Convex analysis and optimisation

September 27, 2021

Contents

1 Topology and functional analysis

1.1 Point-set topology

We briefly recall some elementary notions from topology, even if we assume some familiarity with them.

Definition 1 (Topology)**.** A *topology* on a set X is a family of subsets τ of X, called *open sets*, which satisfy the following axioms:
(i) the empty set is open, i.e. $\emptyset \in \tau$;
(ii) the intersection of a finite family of open subsets is open: if $\omega_1, \ldots, \omega_k \in \tau$, then $\omega_1 \cap \ldots \cap \omega_k \in \tau$.
(iii) any union of open subsets if open: if $(\omega_i)_{i\in I}$ belongs to τ, then so does $\cup_i \omega_i$.
The space (X, τ) is then called a topological space. The complement of an open set is called a *closed set*.

Example 1 (Metric topology). Let (X, d) be a metric space, i.e. $d : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ satisfies the assumptions
(i) $d(x, y) \geq 0$ with equality if and only if $x = y$;
(ii) $d(x, y) = d(y, x)$ for all $x, y \in X$;
(iii) $d(x, z) \leq d(x, y) + d(y, z)$ for all $x, y, z \in X$.
We denote $\mathrm{B}(x, r) = \{y \in \mathcal{X} \mid d(x, y) \leq r\}$ the closed ball. We call a set ω open if for any $x \in \Omega$, there exists $r > 0$ such that $B(x, r) \subseteq \omega$. Then, the family τ_d of such open sets satisfies the axioms of a topology.

Definition 2. A topological space $(\mathcal{X}, \tau)$ is called
(i) *separated* or *Hausdorff* if for any distinct points $x, y \in \mathcal{X}$ have neighborhoods O_x, O_y such that $O_x \cap O_y = \emptyset$;
(ii) *metrizable* if the taupology τ is induced by a metric;
(iii) *separable* if $\mathcal{X}$ contains a countable dense subset.

Metric spaces are automatically separated. For a metric space, and hence for a metrizable space, all topological notions may be recovered through sequences, e.g. a set $C \subseteq \mathcal{X}$ is closed iff C contains the limit of every converging sequence in C.

Example 2. Examples of separable spaces include

- the space $(\mathcal{C}^0(X), \|\cdot\|_\infty)$ of continuous functions over a compact subset X of $\mathbb{R}^d$ endowed with the norm of uniform convergence. This is a consequence of the Stone-Weierstrass theorem, which guarantees that continuous functions over a compact subset of $\mathbb{R}^d$ can be uniformly approximated by polynomials ;
- the spaces $(\ell^p, \|\cdot\|_p)$ of p-summable sequences and the spaces $(\mathrm{L}^p(\Omega), \|\cdot\|_p)$ of p-integrable functions over a bounded domain of $\mathbb{R}^d$, for $p \in [1, +\infty)$. Note that ℓ^∞ and $\mathrm{L}^\infty(\Omega)$ are *not* separable.

Definition 3 (Neighborhood)**.** Let (X, τ) be a topological space. A neighborhood of $x \in \mathcal{X}$ is any set that contains an open set, which itself contains x. A *neighborhood basis* at x is a family $\mathcal{B}_x$ of subsets of $\mathcal{X}$ sucht that (i) every set $B \in \mathcal{B}_x$ is a neighborhood of x and (ii) for any neighborhood N of x, there exists $B \in \mathcal{B}_x$ such that $B \subseteq N$.

Example 3. In a metric space, the sets $\mathcal{B}_x = \{B(x, \delta) \mid \delta > 0\}$ and $\mathcal{B}'_x = \{B(x, \frac{1}{n}) \mid n \in \mathbb{N}\}$ form two neighborhood bases at x.

Definition 4 (Continuity at a point)**.** A function $f : (\mathcal{X}, \sigma) \to (\mathcal{Y}, \tau)$ between topological spaces is *continuous* at x if for any neighborhood B of $f(x)$, the inverse image $f^{-1}(B)$ is a neighborhood of x. Equivalently, f is continuous at $x \in \mathcal{X}$ iff there are neighborhood bases $\mathcal{B}_x$ at x and $\mathcal{B}_{f(x)}$ at $f(x)$ such that

$$\forall C \in \mathcal{B}_{f(x)}, \exists B \in \mathcal{B}_x \text{ s.t. } f^{-1}(C) \subseteq B.$$

In the case of metric spaces, one recovers the definition of continuity using ε and δ.

The notion of continuity depends on the choice of the topologies σ and τ. This is important to keep in mind, because we will sometimes consider two topologies over the same space.

Definition 5 (Continuity)**.** The function $f : (\mathcal{X}, \sigma) \to (\mathcal{Y}, \tau)$ is *continuous* on $\mathcal{X}$ if it is continuous at every $x \in \mathcal{X}$. Equivalently, f is continuous on $\mathcal{X}$ if the inverse image of any open set under f is open, i.e. $\forall \omega \in \tau$, $f^{-1}(\omega) \in \sigma$.

1.2 Topological vector spaces

Most of the statements of this course hold when $\mathcal{X}$ is a normed space i.e. a vector space endowed with a norm $\|\|$ which endows $\mathcal{X}$ with a topology. The closed ball of radius around a point x is then denoted $\mathrm{B}(x, r) = \{y \in \mathcal{X} \mid \|x - y\| \leq r\}$. In particular, they can be applied to Hilbert spaces, Banach spaces, and even $\mathbb{R}^d$.

However, it is often the case that our statement hold in the more setting of (locally convex) topological vector spaces, which is particularly adequate when one wants to consider weak/weak$*$ topologies, which are important in some applications.

Definition 6 (Topological vector space)**.** A *topological vector space* is a vector space $\mathcal{X}$ endowed with a topology τ which makes the addition of vectors $(x, y) \in \mathcal{X}^2 \mapsto$

$x+y \in \mathcal{X}$ and the multiplication by a scalar $(\lambda, x) \in \mathbb{R} \times \mathcal{X} \mapsto \lambda x \in \mathcal{X}$ continuous. A topological vector space is called *locally convex* if in addition the origin admits a neighborhood basis made of convex sets.

In order to define the continuity of the maps $\mathcal{X}^2 \to \mathcal{X}$ and $\mathbb{R} \times \mathcal{X} \to \mathcal{X}$, we need to define a topology on these product spaces, called the product topology.

Definition 7 (Product topology)**.** Let (X_1, τ_1) and (X_2, τ_2) be topological spaces. A set $O \subseteq X_1 \times X_2$ is open for the *product topology* $\tau_1 \otimes \tau_2$ if for any $(x_1, x_2) \in O$, there exists neighborhoods N_1 and N_2 of x_1 and x_2 such that $N_1 \times N_2 \subseteq O$.

One can check that if O is an open subset of a topological vector space, then for all $x \in X$, $O + x = \{x + y \mid y \in O\}$ is also open. In particular, if $\mathcal{B}$ is a basis of neighborhood of the origin, then $\{B + x \mid B \in \mathcal{B}\}$ is a basis of neighborhood at x.

Definition 8 (Dual space)**.** The *topological dual* of $\mathcal{X}$ is the space of all continuous linear functions on $\mathcal{X}$, and is denoted $\mathcal{X}^*$.

The elements of $\mathcal{X}^*$ are denoted with a star, e.g. $x^* \in \mathcal{X}^*$, and we denote the pairing between $\mathcal{X}^*$ and X using the notation for scalar product, i.e. $\langle x^*|x\rangle := x^*(x)$, to underline the similarity with the case of Hilbert spaces. When $(\mathcal{X}, \|\cdot\|)$ is a normed vector space, the topological dual $\mathcal{X}^*$ may be endowed with the dual norm

$$\|x^*\| = \sup_{x \in \mathcal{X}} \langle x^*|x\rangle.$$

One could construct a “strong” topology on $\mathcal{X}^*$ even when $\mathcal{X}$ is not a normed space, but this will not be necessary in this course.

We recall the following characterization of continuous linear forms on a topological vector space, which generalizes a similar criterion for normed spaces. Its proof is left as an exercise.

Lemma 1. *(Continuous linear forms) Let $x^* : \mathcal{X} \to \mathbb{R}$ be a linear form on a topological vector space (X, σ). Then, the following are equivalent*

(i) $\langle x^*|\cdot\rangle$ *is continuous,*
(ii) $\langle x^*|\cdot\rangle$ *is continuous near the origin (i.e. for all $\varepsilon > 0$, there exists $O \in \sigma$ containing 0 such that $\forall x \in O, \langle x^*|x\rangle$ belongs to $(-\varepsilon, \varepsilon)$),*
(iii) $\langle x^*|\cdot\rangle$ *is bounded (from above or from below) in a neighborhood of the origin,*
(iv) $\langle x^*|\cdot\rangle$ *is bounded (from above or from below) on a non-empty open subset of $\mathcal{X}$.*

Lemma 2. *(Closed halfspaces) Let x^* be a (not necessarily continuous) linear form over a topological vector space $\mathcal{X}$. Then, the following are equivalent:*

(i) x^* *is continuous,*
(ii) the set $H = \{x \in \mathcal{X} \mid \langle x^*|x\rangle \leq \alpha\}$ *is closed,*
(iii) the set $H' = \{x \in \mathcal{X} \mid \langle x^*|x\rangle = \alpha\}$ *is closed,*

Proof. If x^* is continuous, then the halfspace H (resp. hyperplane H') is closed as the inverse image of $(-\infty, \alpha]$ (resp. $\{\alpha\}$) under x^*. Conversely, if H is closed (resp. H' is closed), then $\langle x^*|\cdot\rangle > \alpha$ on the non-empty open set $\mathcal{X} \setminus H$ (resp. $\mathcal{X} \setminus H'$), so that x^* is continuous by the previous lemma. □

1.3 Weak and weak∗ topologies

In many cases, the topology associated with the norm is too fine, in the sense that it has too few compact sets, making it difficult to to prove existence to optimization problems. This is why we will try to define coarser topologies, i.e. topologies with less open sets and thus more compact sets. One standard way to do this is to start with a family of maps, and to construct the *coarsest* (i.e. smallest with respect to inclusion) topology that makes all these maps continuous.

We will explain the construction of the coarsest topology making a family of linear forms $\mathcal{Y}$ over a vector space $\mathcal{X}$ continuous.

Lemma 3. *Let $\mathcal{X}$ be a vector space and let $\mathcal{Y}$ be a vector space of linear forms over $\mathcal{X}$. Assume that σ is a topology over $\mathcal{X}$ and that every linear form in $\mathcal{Y}$ is continuous with respect to $\mathcal{X}$. Then, all sets of the following form are open with respect to σ:*

$$B_{x,y_1,\dots,y_N,\varepsilon} := \{z \in \mathcal{X} \mid \forall i \in \{1,\dots,N\}, |\langle y_i|z-x\rangle| < \varepsilon\},$$

with $x \in \mathcal{X}$, $y_1,\dots,y_N \in \mathcal{Y}$ and $\varepsilon > 0$.

Proof. Let $y_1,\dots,y_N \in \mathcal{Y}$. Then,

$$B_{x,y_1,\dots,y_N,\varepsilon} = \bigcap_{i=1}^{N} y_i^{-1}((\langle y_i|x\rangle - \varepsilon, \langle y_i|x\rangle + \varepsilon)).$$

Thus, $B_{x,y_1,\dots,y_N,\varepsilon}$ is open with respect to σ as a finite intersection of inverse images of the open sets $(\langle y_i|x\rangle - \varepsilon, \langle y_i|x\rangle + \varepsilon)$ under the maps $y_1,\dots,y_N$ which we have assumed continuous with respect to σ. □

This lemma suggests the following definition.

Definition 9 (Topology generated by linear forms)**.** Let $\mathcal{X}$ be a vector space and let $\mathcal{Y}$ be a vector space of linear forms over $\mathcal{X}$. A set O is open with respect to the topology generated by $\mathcal{Y}$, denoted $\sigma(\mathcal{X},\mathcal{Y})$ if for any $x \in O$, there exists $y_1,\dots,y_N \in \mathcal{Y}$ and $\varepsilon > 0$ such that $B_{x,y_1,\dots,y_N,\varepsilon} \subseteq O$.

The sets $B_{x,y_1,\dots,y_N,\varepsilon}$ are open with respect to $\sigma(\mathcal{X},\mathcal{Y})$ (exercise!) and a neighborhood basis of the point $x \in \mathcal{X}$ for $\sigma(\mathcal{X},\mathcal{Y})$ is given by

$$\mathcal{B}_x = \{B_{x,y_1,\dots,y_N,\varepsilon} \mid N \in \mathbb{N}, y_1,\dots,y_N \in \mathcal{Y}, \text{ and } \varepsilon > 0\}.$$

In some sense, the sets $B_{x,y_1,\dots,y_N,\varepsilon}$ play the role of balls for the weak topology. However, this topology has a rather surprising feature: if $\mathcal{X}$ is infinite dimensional, then any set with non-empty $\sigma(X,Y)$-interior must contain a set of the above form and must therefore be unbounded. In particular, in infinite dimension the unit ball $\mathrm{B}(0,1)$ has empty $\sigma(X,Y)$-interior.

Proposition 4. *Let $\mathcal{X}$ be a vector space and let $\mathcal{Y}$ be a vector space of linear forms over $\mathcal{X}$. Then,* *(i)* $(\mathcal{X}, \sigma(\mathcal{X}, \mathcal{Y}))$ *is a locally convex topological vector space;*

(ii) A linear form $y : \mathcal{X} \to \mathbb{R}$ *is continuous with respect to the topology* $\sigma(\mathcal{X}, \mathcal{Y})$ *if and only if it belongs to* $\mathcal{Y}$*;*

Corollary 5. *Given a linear form* $y : \mathcal{X} \to \mathbb{R}$ *and* $\alpha \in \mathbb{R}$*, the hyperplane* $y^{-1}(\alpha)$ *(resp. the halfspace* $y^{-1}((-\infty, \alpha])$*) is closed w.r.t.* $\sigma(\mathcal{X}, \mathcal{Y})$ *if and only if* $y \in \mathcal{Y}$*.*

Remark 1 (Coarsest topology). Lemma 3 shows that if σ is a topology on $\mathcal{X}$ such that all the linear forms in $\mathcal{Y}$ are continuous, then it contains all the sets of the form $B_{x,y_1,\ldots,y_N,\varepsilon}$ with $y_1, \ldots, y_N \in \mathcal{Y}$. Thus, any such topology σ must contain $\sigma(\mathcal{X}, \mathcal{Y})$, implying that $\sigma(\mathcal{X}, \mathcal{Y})$ is the coarsest topology on $\mathcal{Y}$ making all the linear forms in $\mathcal{Y}$ are continuous. (This also means that it has the most compact subsets, or the most converging sequences...)

Remark 2 (Convergence of sequences). One can show that a sequence $(x_n)_{n\geq 1}$ of points of $\mathcal{X}$ converges to $x \in \mathcal{X}$ with respect to the topology of $\sigma(\mathcal{X}, \mathcal{Y})$ — meaning that for all neighborhood O of x one has $x_n \in O$ for n large enough — if and only if

$$\forall y \in \mathcal{Y}, \lim_{n\to+\infty} \langle y|x_n\rangle = \langle y|x\rangle.$$

Proof. (i) The family $\sigma(\mathcal{X}, \mathcal{Y})$ contains $\emptyset$, is (easily) stable under arbitrary unions, and is finite under finite intersections because

$$B_{x,y_1,\ldots,y_N,y'_1,\ldots,y'_M,\min(\varepsilon,\varepsilon')} \subseteq B_{x,y_1,\ldots,y_N,\varepsilon} \cap B_{x,y'_1,\ldots,y'_M,\varepsilon'}.$$

This proves that $\sigma(\mathcal{X}, \mathcal{Y})$ is a topology. We will prove that the map $S : \mathcal{X} \times \mathcal{X} \to \mathcal{X}$ defined by $S(x, y) = x + y$ is continuous with respect to $\sigma(\mathcal{X}, \mathcal{Y})$. To see this, take $x, y \in \mathcal{X}$ and consider a neighborhood N_z of $z = S(x, y) = x + y$. We want to prove that there exists neighborhoods N_x, N_y of x and y so that $S(N_x, N_y) \subseteq N_z$. By definition, the neighborhood N_z contains a set of the form $B_z := B_{z,y_1,\ldots,y_N,\varepsilon}$. Define $B_x := B_{x,y_1,\ldots,y_N,\frac{1}{2}\varepsilon}$ and $B_y = B_{y,y_1,\ldots,y_N,\frac{1}{2}\varepsilon}$. For any $x' \in B_x$ and $y' \in B_y$, let $z' = S(x', y') = x' + y')$. We have

$$|\langle y_i|z' - z\rangle| = \frac{1}{2} |\langle y_i|x' + y' - (x + y)\rangle| \leq \frac{1}{2}(|\langle y_i|x' - x\rangle| + |\langle y_i|x' - x\rangle|) < \varepsilon,$$

thus ensuring that $z' \in B_z$. In other words, $S(B_x, B_y) \subseteq B_z$, and B_x and B_y are neighborhoods of x and y respectively: this shows the continuity of S at (x, y). We would prove similarly that the map $P : (\lambda, x) \in \mathbb{R} \times \mathcal{X} \mapsto \lambda x \in \mathcal{X}$, thus proving that $(\mathcal{X}, \sigma(\mathcal{X}, \mathcal{Y}))$ is a topological vector space. Finally, the space is locally convex because all the sets $B_{x,y_1,\ldots,y_N,\varepsilon}$ are convex.

(ii) Let y be a linear form over $\mathcal{X}$ and assume that it is continuous with respect to the topology $\sigma(\mathcal{X}, \mathcal{Y})$. Thus, there exists a neighborhood N of the origin on which the linear form y is bounded by 1. By construction of the topology, there exists $y_1, \ldots, y_N$ and $\varepsilon > 0$ such that $B_{0,y_1,\ldots,y_N,\varepsilon} \subseteq N$. Thus, we have

$$(\forall i \in \{1, \ldots, N\}, \quad \langle y_i|x\rangle < \varepsilon) \Longrightarrow |\langle y|x\rangle| < 1,$$

which implies in particular that if $\mathrm{Ker}(y_1) \cap \ldots \mathrm{Ker}(y_n) \subseteq \mathrm{Ker}(y)$. By a the linear algebra lemma below, this implies that $y \in \mathrm{vect}(y_1, \ldots, y_N) \subseteq \mathcal{Y}$. □

Lemma 6. *Let $\mathcal{X}$ be a vector space, and let $y_1, \ldots, y_N, y$ be linear forms on $\mathcal{X}$ such that* $\mathrm{Ker}(y_1) \cap \ldots \mathrm{Ker}(y_n) \subseteq \mathrm{Ker}(y)$. *Then* $y \in \mathrm{vect}(y_1, \ldots, y_N)$.

Proof. Let $F : \mathcal{X} \to \mathbb{R}^{N+1}$ be defined by $F(x) = (y(x), y_1(x), \ldots, y_N(x))$. The assumption implies that the subspace $L = F(\mathcal{X})$ and the point $z = (1, 0, \ldots, 0)$ are disjoint. Denoting $p \in L$ the orthogonal projection of z on L and $p - z = (\lambda, \lambda_1, \ldots, \lambda_N) \neq 0$, the characterization of the orthogonal projection gives

$$\forall x \in \mathcal{X}, \langle F(x) | \lambda \rangle = 0,$$

i.e. $\lambda y + \sum_i \lambda_i y_i = 0$. Moreover, $\langle p | z - p \rangle > 0$, giving $\lambda \neq 0$. □

Definition 10 (Weak-topology)**.** Let $\mathcal{X}$ be a topological vector space, and let $\mathcal{X}^*$ be the space of continuous linear forms over $\mathcal{X}$. The topology over $\mathcal{X}$ generated by the linear forms $\mathcal{X}^*$, namely $\sigma(\mathcal{X}, \mathcal{X}^*)$, is called the weak topology.

Remark 3 (Relation to the strong topology). We will call the original topology on $\mathcal{X}$ the *strong* topology to distinguish it from the weak topology $\sigma(\mathcal{X}, \mathcal{X}^*)$. The sets $B_{x, x_1^*, \ldots, x_N^*, \varepsilon}$, which form a basis of the weak topology, are open with respect to the original topology. This directly implies that weakly open sets are strongly open, and similarly that weakly closed sets are strongly closed. Maybe unintuitively, this also implies that if a function $f : \mathcal{X} \to \mathbb{R}$ is weakly continuous, then it is strongly continuous (Proof: assume that f is weakly continuous. Then, for any open subset O of $\mathbb{R}$, the set $f^{-1}(O)$ is weakly open, therefore strongly open. Thus, f is strongly continuous). The converse implications are false in general, but we will show that they are true when convexity hypotheses are added (see e.g Corollary 21).

Definition 11 (Weak$*$-topology)**.** Let $\mathcal{X}$ be a topological vector space, and consider the canonical injection

$$i : \mathcal{X} \to \mathcal{X}^{**}, x \mapsto (x^* \mapsto \langle x^* | x \rangle).$$

Thanks to this injection, $\mathcal{X}$ can be identified with the set $i(\mathcal{X})$ of linear forms over $\mathcal{X}^*$. The topology over $\mathcal{X}^*$ generated by the linear forms $i(\mathcal{X})$ is called the weak topology and often denoted $\sigma(\mathcal{X}^*, \mathcal{X}) := \sigma(\mathcal{X}, i(\mathcal{X}))$.

We refer to [Bre10]) for a more thorough treatment of the weak/weak* topology.

Remark 4 (Weak$*$ and pointwise convergence). A good exercise is to show that a sequence (x_n^*) of elements of $\mathcal{X}^*$ converges to x^* with respect to $\sigma(\mathcal{X}^*, \mathcal{X})$ (for short we will say that (x_n^*) *weak-$*$ converges to* x^*) if and only if

$$\forall x \in \mathcal{X}, \quad \lim_{n \to +\infty} \langle x_n^* | x \rangle = \langle x^* | x \rangle.$$

In plain words, (x_n^*) weak-$*$ converges to x^* if only if (x_n^*) converges *pointwise* to x^*, when these linear forms are seen as functions on $\mathcal{X}$.

Remark 5 (Separation). The space $(\mathcal{X}^*, \sigma(\mathcal{X}^*, \mathcal{X}))$ is separated, i.e. Hausdorff. Indeed, let $x^*, y^* \in \mathcal{X}^*$ be two distinct linear forms. Then, there exists $x \in \mathcal{X}$ such that $\langle x^*|x\rangle \neq \langle y^*|x\rangle$, for instance $\langle x^*|x\rangle < \langle y^*|x\rangle$. Thus, denoting $r = \langle \frac{1}{2}(x^* + y^*)|x\rangle$, the two weak$*$ open sets $O_- = \{\langle \cdot|x\rangle < r\}$ and $O_+ = \{\langle \cdot|x\rangle > r\}$ separate the points x^* and y^*: they are disjoint and $x^* \in O_-$ and $y^* \in O_+$.

Example 4 (Measures). Let $K \subseteq \mathbb{R}^d$ be compact and let $\mathcal{X} = \mathcal{C}^0(K)$ be the space of continuous functions over K. What are examples of linear functionals over $\mathcal{X}$? Let μ be a finite (Radon) measure over $\mathcal{X}$: then, the functional $f \in \mathcal{C}^0(K) \mapsto \int f \mathrm{d}\mu$ is linear. Conversely, Riesz-Markov's theorem asserts that any linear functional over $\mathcal{X}$ is induced by a finite (Radon) measure. Thus, in this course, we will define the space of Radon measures as

$$\mathcal{M}(K) := \mathcal{C}^0(K)^*.$$

An example of Radon measure is the Dirac mass $\delta_x \in \mathcal{M}(K)$, defined by

$$\forall \phi \in \mathcal{C}^0(K), \langle \delta_x|\phi\rangle = \delta_x(\phi) := \phi(x).$$

The dual norm on $\mathcal{M}(K)$ is called the *total variation* of the measure μ. It is defined by $\|\nu\|_{TV} = \sup\{\langle \nu|\phi\rangle \mid \|\phi\|_\infty \leq 1\}$. Note that if x, y are distinct points in K, it is possible to construct a function $\phi \in \mathcal{C}^0(X)$ so that $\phi(x) = 1$ and $\phi(y) = -1$ and $\|\phi\|_\infty \leq 1$. Thus,

$$\|\delta_x - \delta_y\|_{TV} \geq \int \phi \mathrm{d}(\delta_x - \delta_y) = \phi(x) - \phi(y) = 2.$$

In other words, even if (x_n) converges to $x \in K$, the Dirac mass δ_{x_n} does *not* converge to δ_x with respect to the total variation $\|\cdot\|_{TV}$. On the other hand,

$$\forall \phi \in \mathcal{C}^0(X), \langle \phi|\delta_{x_n}\rangle = \phi(x_n) \longrightarrow_{n\to+\infty} \langle \phi|\delta_x\rangle,$$

so that (δ_{x_n}) weak$*$-converges to δ_x, i.e. with respect to the topology $\sigma(\mathcal{M}(X), \mathcal{C}^0(X))$. This is an illustration of the fact that the weak-$*$ topology $\sigma(\mathcal{M}(X), \mathcal{C}^0(X))$ has more converging sequences, and thus more compact sets as well.

Remark 6 (Continuous linear forms over $(\mathcal{X}^*, \sigma(\mathcal{X}^*, \mathcal{X}))$). By Proposition 4, we know that a linear form over $\mathcal{X}^*$ is continuous if and only if is induced by an element to $\mathcal{X}$, i.e. $(\mathcal{X}^*, \sigma(\mathcal{X}^*, \mathcal{X}))$ can be identified with $\mathcal{X}$. This implies, for instance, that the dual of the space of measures endowed with its weak$*$ topology is $(\mathcal{M}(K), \sigma(\mathcal{M}(K), \mathcal{C}^0(K)))^* \simeq \mathcal{C}^0(K)$, while the dual $(\mathcal{M}(K), \|\cdot\|_*)^*$ is a much larger space and much less understood (there is a book devoted to this topic [Kap11]).

Theorem 7 (Metrizability of the dual ball). *Let $\mathcal{X}$ be a separable normed vector space, and let $(x_n)_{n\geq 1}$ be a dense sequence in $\mathcal{X}$. Then the weak$*$ topology on a bounded subset S of $\mathcal{X}^*$ is induced by the distance*

$$\mathrm{d}(x^*, y^*) = \sum_{n\geq 1} 2^{-n} \left| \langle x^* - y^* | \frac{x_n}{\|x_n\|}\rangle \right|.$$

Note that the open sets of the weak$*$ topology on S are of the form $O' = O \cap S$, where O is an weak$*$ open sets of $\mathcal{X}^*$.

Proof. The series defining $\mathrm{d}(x^*, y^*)$ is converging. Indeed, by definition of $\|\cdot\|_*$,

$$\left|\langle x^* - y^* | \frac{x_n}{\|x_n\|}\rangle\right| \leq \|x^* - y^*\|_* .$$

The symmetry of d, and the triangle inequality are clear. If $\mathrm{d}(x^*, y^*) = 0$, then for all x_n we have $\langle x^*|x\rangle = \langle y^*|x\rangle$. Thus, the continuous linear forms x^* and y^* agree on the dense subset $(x_n)_{n\geq 1}$, and therefore they agree on $\mathcal{X}$, i.e. $x^* = y^*$.

Let S be a bounded set in $\mathcal{X}^*$; we assume that $S \subseteq \mathrm{B}(0^*, R)$ for some $R > 0$. Let $O' = O \cap S$ be an open set with respect to the distance d. Our goal is to prove that $O \cap S$ is also weak$*$ open. Let $x^* \in O \cap S$: by openness of O with respect to the distance d, there exists $\varepsilon > 0$ such that $\mathrm{B}(x^*, \varepsilon) \subseteq O$. For any $y^*, z^* \in S$, we have $\|y^* - z^*\| \leq 2R$, so that there exists some $N \in \mathbb{N}$ such that

$$\forall y^*, z^* \in S, \quad \sum_{n\geq N} 2^{-n} \left|\langle x^* - y^* | \frac{x_n}{\|x_n\|}\rangle\right| \leq \frac{\varepsilon}{2}.$$

If we consider some point $z^* \in S \cap B_{x^*, y_1, \ldots, y_N, \frac{\varepsilon}{4}}$, then

$$\forall i \in \{1, \ldots, N\} \, |\langle x^* - z^* | y_i\rangle| < \frac{\varepsilon}{4}.$$

This implies that such a point belongs to the ball $\mathrm{B}(x^*, \varepsilon)$ because

$$d(x^*, z^*) = \sum_{1\leq n<N} 2^{-n} \left|\langle x^* - y^* | \frac{x_n}{\|x_n\|}\rangle\right| + \sum_{n\geq N} 2^{-n} \left|\langle x^* - y^* | \frac{x_n}{\|x_n\|}\rangle\right| \leq \varepsilon.$$

To summarize, the set $S \cap B_{x^*, y_1, \ldots, y_N, \frac{\varepsilon}{4}}$ is included in $S \cap \mathrm{B}(x^*, \varepsilon)$, itself included in O'. By definition, this implies that O' is weak$*$ open in S.

