that the function $L \mapsto y_L$ is a linear isometry. □

As a consequence, the inner product $\langle \cdot, \cdot \rangle_* : H^* \times H^*$ defined by

$$\langle L_1, L_2 \rangle_* = L_1(y_{L_2}) = \langle y_{L_1}, y_{L_2} \rangle$$

turns H^* into a Hilbert space, which is isometrically isomorphic to H. The norm associated with $\langle \cdot, \cdot \rangle_*$ is precisely $\| \cdot \|_*$.

Corollary 2.9. Hilbert spaces are reflexive.

Proof. Given $\mu \in H^{**}$, use the Riesz Representation Theorem 2.8 twice to obtain $L_\mu \in H^*$ such that $\mu(\ell) = \langle L_\mu, \ell \rangle_*$ for each $\ell \in H^*$, and then $y_{L_\mu} \in H$ such that $L_\mu(x) = \langle y_{L_\mu}, x \rangle$ for all $x \in H$. It follows that $\mu(\ell) = \langle L_\mu, \ell \rangle_* = \ell(y_{L_\mu})$ for each $\ell \in H^*$ by the definition of $\langle \cdot, \cdot \rangle_*$. □

Remark 2.10. Theorem 2.8 also implies that a sequence (x_n) on a Hilbert space H converges weakly to some $x \in H$ if, and only if, $\lim_{n\to\infty} \langle x_n - x, y \rangle = 0$ for all $y \in H$.

Strong and weak convergence can be related as follows:

Proposition 2.11. A sequence in (x_n) converges strongly to $\bar{x}$ if, and only if, it converges weakly to $\bar{x}$ and $\limsup_{n\to\infty} \|x_n\| \leq \|\bar{x}\|$.

Proof. The only if part is immediate. For the if part, notice that $0 \leq \limsup_{n\to\infty} \|x_n - \bar{x}\|^2 = \limsup_{n\to\infty} \left[\|x_n\|^2 + \|\bar{x}\|^2 - 2 \langle x_n, \bar{x} \rangle \right] \leq 0$. □

Remark 2.12. Another consequence of Theorem 2.8 is that we may interpret the gradient of a Gâteaux-differentiable function f as an element of H instead of H^* (see Subsection 1.4).

Bibliography

[1] Brézis H. Functional analysis, Sobolev spaces and partial differential equations. Springer; 2011.

[2] Cartan H. Differential calculus. London: Kershaw Publishing Co; 1971.

[3] Dunford N, Schwartz J. Linear operators. Part I. General theory. John Wiley & Sons, Inc; 1988.

[4] Opial Z. Weak Convergence of the sequence of successive approximations for nonexpansive mappings. Bull Amer Math Soc. 1967;73:591-7.

[5] Peypouquet, J. Convex optimization in normed spaces : Theory, methods and examples. Cham : Springer, 2015. 117 p. (SpringerBriefs in Optimization).

[6] Rudin W. Functional analysis. McGraw-Hill Inc; 1991.

[7] Schaefer HH, with Wolff MP. Topological vector spaces. Springer; 1999.