To conclude that the weak$*$ and the metric topology agree on S, we need to prove that conversely, any weak$*$ open set of S is open with respect to the distance d. The proof of this fact is similar, and is left as an exercise. □

Theorem 8 (Banach-Alaoglu)**.** *Let $\mathcal{X}$ be a separable normed space. Then, any bounded sequence in $\mathcal{X}^*$ admit a weak*–converging subsequence.*

Proof. Let A be a dense countable subset of $\mathcal{X}$. Let (x_n^*) be a bounded sequence in $\mathcal{X}^*$. Then, for all $x \in A$, the sequence $(\langle x_n^*|x\rangle)_{n\in\mathbb{N}}$ is bounded and admits a converging subsequence. By a diagonal argument, and taking subsequences where necessary, we can assume that for all $x \in A$ there exists $f_a \in \mathbb{R}$ such that

$$\lim_{n\to+\infty} \langle x_n^*|x\rangle = \lim_{n\to+\infty} x_n^*(x) = f_a.$$

Thus, the sequence of functions $x_n^*|_A$ converges pointwise to the function $f : A \to \mathbb{R}$ defined by $f(x) = f_a$. By boundedness of the sequence (x_n^*), there exists $R > 0$ such that $\|x_n^*\|_* \leq R$ for all $n \in \mathbb{N}$. Then,

$$\forall x, y \in \mathcal{X}, |\langle x_n^*|x\rangle - \langle x_n^*|y\rangle| \leq R \|x - y\|.$$

Passing to the limit as $n \to +\infty$, this proves that the function f is R-Lipschitz on the dense set $A \subseteq \mathcal{X}$ and can therefore be extended uniquely into a R-Lipschitz function function $\hat{f} : \mathcal{X} \to \mathbb{R}$. We will now show that x_n^* converges pointwise to $\hat{f}$ on $\mathcal{X}$. Let x be an arbitrary point $\mathcal{X}$, and let $\varepsilon > 0$. By density of A, there exists $y \in A$ such that $\|x - y\| \leq \frac{1}{3R}\varepsilon$; by convergence of $x_n^*(y)$ to $\hat{f}(y)$, there exists $N \in \mathbb{N}$ such that for $n \geq N$, $\left|\hat{f}(y) - x_n^*(y)\right| \leq \frac{1}{3}\varepsilon$. Then,

$$\left|\hat{f}(x) - x_n^*(x)\right| \leq \left|\hat{f}(x) - \hat{f}(y)\right| + \left|\hat{f}(y) - x_n^*(y)\right| + |x_n^*(y) - x_n^*(x)| \leq \varepsilon.$$

This shows that $\lim_{n\to+\infty} x_n^*(x) = \hat{f}(x)$, and that weak$*$ converges to $\hat{f}$. Finally, we note that a pointwise limit of linear functions is linear to conclude that $\hat{f} \in \mathcal{X}^*$. □

Theorem 9 (Banach-Alaoglu)**.** *Let $\mathcal{X}$ be a separable normed space. Then, the unit ball of $\mathcal{X}^*$ is weakly*–closed.*

Proof. The unit ball B of $\mathcal{X}^*$ is bounded, so that its weak$*$ topology is metrizable by Theorem 7. Thus, B is weak$*$ compact if and only if it is sequentially weak$*$ compact, i.e. iff any sequence of elements of B admits a weak$*$ converging subsequence. One concludes by invoking Theorem 9. □

Example 5 (Probability measures). Let K be a compact subset of $\mathbb{R}^d$, and let $\mathcal{X} = \mathcal{C}^0(K)$. Then, the dual space $\mathcal{X}^*$ is the space of Radon measures over K, i.e. $\mathcal{X}^* = \mathcal{M}(K)$, and its unit ball is

$$B = \{\mu \in \mathcal{M}(K) \mid \|\mu\|_{TV} \leq 1\},$$

and by Banach-Alaoglu's theorem, B is weak$*$ compact. Now let $\mathcal{M}^+(K)$ be the set of positive measures on K, i.e.

$$\mathcal{M}^+(K) = \{\mu \in \mathcal{M}(K) \mid \forall \phi \in \mathcal{C}^0(K), \phi \geq 0 \Longrightarrow \langle \mu|\phi\rangle \geq 0\}.$$

Then, $\mathcal{M}^+(K)$ is a weak$*$ closed set (which is also convex). Thus, the set of probability measures

$$\mathcal{P}(X) = \{\mu \in \mathcal{M}^+(K) \mid \langle \mu|\mathbf{1}_K\rangle = 1\} = \mathcal{M}^+(K) \cap B,$$

is a weak$*$-compact set.

We finish by an application of the Banach-Alaoglu theorem to reflexive spaces, a class of spaces that include Hilbet spaces.

Definition 12 (Reflexive space)**.** A normed space $\mathcal{X}$ is *reflexive* if and only if $\mathcal{X}^{**}$, the dual space to $\mathcal{X}^*$, can be identified with $\mathcal{X}$, meaning that the *canonical injection*

$$\begin{aligned} i : \mathcal{X} &\to \mathcal{X}^{**}, \\ x &\mapsto (\phi_x : x^* \in \mathcal{X}^* \mapsto \langle x^* | x \rangle) \end{aligned}$$

is a bijection. In other words, a space $\mathcal{X}$ is reflexive if every continuous linear form on its dual $\mathcal{X}^*$, i.e. $\phi : \mathcal{X}^* \to \mathbb{R}$, can be written under the form $\phi(x^*) = \langle x^* | x \rangle$ for some $x \in \mathcal{X}$.

Note that the canonical injection is an isometry, i.e. $\|i(x)\| = \|x\|$. In this course, it will be sufficient to know a few facts about reflexive spaces, such as

- finite-dimensional spaces are reflexive;
- Hilbert spaces are reflexive;
- L^p spaces and Sobolev spaces $W^{1,p}$ are separable and reflexive for $1 < p < +\infty$.

Corollary 10. *In a separable and reflexive space, the unit ball is weakly compact.*

Proof. We only prove sequential weak compactness.Let K be a bounded closed convex set and let $(x_n)_{n\in\mathbb{N}}$ be a sequence of elements of $\mathcal{X}$ with $\|x_n\| \leq 1$. We regard these points as elements of $\mathcal{X}^{**}$ through the canonical injection, i.e. $x_n^{**} := i(x_n^*)$. The sequence $(x_n^{**})_{n\in\mathbb{N}}$ is also bounded (since $\|i(x)\| = \|x\|$), ensuring by Banach-Alaoglu's theorem the existence of a weak*–converging subsequence. Relabeling if necessary, we may therefore assume that there exists $x^{**} \in \mathcal{X}^{**}$ s.t.

$$\forall x^* \in \mathcal{X}^*, \langle x_n^{**} | x^* \rangle \xrightarrow{n\to+\infty} \langle x^{**} | x^* \rangle.$$

Since the canonical injection is surjective (by reflexivity of $\mathcal{X}$), there exists some x such that $x^{**} = i(x)$. The previous convergence then yields

$$\forall x^* \in \mathcal{X}^*, \langle x^* | x_n \rangle \xrightarrow{n\to+\infty} \langle x^* | x \rangle,$$

i.e. $(x_n)_{n\in\mathbb{N}}$ weakly converges to x. Since K is closed and convex, it is sequentially weakly closed, from which we deduce that $x \in K$. $\square$

2 Convex sets

2.1 Convex sets

Definition 13 (Convex set)**.** A subset K of $\mathcal{X}$ is called *convex* if and only if for any $x, y \in K$ and every $t \in [0,1]$, $(1-t)x + ty$ belongs to K.

Example 6. Basic examples of convex sets include

- open and closed balls in normed vector spaces,
- hyperplanes, i.e. sets of the form $\{x \in \mathcal{X} \mid \langle x^* | x \rangle = \alpha\}$ for some linear function x^* on $\mathcal{X}$ and some $\alpha \in \mathbb{R}$,
- affine subspaces, i.e. intersection of hyperplanes,
- halfspaces, i.e. sets of the form $\{x \in \mathcal{X} \mid \langle x^* | x \rangle \leq \alpha\}$ for some linear function x^* on $\mathcal{X}$ and some $\alpha \in \mathbb{R}$,
- intersection of halfspaces, which are called *polyhedra* if one take a *finite* intersection of halfspaces,
- the space of symmetric positive definite matrices,
- sublevel sets $\{x \in \mathcal{X} \mid f(x) \geq \alpha\}$ or epigraphs $\{(x,t) \mid t \geq f(x)\}$ where $f : \mathcal{X} \to \mathbb{R}$ is a convex function and $\alpha \in \mathbb{R}$,
- the *Minkowski sum* (or difference) of two convex sets $A, B \subseteq \mathcal{X}$, i.e.

$$A \oplus B := \{x + y \mid x \in A, y \in B\}, \quad A \ominus B := \{x + y \mid x \in A, y \in B\}.$$

Definition 14 (Convex hull)**.** The *convex hull* of a set $A \subseteq \mathcal{X}$, denoted $\operatorname{conv}(A)$, is the smallest convex set containing A.

Proposition 11. *The following properties hold*

(i) If $(K_\alpha)_{\alpha \in A}$ is family of convex sets, $\cap_{\alpha \in A} K_\alpha$ is convex
(ii) If $K \subseteq \mathcal{X}$ is convex and $L : \mathcal{X} \to \mathcal{Y}$ is linear, then $L(K)$ is convex.
(iii) If $K \subseteq \mathcal{X}$ is convex and $L : \mathcal{Y} \to \mathcal{X}$ is linear, then $L^{-1}(K)$ is convex.
(iv) If $K \subseteq \mathcal{X}, L \subseteq \mathcal{Y}$ are convex, then $K \times L$ is convex.
(v) If $K, L \subseteq \mathcal{X}$ are convex, then $K \oplus L$ is convex.
(vi) $\operatorname{conv}(A)$ is the intersection of all convex sets containing A.
(vii) $\operatorname{conv}(A) = \left\{ \sum_{i=1}^k \alpha_i x_i \mid k \geq 1, x_i \in A, \ \alpha_i \geq 0, \sum_{i=1}^k \alpha_i = 1 \right\}$.

Proof. Exercise. □

2.2 Linear separation

The most important tool of this course is the representation of closed convex sets as intersection of closed half-spaces, which is the basis of subdifferential calculus, of the duality theory of Fenchel and Rockafellar, etc. This representation relies mainly on the Hahn-Banach theorem.

Definition 15 (Sublinearity)**.** A function $p : \mathcal{X} \to \mathbb{R}$ is *sublinear* if it is

(i) 1-*homogeneous*: for any $x \in \mathcal{X}$ and $\lambda \geq 0$, $p(\lambda x) = \lambda p(x)$;
(ii) *subadditive*: for any $x, y \in \mathcal{X}$, $p(x + y) \leq p(x) + p(y)$.

Theorem 12 (Hahn-Banach)**.** *Let $\mathcal{X}$ be a topological vector space and let p be a continuous sublinear function on $\mathcal{X}$. If E is a linear subspace of $\mathcal{X}$ and if f is a linear function on E satisfying $f \leq p$, then f can be extended into a continuous linear form $\hat{f} \in \mathcal{X}^*$ which also satisfies $\hat{f} \leq p$.*

We refer to Brézis [Bre10] for a proof of the Hahn-Banach theorem in the case of general vector spaces (which also does not require that p is continuous), while we prove it only in the case where $\mathcal{X}$ is separable, which makes the proof more constructive.

Proof. Let $(v_i)_{i\geq 1}$ be a countable dense subset of $\mathcal{X}$, and define by induction $E_0 = E$ and $E_i = \mathrm{span}(E_{i-1}, v_i)$.

Step 1: Extension of f to E_i: We first show how to construct a linear extension f_i of f on E_i satisfying $f_i \leq p$, assuming that f_{i-1} is already constructed. If v_i belongs to E_{i-1}, one has $E_i = E_{i-1}$ and there is nothing to do. If not, for any $\alpha \in \mathbb{R}$ we can define a linear function $f_{i,\alpha}$ on E_i by setting

$$\forall x \in E_{i-1}, \forall t \in \mathbb{R}, f_{i,\alpha}(x + tv_i) = f_{i-1}(x) + t\alpha.$$

We now need to choose α so that $f_{i,\alpha} \leq p$ or equivalently so that

$$\begin{aligned}
&\forall x \in E_{i-1}, \forall t \in \mathbb{R}, f_{i-1}(x) + t\alpha \leq p(x + te_i) \\
&\iff \forall x \in E_{i-1}, \forall t \in \mathbb{R}\setminus\{0\}, f_{i-1}\left(\frac{x}{|t|}\right) + \frac{t}{|t|}\alpha \leq p\left(\frac{x}{|t|} + \frac{t}{|t|}e_i\right) \\
&\iff \forall x \in E_{i-1}, f_{i-1}(x) \pm \alpha \leq p(x \pm e_i).
\end{aligned}$$

We used the homogeneity of p to go from the first to the second line. This is again equivalent to α satisfying

$$\sup_{x \in E_{i-1}} f_{i-1}(x) - p(x - e_i) \leq \alpha \leq \inf_{y \in E_{i-1}} p(y + e_i) - f_{i-1}(y).$$

Thus, such an α exists if the supremum is less than the infimum, i.e. if

$$\begin{aligned}
&\forall (x, y) \in E_{i-1}, f_{i-1}(x) - p(x - e_i) \leq p(y + e_i) - f_{i-1}(y) \\
&\iff \forall (x, y) \in E_{i-1}, f_{i-1}(x + y) \leq p(y + e_i) + p(x - e_i) \\
&\Longleftarrow \forall (x, y) \in E_{i-1}, f_{i-1}(x + y) \leq p(x + y),
\end{aligned}$$

where the last implication is deduced from the subadditivity of p. Since $f_{i-1} \leq p$, we can backtrack the chain of implications to deduce the existence of $\alpha \in \mathbb{R}$ such that $f_{i,\alpha} \leq p$. In practice, one can set

$$\alpha = \inf_{y \in E_{i-1}} p(y + e_i) - f(y) \geq -p(-e_i) > -\infty.$$

Step 2: Extension of f to $\mathcal{X}$ The previous constructions allows to construct a linear function $\tilde{f} : \hat{E} \to \mathbb{R}$ on the linear subspace $\hat{E} = \cup_i E_i$ satisfying $\tilde{f} \leq p$ and

$\tilde{f}\Big|_E = f$. The subspace $\hat{E}$ contains the vectors $(v_i)_{i\geq 1}$ and is therefore dense in $\mathcal{X}$. To extend $\tilde{f}$ to $\mathcal{X}$, we will use the continuity of p. We first notice that by assumption

$$-p(y-x) \leq \tilde{f}(x-y) \leq p(x-y),$$

implying that

$$\left|\tilde{f}(x-y)\right| \leq \max |p(x-y)|, |p(y-x)|.$$

One can first use this inequality to show that if (x_n) is a sequence of points of $\hat{E}$ converging to some $x \in \mathcal{X}$, then $(\tilde{f}(x_n))$ is a Cauchy sequence in $\mathbb{R}$, thus converging to some value, denoted $\hat{f}(x)$. We can use again the same inequality to prove that if two sequences $(x_n), (y_n)$ of elements of $\hat{E}$ converge to the same point $x \in \mathcal{X}$, then $\lim_{n\to+\infty} \tilde{f}(x_n) = \lim_{n\to+\infty} \tilde{f}(y_n)$, proving that $\hat{f}$ is well-defined. Finally, $\hat{f}$ also satisfies $\left|\hat{f}(x-y)\right| \leq p(x-y)$ for all $x, y \in \mathcal{X}$, proving its continuity. □

Definition 16 (Separation)**.** Two sets $K, L \subseteq X$ are *continuously linearly separated*, or separated for short, if there exists a non-trivial continuous linear form $x^* \in X^* \backslash \{0\}$ such that $\sup_K \langle x^* | \cdot \rangle \leq \inf_L \langle x^* | \cdot \rangle$. If the inequality is strict, we will say that the sets are *strongly separated*.

Corollary 13. *Let $\mathcal{X}$ be a topological vector space, K an open convex subset of $\mathcal{X}$ and $z \in \mathcal{X} \setminus K$. Then K and $\{z\}$ are separated.*

Definition 17 (Gauge)**.** The *gauge* of a convex $K \subseteq \mathcal{X}$ convex is the function j_K defined on $\mathcal{X}$ by $j_K(x) = \inf\{r > 0 \mid x \in rK\}$.

Lemma 14. *Let $K \subseteq \mathcal{X}$ be a convex neighborhood of the origin. Then the gauge j_K is sublinear and continuous, and $K \subseteq \{j_K \leq 1\}$.*

Proof. See Exercise 2 □

Proof of Corollary 13. Translating if necessary, we assume that K contains the origin in its interior, and we consider its gauge j_K. Let $E = \mathbb{R}z$, and define $f(tz) := t$ on E. Since $z \notin K$, $j_K(z) \geq 1 = f(z)$, from which we deduce by positive homogeneity that $f(tz) \leq j_K(tz)$ for $t \geq 0$. For $t \leq 0$ we have $f(tz) = t \leq 0 \leq j_K(tz)$. Thus, $f \leq j_K$ on E. By Hahn-Banach's theorem, we get the existence of continuous linear extension $\hat{f}$ of f such that $\hat{f} \leq j_K$ on $\mathcal{X}$. Moreover, for all $x \in K$, $\hat{f}(x) \leq j_K(x) \leq 1 = \hat{f}(z)$ as desired. (We note that the last equality implies that f is non-trivial.) □

Corollary 15. *Let $\mathcal{X}$ be a topological vector space and let K and L be two disjoint convex subsets of $\mathcal{X}$, such that L is open. Then, K and L are separated.*

Proof. We consider the Minkowski difference between K and L,

$$M := K \ominus L = \{x - y \mid x \in K, y \in L\} = \bigcup_{y\in L} (K - y),$$

which is open (as a union of open sets) and convex as already noted. Since K and L are disjoint, M does not contain $z := 0$ in its interior. By the previous corollary, we get the existence of $x^* \in \mathcal{X}^*$ such that

$$\forall x \in M = K \ominus L, \langle x^*|x\rangle \leq \langle x^*|z\rangle = 0$$

This directly implies that K and L are separated by x^*. □

Corollary 16. *Let $\mathcal{X}$ be a topological vector space and let K and L be two convex subsets of $\mathcal{X}$. Assume that there exists an open convex neighborhood V of 0 such that $(K \oplus V) \cap L = \emptyset$. Then, K and L are strongly separated.*

Proof. We first note that the set $K \oplus V$ is convex (at the sum of convex sets) and open (as in the proof of the previous corollary). Thus, by the previous corollary, we can separate $K \oplus V$ from L, i.e. there exists $x^* \in \mathcal{X}^* \setminus \{0\}$ so that

$$\sup_{x \in K \oplus V} \langle x^*|x\rangle \leq \inf_{y \in L} \langle x^*|y\rangle.$$

In addition, one easily sees that

$$\sup_{x \in K \oplus V} \langle x^*|x\rangle = \sup_{x \in K} \langle x^*|x\rangle + \sup_{y \in V} \langle x^*|y\rangle.$$

Since $x^* \neq 0$, there exists $v \in \mathcal{X}$ such that $\langle x^*|v\rangle \neq 0$, and replacing v by $-v$ if necessary, we can assume that $\langle x^*|v\rangle > 0$. Since the set V is open and since 0 belongs to V, for $t > 0$ sufficiently small, tv must also belongs to V. Thus, $\sup_{y \in V} \langle x^*|y\rangle > t\langle x^*|v\rangle > 0$: this shows the strong separation. □

Corollary 17. *Let $\mathcal{X}$ be a locally convex topological vector space and let K and L be two convex subsets of $\mathcal{X}$, and assume that K is compact and L is closed. Then, K and L are strongly separated.*

Lemma 18. *In a topological vector space, if X and Y are respectively compact and closed, then $X \oplus Y$ is closed.*

Proof of Lemma 18, Normed spaces. Let z_n be a converging sequence of points in $Z = X \oplus Y$, so that $z_n = x_n + y_n$ with $x_n \in X$ and $y_n \in Y$. We denote z the limit of $(z_n)_{n \geq 1}$. By compactness of X, taking a subsequence if necessary, we can assume that the sequence $(x_n)_{n \geq 1}$ converges to some $x \in X$. The relation $z_n = x_n + y_n$ then implies that the sequence $(y_n)_{n \geq 1}$ converges to some point $y = z - x$, and by closedness of L, the point y belongs to L. In conclusion, $z = x + y$ belongs to $Z = X \oplus Y \oplus (-L)$ and M is closed. □

Proof of Lemma 18, Topological vector spaces. Let $z \notin X \oplus Y$. Then, for every $x \in X$ we have $z - x \notin Y$, implying by closedness of Y that there exists an open set O_x so that $z - x \in O_x$ and $O_x \cap Y = \emptyset$. By continuity of the substraction, there exists opens set B_x containing x and U_x containing z such that $(U_x \ominus B_x) \subseteq O_x$, so that $(U_x \ominus B_x) \cap Y = \emptyset$. The sets $(B_x)_{x \in X}$ cover the compact set X, so that by compactness, one may extract a finite family $x_1, \ldots, x_N \in X$ s.t. $\bigcup_i B_{x_i}$ contains X. Then $U = \cap_i U_{x_i}$ is an open set containing z and such that

$$(U \ominus X) \cap Y \subseteq \left(U \ominus \bigcup_i B_{x_i}\right) \cap Y \subseteq \bigcup_i (U_{x_i} \ominus B_{x_i}) \cap Y = \emptyset,$$

thus implying that $U \cap (X \oplus Y) = \emptyset$. This shows that $(X \oplus Y)$ is closed. □

Proof of Corollary 17. We first notice that thanks to the previous lemma, $K \ominus L = K \oplus (-L)$ is closed, and by hypothesis it does not contain the origin. Therefore, there exists an open set V such that $V \cap (K \ominus L) = \emptyset$. Moreover, since we assumed that $\mathcal{X}$ is locally convex, we may assume that V is convex. Thus, we can apply Corollary 16 to conclude that K and L are strongly separated.

□

Remark 7 (Separation of points). The previous corollary implies that if $\mathcal{X}$ is a separated (Hausdorff) locally convex topological vector space, then for every distinct points $x, y \in \mathcal{X}$, there exists $x^* \in \mathcal{X}^*$ such that $\langle x^* | x \rangle > 0 > \langle x^* | y \rangle$.

2.3 Closed convex sets

Definition 18. A *closed half-space* is a set of the form $H = \{x \in \mathcal{X} \mid \langle x^* | x \rangle \leq \alpha\}$ where $x^* \in \mathcal{X}^*$ and $\alpha \in \mathbb{R}$.

Proposition 19. *Let K be a closed convex set of a locally convex topological vector space $\mathcal{X}$. Then, K is the intersection of all the closed half-spaces containing K.*

Proof. Denote L the intersection of all the closed half-spaces containing K. The inclusion $K \subseteq L$ is obvious, so let us prove that $(\mathcal{X} \setminus K) \subseteq (\mathcal{X} \setminus L)$. To do that, we consider some point $x \in \mathcal{X} \setminus K$. By the strong separation theorem (Corollary 17) applied to the closed convex set K and the compact convex set $\{x\}$, there exists $x^* \in \mathcal{X}^*$ and $\varepsilon > 0$ so that

$$\forall z \in K, \quad \langle x^* | z \rangle + \varepsilon \leq \langle x^* | x \rangle.$$

Thus, the closed half-space $H = \{z \in E \mid \langle x^* | z \rangle + \frac{\varepsilon}{2} \leq \langle x^* | x \rangle\}$ contains K, so that by definition $L \subseteq H$. Since x does not belong to H, this shows that $x \notin L$. □

Definition 19 (Closed convex hull)**.** The *closed convex hull* of a subset A of a normed space is the smallest closed convex set containing A, or equivalently the intersection of all the closed convex sets containing A. We denote it $\overline{\text{conv}} A$.

Corollary 20. *Let $\mathcal{X}$ be a locally convex topological vector space. Then, for all subsete $A \subseteq \mathcal{X}$, the closed convex hull of A equals the intersection of all the half-spaces containing A.*

Example 7. Let $A = \{(x, y) \in \mathbb{R}^2 \mid y \geq \frac{1}{1+x^2}\}$. Then, $\overline{\text{conv}} A = \{(x, y) \in \mathbb{R}^2 \mid y \geq 0\}$.

Corollary 21. *Let $\mathcal{X}$ be a locally convex topological vector space, and let $K \subseteq \mathcal{X}$ be convex. Then, K is weakly closed if and only if it is closed.*

Proof. We already know that (Remark 3) that if K is weakly closed, then K is closed. Conversely, assume that K is closed and convex. By the Proposition 19, K is a (possibly uncountable) intersection of closed halfspaces. Since every closed halfspace is weakly closed, we deduce that K is also weakly closed. □

2.4 Support function and normal cone

Definition 20 (Support function, support hyperplane)**.** The *support function* of a non-empty set $A \subseteq \mathcal{X}$ is the fonction σ_A on $\mathcal{X}^*$ defined by

$$\sigma_A : x^* \in \mathcal{X}^* \mapsto \sup_{x \in A} \langle x^* | x \rangle \in \mathbb{R} \cup \{+\infty\} \tag{1}$$

A (closed) hyperplane $H = \{\langle x^* | \cdot \rangle = \alpha\}$ is a *support hyperplane* to A at at a point $x \in A$, or *supports A at x* if $\langle x^* | x \rangle = \alpha$ and if $A \subseteq \{\langle x^* | \cdot \rangle \leq \alpha\}$. The linear form x^* is then called an *(exterior) normal* to A at x.

Remark 8. Note that H supports A at x if and only if $\sigma_A(x^*) = \langle x^* | x \rangle$, explaining why σ_A is called the support function of A.

Dual norm

Sup over extremal points and relation to Krein-Milman

The support function of a singleton ($A = \{x\}$) is linear, while in general the support function of a set is sublinear (as in Definition 15). Support functions were introduced by Minkowski in the finite-dimensional setting. A nice feature of the embedding $K \mapsto \sigma_K$ is that it preserves much of the structure of the set of closed convex subsets of $\mathcal{X}$. This gives a geometric interpretation of some rules of subdifferential calculus. We refer to Hörmander [Hör55] for a short summary of the properties of support functions of closed convex sets.

Our first propositions shows that the support function of a closed convex set can be used to reconstruct it.

Proposition 22. *Let $\mathcal{X}$ be a locally convex topological vector space and let $K \subseteq \mathcal{X}$ be closed and convex. Then,*

$$K = \bigcap_{x^* \in \mathcal{X}^*} \{x \in \mathcal{X} \mid \langle x^* | x \rangle \leq \sigma_K(x^*)\}. \tag{2}$$

Proof. Denote L the intersection of halfspaces $H_{x^*} = \{\langle x^* | \cdot \rangle \leq \sigma_K(x^*)\}$. Since every halfspace in the intersection contains K, so does L. We now prove the converse inclusion $\mathcal{X} \setminus K \subseteq \mathcal{X} \setminus L$. Given a point $x \in \mathcal{X} \setminus K$, Proposition 19 shows that there exists a closed convex hyperplane $H = \{\langle x^* | \cdot \rangle \leq \alpha\}$ containing K but not containing the point x. Then, using

$$\alpha \geq \sup_{x \in K} \langle x^* | x \rangle = \sigma_K(x^*),$$

we see that $H_{x^*} \subseteq H$. This implies that $x \notin H_{x^*}$, so that x does not belong to L either. □

Proposition 23. *Let $\mathcal{X}$ be a locally convex topological vector space.*

(i) Let $K, L \subseteq \mathcal{X}$ be closed and convex. Then, $K \subseteq L$ iff $\sigma_K \leq \sigma_L$. (In particular $\sigma_K = \sigma_L$ if and only if $K = L$.)

(ii) Let $K_1, \ldots, K_\ell$ be closed and convex and let $\lambda_1, \ldots \lambda_\ell > 0$. Then

$$\sigma_{\bigoplus_i \lambda_i K_i} = \sum_i \lambda_i \sigma_{K_i};$$

(iii) Let $K_1, \ldots, K_\ell$ be closed and convex. Then, $\sigma_{\bigcup_i K_i} = \max_i \sigma_{K_i}$. Then,

$$\sigma_{\bigcup_i K_i} = \max_i \sigma_{K_i}.$$

Remark 9. This proposition is sometimes used to construct algorithm for computing the Minkowski sums of convex sets and especially convex polyhedra.

isometric embedding

Proof. (i) rewrite using previous prop The direct implication is easy, since if $K \subseteq L$, then

$$\sigma_K(x^*) = \sup_{x \in K} \langle x^* | x \rangle \leq \sup_{x \in L} \langle x^* | x \rangle = \sigma_L(x^*).$$

We assume conversely that $\sigma_K \leq \sigma_L$ and we will prove that $(\mathcal{X} \setminus K) \supseteq (\mathcal{X} \setminus L)$, so that $K \subseteq L$. Consider $x \in \mathcal{X} \setminus L$. Then, by the strong separation theorem Corollary 17, the compact convex set $\{x\}$ is strongly separated from the closed convex set L: there exists a linear form $x^* \in \mathcal{X}^*$ such that

$$\sigma_L(x^*) = \sup_L \langle x^* | \cdot \rangle < \langle x^* | x \rangle.$$

In particular, $\sigma_K(x^*) \leq \sigma_L(x^*) < \langle x^* | x \rangle$, implying by (2) that $x \notin K$.
(ii) We have for every $x^* \in \mathcal{X}$,

$$\begin{aligned}\sigma_K(x^*) = \sup_{x \in K} \langle x^* | x \rangle &= \sup_{x \in \bigoplus_i \lambda_i K_i} \langle x^* | x \rangle \\ &= \sup_{(x_1, \ldots, x_\ell) \in K_1 \times \ldots \times K_\ell} \langle x^* | \sum_i \lambda_i x_i \rangle = \sum_i \lambda_i \sigma_{K_i}(x^*).\end{aligned}$$

(iii) is proven similarly. $\square$

Definition 21 (Normal cone)**.** Let $K \subseteq \mathcal{X}$ be a convex set. The *normal cone* to K at a point x in K is defined by $\operatorname{Nor}_x K = \{x^* \in \mathcal{X}^* \mid \forall y \in K, \langle x^* | x - y \rangle \geq 0\}$. When x is outside of K, we define $\operatorname{Nor}_x K = \emptyset$.

Informally, the normal cone $\operatorname{Nor}_x K$ is the set of "slopes" $x^* \in \mathcal{X}^*$ of all support hyperplanes $\{\langle x^* | \cdot \rangle = \alpha\}$ to K at the point x.

Example 8 (Smooth sublevel set). Let $\mathcal{X} = \mathbb{R}^d$, let $g : \mathbb{R}^d \to \mathbb{R}$ be a $\mathcal{C}^1$ convex function such that $\inf_{\mathbb{R}^d} g < 0$, and let $K = g^{-1}((-\infty, 0])$. Note in particular that K has non-empty interior. Then,

$$\mathrm{Nor}_x K = \begin{cases} \mathbb{R}_+ \nabla g(x) & \text{if } x \in \partial K \\ 0 & \text{if } x \in \mathrm{int}(K) \\ \emptyset & \text{if } x \notin K \end{cases}$$

The notion of normal cone allows to generalize this construction for convex sets that are not differentiable, e.g. polyhedra.

Example 9 (Hilbertian setting). In $\mathcal{X}$ is a Hilbert space, the dual space $\mathcal{X}^*$ can be identified by $\mathcal{X}$ itself thanks to Riesz' theorem (i.e. any continuous linear form on $\mathcal{X}$ is of the form $\langle x|\cdot\rangle$, where $x \in \mathcal{X}$). Then, the normal cone can be characterized by

$$\mathrm{Nor}_x K = \{v \in \mathcal{X} \mid \exists t > 0,\ \mathrm{p}_K(x + tv) = v\},$$

where p_K is the orthogonal projection on K.

Proposition 24. *Let $\mathcal{X}$ be a topological vector space and let $K \subseteq \mathcal{X}$ be closed and convex. Then,*

(i) The normal cone $\mathrm{Nor}_x K$ is convex and weak∗ closed for any $x \in K$, and can be characterized by $x^* \in \mathrm{Nor}_x K \iff \sigma_K(x^*) = \langle x^*|x\rangle$.
(ii) If K has non-empty interior, then for any $x \in \partial K$, $\mathrm{Nor}_x K \neq \emptyset$.

Proof. (i) The convexity and weak∗ closedness of $\mathrm{Nor}_x K$ follows from the fact that this set can be written as an intersection of weak∗ closed halfspaces:

$$\mathrm{Nor}_x K = \{x^* \in \mathcal{X}^* \mid \forall y \in K, \langle x^*|x - y\rangle \geq 0\} = \bigcap_{y \in K} \{\langle x^*|\cdot\rangle \leq \langle x^*|x\rangle\}.$$

The characterization of the normal cone follows from the definitions.
(ii) Let $L = \mathrm{int}\, K$ and let $x \in \partial K$. Since $x \notin L$, the open convex set L and the point x can be linearly separated (Corollary 13): there exists $x^* \neq 0$ such that

$$\sup_{y \in L} \langle x^*|y\rangle \leq \langle x^*|x\rangle.$$

This implies, by definition, that x^* belongs to $\mathrm{Nor}_x K$. □

3 Convex functions

In convex analysis it is very common to encounter functions taking values in the (half) extended real line $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$. They arise for instance when one includes constraints directly in the minimized functional, i.e. if one replaces the problem

$$\min_K f,$$

where $K \subseteq \mathcal{X}$ is convex, by the unconstrained minimization problem

$$\min_{\mathcal{X}} f + \mathrm{i}_K,$$

where i_K is the convex indicator function of K, i.e.

$$\mathrm{i}_K(x) = \begin{cases} 0 \text{ if } x \in K \\ +\infty \text{ if not.} \end{cases}$$

This may seem like a formal trick, but this formulation is often of practical interest, because it allows to solve constrained convex optimization problem using general algorithms for minimizing the sum of two convex functions. More importantly perhaps, convex functions taking the value $+\infty$ appear naturally one when considers the Legendre-Fenchel conjugate. For instance, we will see that the Legendre-Fenchel conjugate of a norm is the $0/+\infty$-valued indicator of the dual unit ball.

The set $\mathbb{R} \cup \{\pm\infty\}$ comes with the intuitive calculus rules such $a + (+\infty) = +\infty$ for $a \in \mathbb{R}$, $a \times (+\infty) = +\infty$ for $a > 0$. Often, one adopts the convention $0 \times +\infty = 0$, so that $0 \times \mathrm{i}_K$ is the zero function.

3.1 Definition and first properties

Definition 22 (Domain and epigraph)**.** Let $f : \mathcal{X} \to \overline{\mathbb{R}}$. We call

(i) *domain of* f, denoted $\mathrm{dom}(f)$, the subset of $\mathcal{X}$ where f take finite values:

$$\mathrm{dom}(f) = \{x \in \mathcal{X},\ f(x) \neq +\infty\};$$

(ii) *epigraph of* f the subset of $\mathcal{X} \times \mathbb{R}$ above the graph of f, i.e.

$$\mathrm{epi}(f) = \{(x, t) \in \mathcal{X} \times \mathbb{R};\, t \geq f(x)\}.$$

A function $f : \mathcal{X} \to \overline{\mathbb{R}}$ is *proper* iff $\mathrm{dom}(f) \neq \emptyset$, i.e. if $f \not\equiv +\infty$.

Definition 23. A function $f : \mathcal{X} \to \overline{\mathbb{R}}$ is *convex* if $\mathrm{epi}(f)$ is a convex subset of $\mathcal{X} \times \mathbb{R}$.

The relationship between this definition and the usual definition of convexity is given in the next proposition.

Proposition 25. *A function* $f : \mathcal{X} \to \overline{\mathbb{R}}$ *is convex iff* $\mathrm{dom}(f)$ *is convex and if for all* x, y *in* $\mathrm{dom}(f)$ *and all* $\alpha \in [0, 1]$,

$$f\left((1-\alpha)x + \alpha y\right) \leq (1-\alpha)f(x) + \alpha f(y). \tag{3}$$

Proof. Assume f is convex. Defining $\Pi_{\mathcal{X}} : \mathcal{X} \times \mathbb{R} \to \mathbb{R}$ as the (linear) application $\Pi_{\mathcal{X}}(x,t) = x$, one has

$$\mathrm{dom}(f) = \{x \in \mathbb{R} \mid \exists t \in \mathbb{R}, (x,t) \in \mathrm{epi}(f)\} = \Pi_{\mathcal{X}}(\mathrm{epi}(f)),$$

Thus, $\mathrm{dom}(f)$ is convex as the image of a convex set under an affine application. Moreover, for all $x, y \in \mathrm{dom}(f)$, the points $x' := (x, f(x))$ and $y' := (y, f(y))$ belong to $\mathrm{epi}(f)$. Thus, by convexity of $\mathrm{epi}(f)$, one gets that for all $\alpha \in [0,1]$,

$$z' = (1-\alpha)x' + \alpha y' = ((1-\alpha)x + \alpha y, (1-\alpha)f(x) + \alpha f(y))$$

belongs to the epigraph of f, giving (3). The converse is left as an exercise. □

The proposition belows give some examples of how to construct convex functions.

Proposition 26. *(i) If $(f_i)_{i\in I}$ is a (possibily uncountable) family of convex functions, the function $x \mapsto \sup_{i\in I} f_i(x)$ is also convex.*
(ii) Let $A : \mathcal{X} \to \mathcal{Y}$ be affine, and let f be convex on $\mathcal{X}$. Then, $f \circ A$ is convex.
(iii) If $f_1, \dots, f_N$ are convex and if $\lambda_1, \dots, \lambda_N \geq 0$, then $\sum_{i=1}^N \lambda_i f_i$ is convex.

Example 10. a. Let $K \subseteq \mathcal{X}$ and let $\mathrm{i}_K : E \to \mathbb{R}$ be its indicatrix function, defined by $\mathrm{i}_K(x) = 0$ if $x \in K$ and $+\infty$ if $x \notin K$. Then, i_K is convex iff K is convex.
b. Linear forms (even discontinuous) are convex.
c. Sublinear functions (e.g. norms, gauges, support functions) are convex.
d. The sublevel sets of a convex functions are convex sets. The reciprocal is false: there exists non-convex functions whose sublevel sets are convex. For instance, the sublevel sets of any mononotone function on $\mathbb{R}$ are convex. Functions with convex sublevel sets are usually called *quasi-convex.*
e. Let $A \subseteq \mathcal{X}$ be a bounded subset of a normed space. The function $x \mapsto \sup_{p\in A} \|x - p\|$ is convex.
f. If f is convex and if K is a convex set, the function $g = f + \mathrm{i}_K$ is convex.
g. The composition of convex functions is not necessarily convex: e.g. $f(x) = x^2$ and $g(x) = -x$ are convex on $\mathbb{R}$, but $g \circ f$ is not.

Definition 24 (Strict convexity)**.** A function $f : \mathcal{X} \to \overline{\mathbb{R}}$ is *strictly convex* if it is convex and if for all $x \neq y \in \mathrm{dom}\, f$ and all $t \in]0,1[$,

$$f((1-t)x + ty) < (1-t)f(x) + tf(y) \tag{4}$$

Remark 10 (Strong and semi-convexity). On a normed vector space, we will say that a function f has convexity modulus $\omega : \mathbb{R} \to \mathbb{R}$ if for all $x, y \in \mathrm{dom}\, f$ and $t \in [0,1]$,

$$f((1-t)x + ty) + t(1-t)\omega(\|x-y\|) \leq (1-t)f(x) + tf(y) \tag{5}$$

When $\omega(t) = \frac{\alpha}{2}t^2$ with $\alpha > 0$, f is called *α-strongly convex.* If $\alpha < 0$, f is called *α-semiconvex.* Note that a semi-convex function is usually not convex.

3.2 Lower semicontinuous convex functions

Definition 25 (Lower semi-continuity)**.** Let $\mathcal{X}$ be a topological vector space. A function $f : \mathcal{X} \to \overline{\mathbb{R}}$ is called *lower semicontinuous* or *lsc* if for all $a \in \mathbb{R}$ the sublevel set $\mathrm{lev}_{\leq a} f$ is closed, where

$$\mathrm{lev}_{\leq a} f = \{x \in \mathcal{X} \mid f(x) \leq a\}.$$

Equivalently, f is lsc if the strict superlevel set $\mathrm{lev}_{>a} f$ is open for all $a \in \mathbb{R}$. Lower semi-continuous functions are sometimes called *closed* in the literature.

Remark 11 (Continuity). A function $f : \mathcal{X} \to \overline{\mathbb{R}}$ is continuous if the sublevel sets $\mathrm{lev}_{\leq a}$ and superlevel sets $\mathrm{lev}_{\geq a}$ are closed for all $a \in \mathbb{R}$, or equivalently if the strict sublevel sets $\mathrm{lev}_{<a}$ and strict superlevel sets $\mathrm{lev}_{>a}$ are open. Indeed, this openness assumption implies that for all $a \leq b$,

$$f^{-1}((a, b)) = \mathrm{lev}_{>a} f \cap \mathrm{lev}_{<b} f$$

is open, as the intersection of a finite number of open sets. Recalling that every open set O of $\mathbb{R}$ is a (possibly uncountable) union of open intervals, we deduce that $f^{-1}(O)$ is open, proving the continuity of f.

Proposition 27. *Let $\mathcal{X}$ be a topological space, let $f_i : \mathcal{X} \to \overline{\mathbb{R}}$ be lower semicontinuous functions indexed by some set I. Then,*

(i) the function $\sup_{i\in I} f_i$ is lower semicontinuous;

(ii) if I is finite and if $(\lambda_i)_{i\in I}$ are positive, then $\sum_{i\in I} \lambda_i f_i$ is lower semicontinuous.

Proof. (i) Let $f = \sup_{i\in I} f_i$. Then its sublevel sets

$$\mathrm{lev}_{\leq a} f = \{x \in \mathcal{X} \mid \forall i, f_i(x) \leq a\} = \bigcap_{i\in I} \mathrm{lev}_{\leq a} f_i,$$

are closed as intersections of closed sets.

(ii) We treat the case $\mathrm{Card}\, I = 2$: let f_1, f_2 be two lsc functions. Then,

$$\mathrm{lev}_{>\alpha} f_1 + f_2 = \{x \in \mathcal{X} \mid f_1(x) + f_2(x) > \alpha\} = \bigcup_{t+s>\alpha} \mathrm{lev}_{>t} f_1 \cap \mathrm{lev}_{>s} f_2,$$

is open as an (arbitrary) union of finite intersections of open sets. □

We have the following alternative characterizations of lower semicontinuity. In the following the product space $\mathcal{X} \times \mathbb{R}$ is always endowed with the product topology (Definition 7), i.e. a subset $O \subseteq \mathcal{X} \times \mathbb{R}$ is open if for all point $(x, t) \in O$, there exists an neighborhood N_x of x in $\mathcal{X}$ and $\varepsilon > 0$ such that $N_x \times (t - \varepsilon, t + \varepsilon) \subseteq O$.

Proposition 28. *The following properties are equivalent:*

(i) the function f is lower-semicontinuous ;

(ii) the set $\mathrm{epi}(f) \subseteq \mathcal{X} \times \mathbb{R}$ is closed ;

In a metric space $\mathcal{X}$, a function $f : \mathcal{X} \to \overline{\mathbb{R}}$ is lower semi-continuous if for all $x \in \mathcal{X}$ and every sequence $(x_n)_{n\geq 1}$ converging to x one has

$$f(x) \leq \liminf_{n\to+\infty} f(x_n).$$

The limit inferior of a real-valued sequence $(f_n) \in \mathbb{R}^{\mathbb{N}}$, denoted $\liminf_{n\to+\infty} f_n$, is the smallest cluster point of the sequence.

Proof. (ii) $\Longrightarrow$ (i): Assume that $\mathrm{epi}(f)$ is closed. Then, for all $a \in \mathbb{R}$, the sublevel set of f can be written as

$$\mathrm{lev}_{\leq a} f = \Pi_{\mathcal{X}}(\mathrm{epi}(f) \cap \mathcal{X} \times [a, +\infty)),$$

i.e. the image under the continuous projection $\Pi_{\mathcal{X}}$ of the intersection between two closed subsets of $\mathcal{X} \times \mathbb{R}$. Thus, $\mathrm{lev}_{\leq a} f$ is closed for every a and f is lsc.

(i) $\Longrightarrow$ (ii): We will show that the complement $(\mathcal{X} \times \mathbb{R}) \setminus \mathrm{epi}(f)$ is open. Consider a point $(x, t) \notin \mathrm{epi}(f)$, i.e. $t < f(x)$, and let $\alpha \in \mathbb{R}$ be such that $t < \alpha < f(x)$. Then, x does not belong to $\mathrm{lev}_{\leq\alpha} f$, and by closedness of this set there exists a small neighborhood N_x around x so that $N_x \cap \mathrm{lev}_{\leq\alpha} f = \emptyset$. The set $N_x \cap (-\infty, \alpha]$ is a neighborhood of (x, t) in $\mathcal{X} \times \mathbb{R}$ and it does not intersect $\mathrm{epi}(f)$, proving that $(\mathcal{X} \times \mathbb{R}) \setminus \mathrm{epi}(f)$ is closed.

The characterization in the metric setting is left as an exercise. □

Remark 12 (Lsc envelope). Let $\mathcal{X}$ be a topological space and let $f : \mathcal{X} \to \overline{\mathbb{R}}$. The previous propositions shows that the function $\hat{f}$ defined by

$$\hat{f}(x) = \sup\{g(x) \mid g : \mathcal{X} \to \overline{\mathbb{R}} \text{ is lsc and } g \leq f\}.$$

is lower semicontinuous; it is in fact the largest lower semicontinuous function below f, called the lsc envelope of f.

Example 11 (Indicator function). The $0/+\infty$-valued indicator function i_A of a set $A \subseteq \mathcal{X}$ is lower semi-continuous iff A is closed.

Example 12 (Maximum of continuous linear functions). Any continuous linear form $x^* \in \mathcal{X}^*$ (and more generally any continuous function) is lower semicontinuous. Since any supremum of lsc function remains lsc (Proposition 27), one deduces that any function of the form

$$\sup_{i\in I} \langle x_i^* | \cdot \rangle + t_i,$$

is lower semicontinuous, where $(x_i^*)_{i\in I}$ is a (possibly uncountable) family of continuous linear forms and where $t_i \in \mathbb{R}$.

The next proposition asserts that for convex functions, weak and strong lower-semicontinuity coincide. This is useful because typically much easier to prove that a function is lower semi-continuous than to prove that it is weakly lower-semicontinuous.

Proposition 29. *Let $\mathcal{X}$ be a locally convex topological vector space, and let $f : \mathcal{X} \to \overline{\mathbb{R}}$ be convex. Then f is lower semi-continuous if and only if f is weakly lower semi-continuous.*

Proof. This follows from Corollary 21: since the sublevel sets of a convex function are convex, they are closed iff they are weakly closed. □

Example 13 (Norm). Let $\phi : \mathbb{R} \to \overline{\mathbb{R}}$ be a convex non-decreasing function, which we assume to be lower semi-continuous. Then, $f(x) := \phi(\|x\|)$ is convex and strongly lower semi-continuous, and therefore sequentially weakly lower-semicontinuous.

Example 14 (An integral functional). Let $\Omega \subseteq \mathbb{R}^d$ be a bounded open set, let $\phi \in \mathcal{C}^0(\mathbb{R}^d)$ and assume that $\phi(x) \geq a\|x\|^p$ for some $a \in \mathbb{R}$ and $p > 1$. Consider $\mathcal{X} = \mathrm{L}^p(\Omega)$, and for any $v \in \mathcal{X}$ define

$$f(v) = \int_\Omega \phi(v(x))\mathrm{d}x \in \overline{\mathbb{R}}.$$

We first prove that f is (strongly) lower semi-continuous. In a normed space, f is lower semicontinuous at $v \in \mathcal{X}$ iff for every sequence (v_n) converging (in norm) to $v \in \mathcal{X}$ one has $f(v) \leq \liminf_{n\to+\infty} f(v_n)$. By a known result on L^p spaces, taking a subsequence if necessary, we may assume that v_n converges (Lebesgue-)almost everywhere to v. Applying Fatou's lemma to $x \mapsto \phi(v_n(x)) - a\|v_n(x)\|^p \geq 0$ we get

$$\begin{aligned} f(v) - a\|v\|_{\mathcal{X}}^p &= \int_\Omega \lim_{n\to+\infty} \phi(v_n(x)) - a\|v_n(x)\|^p \,\mathrm{d}x \\ &\leq \liminf_{n\to+\infty} \int_\Omega \phi(v_n(x)) - a\|v_n(x)\|^p \,\mathrm{d}x \\ &= \liminf_{n\to+\infty} f(v_n) - a \lim_{n\to+\infty} \|v_n\|_{\mathcal{X}}^p, \end{aligned}$$

thus showing that $f(v) \leq \liminf_{n\to+\infty} f(v_n)$, and establishing that f is strongly lsc. If one assumes ϕ to be convex, then f is also convex, and by the previous proposition we deduce that it is weakly lower-semicontinuous.

Definition 26 (Convex lsc functions)**.** The set of lower-continuous convex functions is denoted $\Gamma(\mathcal{X})$, and the set of *proper* lsc convex functions is denoted

$$\Gamma_0(\mathcal{X}) = \{f \in \Gamma(\mathcal{X}) \mid \mathrm{dom}(f) \neq \emptyset\}.$$

Proposition 30. *Let $\mathcal{X}$ be a locally convex topological vector space. If $f \in \Gamma(\mathcal{X})$ is convex lsc, then f is equal to the supremum of its affine minorant, i.e.*

$$\forall x \in \mathcal{X}, f(x) = \sup\{\langle x^*|x\rangle + b \mid x^* \in \mathcal{X}, b \in \mathbb{R} \textit{ s.t. } f \geq \langle x^*|\cdot\rangle + b\}.$$

Example 15. Let $f(x) = \|x\|$ on a normed space. Then,

$$f(x) = \sup\{\langle x^*|x\rangle \mid x^* \in \mathcal{X}^* \text{ and } \|x^*\| \leq 1\}.$$

Proof. If $\mathrm{dom}\, f = \emptyset$, f is constant and equal to $+\infty$, so that there is nothing to prove. Let us suppose that $\mathrm{dom}\, f \neq \emptyset$, or equivalently that $K = \mathrm{epi}(f)$ is a closed convex subset of $\mathcal{X} \times \mathbb{R}$. Define

$$g(x) = \sup\{\langle x^*|x\rangle + b \mid x^* \in \mathcal{X}^*, b \in \mathbb{R} \text{ s.t. } f \geq \langle x^*|\cdot\rangle + b\}.$$

The function g is a supremum of functions below f, so that $g \leq f$. We now prove that $g \geq f$. For this purpose, we will prove that for all x_0 in $\mathcal{X}$ and all $t_0 < f(x_0)$, one has $g(x_0) \geq t_0$. Since $t_0 < f(x_0)$, the point (x_0, t_0) does not belong to K, so that we can strongly separate the closed convex set K from the compact convex set $\{(x_0, t_0)\}$ (Corollary 17). This means that there exists a continuous linear form on $\mathcal{X} \times \mathbb{R}$, which can be described by $(x^*, v) \in \mathcal{X}^* \times \mathbb{R}$ such that

$$\inf_{(x,t)\in \mathrm{epi}(f)} \langle x^* | x\rangle + tv > \langle x^* | x_0\rangle + t_0 v$$

Taking $(x,t) = (x_0, f(x_0))$, we get in particular $vf(x_0) > t_0 v$, i.e. $v(f(x_0) - t_0) > 0$. Since $f(x_0) > t_0$, we deduce that $v > 0$. Therefore, taking $(x, f(x)) \in \mathrm{epi}(f)$ in the previous inequality, we deduce

$$\forall x \in \mathcal{X}, \quad f(x) \geq \langle \frac{x^*}{v} | x_0 - x\rangle + t_0.$$

By definition of g as the supremum of affine minorant of f, we get

$$g(x) \geq \langle \frac{x^*}{v} | x_0 - x\rangle + t_0,$$

In particular, $g(x_0) \geq t_0$ for any $t_0 < f(x_0)$ so that $g(x_0) \geq f(x_0)$. □

Since an intersection of closed convex sets is convex and closed, we deduce that a supremum of convex lower semi-continuous functions is convex lower semi-continuous. This leads to the following definition.

Definition 27 (Lsc convex envelope)**.** Let $f : \mathcal{X} \to \overline{\mathbb{R}}$ be a function. Its convex lower-semicontinuous envelope is defined by

$$\overline{\mathrm{conv}} f = \sup\{g \mid g \leq f \text{ and } g \in \Gamma(\mathcal{X})\}.$$

By the previous proposition, we can also characterize $\overline{\mathrm{conv}} f$ by

$$\overline{\mathrm{conv}} f(x) = \sup\{\langle x^* | x\rangle + b \mid x^* \in \mathcal{X}, b \in \mathbb{R} \text{ s.t. } f \geq \langle x^* | \cdot\rangle + b\}.$$

3.3 Application: Existence of minimizers

Examples in functional spaces For the next two examples, we will use the Sobolev space $W^{1,p}(\Omega)$, where Ω is a bounded open set of $\mathbb{R}^d$ and $p > 1$. Roughly speaking, an element of $W^{1,p}(\Omega)$ is a function $u \in \mathrm{L}^p(\Omega)$ such that for all $1 \leq i \leq d$, the partial derivative $\partial_i u$ belongs to $\mathrm{L}^p(\Omega)$ in the sense of distributions.[1] This Sobolev space is a Banach space, when endowed with the norm

$$\|u\|^p_{W^{1,p}} = \|u\|^p_{\mathrm{L}^p(\Omega)} + \|\nabla u\|^p_{\mathrm{L}^p(\Omega, \mathbb{R}^d)},$$

[1] This means that there exists a function $v_i \in \mathrm{L}^p(\Omega)$ such that for all smooth and compactly supported test function $\phi \in \mathcal{C}^\infty_c(\Omega)$ one has $\int_\Omega (\partial_i \phi) u = -\int_\Omega u v_i$. One then defines $\partial_i u := v_i$.

and a reflexive space when $1 < p < +\infty$. We consider $W_0^{1,p}(\Omega)$ to be the space of elements of Ω who vanish on $\partial\Omega$, formally defined as the closure of $\mathcal{C}_c^\infty(\Omega)$ in $W^{1,p}(\Omega)$. Poincaré's inequality asserts that there exists a constant C such that

$$\forall u \in W_0^{1,p}(\Omega), \|u\|_{W^{1,p}(\Omega)} \leq C \|\nabla u\|_{\mathrm{L}^p(\Omega,\mathbb{R}^d)}.$$

When $p = 2$, these spaces are Hilbert spaces, and one denotes $H^1(\Omega) = W^{1,2}(\Omega)$ and $H_0^1(\Omega) = W_0^{1,2}(\Omega)$.

Example 16 (Obstacle problem). Let $\mathcal{X} = H_0^1(\Omega)$, $K = \{u \in \mathcal{X} \mid u \geq \Psi a.e.\}$ for some $\Psi \in \mathcal{C}_c^\infty(\Omega)$ and consider the minimization problem

$$\min_{u \in K} f(u) \text{ with } f(u) := \int_\Omega \|\nabla u(x)\|^2 \,\mathrm{d}x.$$

Following Exercise 5, one can prove that K is closed, and this set is obviously convex. In addition, the function f is strictly convex, so that the problem $\min_K f$ admits *at most* one solution. Let $(u_k)_{k\geq 1}$ be a minimizing sequence, that is a sequence of elements of K such that

$$\inf_K f = \lim_{k\to+\infty} f(u_k).$$

We can assume, without loss of generality that $(f(u_k))_{k\geq 1}$ is decreasing. Then, Poincaré's inequality implies that $(u_k)_{k\geq 1}$ is bounded, so that by Banach-Alaoglu, this sequence admits a weakly converging subsequence, with limit u. Since K is convex and closed, it is weakly closed, so that u belongs. Finally, since f is lower semicontinuous we have

$$f(u) \leq \liminf_{k\to+\infty} f(u_k) = \inf_K f,$$

thus showing the existence of a minimizer

This example can be generalized as follows.

Definition 28 (Coercivity)**.** A function $f : \mathcal{X} \to \overline{\mathbb{R}}$ is *coercive* on a subset K of a normed vector space $\mathcal{X}$ if for every sequence $(x_n)_{n\in\mathbb{N}}$ of elements of K satisfying $\lim_{n\to+\infty} \|x_n\| = +\infty$, one has $\lim_{n\to+\infty} f(x_n) = +\infty$.

Proposition 31. *Let $\mathcal{X}$ be a separable and reflexive space, let $f : \mathcal{X} \to \overline{\mathbb{R}}$ be convex, proper, lsc, let $K \subseteq \mathcal{X}$ be closed and convex, and consider the minimization problem*

$$\inf_{x\in K} f(x).$$

Then the minimum is attained, provided that one of the following conditions hold

a. the set K is bounded ;

b. or the function f is coercive on K.

The minimizer is unique if f is strictly convex on K.

Proof. Exercise. □

Example 17 (A variational problem). On the space $\mathcal{X} := W_0^{1,p}(\Omega)$ we consider the integral functional

$$f(u) = \int_\Omega \phi(\nabla u(x))\mathrm{d}x,$$

with $\phi : \mathbb{R} \to \mathbb{R}$ a convex continuous function satisfying $\phi(v) \geq a \|v\|^p$ and $a > 0$. Then, the minimization problem

$$\inf_{u\in\mathcal{X}} f(u),$$

admits a solution. To see this, consider a minimizing sequence, i.e. a sequence (u_n) of elements of $\mathcal{X}$ such that $\lim_{n\to+\infty} f(u_n) = \inf_{\mathcal{X}} f$. In particular, $f(u_n) \leq M$ for some constant $M > 0$. By the assumption on f and Poincaré's inequality, we deduce that the sequence $(u_n)_{n\in\mathbb{N}}$ is bounded,

$$\|u_n\|^p_{W^{1,p}(\Omega)} \leq C^p \|\nabla u_n\|^p_{\mathrm{L}^p(\Omega,\mathbb{R}^d)} \leq \frac{C^p}{a} f(u_n) \leq \frac{MC^p}{a}$$

By Banach-Alaoglu's theorem, taking a subsequence if necessary, we may assume that the sequence $(u_n)_{n\in\mathbb{N}}$ admits a weakly converging subsequence, with weak limit u. Then, as in Example 14, we know that the functional f is weakly lower semi-continuous on $\mathcal{X}$, so that

$$f(u) \leq \liminf_{n\to+\infty} f(u_n) = \lim_{n\to+\infty} f(u_n) = \inf_{\mathcal{X}} f,$$

thus showing that u minimizes f over $\mathcal{X}$.

Example in spaces of measures In the next example, we work in the space of Radon measures over a compact subset $\Omega \subseteq \mathbb{R}^d$, i.e.

$$\mathcal{M}(\Omega) := \mathcal{C}^0(\Omega)^*.$$

Example 18 (Sparse spikes deconvolution). Consider a measure on the circle $\mathbb{T} := \mathbb{R}/\mathbb{Z}$ made of finitely many Dirac mass (called "spikes" in this setting), i.e.

$$\mu = \sum_{1\leq i\leq N} \alpha_i \delta_{x_i}.$$

The measure μ is not known, but one has access to a blurry version of it, $y = \Phi\mu \in \mathrm{L}^2(\mathbb{T})$. More precisely, we are given a linear operator (typically a convolution operator) $\Phi : \mathcal{M}(\Omega) \to \mathrm{L}^2(\mathbb{T})$, which we assume to be sequentially weak*-continuous. The Beurling LASSO problem is the following optimization problem, for $\lambda > 0$:

$$\inf_{\nu\in\mathcal{M}(\mathbb{T})} \|\Phi\nu - y\|^2_{\mathrm{L}^2(\mathbb{T})} + \lambda \|\nu\|_{TV},$$

We now show that this problem admits a solution. First note that $g = \|\cdot\|_{TV}$ and $h = \|\Phi(\cdot) - y\|^2_{\mathrm{L}^2(\mathbb{T})}$ are sequentially weak* lsc : for g this follows from the definition of the dual norm as a supremum, and for h this is by assumption. In addition, any minimizing sequence $(\nu_n)_{n\in\mathbb{N}}$ is bounded, ensuring by Banach-Alaoglu the existence of a (not relabeled) weak*-converging subsequence, with limit ν. Thus,

$$(g+h)(\nu) \le \liminf_{n\to+\infty}(f+g)(\nu_n) = \inf_{\nu} g+h.$$

Note also that $g+h$ is a convex function, which is crucial for being able to solve the minimization problem numerically, but we did not use this fact for proving existence. In fact, since the space $\mathcal{M}(X)$ is not reflexive, we cannot apply Proposition 31.

3.4 Continuity of convex functions

Proposition 32. *Let $f:\mathcal{X}\to\overline{\mathbb{R}}$ be convex be convex on a topological vector space, and assume that f is upper bounded on a neighborhood of x. Then, f is continuous at x.*

Remark 13. Proposition 32 implies that if f is linear, then it is continuous at x if and only if it is upper bounded (or lower bounded) on a neighborhood of x. Proposition 32 may be seen as a generalization of the well-know characterization of continuity for linear forms (Lemma 1).

Proof. Without loss of generality, we assume that $x=0$ and $f(x)=0$, and we let O be a neighborhood of the origin and $M\ge 0$ so that $f\le M$ on O. Replacing O by the intersection $O\cap(-O)$ if necessary, we may assume that O is symmetric, i.e. $O=-O$. Let $\varepsilon\in(0,1)$ and let $x\in O_\varepsilon := \varepsilon O$. Then, since $x=(1-\varepsilon)0+\frac{1}{\varepsilon}x$ we get by convexity of f

$$f(x)\le(1-\varepsilon)f(0)+\varepsilon f(x/\varepsilon)\le\varepsilon M.$$

Using $0=1/(1+\varepsilon)x+\varepsilon/(1+\varepsilon)(-x/\varepsilon)$ we get by convexity of f

$$f(0)\le\frac{1}{1+\varepsilon}f(x)+\frac{\varepsilon}{1+\varepsilon}f(-x/\varepsilon),$$

so that, using $-x/\varepsilon\in O$ and $f\le M$ on O we get

$$f(x)\ge(1+\varepsilon)f(0)-\varepsilon f(-x/\varepsilon)\ge -M\varepsilon.$$

We have proven that $|f|\le M\varepsilon$ on O_ε, thus showing continuity of f at 0. □

From now on, we denote

$$\operatorname{cont} f=\{x\in\operatorname{dom} f\mid f\text{ is continuous at }x\}.$$

Proposition 33. *Let $f:\mathcal{X}\to\mathbb{R}$ be convex on a topological vector space. If there exists an open set on which f is upper bounded, then* $\operatorname{cont} f=\operatorname{int}\operatorname{dom} f$.

Proof. Let O be an open set on which $f\le M$, and let $x\in O$. We will use this hypothesis to prove that f is locally upper bounded near any point $y\in\Omega=\operatorname{int}\operatorname{dom}(f)$.

By openness of Ω, there exists $t > 0$ such that $z := y + t(y - x)$ belongs to Ω. The point y belongs to the segment $[x, z]$, and more precisely

$$y = (1 - \alpha)x + \alpha z$$

with $\alpha = 1/(1+t)$. The set $O_y = (1-\alpha)O + \alpha z$ is therefore open (as the Minkowski sum of an open set with a non-empty set) and it contains y. We now prove that $f \leq \max(M, f(z))$ on $B := \mathrm{B}(y, (1 - \alpha)\delta)$. By definition of the Minkowski sum, for all point $w_y \in O_y$, there exists $w_x \in O$ such that $w_y = (1 - \alpha)w_x + \alpha z$. Thus, using $f \leq M$ on O,

$$f(w_y) \leq (1 - \alpha)f(w_x) + \alpha f(z) \leq \max(M, f(z)).$$

The function is then upper bounded in a neighborhood of y, and by the previous proposition, it is continuous at y. □

Remark 14. The hypothesis that f is upper bounded is crucial. For instance, take f a discontinuous linear form on an infinite-dimensional space $\mathcal{X}$. Then, $\mathrm{dom}(f) = \mathcal{X}$, f is convex, but f is discontinuous at every point of $\mathcal{X}$.

Example 19. Let $\mathcal{X}$ be a normed space, let $K \subseteq \mathcal{X}$ and assume that K contains the ball $B(0, r)$ for some $r > 0$. Let $f = j_K$ be the gauge of K. Then, $|f| \leq 1/r$ on the unit ball. The previous proposition implies that f is continuous on $\mathcal{X}$.

Corollary 34. *If* $f : \mathcal{X} \to \overline{\mathbb{R}}$ *is convex and* $\mathrm{cont}\, f \neq \emptyset$*, then* $\mathrm{cont}\, f = \mathrm{int}\,\mathrm{dom}\, f$.

Corollary 35. *If* $\mathcal{X}$ *is finite dimensional, and* $f : \mathcal{X} \to \overline{\mathbb{R}}$ *is convex, then* $\mathrm{cont}\, f = \mathrm{int}\,\mathrm{dom}\, f$.

Proof. Take $f : \mathcal{X} \to \overline{\mathbb{R}}$ convex. If $\mathrm{dom}(f)$ has non-empty interior, then there exists points $x_1, \ldots, x_n$ in the interior so that $K = \mathrm{conv}(\{x_1, \ldots, x_n\})$ also has non-empty interior (take e.g. $x_1 \in \mathrm{int}\,\mathrm{dom}(f)$ and $x_2, \ldots, x_n \in \mathrm{int}\,\mathrm{dom}(f)$ so that $\mathrm{vect}(\{x_2 - x_1, \ldots, x_n - x_1\}) = \mathcal{X}$). Moreover, since every element of K is a convex combination of the points $x_1, \ldots, x_n$,

$$\forall x \in K, f(x) \leq M = \max_i f(x_i).$$

We conclude using the previous corollary. □

In the remainder of this paragraph, we assume that $\mathcal{X}$ is a normed space. Then, the continuity of convex functions can be improved into local Lipschitzness.

Definition 29 (Locally Lipschitz)**.** A function $f : \Omega \subseteq \mathcal{X} \to \mathbb{R}$ is *locally Lipschitz* on the open set Ω if every point in Ω has a neighborhood on which f is Lipschitz:

$$\forall x_0 \in \Omega, \exists \delta > 0, \exists M \in \mathbb{R}, \quad \forall x, y \in \mathrm{B}(x_0, \delta), |f(x) - f(y)| \leq M \|x - y\|.$$

We start with the following intermediary result.

Lemma 36. *Let* $f : \mathcal{X} \to \overline{\mathbb{R}}$ *be convex and assume that* $|f| \leq M$ *on* $\mathrm{B}(x_0, 2\delta)$*. Then,* f *is* $\frac{2M}{\delta}$*-Lipschitz on* $\mathrm{B}(x_0, \delta)$*.*

Proof. Let $x, y \in \mathrm{B}(x_0, \delta)$, and define $\alpha = \|x - y\|$ and $z := y + \frac{\delta}{\alpha}(y - x)$ which by construction belongs to $\mathrm{B}(x_0, 2\delta)$. The point y is a convex combination of z and x,

$$y = \frac{\delta/\alpha}{1 + \delta/\alpha} x + \frac{1}{1 + \delta/\alpha} z,$$

so that by convexity of f,

$$\begin{aligned} f(y) &\leq \frac{\delta/\alpha}{1 + \delta/\alpha} f(x) + \frac{1}{1 + \delta/\alpha} f(z) \\ \text{i.e.} \quad f(y) - f(x) &\leq \frac{-1}{1 + \delta/\alpha} f(x) + \frac{1}{1 + \delta/\alpha} f(z) \end{aligned}$$

We now use the upper bound on $|f|$:

$$f(y) - f(x) \leq \frac{2M}{1 + \delta/\alpha} \leq \frac{2M}{\delta} \alpha = \frac{2M}{\delta} \|x - y\|.$$

We conclude by inversing the role of x et y to get the desired Lipschitz property. □

Combining Lemma 36 and the previous continuity results, we get:

Proposition 37. *Let* $f : \mathcal{X} \to \overline{\mathbb{R}}$ *be convex on a normed space. If* $\operatorname{cont} f \neq \emptyset$*, then* f *is locally Lipschitz in* $\operatorname{int}\operatorname{dom}(f)$*.*

Remark 15. The assumption that Ω is open cannot be removed. Take for instance $f : x \in [0, 1] \mapsto -\sqrt{x}$. Then, $\lim_{x \to 0, x \neq 0} f'(x) = -\infty$, so that f is not locally Lipschitz at 0.

4 Subdifferential

4.1 Directional derivatives

In this section we study the "algebraic" properties of the directional derivatives of convex functions. All the properties hold for any vector space $\mathcal{X}$, even without a topology. Since we are only using the linear properties of the space, one shouldn't hope to get any information about regularity (even continuity) of the functions.

Definition 30. Given a function $f : \mathcal{X} \to \overline{\mathbb{R}}$, $x \in \mathrm{dom}(f)$ and $v \in \mathcal{X}$, we define the directional derivative as the following limit (if it exists):

$$f^+(x; v) = \lim_{\varepsilon \to 0^+} \frac{f(x + \varepsilon v) - f(x)}{\varepsilon} \tag{6}$$

Proposition 38. *Let $f : \mathcal{X} \to \overline{\mathbb{R}}$ be convex and let $x \in \mathrm{dom}(f)$. Then, the directional derivative $f^+(x; v) \in \mathbb{R} \cup \{\pm\infty\}$ is well defined for any $v \in \mathcal{X}$ and moreover,*

$$f^+(x; v) = \inf_{\varepsilon > 0} \frac{f(x + \varepsilon v) - f(x)}{\varepsilon} \tag{7}$$

Remark 16. The limit defining $f^+(x; v)$ can take the values $\pm\infty$.

(i) Since the limit (6) can be replaced by an infimum (7), one has $f^+(x; v) = +\infty$ if and only if the half-line $\{x + tv \mid t > 0\}$ does not intersect $\mathrm{dom}\, f$. As a consequence, if x belongs to the interior of $\mathrm{dom}(f)$, then $f^+(x; v) < +\infty$.

(ii) It is easy to build examples of convex functions on $\mathbb{R}$ such that $f^+(x; v) = -\infty$ for some $x \in \mathrm{dom}(f)$. (Exercise: find one.)

The proof is deduced directly from the following lemma, which is a well-known property of 1D convex functions (slopes are increasing).

Lemma 39. *The function $\varepsilon \in \mathbb{R}_+ \mapsto \frac{1}{\varepsilon}(f(x + \varepsilon v) - f(x))$ is increasing.*

Proof. Let $\varepsilon_2 > \varepsilon_1 \geq 0$. Since $x + \varepsilon_1 v = (1 - \varepsilon_1/\varepsilon_2)x + \varepsilon_1/\varepsilon_2(x + \varepsilon_2 v)$, one has using the convexity of f,

$$f(x + \varepsilon_1 v) \leq (1 - \varepsilon_1/\varepsilon_2) f(x) + \varepsilon_1/\varepsilon_2 f(x + \varepsilon_2 v),$$

which gives the desired inequality. □

Proposition 40. *Let $f : \mathcal{X} \to \overline{\mathbb{R}}$ be convex. Then,*

(i) If $x, y \in \mathrm{dom}(f)$, then the following monotonicity property holds

$$\forall x, y \in \mathrm{dom}(f), \quad f^+(x; y - x) \leq f^+(y; y - x). \tag{8}$$

(ii) If $x \in \mathrm{int}(\mathrm{dom}(f))$, then $f^+(x, \cdot)$ takes finite values only and is sublinear.

(iii) If $x \in \mathrm{cont}(f)$, then $f^+(x, \cdot)$ is a continuous sublinear function.

Proof. (i) This properties directly corresponds to the fact that the slopes of convex function on $\mathbb{R}$ are increasing.
(ii) We first prove finiteness. Let $v \in \mathcal{X}$. Since $x \in \text{int}(\text{dom}(f))$, there exists $\varepsilon > 0$ such that $x \pm \varepsilon v$ belong to $\text{dom}(f)$. Using (7), we deduce that $f^+(x, v) \leq \frac{1}{\varepsilon}(f(x+\varepsilon v) - f(x)) < +\infty$. Similarly, we have $f^+(x; -v) < +\infty$. Now, by convexity of f we have

$$f(x) = f\left(\frac{x+\varepsilon v}{2} + \frac{x-\varepsilon v}{2}\right) \leq \frac{1}{2}f(x+\varepsilon v) + \frac{1}{2}f(x-\varepsilon v),$$

and all terms are finite. Therefore,

$$\frac{f(x+\varepsilon v) - f(x)}{\varepsilon} \geq -\frac{f(x-\varepsilon v) - f(x)}{\varepsilon},$$

Taking the limit $\varepsilon \to 0$, we get $f^+(x; v) \geq -f^+(x; -v) > +\infty$, i.e $f^+(x; \cdot)$ is finite.

We now prove sublinearity of $f^+(x; \cdot)$. The definition directly implies 1-homogeneity of $f^+(x; \cdot)$, we therefore show subadditivity. Let $u, v \in \mathcal{X}$, and $\varepsilon > 0$. Then,

$$x + \varepsilon(u+v) = \frac{x+2\varepsilon u}{2} + \frac{x+2\varepsilon v}{2}$$

so that by convexity

$$\frac{1}{\varepsilon}(f(x+\varepsilon(u+v)) - f(x)) \leq \frac{1}{2\varepsilon}(f(x+2\varepsilon u) - f(x)) + \frac{1}{2\varepsilon}(f(x+2\varepsilon v) - f(x)).$$

Taking the limit as $\varepsilon \to 0$ gives the desired inequality.

(iii) As f is continuous at x, it is bounded in a neighborhood O_x of x. Then, using (7), and setting $O = O_x - x$, we have

$$\forall v \in O, f^+(x; v) \leq f(x+v) - f(x) \leq \max_{O_x} f - f(x).$$

This shows that $f^+(x; \cdot)$ is bounded in a neighborhood of the origin. Since $f^+(x; \cdot)$ is convex (by sublinearity), we deduce from Corollary 34 that $f^+(x; \cdot)$ is continuous on the interior of $\text{dom}(f^+(x; \cdot)$, i.e. on the whole space since $f^+(x; \cdot)$ takes finite values only. □

4.2 Gâteaux and Fréchet differentiability

Definition 31 (Gâteaux-differentiability)**.** A function $f : \mathcal{X} \to \overline{\mathbb{R}}$ on a topological vector space $\mathcal{X}$ is called *Gâteaux-differentiable* at $x \in \text{dom} f$ if and only if the directional derivative $v \in \mathcal{X} \mapsto f^+(x; v)$ is a continuous linear form.

Remark 17. The notion of of Gâteaux-differentiability is quite weak. For instance, consider$f : \mathbb{R}^2 \to \mathbb{R}$ defined

$$f(x_1, x_2) = \begin{cases} 1 \text{ if } x_1 \neq 0 \text{ and } x_2 = x_1^2 \\ 0 \text{ if not} \end{cases}$$

Then, $f^+(0; \cdot) \equiv 0$ so that f is Gâteaux-differentiable at the origin. On the other hand, f is not even continuous at $(0, 0)$!

Corollary 41. *Let $f : \mathcal{X} \to \overline{\mathbb{R}}$ be convex and let $x \in \operatorname{cont} f$. Then, f is Gâteaux-differentiable at x if and only if $f^+(x; v) \leq -f^+(x; -v)$ for all $v \in \mathcal{X}$.*
($\simeq$ iff the left and right derivatives at x coincide for any direction v)

Remark 18. The function $v \mapsto f^+(x; v)$ can be linear and discontinuous, e.g. if f is a discontinuous linear form. Note also that the inequality $f^+(x; v) \geq -f^+(x; -v)$ always holds, so that the hypothesis could be replaced by $f^+(x; v) \leq -f^+(x; -v)$.

Example 20. Let $f(x) = |x|$ on $\mathbb{R}$, then $f^+(0; 1) = 1$ and $f^+(0; -1) = 1 \neq -f^+(0, 1)$.

This proposition follows directly from Proposition 40 and the following lemmas.

Lemma 42. *A sublinear function $g : \mathcal{X} \to \mathbb{R}$ is linear if and only if $g(v) \leq -g(-v)$ for all $v \in \mathcal{X}$.*

Proof. The direct implication is obvious, let us prove the converse. By sublinearity, we have $0 \leq g(v) + g(-v)$, i.e. $g(v) \geq -g(-v)$, so that the assumption implies $g(v) = -g(-v)$. Then,

$$\begin{aligned} &g(v + w) \leq g(v) + g(w) \\ &g(-(v + w)) \leq g(-v) + g(-w) = -g(v) - g(w), \end{aligned}$$

where we used the assumption on the last line. Thus,

$$g(v) + g(w) \leq -g(-(v + w)) = g(v + w) \leq g(v) + g(w),$$

and all inequalities must be equalities; in particular $g(v + w) = g(v) + g(w)$. Since in addition we know from Proposition 40 that $g(\lambda v) = \lambda g(v)$, g is linear. ☐

Definition 32 (Fréchet-differentiability)**.** A function $f : \mathcal{X} \to \overline{\mathbb{R}}$ is *Fréchet-différentiable* at $x \in \operatorname{dom}(f)$ if it is Gâteaux-differentiable at x and if

$$\lim_{v \to 0, v \neq 0} \frac{|f(x + v) - f(x) - f^+(x; v)|}{\|v\|} = 0, \tag{9}$$

i.e. $f(x + v) = f(x) + f^+(x; v) + \mathrm{o}(\|v\|)$.

Remark 19. Fréchet-differentiability is the usual differentiability (and implies continuity). In general, Fréchet différentiability $\Longrightarrow$ Gâteaux-différentiability $\Longrightarrow$ linearity of $v \mapsto f^+(x; v)$. The converse implications are false in general, but is true when f is a convex function on a finite-dimensional space, see Section 4.6.2.

4.3 Definition of the subdifferential and first properties

Definition 33 (Subgradient and subdifferential)**.** Let $f : \mathcal{X} \to \overline{\mathbb{R}}$ and $x_0 \in \operatorname{dom} f$. A linear form $x^* \in \mathcal{X}^*$ is called a *subgradient* of f at x_0 if

$$\forall y \in \mathcal{X}, \ f(y) \geq f(x_0) + \langle x^* | y - x_0 \rangle.$$

The set of subgradients of f at x_0 is called the *subdifferential* of f at x_0 and is denoted $\partial f(x_0)$. When $x_0 \notin \operatorname{dom} f$, we set $\partial f(x_0) := \emptyset$.

Remark 20. When $\mathcal{X}$ is a Hilbert space or $\mathbb{R}^d$, we will often consider the subdifferential of a function at a point as a subset of $\mathcal{X}$, using the isomorphism $\mathcal{X}^* \simeq \mathcal{X}$.

Example 21. (i) Let $f(x) = |v| = \max(x, -x)$. Then,

$$\partial f(x) = \begin{cases} \{-1\} & \text{if } x < 0 \\ \{1\} & \text{if } x > 0 \\ [-1, 1] & \text{if x = 0} \end{cases}$$

(ii) If f is defined on $\mathbb{R}^d$ by $f(x) = \sum_{i=1}^d g(x_i)$ with $g : \mathbb{R} \to \overline{\mathbb{R}}$ convex,

$$\partial f(x) = \prod_i \partial g(x_i).$$

Thus, if $f(x) = \sum_{1 \le i \le N} |x_i|$. Then, denoting $s(x) = \partial |\cdot| (x)$,

$$\partial f(x) = \prod_{1 \le i \le d} s(x_i).$$

(iii) Let $\mathcal{X}$ be a normed space and let $f(x) = \|x\|$. Then,

$$x^* \in \partial f(0) \iff \forall x \in \mathcal{X}, \langle x^* | x \rangle \le \|x\| \iff \|x^*\|_* \le 1$$

In other words, $\partial f(0)$ is the dual unit ball.

(iv) Let $f(x) = -\sqrt{x}$ on $[0, +\infty[$. Then, $\partial f(0) = \emptyset$ even though 0 belongs to $\operatorname{dom} f$.

The last example, in the form of a proposition, makes the connection between the notion of subdifferential and the notion of normal cone:

Proposition 43 (Subdifferential of indicator function)**.** *Let $K \subseteq \mathcal{X}$ be a convex set and let $x \in K$. Then, $\partial \mathrm{i}_K(x) = \mathrm{Nor}_x K$.*

Proposition 44 (General properties)**.** *Let $f : \mathcal{X} \to \overline{\mathbb{R}}$. Then,*

(i) (Fermat's rule) f attains its minimum at $x_0 \in \mathcal{X}$ if and only if $0 \in \partial f(x_0)$.

(ii) (Convexity and closedness) $\partial f(x)$ is weak-closed and convex.*

(iii) (Monotonicity) $\forall x, y \in \mathcal{X}, x^* \in \partial f(x)$ and $y^* \in \partial f(y)$, $\quad \langle x^* - y^* | x - y \rangle \ge 0$.

(iv) (Upper semi-continuity) Assume f is lower semicontinuous. Then, if $(x_n, x_n^*)_{n \in \mathbb{N}}$ is such that $x_n^* \in \partial f(x_n)$, and if

$$x_n \xrightarrow[n \to +\infty]{\text{strong}} x \text{ and } x_n^* \xrightarrow[n \to +\infty]{\text{weak*}} x^*,$$

then $x^* \in \partial f(x)$.

(The same conclusion holds if instead $x_n \xrightarrow[n \to +\infty]{\text{weak}} x$ and $x_n^* \xrightarrow[n \to +\infty]{\text{strong}} x^*$.)

Proof. (ii) This follows from the following representation of $\partial f(x)$ as an intersection of weak* closed halfspaces:

$$\partial f(x) = \bigcap_{y \in \mathcal{X}} \{x^* \in \mathcal{X}^* \mid f(y) \ge f(x) + \langle x^* | y - x \rangle\}.$$ □

(iii) By definition of $x^* \in \partial f(x)$ and $y^* \in \partial f(y)$ we have

$$f(y) \geq f(x) + \langle x^* | y - x \rangle \text{ and } f(x) \geq f(y) + \langle y^* | x - y \rangle.$$

We conclude by summing these inequalities.
(iv) Under the assumptions, we have for all $y \in \mathcal{X}$

$$\begin{aligned} f(y) &\geq f(x_n) + \langle x_n^* | y - x_n \rangle \\ &= f(x_n) + \langle x_n^* | y - x \rangle + \langle x^* | x - x_n \rangle + \langle x_n^* - x^* | x - x_n \rangle \end{aligned}$$

Taking the limit (or liminf) as $n \to +\infty$, we get $f(y) \geq f(x) + \langle x^* | y - x \rangle$, thus showing that $x^* \in \partial f(x)$.

Theorem 45 (Subdifferential of convex functions)**.** *Let $\mathcal{X}$ be a locally convex topological vector space and let $f : \mathcal{X} \to \overline{\mathbb{R}}$ be convex. Then,*

(i) if $x \in \mathrm{dom}(f)$, $\partial f(x) = \{x^* \in \mathcal{X}^* \mid f^+(x; \cdot) \geq \langle x^* | \cdot \rangle\}$.
(ii) if $x \in \mathrm{cont}\, f$, $\partial f(x)$ is non-empty.
(iii) if $x \in \mathrm{cont}\, f$ and if $\mathcal{X}$ is a normed space, $\partial f(x)$ is bounded w.r.t $\|\cdot\|_*$.
(iv) if $x \in \mathrm{cont}\, f$, $f^+(x; v) = \sup\{\langle x^* | v \rangle \mid x^* \in \partial f(x)\}$.
(v) if $x \in \mathrm{cont}\, f$, then f is Gâteaux-differentiable at x if and only if $\partial f(x)$ is a singleton $\{x^*\}$, and then $f^+(x; \cdot) = \langle x^* | \cdot \rangle$.

Remark 21 (Support function of the subdifferential). Let $f \in \Gamma_0(\mathcal{X})$ and $x \in \mathrm{cont}\, f$. Note that we endowed $\mathcal{X}^*$ with the weak$*$ topology and that any weak$*$ continuous linear form over $\mathcal{X}^*$ is induced by a vector of $\mathcal{X}$, i.e. $(\mathcal{X}^*)^* \simeq \mathcal{X}$. The support function of $\partial f(x)$ can therefore be defined as

$$\sigma_{\partial f(x)} : y \in \mathcal{X} \mapsto \sup_{x^* \in \partial f(x)} \langle x^* | y \rangle \in \mathbb{R}$$

Then Theorem 45 immediately implies that $\sigma_{\partial f(x)} = f^+(x, \cdot)$, i.e. the directional derivative can be thought of as the support function of the subdifferential.

Proof. (i) Let A be the second member of the equality. If $x^* \in A$ we have,

$$\forall v \in \mathcal{X}, \quad \langle x^* | v \rangle \leq f^+(x; v) = \inf_{t > 0} \frac{f(x + tv) - f(x)}{t} \leq f(x + v) - f(x)$$

Thus, given $y \in \mathcal{X}$ and letting $v = y - x$, we get $\langle x^* | y - x \rangle \leq f(y) - f(x)$, implying that x^* belongs to $\partial f(x)$. Reciprocally, assume that $x^* \in \partial f(x)$. Then, for all $v \in \mathcal{X}$, $t > 0$ and $y = x + tv$ we have $\langle x^* | y - x \rangle \leq f(x + tv) - f(x)$. Dividing by t and using $y - x = tv$, we get

$$\langle x^* | v \rangle \leq \frac{f(x + tv) - f(x)}{t}.$$

Since this is true for all $t > 0$, we have as desired $\langle x^* | v \rangle \leq f^+(x; v)$, i.e. $x \in A$.

(ii) By Proposition 40, the function $g = f^+(x; \cdot)$ is convex and continuous, so that by Proposition 30, g is equal to the supremum of its affine minorant. In particular,

g admits at least a affine minorant: there exists $x^* \in \mathcal{X}^*$ and $\alpha \in \mathbb{R}$ such that $g(x) \geq \langle x^*|x\rangle + \alpha$ for all $x \in \mathcal{X}$. Letting $x = ty$, we use the homogeneity of g to get

$$\forall y \in \mathcal{X}, \forall t > 0, \quad t(\langle x^*|y\rangle + \alpha/t) \leq tg(y).$$

Letting $t \to +\infty$ we get $g \geq \langle x^*|\cdot\rangle$, so that, using (i), $x^* \in \partial f(x) \neq \emptyset$.

(iii) To show that $\partial f(x)$ is bounded, we use that since f is continuous at x, it is L-Lipschitz near x for some $L \geq 0$. This directly implies that $f^+(x; v) \leq L\|v\|$, so that if $x^* \in \partial f(x)$ we have $\langle x^*|v\rangle \leq f^+(x; v) \leq L\|v\|$. By definition of the dual norm, we get $\|x^*\| \leq L$.

(iv) If $x^* \in \partial f(x)$, one has $f^+(x; \cdot) \geq \langle x^*|\cdot\rangle$, giving

$$\forall v \in \mathcal{X}, \quad f^+(x; v) \geq g(v) := \sup_{x^* \in \partial f(x)} \langle x^*|x\rangle.$$

We now prove the converse inequality. Let $v \in \mathcal{X} \setminus \{0\}$ and define a linear form on the 1D vector subspace $V = \mathbb{R}v \subseteq \mathcal{X}$ by setting $\phi(v) = f^+(x; v)$. The linear form $\phi : V \to \mathbb{R}$ is bounded by the sublinear function $f^+(x; \cdot)$ (by homogeneity), so that by Hahn-Banach theorem ϕ can be extended into a linear form on $\mathcal{X}$ such that $\phi \leq f^+(x; \cdot)$. As in (ii) one deduces that ϕ is continuous, and by (i) we get $\phi \in \partial f(x)$. By definition of g we conclude that $f^+(x; v) = \langle x^*|v\rangle \leq g(v)$.

(v) If f is Gâteaux-differentiable at x, then $\phi := f^+(x; \cdot)$ is linear continuous, and (i) implies that $\partial f(x) = \{\phi_0\}$. Conversely, if $\partial f(x) = \{x^*\}$ is a singleton, (iv) directly implies that $f^+(x; \cdot) = \langle x^*|\cdot\rangle$ is linear continuous. □

4.4 Subdifferential calculus

In applications, one often encounters minimization of functions which are defined as a sum, maximum, or composition of other functions with linear operators. It is therefore of prime interest to study the effect of these operations on the subdifferential. We start with a very general and easy inclusion result. We recall the definition of the adjoint operator:

Definition 34 (Adjoint)**.** Let $\mathcal{L}(\mathcal{Y}, \mathcal{X})$ be the set of continuous linear operators between two topological vector spaces $\mathcal{Y}$ and $\mathcal{X}$. The adjoint of $A \in \mathcal{L}(\mathcal{Y}, \mathcal{X})$ is a linear operator between the dual spaces $\mathcal{X}^*$ and $\mathcal{Y}^*$ defined by

$$A^* : x^* \in \mathcal{X}^* \mapsto (y \in \mathcal{Y} \mapsto \langle x^*|Ay\rangle) \in \mathcal{Y}^*.$$

It is characterized by the property $\langle A^*x^*|y\rangle = \langle x^*|Ay\rangle$.

Proposition 46 (Inclusions)**.** *(i) for all function $f, g : \mathcal{X} \to \overline{\mathbb{R}}$ and all $x \in \mathcal{X}$,*

$$\partial f(x) + \partial g(x) \subseteq \partial(f + g)(x).$$

(ii) if $(f_i)_{i\in I}$ is a family of functions from $\mathcal{X} \to \overline{\mathbb{R}}$, and $f = \sup_{i\in I} f_i$, then for all $x \in \mathcal{X}$, and $I_x := \arg\max_{i\in I} f_i(x)$,

$$\overline{\mathrm{conv}} \bigcup_{i \in I_x} \partial f_i(x) \subseteq \partial f(x).$$

(iii) if $f : \mathcal{X} \to \overline{\mathbb{R}}$, $A \in \mathcal{L}(\mathcal{Y}, \mathcal{X})$, and $g = f \circ A$ then for $y \in \mathcal{Y}$,

$$A^*\partial f(Ay) \subseteq \partial g(y)$$

Proof. Exercise. □

Remark 22 (A counterexample). We note that these inclusions may sometimes be strict. For instance, consider $f = \mathrm{i}_{[-\infty,0)}$ and $g(x) = -\sqrt{x}$ on for $x \geq 0$ and $g(x) = +\infty$ for $x < 0$. Then, $f + g$ takes value $+\infty$ everywhere expect at $x = 0$, so that $\partial(f + g)(0) = \mathbb{R}$. On the other hand, $\partial g(0) = \emptyset$, so that $\partial f(0) + \partial g(0) = \emptyset$.

In order to avoid this counterexample, we need to assume that the domains of f and g intersect non-trivially. Such assumptions are called *qualification conditions* in the literature. The next theorem gives an example of such a qualification condition, but there exists weaker conditions on Banach spaces (see e.g. Brézis-Attouch [AB86]) or when $\mathcal{X} = \mathbb{R}^d$ (see [Roc70, Chapter 23]).

Theorem 47 (Moreau-Rockafellar)**.** *Let $\mathcal{X}$ be a locally convex topological vector space, let $f, g \in \Gamma_0(\mathcal{X})$ and assume the* qualification condition *holds:*

$$\operatorname{dom} f \cap \operatorname{cont} g \neq \emptyset.$$

Then, for all $x \in \mathcal{X}$ one has

$$\partial f(x) + \partial g(x) = \partial(f + g)(x).$$

Proof. Let $x \in \mathcal{X}$ and consider $x^* \in \partial(f + g)(x)$. Thus,

$$\forall y \in \mathcal{X}, f(y) + g(y) \geq f(x) + g(x) + \langle x^*|y - x\rangle.$$

Setting $\tilde{f}(y) = f(y) - (f(x) + g(x)) - \langle x^*|y - x\rangle$, this is equivalent to

$$\forall y \in \mathcal{X}, \tilde{f}(y) \geq -g(y).$$

Consider the set $K = \operatorname{epi}(\tilde{f})$ and $L = -\operatorname{epi}(g)$ in $\mathcal{X} \times \mathbb{R}$. Since $\operatorname{cont}(g) \neq \emptyset$, L has non-empty interior, ensuring by Exercise 4, that L is the closure of its interior $M := \operatorname{int} L$. As a consequence of the Hahn-Banach theorem (Corollary 15), there exists a continuous affine function on $\mathcal{X} \times \mathbb{R}$ separating the disjoint convex sets K and M, with M open. This means that there exists $z^* \in \mathcal{X}^*$ and $\alpha \in \mathbb{R}$ such that

$$\inf_{(y,t)\in K} \langle z^*|y\rangle + \alpha t \geq \sup_{(y,s)\in M} \langle z^*|y\rangle + \alpha s = \sup_{(y,s)\in L} \langle z^*|y\rangle + \alpha s, \tag{10}$$

where the last equality holds because L is the closure of M. Taking $y = x_0$, $t > \tilde{f}(x_0)$ and $s < -g(x_0)$ in the previous inequality (so that $t - s > \tilde{f}(x_0) - g(x_0) \geq 0$) gives

$$0 \leq \alpha(t - s),$$

which is only possible if $\alpha \geq 0$. We now rule out the possibility that $\alpha = 0$. By contradiction, assume that $\alpha = 0$. Then, the separation inequality gives

$$\inf_{y\in\operatorname{dom} f} \langle z^*|y\rangle \geq \sup_{y\in\operatorname{dom} g} \langle z^*|y\rangle,$$

i.e. $\operatorname{dom} f$ and $\operatorname{dom} g$ are linearly separated (note that $z^* \neq 0$ since $(z^*, \alpha) \neq 0$ and $\alpha = 0$). This contradicts the assumption $\operatorname{dom} f \cap \operatorname{cont} g = \emptyset$. Thus, $\alpha > 0$. Replacing z^* by $\frac{1}{\alpha} z^*$ if necessary, we may therefore assume that $\alpha = 1$ in the separation inequality (10). With $(x, \tilde{f}(x)) \in K$ and $(y, -g(y)) \in L$, this inequality gives

$$\forall y \in \mathcal{X}, \quad \langle z^* | x \rangle + \tilde{f}(x) \geq \langle z^* | y \rangle - g(y),$$

and since $\tilde{f}(x) = -g(x)$,

$$\forall y \in \mathcal{X}, g(y) \geq g(x) + \langle z^* | y - x \rangle.$$

In other words, $z^* \in \partial g(x)$. Applying again the separation equality (10) but with $(y, \tilde{f}(y)) \in K$ and $(x, -g(x)) \in L$ we get

$$\forall y \in \mathcal{X}, \quad \langle z^* | y \rangle + \tilde{f}(y) \geq \langle z^* | x \rangle - g(x),$$

Using $\tilde{f}(y) = f(y) - (f(x) + g(x)) - \langle x^* | y - x \rangle$ we get

$$\forall y \in \mathcal{X}, \quad f(y) \geq \langle x^* - z^* | y - x \rangle + f(x),$$

so that $x^* - z^* \in \partial f(x)$. Thus, $x^* \in \partial (f + g)(x)$ can be written as the sum of $z^* \in \partial g(x)$ and $x^* - z^* \in \partial f(x)$. ☐

Theorem 48 (Moreau-Rockafellar). *Let $f \in \Gamma_0(\mathcal{X})$ and $A : \mathcal{Y} \to \mathcal{X}$ be a continuous linear operator between two topological vector spaces. Assume the following qualification hypothesis holds:*

$$\operatorname{cont} f \cap A(\mathcal{Y}) \neq \emptyset.$$

Then, for all $y \in \mathcal{Y}$ one has

$$\partial (f \circ A)(y) = A^* \partial f(Ay).$$

Proof. Let $y \in \mathcal{Y}$, and y^* be an element of $\partial (f \circ A)(y)$. This reads

$$\forall z \in \mathcal{Y}, f(Az) \geq \ell(z) := f(Ay) + \langle y^* | z - y \rangle.$$

We introduce $K = \operatorname{epi} f$ and $L = \{(Az, \ell(z) \mid z \in \mathcal{Y}\}$ which are two convex subsets of $\mathcal{X} \times \mathbb{R}$. Since the function f is continuous at some point in $\mathcal{X}$, the set K has non-empty interior and is therefore equal to the closure of its interior $M := \operatorname{int} K$. The convex sets M and L are disjoint thanks to the subdifferential inequality above and the set M is open, so that by Corollary 15 there exists a non-zero continuous linear form $(x^*, \alpha) \in \mathcal{X} \times \mathbb{R}$ such that

$$\inf_{(x,t) \in K} \langle x^* | z \rangle + \alpha t = \inf_{(x,t) \in M} \langle x^* | z \rangle + \alpha t \geq \sup_{(x,s) \in L} \langle x^* | x \rangle + \alpha s.$$

We argue similarly as before to prove that $\alpha > 0$. By assumption, there exists a point $y_0 \in \mathcal{Y}$ such that $Ay_0 \in \operatorname{cont} f$. Taking $x = Ay_0$ in the infimum and supremum and choosing $t > f(Ay_0)$ and $s = \ell(y_0)$, so that $t > s$, we obtain that $\alpha \geq 0$. Second, we note that $\alpha = 0$ implies that $\operatorname{dom} f$ and $A\mathcal{Y}$ are linearly separated, a contradiction with $\operatorname{cont} f \cap A\mathcal{Y} \neq \emptyset$. Dividing z^* by $\alpha > 0$ if necessary, we now assume that $\alpha = 1$. The separation inequality can then be rewritten as

$$\forall x \in \mathcal{X}, \forall z \in \mathcal{Y}, \quad \langle x^*|x\rangle + f(x) \geq \langle x^*|Az\rangle + \ell(z) = \langle x^*|Az\rangle + f(Ay) + \langle y^*|z - y\rangle.$$

Taking $z = y$ in this inequality inequality, we get

$$\forall x \in \mathcal{X}, \langle x^*|x\rangle + f(x) \geq \langle x^*|Ay\rangle + f(Ay),$$

i.e. $-x^* \in \partial f(Ay)$. Taking $x = Ay$ on the other hand, we obtain

$$\forall z \in \mathcal{Y}, \quad \langle A^*x^* + y^*|y - z\rangle \geq 0,$$

showing that $-A^*x^* = y^*$. Thus, $y^* = A^*(-x^*)$ with $-x^* \in \partial f(Ay)$ as desired. □

Theorem 49 (Dubovitskii-Milyutin)**.** *Let $\mathcal{X}$ be a topological vector space, let $f_1, \ldots, f_N \in \Gamma_0(\mathcal{X})$ and define $f = \sup_{1 \leq i \leq N} f_i$. Then for all $x \in \operatorname{cont}(f_1) \cap \ldots \cap \operatorname{cont}(f_N)$, and $I_x := \arg\max_{i \in I} f_i(x)$,*

$$\partial f(x) = \overline{\operatorname{conv}} \left(\bigcup_{i \in I_x} \partial f_i(x) \right).$$

A more general version of this theorem, where the index set is not finite but compact is proven in [Zal02, Theorem 2.4.18].

Proof. We let $K = \overline{\operatorname{conv}} \left(\bigcup_{i \in I_x} \partial f_i(x) \right) \subseteq \mathcal{X}^*$. For proving that $\partial f(x) = K$ we note

$$\begin{aligned} f^+(x; v) &= \lim_{t \to 0, t > 0} \frac{f(x + tv) - f(x)}{t} \\ &= \lim_{t \to 0, t > 0} \max_{i \in I_x} \frac{f_i(x + tv) - f_i(x)}{t} \\ &= \max_{i \in I_x} \lim_{t \to 0, t > 0} \frac{f_i(x + tv) - f_i(x)}{t} \\ &= \max_{i \in I_x} f_i^+(x; v). \end{aligned}$$

Since the f_i are continuous at x, the function f is also continous at x. Using Remark 21, the previous computation thus yields

$$\sigma_{\partial f(x)} = \max_{i \in I_x} \sigma_{\partial f_i(x)}.$$

By Proposition 23, we deduce that $\sigma_{\partial f(x)} = \sigma_K$, and since both K and $\partial f(x)$ are convex and weak*-closed, we deduce from Proposition 23 that $K = \partial f(x)$. □

4.5 Application: optimality conditions

4.5.1 Examples

Example 22 (Fermat's problem). We consider the following minimization problem

$$\min_{x\in\mathbb{R}^d} \alpha_i \|x - x_1\| + \ldots + \alpha_N \|x - x_N\|,$$

where $\alpha_1, \ldots, \alpha_N \geq 0$ and where $x_1, \ldots, x_N \in \mathbb{R}^d$ are *distinct points*. The minimizer of this problem is called a Fermat point. Putting $f_i = \|\cdot - x_i\|$, we have

$$\partial f_i(x) = \begin{cases} \frac{x-x_i}{\|x-x_i\|} & \text{if } x \neq x_i \\ \mathrm{B}(0,1) & \text{if } x = x_i \end{cases}.$$

Since all the functions f_i are convex and continuous, by Theorem 47

$$\partial(\alpha_1 f_1 + \ldots + \alpha_1 f_N) = \begin{cases} \{\sum_i \alpha_i \frac{x-x_i}{\|x-x_i\|}\} & \text{if } x \notin \{x_1, \ldots, x_N\} \\ B(0, \alpha_j) + \sum_{i\neq j} \alpha_i \frac{x-x_i}{\|x-x_i\|} & \text{if } x = x_j. \end{cases}$$

Thus, x is a Fermat point iff

$$(x = x_j \text{ and } \left\| \sum_{i\neq j} \alpha_i \frac{x_j - x_i}{\|x_j - x_i\|} \right\| \leq \alpha_j)$$

$$\text{or } (x \notin \{x_1, \ldots, x_N\} \text{ and } \sum_i \alpha_i \frac{x - x_i}{\|x - x_i\|} = 0.)$$

Example 23 (Lasso problem). Let A a m-by-d matrix, $y \in \mathbb{R}^m$ and $\gamma > 0$. We consider the following modification of the least-squared problem

$$\min_{x\in\mathbb{R}^d} \frac{1}{2} \|Ax - y\|_2^2 + \gamma \|x\|_1.$$

Set $f(x) = \frac{1}{2} \|Ax - y\|_2^2$ and $g(x) = \|x\|_1$. Setting $h = |\cdot|$ we have

$$\partial f(x) = \{\nabla f(x)\} = \{A^T(Ax - y)\},$$
$$\partial g(x) = \partial h(x_1) \times \ldots \times \partial h(x_d)$$

Since these functions are continuous and convex, Theorem 47 gives $\partial(f + \gamma g)(x) = \partial f(x) + \gamma \partial g(x)$. Thus, x is a minimizer of the problem if and only if

$$A^T(Ax - y) = -\gamma v \text{ where } v \text{ satisfies } \begin{cases} v_i = \mathrm{sgn}(x_i) & \text{if } x_i \neq 0 \\ v_i \in [-1, 1] & \text{if } x_i = 0 \end{cases}$$

or in other words, letting A_i be the ith column of A, x is a minimizer iff

$$\begin{cases} |\langle A_i | Ax - y\rangle| \leq \gamma & \text{if } x_i = 0 \\ \langle A_i | Ax - y\rangle = \gamma\, \mathrm{sgn}(x_i) & \text{otherwise} \end{cases}$$

$$\iff \begin{cases} |\langle A_i | Ax - y\rangle| \leq \gamma & \text{if } x_i = 0 \\ \langle A_i | Ax - y\rangle = \gamma\, \mathrm{sgn}(x_i) & \text{otherwise.} \end{cases}$$

One can see from this example that the term $\|x\|_1$ induces sparsity of the solution

4.5.2 Karush-Kuhn-Tucker theorem

Theorem 50 (Variational characterization of optimality). *Let $K \subseteq \mathcal{X}$ be a closed convex subset and $f : \mathcal{X} \to \overline{\mathbb{R}}$ be convex lsc, and assume that*

$$(\operatorname{int} K \cap \operatorname{dom} f \neq \emptyset) \text{ or } (K \cap \operatorname{cont} f \neq \emptyset)).$$

Then, the following are equivalent :

(i) x is a minimizer of f on K ;

(ii) $-\partial f(x) \cap \operatorname{Nor}_x K \neq \emptyset$.

(iii) $\exists x^* \in \partial f(x)$ s.t.

$$\forall y \in K, \langle x - y | - x^* \rangle \geq 0.$$

Proof. The point $x \in \mathcal{X}$ minimizes f on K if and only if it minimizes $f + g$ with $g = \mathrm{i}_K$. This is also equivalent to $0 \in \partial(f + g)(x)$. Since $\operatorname{dom} g = K$ and $\operatorname{cont} g = \operatorname{int} K$, the first hypothesis is equivalent to $\operatorname{cont} g \cap \operatorname{dom} f \neq \emptyset$ and the second one is equivalent to $\operatorname{dom} g \cap \operatorname{cont} f \neq \emptyset$. In either case, one can apply Theorem 47 to get

$$\partial(f + g) = \partial f(x) + \partial \mathrm{i}_K = \partial f(x) + \operatorname{Nor}_x K.$$

We deduce that x is a minimizer if and only if there exists $x^* \in \partial f(x)$ such that $-x^* \in \operatorname{Nor}_x K$ (this is (ii)) and the equivalence with (iii) comes from the definition of the normal cone. □

Proposition 51 (Normal cone of sublevel set). *Let $g \in \Gamma_0(\mathcal{X})$, let $K = \operatorname{lev}_{\leq 0} g$ and assume Slater's condition:*

$$\begin{cases} K \subseteq \operatorname{cont} g \\ \operatorname{lev}_{<0} g \neq \emptyset \end{cases}$$

Then,

$$\operatorname{Nor}_x K = \begin{cases} 0 & \text{if } g(x) < 0 \\ \mathbb{R}^+ \partial g(x) & \text{if } g(x) = 0 \\ \emptyset & \text{if } g(x) > 0 \end{cases}.$$

Proof. First we note that if $g(x) < 0$, then by assumption $x \in \operatorname{cont} g$. Thus, $g \leq 0$ on a neighborhood of x, which implies that i_K is locally equal to zero. From this we deduce that $\operatorname{Nor}_x K = \partial \mathrm{i}_K(x) = \{0\}$.

From now on we assume that $g(x) = 0$. We first prove the inclusion $\mathbb{R}^+ \partial g(x) \subseteq \operatorname{Nor}_x K$. Let $x^* \in \partial g(x)$, so that

$$\forall y \in \mathcal{X}, g(y) \geq \langle x^* | y - x \rangle + g(x)$$

Multiplying this inequality by $\lambda \geq 0$ we get

$$\forall y \in K, 0 \geq \lambda g(y) \geq \langle \lambda x^* | y - x \rangle + \underbrace{g(x)}_{=0},$$

thus showing that $\lambda x^* \in \operatorname{Nor}_x K$.

Let us prove the converse inclusion $\operatorname{Nor}_x K \subseteq \mathbb{R}^+\partial g(x)$. Let $x^* \in \operatorname{Nor}_x K$. By definition of the normal cone, for all $y \in K$ one has $\langle x^*|x\rangle \geq \langle x^*|y\rangle$, implying that K is included in the complement of the closed half-space $H = \{y \in X \mid \langle x^*|y\rangle \geq \langle x^*|x\rangle \geq 0\}$. This in turns implies that $g \geq 0$ on H, so that x minimizes g over H. Since $x \in \operatorname{cont} g \cap \operatorname{dom}(\mathrm{i}_H)$, we can apply the subdifferential sum rule:

$$0 \in \partial(g + \mathrm{i}_H) = \partial g(x) + \mathbb{R}^+\{-x^*\}.$$

In other words, there exists $y^* \in \partial g(x)$ and $\lambda \geq 0$ such that $\lambda x^* = y^*$. Note that λ cannot be zero since x is not a minimizer of g. Then, $x^* = \frac{1}{\lambda} y^*$ as desired. □

Theorem 52 (Karush-Kuhn-Tucker). *Let $f, g_1, \ldots, g_N \in \Gamma_0(\mathcal{X})$ let $K = \bigcap_{1\leq i\leq N} \operatorname{lev}_{\leq 0} g_i$, and assumes that Slater's condition*

$$\begin{cases} \operatorname{lev}_{\leq 0} g_i \subseteq \operatorname{int}\operatorname{dom} g_i & \forall i \in \{1, \ldots, N\} \\ \operatorname{dom} f \cap \operatorname{lev}_{<0} g_1 \cap \ldots \cap \operatorname{lev}_{<0} g_N \neq \emptyset \end{cases} \tag{11}$$

hold. Then the following are equivalent:

(i) x is a minimizer of f on K ;

(ii) there exists $\lambda_1, \ldots, \lambda_N \geq 0$ such that

$$\begin{cases} 0 \in \partial f(x) + \lambda_1 \partial g_1(x) + \ldots + \lambda_N \partial g_N(x) \\ \lambda_i g_i(x) = 0 & \forall i \in \{1, \ldots N\} \end{cases}$$

Remark 23. If the functions g_i are continuous, the first hypothesis in Slater's condition automatically holds.

Remark 24 (Lagrange multipliers). The scalars $\lambda_1, \ldots, \lambda_N$ whose existence is given by the theorem are called *Lagrange multipliers*. If $\lambda_1, \ldots, \lambda_N$ are as in the theorem, we directly obtain that x minimizes $f + \lambda_1 g_1 + \ldots + \lambda_N g_N$ over $\mathcal{X}$. Thus, knowing the Lagrange multipliers allows to replace the constrained optimization problem $\min_K f$ by an unconstrained problem.

Proof. We consider $K_i = \operatorname{lev}_{\leq 0} g_i$. The second assumption's in Slater's condition ensures that there exists a point in $\operatorname{dom} f$ in the interior of the sets K_i, i.e.

$$\operatorname{dom} f \cap \operatorname{cont}(i_{K_1}) \cap \ldots \cap \operatorname{cont}(i_{K_N}) \neq \emptyset.$$

Applying Theorem 47 recursively, we obtain

$$\begin{aligned} \partial(f + \mathrm{i}_{K_1} + \ldots + \mathrm{i}_{K_N})(x) &= \partial f(x) + \partial \mathrm{i}_{K_1}(x) + \ldots + \partial \mathrm{i}_{K_N}(x) \\ &= \partial f(x) + \operatorname{Nor}_x K_1 + \ldots + \operatorname{Nor}_x K_N, \end{aligned}$$

so that $x \in K$ minimizes f on K iff there exists $x^* \in \partial f(x)$ and $x_i^* \in \operatorname{Nor}_x K_i$ such that $-x^* = x_1^* + \ldots + x_i^*$. Since in addition for $x \in K$, we have

$$\operatorname{Nor}_x K_i = \begin{cases} \mathbb{R}^+\partial g_i(x) & \text{if } g_i(x) < 0 \\ \{0\} & \text{if not} \end{cases},$$

we get that $x_i^* = \lambda_i y_i^*$ with $y_i^* \in \partial g_i(x)$ and $\lambda_i = 0$ if $g_i(x) < 0$ and $\lambda_i \geq 0$ if not, as desired. □

4.6 Differentiability almost everywhere

Motivation Given a compact convex domain $K \subseteq \mathcal{X}$ and $x^* \in \mathcal{X}^*$, we consider the following linear programming problem:

$$\sup_{x \in K} \langle x^* | x \rangle \tag{12}$$

We denote $f : x^* \in \mathcal{X}^* \to \mathbb{R}$ the *value function* of this problem, i.e. $f(x^*)$ is the value of the maximum in the previous definition. In other words, f is the support function of K. Then, f is convex and if x is a solution to (12), i.e. if $x \in K$ and $f(x^*) = \langle x^* | x \rangle$, one has

$$\begin{aligned} f^+(x^*, v^*) &= \lim_{\varepsilon \to 0} \frac{1}{\varepsilon}(f^+(x^* + \varepsilon v^*) - f(x^*)) \\ &\geq \lim_{\varepsilon \to 0} \frac{1}{\varepsilon}(\langle x^* + \varepsilon v^* | x \rangle - \langle x^* | x \rangle) \geq \langle v^* | x \rangle \end{aligned}$$

It is often useful to know whether (12) has a unique maximizer. Suppose that there exists $x \neq y \in K$ such that $f(x^*) = \langle x^* | x \rangle = \langle x^* | y \rangle$. Then, as before,

$$f^+(x^*, v^*) \geq \max(\langle v^* | x \rangle, \langle v^* | y \rangle).$$

This shows that the application $v^* \in \mathcal{X}^* \mapsto f^+(x^*; v^*)$ is not linear, so that f is not differentiable at x. Thus Gâteaux-differentiability of f at x^* implies the uniqueness of the minimizer to the linear programming problem (12). If we prove that f is Gâteaux-differentiable "almost everywhere", we get that the problem (12) has a unique solution for "almost every" linear form $x^* \in \mathcal{X}^*$. (Note that this last property is quite intuitive when K is a convex polytope in $\mathbb{R}^d$.)

Remark 25. Characterizing the non-differentiability locus of convex functions also has applications in optimal transport [M+95], in optimal control [CS04], etc. There exists many results on the "size" of the non-differentiability locus of convex functions, both in finite [AAC92] and infinite dimensions [BV+10, §4.6].

On $\mathbb{R}$, convex functions are differentiable almost everywhere thanks to the following proposition.

Proposition 53. *Let $f : \mathbb{R} \to \overline{\mathbb{R}}$ be convex. Then, there are at most a countable number of points in* $\operatorname{dom} f$ *where f is non-differentiable.*

This proposition is false in higher dimension, consider e.g. $f(x_1, x_2) = |x_1|$ on $\mathbb{R}^2$.

Proof. Consider $f'_+(x) = f^+(x; 1)$ et $f'_-(x) = -f^+(x, -1)$ the right and left derivatives. These functions are increasing (exercise) and for all $x < x_0$ in $\operatorname{dom}(f)$,

$$f'_+(x) = \inf_{y > x} \frac{f(y) - f(x)}{y - x} \leq \frac{f(x_0) - f(x)}{x_0 - x} \leq f'_-(x_0),$$

thus showing the inequality

$$\lim_{x \to x_0^-} f'_+(x) \leq f'_-(x_0) \leq f'_+(x_0).$$

The function f is differentiable at x_0 if and only $f'_-(x_0) = f'_+(x_0)$. Thus, if f is not differentiable at x_0, the right derivative f'_+ has a jump at x_0:

$$\lim_{x \to x_0^-} f'_d(x) < f'_d(x_0).$$

One concludes by using that an increasing function can only have a countable number of jumps. □

4.6.1 Gâteaux-différentiability almost everywhere

Theorem 54 (Mazur)**.** *Let $\mathcal{X}$ be a separable Banach space and let $f : \mathcal{X} \to \mathbb{R}$ be continuous and convex. Then, f is Gâteaux-differentiable on a dense subset of $\mathcal{X}$.*

To prove this theorem, we start by considering a dense sequence $(v_n)_{n \geq 0}$ in $\mathcal{X}$, which exist thanks to the separability assumption. Then, we introduce the sets

$$A_{m,n} = \{x \in \mathcal{X} \mid f^+(x, v_n) + f^+(x, -v_n) \geq 1/m\}, \qquad A = \bigcup_{m,n \geq 1} A_{m,n} \tag{13}$$

The sketch of the proof is as follows:

a. First we prove that f is Gâteaux-differentiable on $\mathcal{X} \setminus A$;

b. Second, that all sets $A_{m,n}$ are closed and have empty interior.

Then, by Baire's theorem, we know that A has empty interior (i.e. $\mathcal{X} \setminus A$ is dense).

Lemma 55. *The function f is Gâteaux-differentiable on $\mathcal{X} \setminus A$.*

Proof. As f is continuous on $\mathcal{X}$, by Corollary 41,

$$\begin{aligned}
&f \text{ is not Gâteaux-differentiable at } x \in \mathcal{X} \\
&\quad \Longrightarrow \exists v \in \mathcal{X}, f^+(x, v) + f^+(x, -v) > 0 \\
&\quad \Longrightarrow \exists v \in \mathcal{X}, \exists m > 1, f^+(x, v) + f^+(x, -v) > 2/m \\
&\quad \Longrightarrow \exists m, n \geq 1, f^+(x, v_n) + f^+(x, -v_n) > 1/m \\
&\quad \Longrightarrow x \in A_{m,n} \subseteq A,
\end{aligned}$$

where we used the continuity of the map $v \mapsto f^+(x, v)$. □

Lemma 56. *The set $A_{m,n}$ defined in* (13) *is closed.*

This is a consequence of the following proposition, showing that for any $v \in \mathcal{X}$, the directional derivative $f^+(\cdot, v)$ is upper-semicontinuous. Then, $f^+(\cdot, v) + f^+(\cdot, -v)$ is also usc, and $A_{m,n}$ is closed as a superlevel set of a usc function.

Lemma 57. *Let $f : \mathcal{X} \to \overline{\mathbb{R}}$ be convex and let $x \in \mathrm{cont}(f)$. Then, for any sequence $(x_n)_{n \in \mathbb{N}}$ converging to x, one has $f^+(x, v) \geq \limsup_{k \to \infty} f^+(x_k, v)$.*

Proof. By convexity and continuity, f is M-Lipschitz in a neighborhood of x. Without loss of generality, we assume that the sequence $(x_n)_{n\in\mathbb{N}}$ remains in this neighborhood. Let $\varepsilon > 0$. Using the Lipschitz property, we get

$$\begin{aligned}\frac{1}{\varepsilon}(f(x+\varepsilon v) - f(x)) &\geq \frac{1}{\varepsilon}(f(x_k+\varepsilon v) - f(x_k) - 2L\|x - x_k\|)\\ &\geq f^+(x_k, v) - \frac{2L\|x-x_k\|}{\varepsilon}\end{aligned}$$

Taking the infimum on the left-hand side we get

$$f^+(x; v) \geq \limsup_{k\to\infty} f^+(x_k, v).$$ □

Lemma 58. *The set $A_{m,n}$ defined in* (13) *has empty interior.*

Proof. Assume that the interior of $A_{m,n}$ contains a point x, i.e. there exists $r > 0$ such that $\mathrm{B}(x, r) \subseteq A_{m,n}$. Let $x_t := x + tv_n$ and $g : t \in [0, r] \mapsto f(x_t)$. Then,

$$\begin{aligned}&\forall t \in [0, r], \quad -f^+(x_t, -v_n) + 1/m \leq f^+(x_t, v_n)\\ &\implies \forall t \in [0, r], \quad g \text{ is not differentiable at } t\end{aligned}$$

This contradicts Proposition 53, which shows that the non-differentiability set of g is countable. □

4.6.2 Fréchet-differentiability almost everywhere on $\mathbb{R}^d$

The behaviour of convex functions on $\mathbb{R}^d$ is much simpler than on infinite-dimensional spaces. Let $f : \mathbb{R}^d \to \overline{\mathbb{R}}$ be a convex function and let $x \in \operatorname{cont} f$. We will prove the following chain of implications:

$$\begin{aligned}&f \text{ admits partial derivatives } \left(\frac{\partial f}{\partial e_i}(x)\right)_{1\leq i\leq d}\\ &\implies \text{the application } v \mapsto f^+(x; v) \text{ is linear}\\ &\implies f \text{ is Gâteaux-differentiable at } x\\ &\implies f \text{ is Fréchet-differentiable at } x\end{aligned}$$

From this, we will deduce that a convex function $f : \mathbb{R}^d \to \overline{\mathbb{R}}$ is a.e. differentiable on its domain.

Proposition 59. *Let $f : \mathbb{R} \to \overline{\mathbb{R}}$ be convex. If f is Gâteaux-differentiable at $x \in \operatorname{int}(\operatorname{dom}(f))$, then f is also Fréchet-différentiable at x.*

This proposition follows from the next lemma and from the fact that f is locally Lipschitz around x.

Lemma 60. *Let $f : \mathrm{B}(x, r) \subseteq \mathbb{R}^d \to \mathbb{R}$ be convex and Lipschitz. Then, f is Gâteaux-differentiable at x iff it is Fréchet-différentiable at x.*

Proof. Let S be the unit sphere of $\mathbb{R}^d$. By compactness, for all $\varepsilon > 0$, there exists a finite family of vectors $(v_i)_{1\le i\le N}$ of S such that $S \subseteq \cup_i \mathrm{B}(v_i, \varepsilon)$. By Gâteaux-differentiability of f at x, for all i, there exists δ_i such that

$$\forall t \in [-\delta_i, \delta_i], \ \left\| f(x + tv_i) - (f(x) + tf^+(x; v_i)) \right\| \le \varepsilon\, |t|$$

Consider $\delta := \min_i \delta_i > 0$. By construction of the (v_i), for all v in S, there exists $i \in \{1, \dots, N\}$ such that $\|v_i - v\| \le \varepsilon$. Then, using that f is Lipschitz (which implies that $f^+(x; \cdot)$ is also Lipschitz), we have

$$\begin{aligned} \|f(x + tv_i) - f(x + tv)\| &\le M\, |t|\, \varepsilon \\ \left\|f^+(x; v_i) - f^+(x; v)\right\| &\le M\, |t|\, \varepsilon \end{aligned}$$

Thus, for all $v \in S$ and all $|t| \le \delta$,

$$\begin{aligned} \left\| f(x + tv) - (f(x) + tf^+(x; v)) \right\| &\le \left\| f(x + tv_i) - (f(x) + tf^+(x; v)) \right\| + 2M\varepsilon\, |t| \\ &\le (2M + 1)\varepsilon\, |t| \end{aligned}$$

Equivalently (by homogeneity of $f^+(x; \cdot)$) we have for all $v \in \mathbb{R}^d$, $\|v\| \le \delta$,

$$\left\| f(x + v) - (f(x) + f^+(x; v)) \right\| \le (2M + 1)\varepsilon\, \|v\|\,,$$

thus showing that f is Fréchet-differentiable at x. □

Proposition 61. *Let $f : \mathbb{R}^d \to \overline{\mathbb{R}}$ be convex. Then, f is Gâteaux-differentiable at $x \in \mathrm{int}(\mathrm{dom}(f))$ iff it admits partial derivatives at x.*

Lemma 62. *Let $g : \mathbb{R}^d \to \mathbb{R}$ be sublinear. Then, the set*

$$V = \{v \in \mathbb{R}^d \mid f^+(x; v) = -f^+(x; -v)\}$$

is a linear subspace of $\mathbb{R}^d$.

Proof. The fact that V is stable by multiplication by a scalar follows from the homogeneity of g. By sublinearity, $0 = g(u+(-u)) \le g(u)+g(-u)$, so that $-g(-u) \le g(u)$. Let $v, w \in V$. We have

$$g(v + w) \le g(v) + g(w) = -g(-v) + -g(-w) \le -g(-v - w) \le g(v + w),$$

where we used sublinearity to get the first and second inequality, the definition of V to get the equality, and the property $-g(u) \le g(-u)$ to get the third inequality. This shows that $v + w \in V$, proving that V is a linear subspace. □

Proof of Proposition 61. The function f is locally Lipschitz near x and $g = f^+(x; \cdot)$ is sublinear. Let $V := \{v \in \mathcal{X} \mid g(v) = -g(-v)\}$. By the previous lemma, V is a linear subspace of $\mathcal{X}$. Since the partial derivatives exist, we have $f^+(x; -e_i) = -f^+(x; e_i)$ for all basis vector e_i, thus showing that $e_i \in V$ for all i. Therefore, $V = \mathcal{X}$ and $f^+(x; \cdot)$ is linear. □

Theorem 63. *Let $f : \mathbb{R}^d \to \overline{\mathbb{R}}$ be a convex function. Then f is Fréchet-différentiable at a.e. point in* $\mathrm{int}(\mathrm{dom}(f))$.

Proof. Let A be the non-diffentiability locus of f in $\Omega := \mathrm{int}(\mathrm{dom}(f))$. By Proposition 61, the set A is contained in the union of the sets

$$A_i := \left\{ x \in \Omega \mid \frac{\partial f}{\partial e_i} \text{ does not exist } x \right\}.$$

Therefore, to show that A has zero measure, it suffices to prove that all of the sets A_i have zero measure. Without loss of generality, we assume that $i = n$, and we consider ϕ the indicator function of A_n. By Tonelli's theorem,

$$\lambda(A_n) = \int_{\mathbb{R}^n} \phi(x) \mathrm{d}x = \int_{\mathbb{R}^{n-1}} \int_{\mathbb{R}} \phi(y, x_n) \mathrm{d}x_n \mathrm{d}y$$

However, for all $y \in \mathbb{R}^{n-1}$, $t \mapsto \phi(y, t)$ is the non-differentiability locus of the 1D convex function $t \in \mathbb{R} \mapsto f(y, t)$. By Proposition 53, B_y is countable, and therefore has zero Lebesgue measure, thus concluding the proof. □

Remark 26. In fact, one can prove that the non-differentiability locus of $f : \mathbb{R}^d \to \overline{\mathbb{R}}$ has Hausdorff dimension $d - 1$, and it is in fact a rectifiable set of dimension $\leq d - 1$. This means that S can be covered, up to a set with zero $d - 1$–Hausdorff measure, by a countable union of Lipschitz maps $(\phi_n)_{n \in \mathbb{N}}$ with $\phi_i : \mathbb{R}^{d-1} \to \mathbb{R}$ [AAC92]. This fact is used, in optimal transport, to give a characterization of pairs of measures for which there exists an optimal transport map, see e.g. [M$^+$95].

5 Proximal operator

In this short chapter, we assume that $\mathcal{X}$ is a Hilbert space, which we identify with its dual. We introduce the proximal operator, which is a basic building block of many first-order methods for minimizing non-smooth convex functions. We recall that a function $f : \mathcal{X} \to \overline{\mathbb{R}}$ is coercive if $\lim_{\|x\| \to \infty} f(x) = +\infty$.

5.1 Definition and properties

Definition 35 (Proximal operator)**.** The *proximal operator* associated to a function $f \in \Gamma_0(\mathcal{X})$ is defined by

$$\mathrm{Prox}_{f(x)} = \arg\min_{y \in \mathcal{X}} \frac{1}{2} \|x - y\|^2 + f(y).$$

Example 24 (Projection)*.* If K is a closed convex subset of $\mathcal{X}$ and $f = \mathrm{i}_C$, then for all $\gamma > 0$ one has $\mathrm{Prox}_{\gamma \mathrm{i}_K} x = p_K(x)$. The proximal operator generalizes the projection on a convex set, and shares some of its properties.

Proposition 64. *Let $f \in \Gamma_0(\mathcal{X})$ and $\gamma > 0$. Then,*

(i) The minimization problem

$$\min_{y \in \mathcal{X}} \frac{1}{2\gamma} \|x - y\|^2 + f(y),$$

has a unique solution, implying that $\mathrm{Prox}_{\gamma f}$ is well defined.

(ii) The point $p = \mathrm{Prox}_\gamma f(x)$ is characterized by the relation $x \in (\mathrm{id} + \gamma \partial f)(p)$.

(iii) The point x minimizes f on $\mathcal{X}$ iff $x = \mathrm{Prox}_\gamma f(x)$;

Proof. Let $g = \frac{1}{2\gamma} \|x - \cdot\|^2$ and $h = f + g$. (i) The function h belongs to $\Gamma_0(\mathcal{X})$. Since $f \in \Gamma_0(\mathcal{X})$, f is equal to the supremum of its affine minorants (Proposition 30), so that in particular it admits an affine minorant: there exists $x^* \in \mathcal{X}$ and $\alpha \in \mathbb{R}$ such that $f \geq \langle x^* | \cdot \rangle + \alpha$. Thus,

$$h = f + g \geq \langle x^* | \cdot \rangle + \alpha + \frac{1}{2\gamma} \|x - \cdot\|^2 ,$$

from which we deduce that h is coercive. In addition, h is strictly convex as the sum of a convex and a strictly convex function. By Proposition 31, we deduce that h admits a unique minimizer.

(ii) Applying Theorem 47 on the subdifferential of a sum of two functions, which we can use since $\mathrm{cont}(g) = \mathcal{X}$ and $\mathrm{dom}\, f \neq \emptyset$, we see that $p = \mathrm{Prox}_{\gamma f}(x)$ if and only if

$$0 \in \partial(f + g)(p) = \partial f(p) + \frac{1}{\gamma} \{p - x\} \iff x \in (\mathrm{id} + \gamma \partial f)(p).$$

(iii) The point x minimizes f if and only if 0 belongs to $\partial[\gamma f](x) = \gamma \partial f(x)$, or equivalently if x belongs to $(\mathrm{id} + \gamma \partial f)(x)$. □

Example 25 (Proximal of ℓ^1 norm). Let $h : \mathbb{R} \to \mathbb{R}, x \mapsto |x|$. Then,

$$\partial h(y) = \begin{cases} -1 & \text{if } y < 0 \\ [-1, 1] & \text{if } y = 0 \\ 1 & \text{if } y > 0 \end{cases} \qquad \Longrightarrow \qquad (\mathrm{id} + \gamma\partial h)(y) = \begin{cases} y - \gamma & \text{if } y < 0 \\ [-\gamma, \gamma] & \text{if } y = 0 \\ y + \gamma & \text{if } y > 0 \end{cases}.$$

By the characterization given in the previous proposition, we get

$$\mathrm{Prox}_{\gamma h}(x) = \begin{cases} x - \gamma & \text{if } x \geq \gamma \\ 0 & \text{if } -\gamma \leq x \leq \gamma \\ x + \gamma & \text{if } x \leq -\gamma. \end{cases}$$

Thus, $\mathrm{Prox}_{\gamma h}$ is exactly the so-called soft thresholding operator

$$s_\gamma(r) = \begin{cases} r - \gamma\,\mathrm{sgn}(r) & \text{if } |r| \geq \gamma \\ 0 & \text{otherwise} \end{cases}.$$

already introduced in Example 23. If $\mathcal{X} = \mathbb{R}^n$ and if $f(x) = \|x\|_1 = \sum_i |x_i|$, then

$$\min_{y \in \mathbb{R}^n} \frac{1}{2\gamma} \|x - y\|_2^2 + f(y) = \sum_{1 \leq i \leq n} \min_{y_i \in \mathbb{R}} \frac{1}{2\gamma}(x_i - y_i)^2 + |y_i|,$$

so that $\mathrm{Prox}_\gamma f(x) = (s_\gamma(x_1), \dots, s_\gamma(x_n))$.

5.2 Proximal point algorithm

By Proposition 64, minimizing f is equivalent to finding a fixed point of the proximal operator. This suggests the following minimization algorithm:

$$\begin{cases} x_0 \in \mathcal{X} \\ x_{n+1} = \mathrm{Prox}_\gamma f(x_n) \end{cases} \tag{PPA}$$

This algorithm is called the *proximal point algorithm.* It has been originally introduced by Martinet [Mar70, Mar72] in the 1970s, and was later generalized by Rockafellar [Roc76]. Before proving the convergence of this algorithm, we list a very useful property of the proximal operator.

Definition 36 (Firm non-expansiveness)**.** A map $T : \mathcal{X} \to \mathcal{X}$ is *firmly non-expansive* it it satisfies one of the following equivalent conditions

(i) $\forall x, y \in \mathcal{X},\ \|T(x) - T(y)\|^2 \leq \langle T(x) - T(y) | x - y\rangle$

(ii) $\forall x, y \in \mathcal{X},\ \|T(x) - T(y)\|^2 \leq \|x - y\|^2 - \|(x - T(x)) - (y - T(y))\|^2$

To see that these two conditions are equivalent, it suffices to remark that

$$\|(x - T(x)) - (y - T(y))\|^2 = \|x - y\|^2 + \|T(x) - T(y)\|^2 - 2\langle x - y | T(x) - T(y)\rangle.$$

Note that a firmly non-expansive operator is 1-Lipschitz – but firm non-expansiveness is a stronger property.

Proposition 65. *Let* $f \in \Gamma_0(\mathcal{X})$ *and let* $\gamma > 0$*. Then*

(i) The point $p = \mathrm{Prox}_\gamma f(x)$ *is characterized by the inequality*

$$\forall q \in \mathcal{X}, \ \frac{1}{\gamma}\langle x - p | q - p\rangle \leq f(q) - f(p)$$

(ii) The operator $T : x \mapsto \mathrm{Prox}_\gamma f(x)$ *is firmly non-expansive.*

Note that (i) is a generalization of a well-known characterization of the projection of a point on a convex set.

Proof. (i) The point $p = \mathrm{Prox}_\lambda f(x)$ is characterized by $x \in (\mathrm{id} + \gamma\partial f)(p)$ or equivalently by $\frac{1}{\gamma}(x - p) \in \partial f(p)$. This is equivalent to

$$\forall q \in \mathcal{X}, \ f(q) \geq f(p) + \frac{1}{\gamma}\langle x - p | q - p\rangle.$$

(ii) Let $x_1, x_2 \in \mathcal{X}$ and let $p_i = \mathrm{Prox}_\gamma f(x_i)$. We apply the inequality form (i) using first $x = x_1$, $p = p_1$ et $q = p_2$ and then switching the roles:

$$f(p_2) - f(p_1) \geq \frac{1}{\gamma}\langle x_1 - p_1 | p_2 - p_1\rangle$$
$$f(p_1) - f(p_2) \geq \frac{1}{\gamma}\langle x_2 - p_2 | p_1 - p_2\rangle$$

Summing these inequalities and multipliying by $\gamma > 0$, we obtain

$$\|p_2 - p_1\|^2 \leq \langle p_1 - p_2 | x_1 - x_2\rangle.$$ □

We note that since $T = \mathrm{Prox}_\gamma f$ is merely 1-Lipschitz, we cannot deduce the convergence of the proximal point algorithm from Picard's fixed point theorem (one would need T to be k-Lipschitz for some $k < 1$). However, the property of non-expansiveness allows prove convergence of the PPA.

Theorem 66 (Martinet)**.** *Let* $f \in \Gamma_0(\mathcal{X})$ *and assume that* f *is coercive, meaning that* $\lim_{\|x\| \to +\infty} f(x) = +\infty$*. Then, the sequence of points generated by the proximal point algorithm* (PPA) *weakly converges to a global minimizer of* f*.*

This theorem is a direct consequence of the next theorem about firmly non-expansive operators, where where we have set $T = \mathrm{Prox}_{\gamma f}$. Note that by assumption, f has a minimizer, implying that the set of fixed-points of T is non-empty

$$\mathrm{Fix}(T) = \{x \in \mathcal{X} \mid T(x) = x\} \neq \emptyset.$$

Theorem 67 (Martinet)**.** *Let* T *be firmly non-expansive and such that* $\mathrm{Fix}(T) \neq \emptyset$*. Then the sequence defined by* $x_{n+1} = T(x_n)$ *converges weakly to a fixed point of* T*.*

A modern proof of this theorem placing it in the general setting of fixed point iterations for α-averaged operators can be found in the book of Bauschke and Combettes [BC+11]. We nonetheless present the original proof of Martinet, which is self-contained, for the completeness of these notes.

Proof. Step 1. We first prove that the sequence $(x_n)_{n\in\mathbb{N}}$ is bounded and that $\lim_{n\to\infty}\|x_n - x_{n+1}\| = 0$. Let $c \in \mathrm{Fix}(T)$. Firm non-expansiveness gives:

$$\|Tx_n - Tc\|^2 \leq \|x_n - c\|^2 - \|(x_n - T(x_n)) - (c - T(c))\|^2 .$$

Using $Tx_n = x_{n+1}$ and $T(c) = c$, we deduce that

$$\|x_{n+1} - c\|^2 \leq \|x_n - c\|^2 - \|x_n - x_{n+1}\|^2 .$$

The sequence $\|x_n - c\|_{n\geq 1}$ is therefore decreasing and bounded from below, and thus admits a limit. This gives $\|x_n - x_{n+1}\|^2 \leq \|x_n - c\|^2 - \|x_{n+1} - c\|^2 \underset{n\to+\infty}{\longrightarrow} 0$.

Step 2. We prove that every weak cluster point $\bar{x}$ of $(x_n)_{n\in\mathbb{N}}$ is a fixed point of T. First, we note that

$$\begin{aligned}\|x_n - T(\bar{x})\|^2 &= \|x_n - \bar{x} + \bar{x} - T(\bar{x})\|^2 \\ &= \|x_n - \bar{x}\|^2 + \|\bar{x} - T(\bar{x})\|^2 - 2\langle x_n - \bar{x}|\bar{x} - T(\bar{x})\rangle,\end{aligned}$$

Second, using that $x_{n+1} = T(x_n)$ and that T is 1-Lipschitz we get

$$\begin{aligned}\|x_n - T(\bar{x})\| &= \|x_n - x_{n+1} + x_{n+1} - T(\bar{x})\| \\ &\leq \|x_n - x_{n+1}\| + \|T(x_n) - T(\bar{x})\| \\ &\leq \|x_n - x_{n+1}\| + \|x_n - \bar{x}\| ,\end{aligned}$$

Combining these computations, we therefore get

$$\begin{aligned}\|\bar{x} - T(\bar{x})\|^2 &= \|x_n - T(\bar{x})\|^2 - \|x_n - \bar{x}\|^2 + 2\langle x_n - \bar{x}|\bar{x} - T(\bar{x})\rangle \\ &\leq (\|x_n - x_{n+1}\| + \|x_n - \bar{x}\|)^2 - \|x_n - \bar{x}\|^2 + 2\langle x_n - \bar{x}|\bar{x} - T(\bar{x})\rangle \\ &= \|x_n - x_{n+1}\|^2 + 2\,\|x_n - x_{n+1}\|\,\|x_n - \bar{x}\| + 2\langle x_n - \bar{x}|\bar{x} - T(\bar{x})\rangle \xrightarrow{n\to+\infty} 0.\end{aligned}$$

Note that we have used Step 1. to prove that the first two terms converge to zero and the hypothesis that $\bar{x}$ is a weak cluster point to control the last term.

Step 3. Let c_1 and c_2 be two weak cluster points of the sequence, which by Step 2 belong to $\mathrm{Fix}(T)$. By Step 1, we know that $\|x_{n+1} - c_i\|^2 \leq \|x_n - c_i\|^2$, implying that the sequence

$$\|x_n\|^2 - 2\langle x_n|c_i\rangle = \|x_n - c_i\|^2 - \|c_i\|^2 ,$$

is decreasing and therefore converging. Substrating these sequences for $i = 1$ and 2, we see that

$$(\|x_n\|^2 - 2\langle x_n|c_1\rangle) - (\|x_n\|^2 - 2\langle x_n|c_2\rangle) = 2\langle x_n|c_1 - c_2\rangle$$

is converging. Since c_1 and c_2 are weak cluster points of x_n, we obtain by taking limit over the corresponding subsequences

$$\langle c_1|c_1 - c_2\rangle = \langle c_2|c_1 - c_2\rangle,$$

implying $\|c_1 - c_2\|^2 = 0$, i.e. $c_1 = c_2$. The sequence $(x_n)_{n\geq 1}$ is bounded and has a unique weak cluster point, and is therefore weakly converging. □

6 Convex duality

In this chapter, all spaces are supposed to be locally convex topological vector spaces.

6.1 Convex conjugate

Definition 37 (Convex conjugate)**.** Let $f : \mathcal{X} \to \overline{\mathbb{R}}$ be a proper function on a space $\mathcal{X}$. Its *convex conjugate* or *Legendre-Fenchel transform* is the function $f^* : \mathcal{X}^* \to \overline{\mathbb{R}}$ defined by

$$f^*(x^*) = \sup_{x \in \mathcal{X}} \langle x^* | x \rangle - f(x).$$

Note that since f is proper, there exists $x \in \mathrm{dom}(f)$ implying that $f^* \geq \langle \cdot | x \rangle - f(x)$. In particular, f^* never takes the value $-\infty$.

Example 26. We start of a few examples of conjugate functions.

a. Let A be a non-empty subset of $\mathcal{X}$, and let i_A be its indicator function. Then,

$$\forall x^* \in \mathcal{X}^*, \quad \mathrm{i}_A^*(x^*) = \sup_{x \in \mathcal{X}} \langle x^* | x \rangle - \mathrm{i}_A(x) = \sup_{x \in A} \langle x^* | x \rangle = \sigma_A$$

One can therefore think of the Legendre-Fenchel transform as a generalization of the support function.

b. Let $f = \|.\|$ be the norm on $\mathcal{X}$. Then,

$$\forall x^* \in \mathcal{X}^*, f^*(x^*) = \sup_{x \in \mathcal{X}} \langle x^* | x \rangle - \|x\|$$

If $\|x^*\| > 1$, by definition there exists $x \in \mathcal{X}$ such that $\langle x^* | x \rangle - \|x\| > 0$. Multiplying x by λ, one can see that $f^*(x^*) = +\infty$. On the other hand, if $\|x^*\|_* \leq 1$, $\langle x^* | x \rangle - \|x\| \leq 0$ with equality when $x = 0$. Thus, $f^*(x^*) = 0$. Finally, f^* is the indicator function of the unit ball in $\mathcal{X}^*$.

c. Let $f = \langle z^* | \cdot \rangle$ be a continuous linear form. Then,

$$f^*(x^*) = \sup_{x \in E} \langle x^* - z^* | x \rangle = \begin{cases} 0 & \text{if } x^* = z^* \\ +\infty & \text{if not.} \end{cases}$$

Thus, $f^* = i_{\{z^*\}}$.

d. If $f(x) = \frac{1}{p} |x|^p$ on $\mathbb{R}$ with $p \in (1, +\infty)$, and if we identify the dual space with $\mathbb{R}$, one can verify that $f^*(y) = \frac{1}{q} |y|^q$ where $\frac{1}{p} + \frac{1}{q} = 1$.

Proposition 68 (Basic properties)**.** *Let $f : \mathcal{X} \to \overline{\mathbb{R}}$ be proper. Then, the following properties hold*

(i) [Fenchel-Young]*: $\forall (x, x^*) \in \mathcal{X} \times \mathcal{X}^*, f(x) + f^*(x^*) \geq \langle x^* | x \rangle$,*

(ii) $f^*(0) = -\inf f$,

(iii) f^* is convex and weak lower semicontinuous,*

(iv) if in addition $f \in \Gamma_0(\mathcal{X})$, then $f^* \in \Gamma_0(\mathcal{X}^*)$.

Hypothesis	**Conclusion**	**Remark**
$f(x) \leq g(x)$	$f^*(x^*) \geq g^*(x^*)$	
$g(x) = \mathrm{i}_C(x)$	$g^*(x^*) = \sigma_C(x^*)$	
$g(x) = \|x\|$	$g^*(x^*) = \mathrm{i}_{\mathrm{B}(0,1)}$	
$g(x) = \|x\|_H^2$	$g^*(x^*) = \|x^*\|_H^2$	$H = \text{Hilbert}$
$g(x) = \langle z^*\|x\rangle, z^* \in X^*$	$g^*(x^*) = i_{\{z^*\}}(x^*)$	
$g(x) = f(\lambda x), \lambda \neq 0$	$g^*(x^*) = f^*(x^*/\lambda)$	
$g(x) = \lambda f(x), \lambda > 0$	$g^*(x^*) = \lambda f^*(x^*/\lambda)$	
$g(x) = f(x+b), b \in \mathcal{X}$	$g^*(x^*) = f^*(x^*) - \langle x^*\|b\rangle$	

Table 1: A few examples of convex conjugates.

Example 27. Hölder inequality: with $f(x) = \frac{1}{p}|x|^p$ on $\mathbb{R}$, we have :

$$xy \leq f(x) + f^*(y) = \frac{1}{p}|x|^p + \frac{1}{q}|x|^q$$

Summing, we get for all $x, y \in \mathbb{R}^n$, such that $\|x\|_p = \|y\|_q = 1$,

$$\langle x|y\rangle = \sum_{i=1}^n x_i y_i \leq \frac{1}{p}\|x\|_p + \frac{1}{q}\|y\|_q = \|x\|_p \|y\|_q .$$

By homogeneity, this inequality remains true for all $x, y \in \mathbb{R}^n$.

Proof. (iii) These properties hold because f^* is a supremum of functions of the form $x^* \mapsto \langle x^*|x\rangle - f(x)$, which are convex and weak* continuous.
(iv) If $f \in \Gamma_0(\mathcal{X})$ then by Proposition 30 it is equal to the supremum of its affine minorants. In particular, there exists $x^* \in \mathcal{X}^*$ and $\alpha \in \mathbb{R}$ s.t. $f \geq \langle x^*|\cdot\rangle + \alpha$. Therefore, $f^*(x^*) = \sup_{x\in E}\langle x^*|x\rangle - f(x) \leq \sup_{x\in E}\langle x^* - x^*|x\rangle - \alpha \leq -\alpha$, so that f^* is indeed proper. ☐

Definition 38 (Biconjugate)**.** The *convex biconjugate* of a function $f : \mathcal{X} \to \overline{\mathbb{R}}$ is the function f^{**} on $\mathcal{X}$ defined by

$$f^{**}(x) = \sup_{x^*\in\mathcal{X}^*} \langle x^*|x\rangle - f^*(x^*).$$

(In other words, f^{**} is the restriction of $(f^*)^* : (\mathcal{X}^*)^* \to \overline{\mathbb{R}}$ to the base space $\mathcal{X}$, which one regards as a subset of $\mathcal{X}^{**}$ through the canonical isomorphism.)

Theorem 69 (Fenchel-Moreau)**.** *Let $f : \mathcal{X} \to \overline{\mathbb{R}}$ be a proper function. Then, f^{**} is the lower semicontinuous convex envelope of f (Definition 27):*

$$f^{**} = \overline{\mathrm{conv}} f,$$

In particular one always has $f^{**} \leq f$, with equality if and only $f \in \Gamma_0(\mathcal{X})$.

Proof. We first prove $f^{**} \leq f$. By Fenchel-Young, for all $x \in \mathcal{X}$ and $x^* \in \mathcal{X}^*$ we have $f(x) + f^*(x^*) \geq \langle x^*|x\rangle$. Applying this inequality in the definition of f^{**} we obtain

$$f^{**}(x) = \sup_{x^* \in \mathcal{X}^*} \langle x^*|x\rangle - f^*(x^*) \leq f(x).$$

Since in addition is convex and lower semicontinuous, we deduce from the definition of $\overline{\text{conv}} f$ that $f^{**} \leq \overline{\text{conv}}(f)$. To prove the converse inequality, we use that $\overline{\text{conv}} f$ is equal to the supremum of the affine minorants of f. It is therefore sufficient to prove that f^{**} is larger than any affine minorant of f. Consider $x^* \in \mathcal{X}^*$ and $\alpha \in \mathbb{R}$ s.t. $f \geq \langle x_0^*|\cdot\rangle + \alpha$. Then,

$$f^*(x^*) = \sup_{x \in \mathcal{X}} \langle x^*|x\rangle - f(x) \leq \sup_{x \in \mathcal{X}} \langle x^*|x\rangle - (\langle x^*|x\rangle + \alpha) = -\alpha.$$

Thus,

$$\forall x \in \mathcal{X}, \quad f^{**}(x) = \sup_{z^* \in \mathcal{X}^*} \langle z^*|x\rangle - f^*(z^*) \geq \langle x^*|x\rangle - f^*(x^*) = \langle x^*|x\rangle + \alpha. \qquad \square$$

Theorem 70 (Subdifferential and conjugation). *Let $f : \mathcal{X} \to \overline{\mathbb{R}}$ be proper. Then, for any $x \in \mathcal{X}$ and $x^* \in \mathcal{X}^*$ the following two conditions are equivalent:*
(i) $x^* \in \partial f(x)$;
(ii) $f(x) + f^*(x^*) = \langle x^*|x\rangle$
If in addition f belongs to $\Gamma_0(\mathcal{X})$, then (i) and (ii) are equivalent with
(iii) $x \in \partial f^*(x^*)$;

Proof. (i)$\iff$ (ii) A linear form x^* belongs to $\partial f(x)$ if and only if

$$\begin{aligned}\forall y \in \mathcal{X}, f(y) \geq f(x) + \langle x^*|y - x\rangle &\iff \forall y \in \mathcal{X}, \langle x^*|y\rangle - f(y) \leq \langle x^*|x\rangle - f(x) \\ &\iff f^*(x^*) = \sup_{y \in Y} \langle x^*|y\rangle - f(y) \leq \langle x^*|x\rangle - f(x) \\ &\iff f^*(x^*) + f(x) \leq \langle x^*|x\rangle\end{aligned}$$

Since Fenchel-Young's inequality asserts that $f^*(x^*) + f(x) \geq \langle x^*|x\rangle$ always holds, we get the equivalence between (i) and (iii).
(i)$\iff$ (iii) Since $f \in \Gamma_0(\mathcal{X})$, we know from Fenchel-Moreau that $f^{**} = f$. Therefore,

$$\begin{aligned}x^* \in \partial f(x) &\iff f(x) + f^*(x^*) = \langle x^*|x\rangle \\ &\iff f^{**}(x) + f^*(x^*) = \langle x^*|x\rangle \\ &\iff x \in \partial f^*(x^*),\end{aligned}$$

where we applied the equivalence between (i) and (iii) to the function f^*. $\square$

Example 28 (A sufficient condition for Gâteaux-differentiability). Let $f \in \Gamma_0(\mathcal{X})$. If f^* is strictly convex, then f is Gâteaux-differentiable in $\text{cont}(f)$.

Proof. Let $x \in \text{cont}(f)$. To show that f is Gâteaux-differentiable at x it suffices to establish that $\text{Card}\,\partial f(x) \leq 1$. Assume by contradiction that there exists $x_0^* \neq x_1^*$ such that $x_0^*, x_1^* \in \partial f(x)$. By convexity of $\partial f(x)$ one has $x_t^* = (1-t)x_0^* + tx_1^* \in \partial f(x)$ for all $t \in [0,1]$. Thus, using the previous proposition,

$$\forall t \in [0,1],\ f(x) = \langle x_t^* | x\rangle - f^*(x_t^*) = \langle x_0^* | x\rangle - f^*(x_0^*) = \langle x_1^* | x\rangle - f^*(x_1^*).$$

Therefore, f^* is not strictly convex on $[x_0^*, x_1^*]$. □

Example 29 (Prox of conjugate). Let $\mathcal{X}$ be a Hilbert space and $f \in \Gamma_0(\mathcal{X})$. Then,

$$\forall x \in \mathcal{X}, \text{Prox}_f(x) + \text{Prox}_{f^*}(x) = x. \tag{14}$$

Proof. Let $p = \text{Prox}_f(x)$. Then, by definition,

$$\begin{aligned} x \in (\text{id} + \partial f)(p) &\iff x - p \in \partial f(p) \\ &\iff p \in \partial f^*(x-p) \\ &\iff x \in x - p + \partial f^*(x-p) \\ &\iff x - p = \text{Prox} f^*(x). \end{aligned}$$

□

Remark 27. Eq. (14) generalizes the formula $x = \text{proj}_V x + \text{proj}_{V^\perp} x$, where $V^\perp$ is the orthogonal of the subspace $V \subseteq \mathcal{X}$. See Exercise 26 for a generalization of this formula to convex cones.

6.2 Perturbations of convex problems

This section is inspired by the presentation of Ekeland and Temam [ET99] and by a forthcoming book by Guillaume Carlier. We consider the problem of minimizing a convex function $f \in \Gamma_0(\mathcal{X})$ on a space $\mathcal{X}$

$$P = \inf_{x \in \mathcal{X}} f(x),$$

and we assume that the function f can be written as $f(x) = \Phi(x, 0)$ where $\Phi \in \Gamma_0(\mathcal{X} \times \mathcal{Y})$ is also convex, and where $\mathcal{Y}$ is another space. In other words, we assume that the original problem is a special instance of the following problem, which is parameterized by a vector $y \in \mathcal{Y}$:

$$P_y := \inf_{x \in \mathcal{X}} \Phi(x, y).$$

In the following, we regard Φ^* as a function on $\mathcal{X}^* \times \mathcal{Y}^*$, through the identification $(\mathcal{X} \times \mathcal{Y})^* \simeq \mathcal{X}^* \times \mathcal{Y}^*$. The dual problem to the minimization problem P is then:

$$D = \sup_{y^* \in \mathcal{Y}^*} -\Phi^*(0, y^*).$$

The construction of this dual problem is be found in the proof of the following proposition. Note that there is no uniqueness of the dual problem: there is one dual problem to P associated to each perturbation Φ of f.

Proposition 71 (Weak duality)**.** *The weak duality inequality $P \geq D$ always hold. Moreover, for $(x, y^*) \in \mathcal{X} \times \mathcal{Y}^*$, the following statements are equivalent:*

(i) $\Phi(x, 0) = -\Phi^*(0, y^*)$*,*

(ii) x is a minimizer of P and y^* a maximizer of D and $P = D$,

(iii) $(0, y^*) \in \partial\Phi(x, 0)$*,*

(iv) $(x, 0) \in \partial\Phi^*(0, y^*)$*.*

Proof. Fenchel-Young's formula asserts that

$$\forall (x, y^*) \in \mathcal{X} \times \mathcal{Y}^*, \quad \Phi(x, 0) + \Phi^*(0, y^*) \geq \langle (0, y^*) | (x, 0) \rangle = 0,$$

thus implying $\inf_x \Phi(x, 0) \geq \sup_{y^*} -\Phi^*(0, y^*)$. We deduce at once the equivalence between the statements (i) and (ii). To see the equivalence between (i) and (iii), we use the equality case in Fenchel-Young's inequality (Theorem 70): equality holds if $(0, y^*) \in \partial\Phi(x, 0)$. The equivalence between (i) and (iii) uses the same equality case but applied to Φ^* and $\Phi^{**} = \Phi$. □

Theorem 72 (Strong duality)**.** *Let $\mathcal{X}, \mathcal{Y}$ be two spaces, let $\Phi \in \Gamma_0(X, Y)$ and consider the following minimisation problem: $P = \inf_{\mathcal{X}} \Phi(\cdot, 0)$. Assume:*

- *P is finite ;*
- *the following qualification hypothesis is verified: $\exists x_0 \in \mathcal{X}, \quad 0 \in \mathrm{cont}(\Phi(x_0, \cdot))$.*

Then, the maximum in the the dual problem $D = \max_{y^* \in \mathcal{Y}^*} -\Phi^*(0, y^*)$ is attained and strong duality holds

$$P = D.$$

In addition, the set of the maximizers of D is the subdifferential of the value function

$$v : y \in \mathcal{Y} \mapsto \inf_{x \in \mathcal{X}} \Phi(x, y).$$

The following lemma is central:

Lemma 73. *Under the assumptions of Theorem 72, the value function v satisfies:*

(i) v is convex;

(ii) $v^*(y^*) = \Phi^*(0, y^*)$ *;*

(in particular, $v(0) = P$ and $v^{**}(0) = \sup_{y^* \in \mathcal{Y}^*} -\Phi^*(0, y^*) = D$;

(iii) v is continuous at 0 ;

(iv) $\partial v(0) \neq \emptyset$*.*

Since $P = v(0)$ and $D = v^{**}(0)$, Moreau-Rockafellar's theorem (Theorem 69) would directly imply $P = D$ if v was lower-semicontinuous.

Proof. (i) Let $y_0, y_1 \in \mathcal{Y}$ and let $y_t = (1 - t)y_0 + ty_1$. Then, for $t \in [0, 1]$ and $x \in \mathcal{X}$, one has

$$\Phi(x, y_t) = \Phi(x, (1 - t)y_0 + ty_1) \leq (1 - t)\Phi(x, y_0) + t\Phi(x, y_1).$$

Taking the infimum on both sides gives

$$\begin{aligned} v(y_t) &\le \inf_x (1-t)\Phi(x,y_0) + t\Phi(x,y_1) \\ &\le (1-t)\inf_x \Phi(x,y_0) + t\inf_x \Phi(x,y_1). \\ &= (1-t)v(y_0) + tv(y_1) \end{aligned}$$

(ii) Given $y^* \in \mathcal{Y}^*$ one has

$$\begin{aligned} v^*(y^*) &= \sup_{y\in\mathcal{Y}} \langle y^*|y\rangle - v(y) \\ &= \sup_{y\in\mathcal{Y}} \langle y^*|y\rangle - \inf_{x\in\mathcal{X}} \Phi(x,y) \\ &= \sup_{(x,y)\in\mathcal{Y}\times\mathcal{X}} \langle (0,y^*)|(0,y)\rangle - \Phi(x,y) = \Phi^*(0,y^*). \end{aligned}$$

(iii) By the qualification hypothesis, there exists $x_0 \in \mathcal{X}$ such that $\Phi(x_0,\cdot)$ is continuous near 0. Then,

$$v(y) = \inf_{x\in\mathcal{X}} \Phi(x,y) \le \Phi(x_0,y),$$

is bounded near $y = 0$. Since the function v is convex, boundedness in an open set implies continuity inside that set (Proposition 37).
(iv) From Theorem 45, the continuity of v at 0 implies the non-emptiness of the subdifferential. □

Proof of Theorem 72. By the previous lemma, the subdifferential $\partial v(0)$ of the value function at $y = 0$ contains a linear form $y^* \in \mathcal{Y}$. Then, by the equality case in Fenchel-Young's inequality,

$$\begin{cases} v(0) + v^*(y^*) = 0 \\ v(0) + v^*(z^*) \ge 0 \quad \forall z^* \in \mathcal{Y} \end{cases}.$$

which can be rewritten as $\forall z^* \in \mathcal{Y}^*, \ -v^*(z^*) \le v^*(y)$. Thus,

$$v^{**}(0) = \sup_{z^*\in\mathcal{Y}^*} -v^*(z^*) = -v^*(y^*) = v(0).$$

This shows that $P = D$, and that y^* is a maximizer of the dual problem. □

6.3 Application: Lagrangian duality

In this section, we briefly see how the method of perturbation can be used to recover Lagrangian duality for constrained optimization problems where the constraint set is defined by a family of inequalities. Let $f, g_1, \ldots, g_N \in \Gamma_0(\mathcal{X})$, and consider

$$P = \inf_K f, K = \{x \in \mathcal{X} \mid g_1(x) \le 0, \ldots, g_N(x) \le 0\}.$$

We construct the perturbed problems for $y \in \mathbb{R}^d$ by

$$P_y = \inf\{f(x) \mid g_1(x) \le y_1, \ldots, g_N(x) \le y_N\}.$$

This amounts to introducing the following perturbation function

$$\Phi : \mathcal{X} \times \mathbb{R}^N \to \overline{\mathbb{R}}$$
$$(x, y) \mapsto \begin{cases} f(x) & \text{if } g_i(x) \leq y_i \text{ for } 1 \leq i \leq N \\ +\infty & \text{otherwise} \end{cases}$$

The function Φ can also be expressed as

$$\Phi(x, y) = f(x) + \sum_{1 \leq i \leq N} \mathrm{i}_{\mathrm{epi}(g_i)}(x, y_i),$$

and this expression clearly shows that Φ is convex and lower-semicontinuous. We assume the following condition, similar to (11):

$$\exists x_0 \in \operatorname{dom} f \text{ s.t. } g_i(x_0) < 0 \text{ for } i \in \{1, \ldots, N\}. \tag{15}$$

Thus, $\Phi(x_0, y) = f(x_0)$ as soon as $\|y\|_\infty \leq \min_{1 \leq i \leq N} |g_i(x_0)|$, implying that $\Phi(x_0, y)$ is continuous near $y = 0$. We can therefore apply the strong duality theorem (Theorem 72), which gives us $P = D = \sup_{y^* \in \mathbb{R}^N} -\Phi^*(0, y^*)$. The conjugate $\Phi^*(0, y^*)$ can be computed explicitely in terms of the *Lagrangian* of the problem, i.e.

$$\begin{aligned} L : \mathcal{X} \times \mathbb{R}^N &\to \overline{\mathbb{R}} \\ (x, y) &\mapsto f(x) + \sum_{1 \leq i \leq N} y_i g_i(x), \end{aligned} \tag{16}$$

Lemma 74. *$\Phi^*(0, y^*) = \sup_{x \in \mathcal{X}} L(x, -y^*)$ if $y^* \geq 0$ and $+\infty$ otherwise.*

Proof. By definition,

$$\begin{aligned} \Phi^*(0, y^*) &= \sup_{(x,y) \in \mathcal{X} \times \mathbb{R}^d} \langle y^* | y \rangle - \Phi(x, y) \\ &= \sup\{\langle y^* | y \rangle - f(x) \mid (x, y) \in \mathcal{X} \times \mathbb{R}^d, g_i(x) \leq y_i \forall i\}. \end{aligned}$$

Note that if $y_i^* > 0$ for some i, one can take $x = x_0$ and $y_i^r = r$ (and $y_j^r = 0$ for $j \neq i$), showing that $\Phi^*(0, y^*) = +\infty$. On the other hand, if $y_i^* \leq 0$ for all i, the scalar product $\langle y^* | y \rangle$ is maximized when $y_i = g_i(x)$. Thus, we see that

$$\Phi^*(0, y^*) = \begin{cases} \sup_{x \in \mathcal{X}} \sum_i y_i^* g_i(x) - f(x) & \text{if } y_i^* \leq 0 \\ -\infty & \text{otherwise.} \end{cases}$$

□

This leads to the following expression for the dual problem:

$$D = \sup_{\lambda \in \mathbb{R}_+^N} \inf_{x \in \mathcal{X}} L(x, \lambda).$$

Note that we replaced the variable y by λ to follow the standard notation for Lagrange multipliers.

Theorem 75 (Lagrangian duality)**.** *Let $f, g_1, \ldots, g_N \in \Gamma_0(\mathcal{X})$ satisfying the qualification condition* (15)*, and define the Lagrangian L as in* (16)*. Then,*

$$\inf_{x\in\mathcal{X}} \sup_{\lambda\in\mathbb{R}^N_+} L(x,\lambda) = \max_{\lambda\in\mathbb{R}^N_+} \inf_{x\in\mathcal{X}} L(x,\lambda),$$

if one assumes that the left-hand-side of this equation is finite.

If $\lambda \in \mathbb{R}^N_+$ is a solution of the dual problem, then x is a solution to the primal problem if and only if it satisfies the three Karush-Kuhn-Tucker conditions

- *admissibility: $g_i(x) \leq 0$ for $i \in \{1, \ldots, N\}$*
- *Lagrangian minimization: $x \in \arg\min_{\mathcal{X}} L(\cdot, \lambda)$,*
- *complementary slackness: $\lambda_i g_i(x) = 0$ for $i \in \{1, \ldots, N\}$.*

6.4 Application: Fenchel-Rockafellar' theorem

Theorem 76 (Fenchel-Rockafellar)**.** *Let $\mathcal{X}, \mathcal{Y}$ be two spaces, $g \in \Gamma_0(\mathcal{X})$, $h \in \Gamma_0(\mathcal{Y})$ and $A : \mathcal{X} \to \mathcal{Y}$ be a continuous linear operator. We consider the minimization problem*

$$P = \inf_{x\in\mathcal{X}} g(x) + h(Ax).$$

If P is finite and if the following qualification holds

$$\exists x_0 \in \operatorname{dom} g \text{ s. t. } Ax_0 \in \operatorname{cont} h.$$

Then, $P = D$ where

$$D = \sup_{y^*\in\mathcal{Y}^*} -g^*(A^*y^*) - h^*(-y^*).$$

In addition, the maximum in the dual problem is attained. Moreover, the following statements are equivalent:

(i) x is a minimizer in P and y^* is a maximizer of D ;
(ii) $A^*y^* \in \partial g(x)$ and $-y^* \in \partial h(Ax)$;
(iii) $x \in \partial g^*(A^*y)$ and $Ax \in \partial h^*(-y^*)$.

Remark 28 (Proof by subdifferential calculus). If the minimum in the primal problem P is attained, one can prove this theorem using subdifferential calculus. Assume for simplicity that $\mathcal{Y} = \mathcal{X}$ and $A = \mathrm{Id}$. We first prove the weak duality inequality $P \geq D$: by Fenchel-Young,

$$\forall (y, y^*) \in \mathcal{X} \times \mathcal{X}^*, g(y) + g^*(y^*) \geq \langle y | y^* \rangle$$

$$\forall (y, y^*) \in \mathcal{X} \times \mathcal{X}^*, h(y) + h^*(-y^*) \geq \langle y | -y^* \rangle,$$

thus giving

$$P = \inf_{y\in\mathcal{X}} g(y) + h(y) \geq \sup_{y^*\in\mathcal{X}^*} -g(y^*) - h(-y^*).$$

To prove the converse, we assume that $x \in \arg\min g + h$. Thanks to the qualification hypothesis we can apply Theorem 47 on the subdifferential of the sum, giving

$$0 \in \partial(g + h)(x) = \partial g(x) + \partial h(x).$$

Thus, there exists $x^* \in \partial g(x)$ such that $-x^* \in \partial h(x)$. Therefore, using the equality case in Fenchel-Young (Theorem 70) we get

$$g(x) + g^*(x^*) = \langle x^* | x \rangle, h(x) + h^*(-x^*) = \langle -x^* | x \rangle.$$

Summing these inequalities, we obtain $g(x) + h(x) = -g^*(x^*) - h(-x^*)$, implying the strong duality $P \leq D$.

Example 30 (LASSO). Consider again the LASSO problem

$$\min_{x \in \mathbb{R}^d} \frac{1}{2} \|Ax - b\|_2^2 + \gamma \|x\|_1 ,$$

which is under the desired form by setting $h(y) = \frac{1}{2} \|y - b\|_2^2$ and $g(x) = \gamma \|x\|_1$. To compute the dual problem we need to compute the conjugate functions to f and g:

$$\begin{aligned} g^*(x^*) &= \sup_{x \in \mathbb{R}^d} \langle x^* | x \rangle - \gamma \|x\|_1 \\ &= \sup_{x \in \mathbb{R}^d} \sum_i x_i^* x_i - \gamma \sum_i |x_i| \\ &= \sup_{x \in \mathbb{R}^d} \sum_i x_i (x_i^* - \gamma \operatorname{sgn}(x_i)) \\ &= \begin{cases} 0 & \text{if } \forall i, |x_i^*| \leq \lambda \\ +\infty & \text{otherwise} \end{cases} \\ &= \mathrm{i}_{[-\lambda, \lambda]^d} \end{aligned}$$

Similarly, one can compute h^*:

$$h^*(y^*) = \sup_{y \in \mathbb{R}^d} \langle y^* | y \rangle - \frac{1}{2} \|y - b\|_2^2 = \frac{1}{2} \|y^* + b\|^2 - \|b\|^2 .$$

Thus, the dual problem is

$$D = \max_{x^* \in \mathbb{R}^d} -g^*(A^* y^*) - h^*(-y^*) = \max\{-\frac{1}{2} \|A^* x^* + b\|^2 + \|b\|^2 \mid A^* x^* \in [-\lambda, \lambda]^d\}.$$

The unconstrained non-smooth optimization problem P is tranformed into the constrained smooth optimization problem D.

Example 31 (Rudin-Osher-Fatemi). As a second example, we consider an abstract version of the Rudin-Osher-Fatemi model for image denoising: given some $x_0 \in \mathbb{R}^d$ and a linear operator $D : \mathbb{R}^d \to \mathbb{R}^n$ we consider

$$P = \min_{x \in \mathbb{R}^d} \frac{1}{2} \|x - x_0\|_2^2 + \gamma \|Dx\|_1 .$$

Fenchel-Rockafellar's theorem can be applied by setting $g(x) = \frac{1}{2} \|x - x_0\|^2$, $h(x) = \|\cdot\|_1$ and $A = D$. The dual problem is then

$$D = \max_{y^* \in \mathbb{R}^n} -g^*(D^* y^*) - h^*(-y^*) = \max_{y^* \in [-\gamma, \gamma]^d} -\frac{1}{2} \left\|D^T y^* + x_0\right\|_2^2 + \|x_0\|_2^2 .$$

This problem is much easier than the primal problem, because the optimized function is quadratic and the constraint set is separable (i.e. a product of segments). This problem is, for instance, directly amenable to projected gradient descent. Once a solution y^* to D is found, we know by Theorem 76 that if x^* is a solution of P, then

$$x \in \partial g^*(D^T y^*) = \{\nabla g^*(D^T y^*)\} = \{D^T y^* + x_0\}.$$

This gives us the explicit expression $x = x_0 + D^T y^*$, allowing to recover the primal solution.

Proof of Theorem 76. We introduce $f(x) = g(x) + h(Ax)$ and the perturbation

$$\Phi(x, y) = g(x) + h(Ax - y).$$

The dual problem associated to this perturbation is $D = \sup_{y^*} -\Phi^*(0, y^*)$. To compute $\Phi^*(0, y^*)$, we note that $B(x, y) := (x, Ax-y)$ is an involution, i.e. $B(B(x, y)) = B(x, Ax - y) = (x, Ax - (Ax - y)) = (x, y)$. This gives us

$$\begin{aligned}
\Phi^*(0, y^*) &= \sup_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \langle (0, y^*) | (x, y)\rangle - \Phi(x, y) \\
&= \sup_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \langle (0, y^*) | (x, y)\rangle - (g(x) + h(Ax - y)) \\
&= \sup_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \langle (0, y^*) | B(x, y)\rangle - (g(x) + h(y)) \\
&= \sup_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \langle y^* | Ax - y\rangle - (g(x) + h(y)) \\
&= \sup_{x\in\mathcal{X}} \langle A^* y^* | x\rangle - g(x) + \sup_{y\in\mathcal{Y}} \langle y^* | - y\rangle - h(y) \\
&= g^*(A^* y) + h^*(-y^*).
\end{aligned}$$

Therefore, the dual problem is

$$D = \sup_{y^*} -\Phi^*(0, y^*) = \sup_{y^*} -g^*(A^* y) - h^*(-y^*).$$

By assumption, there exists $x_0 \in \operatorname{dom} g$ such that $Ax_0 \in \operatorname{cont} h$. Thus, $\Phi(x_0, \cdot) = g(x_0) + h(Ax_0 - y)$ is continuous at $y = 0$. By Theorem 72, we get $P = D$ and the existence of a maximizer for D.

Finally we show the equivalence between the three statements using the proposition on weak duality (Proposition 71). Since $P = D$, we get that $x \in \arg\min P$ and $y^* \in \arg\max D$ if and only if

$$\Phi(x, 0) + \Phi^*(0, y^*) = g(x) + h(Ax) + g^*(A^* y) + h^*(-y^*) = 0,$$

if and only if

$$\underbrace{g(x) + g^*(A^* y) - \langle x | A^* y\rangle}_{\geq 0} + \underbrace{h(Ax) + h^*(-y^*) - \langle Ax | - y^*\rangle}_{\geq 0} = 0,$$

where the inequalities are due to Fenchel-Young's inequality. Thus, the sum of the two terms is equal to zero if each of the terms is equal to zero, i.e. if and only if

$$g(x) + g^*(A^*y) = \langle x|A^*y\rangle \text{ and } h(Ax) + h^*(-y^*) = \langle Ax| - y^*\rangle.$$

By Theorem 70, these two equalities holdif and only if $A^*y \in \partial g(x)$ and $-y^* \in \partial h(Ax)$ iff $x \in \partial g^*(A^*y)$ and $Ax \in \partial h(-y^*)$. □

Example 32 (Minimization over a polyhedron). Let A be a $n \times d$ matrix, let $b \in \mathbb{R}^n$ and $g : \mathbb{R}^d \to \mathbb{R}$ be a convex function. We consider the minimization problem

$$P = \inf\{g(x) \mid Ax \leq b\},$$

where the inequality $Ax \leq b$ should be understood coordinatewise, i.e. $(Ax)_i \leq b_i$ for all i. This problem can be recast under the form

$$P = \inf_{x\in\mathbb{R}^d} g(x) + h(Ax)$$

by setting $h = i_L$ where $L = \{y \in \mathbb{R}^n \mid \forall i, y_i \leq b_i\}$. Assume that:

- there exists $x_0 \in \mathbb{R}^d$ such that $Ax_0 \in \operatorname{int} L$. Note that this implies that the polyhedron $K = \{x \in \mathbb{R}^d \mid Ax \leq b\}$ has non-empty interior,
- P is finite.

Then, the assumptions in Fenchel-Rockafellar's theorem are satisfied and we thus get

$$P = D = \max_{y^*\in\mathcal{Y}^*} -g^*(A^*y^*) - h^*(-y^*).$$

Let us compute h^*:

$$h^*(y^*) = \sup_{y\in\mathbb{R}^n} \langle y^*|y\rangle - i_L(y) = \sup_{y\leq b}\langle y^*|y\rangle = \begin{cases} \langle y^*|b\rangle & \text{if } y \leq 0, \\ +\infty & \text{otherwise.} \end{cases}.$$

Thus,

$$D = \max\{-g^*(A^*y^*) + \langle y^*|b\rangle \mid y \in \mathbb{R}^n, y \leq 0\}.$$

If $g(x) = \langle c^*|x\rangle$, P is a linear programming problem, and we recover linear programming duality (under the unnecessary assumption $\operatorname{int} K \neq \emptyset$). Indeed, by Example 26, $g^*(x^*) = i_{c^*}$, so that

$$D = \max\{\langle y^*|b\rangle \mid y \in \mathbb{R}^n, y \leq 0 \text{ and } A^*y^* = c^*\}.$$

7 Exercises

7.1 Chapter 2

Exercise 1. *Convex hull of union.* **1**. Given two convex sets K, L, prove that: $\operatorname{conv}(K \cup L) = \{(1-\alpha)x + \alpha y \mid (x, y, \alpha) \in K \times L \times [0, 1]\}$.

2. Given convex sets $K_1, \ldots, K_n$, prove that

$$\mathrm{conv}(K_1 \cup \ldots \cup K_n) = \left\{ \sum_{1 \leq i \leq n} \alpha_i x_i \mid \alpha \in \Delta_n \text{ and } \forall i,\ x_i \in K_i \right\},$$

where Δ_n is the unit simplex in $\mathbb{R}^n$.

Exercise 2. Prove Lemma 14 when X is a normed space and K contains a ball $\mathrm{B}(0, r)$. Then, prove the same result in the general case of topological vector spaces.

Exercise 3. Prove that the closure of a convex set is convex. Deduce that the closed convex hull of A is equal to the closure of the convex hull of A.

Exercise 4. *Interior of a convex set.* 1. Prove that if $B, C \subseteq K$ are subsets of a convex set K, then $(1-t)B + tC \subseteq K$.

2. Deduce that if x belongs to the interior of K and $y \in K$, then $[x, y)$ lies in the interior of K.
3. Prove that the interior of a convex set is convex.
4. Prove that if K is closed and has non-empty interior, then K is the closure of its interior.

Exercise 5. Let $\mathcal{X} = \mathrm{L}^2([0,1])$ and let $K = \{f \geq 0 \text{ a.e.} \mid f \in \mathcal{X}\}$. Prove that K is convex and closed and therefore sequentially weakly closed. Write explicitely K as an intersection of closed half-spaces.

Exercise 6. *Discontinuous linear form.* Consider $\mathcal{X} = \mathbb{R}[X]$ the space of polynomials, endowed with the sup-norm (if $P = a_n X^n + \ldots + a_0$, then $\|P\| = \max_i |a_i|$), then the linear form $\phi(P) = P'(1)$ is discontinuous everywhere.

Exercise 7. *Mazur's lemma.* Let $(x_n)_{n \in \mathbb{N}}$ be a weakly converging sequence in a normed space $\mathcal{X}$, with weak limit x. Considering the set $K = \mathrm{conv}(\{x_n \mid n \in \mathbb{N}\})$, prove that there exists a sequence y_n of convex combinations of the x_n (i.e. $y_n \in K$) such that y_n converges strongly to x.

Definition 39 (Extreme point)**.** Let $K \subseteq \mathcal{X}$ convex. An *extreme point* is a point $x \in K$ that cannot be obtained by taking a nontrivial convex combination of points in K, i.e. there is no $y, z \in K \setminus \{x\}$ and $\alpha \in (0,1)$ such that $x = (1-\alpha)y + \alpha z$. The set of extremal points of K is denoted $\mathrm{ext}(K)$.

Exercise 8. Let $A \subseteq \mathcal{X}$ and $K = \mathrm{conv}(A)$. Prove that $\mathrm{ext}(K) \subseteq A$.

Exercise 9. *Extreme points and strict convexity of the ball.* We recall that a convex set K is *strictly convex* if

$$\forall x \neq y \in K, \forall \alpha \in (0,1), \quad (1-\alpha)x + \alpha y \in \mathrm{int}(K).$$

1. Prove that the unit ball in a Hilbert space is strictly convex and that all points are extreme.

2. Given $x \in \mathbb{R}^d$, we define $\|x\|_1 = \sum_{1\leq i\leq d} |x_i|$ and $\|u\|_\infty = \max_{1\leq i\leq d} |x_i|$. Are the unit balls of $(\mathbb{R}^d, \|\cdot\|_1)$ and $(\mathbb{R}^d, \|\cdot\|_\infty)$ strictly convex ? What are their extreme points?
3. Same question for $\mathcal{X} = \mathcal{C}^0([0,1])$ endowed with the sup-norm $\|\cdot\|_\infty$ and with $\mathcal{X} = \mathrm{L}^1([0,1])$.

Exercise 10. *Support functions*. Compute the support functions of the following objects:

1. a segment $[a, b]$ in $\mathbb{R}^d$,
2. a square $[0,1]^2$ in $\mathbb{R}^2$,
3. the *unit simplex* $\Delta = \{x \in \mathbb{R}^d_+ \mid \sum_i x_i = 1\}$ in $\mathcal{X} = \mathbb{R}^d$,

Exercise 11. *Hausdorff distance*. Let K, L be two compact convex bodies in $\mathbb{R}^d$, and consider the unit sphere $\mathbb{S}^{d-1} = \{x \in \mathbb{R}^d \mid \|x\| = 1\}$. We denote $\mathrm{d}_H(K, L)$ the Hausdorff distance between K and L, i.e.

$$\mathrm{d}_H(K, L) = \left(\max_{x\in K} \min_{y\in L} \|x - y\|, \max_{y\in L} \min_{x\in K} \|x - y\| \right),$$

1. Prove that $\mathrm{d}_H(K, L) = \min\{\varepsilon \geq 0 \mid K \subseteq L + \mathrm{B}(0, \varepsilon) \text{ and } L \subseteq K + \mathrm{B}(0, \varepsilon)\}$,
2. Prove that $K \subseteq L + \mathrm{B}(0, \varepsilon)$ iff $\sigma_K \leq \sigma_L + \varepsilon$ on $\mathbb{S}^{d-1}$,
3. Conclude that $\mathrm{d}_H(K, L) = \|\sigma_K - \sigma_L\|_{\infty,\mathbb{S}^{d-1}}$.

7.2 Chapter 3

Exercise 12. *Distance functions*. Let $K \subseteq \mathcal{X}$, and define $\mathrm{d}_K(x) = \inf_{p\in K} \|x - p\|$.

1. Prove that if K is convex, then d_K is convex.
2. Prove that if $\mathcal{X}$ is a Hilbert space, then $\|\cdot\|^2 - \mathrm{d}_K^2$ is convex[2] and lsc.
   (nb one does not need to assume that K is convex)

Exercise 13. Suppose that $f : \mathcal{X} \to \mathbb{R}$ is convex and satisfies $f(x) \leq L \|x\|$ for all $x \in \mathcal{X}$. Prove that f is L-Lipschitz.

We recall that a Banach space is a complete normed space. Baire's theorem asserts that if $(F_n)_{n\in\mathbb{N}}$ is a countable family of closed subsets of a Banach space $\mathcal{X}$ (or more generally of a complete metric space), each with empty interior, then $\cup_{n\in\mathbb{N}} F_n$ has empty interior. We use it to deduce a characterization of the continuity set of lsc convex functions.

Exercise 14. Let $f : \mathcal{X} \to \overline{\mathbb{R}}$ be a lower semicontinuous convex function on a Banach space, and assume that $\operatorname{int}\operatorname{dom} f$ contains a point x.

1. Assume that $\mathrm{B}(0, r) \subseteq \operatorname{dom} f$. Letting $F_n = \{y \in \mathrm{B}(0, r) \mid \max(f(y), f(-y)) \leq n\}$, prove that f is bounded on a set with non-empty interior.
2. Conclude that $\operatorname{cont} f = \operatorname{int}\operatorname{dom} f$.

[2] one says that d_K^2 is *semi-concave*

Exercise 15. *Banach-Steinhaus.* Let $(T_\alpha)_{\alpha\in A}$ be a family of continuous linear operators, $T_\alpha : \mathcal{X} \to \mathcal{Y}$ where $\mathcal{X}$ and $\mathcal{Y}$ are Banach spaces, and define $f(x) := \sup_{\alpha\in A} \|T_\alpha(x)\|$.

1. Prove that f is convex and lower-semicontinuous.
2. We now assume that for all $x \in \mathcal{X}$, $\exists M_x \geq 0$ s.t. $\sup_\alpha \|T_\alpha(x)\| < M_x$. Prove that f is continuous on $\mathcal{X}$, and locally Lipschitz near the origin.
3. Deduce the existence of $M \geq 0$ such that $\sup_\alpha \|T_\alpha\| \leq M$.

Exercise 16. *Characterization of support functions.* Let $\sigma : \mathcal{Y} \to \overline{\mathbb{R}}$ be a lower-semicontinuous sublinear (hence convex) function.

1. Let $g(x) = \langle x^*|x\rangle + t$ be an affine minorant of σ. Prove that $g(0) = 0$.
2. Deduce that $\sigma(x) = \sup\{\langle x^*|x\rangle \mid \sigma \geq \langle x^*|\cdot\rangle\}$.
3. Prove that if $\mathcal{Y} = \mathcal{X}^*$ and if $\mathcal{X}^*$ is reflexive, then there exists a convex set $K \subseteq \mathcal{X}$ such that $\sigma = \sigma_K$

7.3 Chapter 4

Exercise 17. *Envelope theorem.* Let $f_i \in \mathcal{C}^1(\mathbb{R}^d)$ be convex functions satisfying

$$\forall i_0 \neq i_1 \in I, \forall x \in \mathbb{R}^n, \nabla f_{i_0}(x) \neq \nabla f_{i_1}(x) \tag{17}$$

We define $f = \max_{i\in I} f_i$ the pointwise maximum of these functions, and we assume that the maximum is attaine at any $x \in \mathbb{R}^d$.

1. Prove that if I is finite, then f is differentiable almost everywhere. *(Hint: invoke the implicit function theorem.)*
2. In the general case, consider the set

$$A = \{x \in \mathbb{R}^n \mid \exists i_0 \neq i_1 \in I,\ f(x) = f_{i_0}(x) = f_{i_1}(x)\}.$$

   Prove that if x belongs to A, then f is not Gâteaux-differentiable at x.
3. Deduce that f that for almost every $x \in \mathbb{R}^d$, f is differentiable at x and there exists a unique $i_x \in I$ s.t. $f(x) = f_{i_x}(x)$ and $\nabla f(x) = \nabla f_{i_x}(x)$.

Exercise 18. *Simplex.* Let $f : \mathbb{R}^d \to \mathbb{R}$ be defined by $f(x_1, \ldots, x_d) = \max(x_1, \ldots, x_d)$. Prove that $\partial f(0) = \{x \in \mathbb{R}^d \mid x_1 + \ldots + x_d = 1 \text{ and } \forall i, x_i \geq 0\}$.

Exercise 19. *Failure of subdifferential sum rule.* Let $A = B((0,0),1)$, $B = B((0,2),1)$ bet closed balls in $\mathbb{R}^2$, and $f = \mathrm{i}_A$, $g = \mathrm{i}_B$. Compute the subdifferentials $\partial f(x), \partial g(x)$ and $\partial(f+g)(x)$ at $x = (0,1)$.

Exercise 20. *Characterization of convexity.* Let $f : \mathcal{X} \to \overline{\mathbb{R}}$ be a proper function so that for all $x \in \operatorname{dom} f$, the subdifferential $\partial f(x)$ is non-empty.

1. Prove that f is equal to the supremum of its affine minorants.
2. Deduce that f is convex lsc.
3. Conversely, prove that if $f : \mathbb{R}^d \to \mathbb{R}$ is convex then $\partial f(x) \neq \emptyset$ for all $x \in \mathbb{R}^d$

Exercise 21. *Subdifferential sum rule.* Prove using support functions (see Remark 21) that $\partial(f+g)(x) = \partial f(x) + \partial g(x)$ when $x \in \operatorname{cont} f \cap \operatorname{cont} g$.

Exercise 22. *Exact penalization using distance*. Let K be a closed non empty convex subset of a Hilbert space $\mathcal{X}$. We will consider subdifferentials, normal cones, etc. as subsets of $\mathcal{X} \simeq \mathcal{X}^*$.

1. Show that $\partial i_K(p) = \{v \in E; \forall q \in K,\ \langle v|p\rangle \geq \langle v|q\rangle\}$. Using the characterization of the orthogonal projection on K, prove that
$$\partial i_K(p) = \{v \in E;\quad \mathrm{p}_K(p+v) = p\}. \tag{18}$$
2. *Application:* Let $f : \mathcal{X} \to \mathbb{R}$ be continuous. Prove that $\bar{p} \in K$ if a minimizer of
$$\min_{p \in K} f(p) \tag{19}$$
iff there exists $w \in \partial f(\bar{p})$ such that $\mathrm{p}_K(\bar{p} - w) = \bar{p}$. (*Indication: use the subdifferential sum rule.*)
3. We now prove that if $p \in K$ and $v \in \mathrm{Nor}_x K = \partial i_K(p)$ then $v/\|v\| \in \partial \mathrm{d}_K(p)$, or equivalently that
$$\forall x \in E,\ \mathrm{d}_K(x) \geq \mathrm{d}_K(p) + \langle \frac{v}{\|v\|}|x - p\rangle. \tag{20}$$
   a. Prove that (20) holds if $\langle v|x\rangle \leq \langle v|p\rangle$.
   b. Let $x \notin H := \{x \in E; \langle v|p\rangle \geq \langle v|x\rangle\}$, and define $x_H = \mathrm{p}_H(x)$. Prove that $x_H = x - \langle \frac{v}{\|v\|}|x - p\rangle \frac{v}{\|v\|}$.
   c. Using $K \subseteq H$, prove that $\mathrm{d}_K(x) \geq \mathrm{d}_H(x) = \|x - x_H\| = \langle \frac{v}{\|v\|}|x - p\rangle$, and deduce that $v/\|v\| \in \partial \mathrm{d}_K$.
4. *Application:* Let $\bar{p} \in K$ be a minimizer of (19).
   a. Prove that there exists $w \in \partial f(\bar{p})$ such that $-w \in \partial i_K(\bar{p})$, so that
$$0 \in \partial(f + \|w\|\, \mathrm{d}_K)(\bar{p}).$$
   b. Deduce the existence of $\lambda \geq 0$ such that $\bar{p}$ minimizes the penalized problem
$$\min_{p \in \mathcal{X}} f(p) + \lambda \mathrm{d}_K(p).$$

Exercise 23. *Limiting subdifferential, Exam 2020*. Let $f \in \mathcal{C}^0(\mathbb{R}^n)$ be convex.

1. Fix a point $x \in \mathbb{R}^n$, and define $K = \overline{\mathrm{conv}} S$ where
$$S = \left\{s \in \mathbb{R}^n \mid \exists x_n \to x, \text{ s.t. } f \text{ is differentiable at } x_n \text{ and } \lim_{n\to\infty} \nabla f(x_n) = s\right\}.$$
   a. Prove that $S \subseteq \partial f(x)$ and $K \subseteq \partial f(x)$.
   The goal of the next questions is to prove the converse inclusion.
   b. Fix some vector $v \in \mathbb{R}^n$. Prove that for all $t_n = 1/n$, there exists $v_n \in \mathbb{R}^n$ such that $\|v_n - v\| \leq t_n$ and such that f is differentiable at $x_n = x + t_n v_n$.
   c. Prove that, taking a subsequence if necessary, one can assume that $\nabla f(x_n)$ converges to a vector $s \in S$. Show that $f^+(x, v) \leq \langle s|v\rangle$.
   d. Deduce that $f^+(x, v) \leq \sigma_K(v)$ where σ_K is the support function of K. Conclude that $\partial f(x) \subseteq K$.
2. *Application.* Assume there exists $G \in \mathcal{C}^0(\mathbb{R}^d, \mathbb{R}^d)$ such that $\forall x \in \mathbb{R}^n, G(x) \in \partial f(x)$. Prove that f belongs to $\mathcal{C}^1(\mathbb{R}^n)$ and that $\nabla f = G$.

subdifferential of TV norm

7.4 Chapter 5

Exercise 24. Let $\mathcal{X} = \mathcal{X}_1 \oplus \mathcal{X}_2$, where $\mathcal{X}_1, \mathcal{X}_2$ are two closed subspaces, let $f_i \in \Gamma_0(\mathcal{X}_i)$ and $f = f_1 \oplus f_2$, i.e. $f : \mathcal{X} \to \overline{\mathbb{R}}$ is defined by

$$\forall (x_1, x_2) \in \mathcal{X}_1 \oplus \mathcal{X}_2 \mapsto f_1(x_1) + f_2(x_2).$$

Prove that $\mathrm{Prox}_{\gamma f}(x_1 + x_2) = \mathrm{Prox}_{\gamma f_1}(x_1) + \mathrm{Prox}_{\gamma f_2}(x_1)$.

Exercise 25. Let $f \in \Gamma_0(\mathcal{X})$ be coercive, let $T = \mathrm{Prox}_f$ and define $x_{n+1} = T(x_n)$.

1. Prove that $\forall x \in \mathcal{X},\ f(x) \geq f(x_{n+1}) + \frac{1}{\gamma}\langle x - x_{n+1} | x_n - x_{n+1}\rangle$.
2. Using this inequality,
   - (i) prove that $(f(x_n))_{n\in\mathbb{N}}$ is decreasing and that $(x_n)_{n\in\mathbb{N}}$ is bounded;
   - (ii) prove that any weak cluster point of $(x_n)_{n\in\mathbb{N}}$ minimizes f globally;
     (Hint: use that $\lim_{n\to+\infty} \|x_{n+1} - x_n\| = 0$*, as in Theorem 67.)*
3. Deduce that $(x_n)_{n\in\mathbb{N}}$ is a minimizing sequence.
4. Conclude that if f is strictly convex, the sequence $(x_n)_{n\in\mathbb{N}}$ converges weakly to the unique minimizer of f.

7.5 Chapter 6

Exercise 26. *Convex cones*. Let K be a non-empty convex cone (i.e. for all $x \in K$ and $\lambda \geq 0$, $\lambda x \in K$) and let $K^* \subseteq \mathcal{X}^*$ be its polar of K

$$K^* = \{x^* \in \mathcal{X}^* \mid \forall x \in K, \langle x^* | x\rangle \leq 0\}.$$

1. Prove that if $f = \mathrm{i}_K$, then $f^* = \mathrm{i}_{K^*}$ and that $\partial f(0) = K^*$.
2. Prove that $K^{**} = \{x \in \mathcal{X} \mid \forall x^* \in K^*, \langle x^* | x\rangle \leq 0\}$ is closed, convex, and contains K.
3. Prove that if K is a closed convex cone, then $K^{**} = K$.
4. Assume that $\mathcal{X}$ is a Hilbert space, let K be a closed convex cone, $K^* \subseteq \mathcal{X} \simeq \mathcal{X}^*$ be its polar. Prove that

$$\forall x \in \mathcal{X}, \quad x = \mathrm{proj}_K(x) + \mathrm{proj}_{K^*}(x).$$

Exercise 27. *Strong convexity and subdifferential*. Let $\mathcal{X}$ be a Hilbert space. A function $f : \mathcal{X} \to \overline{\mathbb{R}}$ is called α-strongly convex if it satisfies one the following equivalent conditions:

$$\forall x, y \in \mathrm{dom}(f), \forall \lambda \in [0,1],$$
$$\frac{\alpha}{2}\lambda(1-\lambda)\|x - y\|^2 + f((1-\lambda)x + \lambda y) \leq (1-\lambda)f(x) + \lambda f(y) \tag{21}$$
$$\forall x_0 \in \mathcal{X},\ f_{x_0} : x \in \mathcal{X} \mapsto f - \frac{\alpha}{2}\|x - x_0\|^2 \text{ is convex} \tag{22}$$

1. Assume that $f \in \Gamma_0(\mathcal{X})$ is α-strongly convex and let $v \in \partial f(x_0)$. Prove that

$$0 \in \partial\left(f - \frac{\alpha}{2}\|\cdot - x_0\|^2 - \langle v|\cdot\rangle\right)(x_0),$$

   and deduce that $\forall x \in \mathcal{X}, \quad f(x) \geq f(x_0) + \langle v | x - x_0\rangle + \frac{\alpha}{2}\|x - x_0\|^2$.

2. Prove that for all $x_0, y_0 \in \mathcal{X}$, $v \in \partial f(x_0)$ and $\forall w \in \partial f(y_0)$ one has

$$\alpha \|x_0 - y_0\| \leq \|v - w\| .$$

3. Deduce that the same inequality holds if $x_0 \in \partial f^*(v)$ and $y_0 \in \partial f^*(w)$.
4. Let $D = \{v \in E \mid \partial f^*(v) \neq \emptyset\}$. Prove that f^* is Gâteaux-differentiable on D, and that the application $v \in D \mapsto \nabla f^*(v)$ is $(1/\alpha)$-Lipschitz.

Exercise 28. *Moreau-Yosida regularization*. Let $\mathcal{X}$ be a Hilbert space and $f \in \Gamma_0(\mathcal{X})$. Given $\tau > 0$, define the Moreau-Yosida regularization of f as

$$f_\tau(x) = \inf_{z \in E} f(z) + \frac{1}{2\tau} \|x - z\|^2$$

1. Prove that f_τ is convex, finite everywhere, and bounded on every bounded subset of $\mathcal{X}$.
2. Prove that f_τ is continuous and that $\partial f_\tau(x) \neq \emptyset$ for all $x \in \mathcal{X}$.
3. Prove that $f_\tau = (f^* + g^*)^*$ where $g(x) = \frac{1}{2\tau} \|x\|^2$.
4. Using the previous exercise, deduce that f_τ is Gâteaux-differentiable at all $x \in \mathcal{X}$ and that ∇f_τ is τ-Lipschitz.
5. Prove that f_τ converges pointwise to f as $\tau \to 0$.
6. Prove that $\mathrm{Prox}_{\tau f}(x) = x - \tau \nabla f_\tau(x)$.
   (Hint: Let $p = \mathrm{Prox}_{\tau f}(x)$, and prove that $\frac{x-p}{\tau} \in \partial f_\tau(x)$ using the equality case in Fenchel-Young's inequality.)

This last question shows that the proximal point algorithm can be interpreted as "usual" gradient descent for the Moreau-Yosida regularization.

For the next exercise, recall that a *measure* on a compact set X is a linear form $\mu : \mathcal{C}^0(X) \to \mathbb{R}$ which is continous for the topology induced by the sup-norm $\|\cdot\|_\infty$. The space of measures is denoted

$$\mathcal{M}(X) = (\mathcal{C}^0(X))^*.$$

A measure μ is *non-negative* if $\forall f \in \mathcal{C}^0(X), (f \geq 0 \implies \mu(f) \geq 0)$. Finally, a probability measure on X is a non-negative measure μ such that $\|\mu\|_{TV} = 1$. The space of non-negative measures is denoted $\mathcal{M}_+(X)$, and the space of probability measures is $\mathcal{P}(X)$.

Exercise 29. *Kantorovich Duality*. Let X, Y be two compact sets, let μ be a probability measure on X, let ν be a probability measure on Y and finally let $c \in \mathcal{C}^0(X \times Y)$. Define a linear form $\Lambda : \mathcal{C}^0(X) \times \mathcal{C}^0(Y) \to \mathcal{C}^0(X \times Y)$ by $\Lambda(\phi, \psi) = \phi \oplus \psi$ where

$$\phi \oplus \psi : (x, y) \in X \times Y \mapsto \phi(x) + \psi(y).$$

We consider the following optimization problem

$$P := \inf\{-(\langle \mu | \phi \rangle + \langle \nu | \psi \rangle) \mid (\phi, \psi) \in \mathcal{C}^0(X) \times \mathcal{C}^0(Y), \phi \oplus \psi \leq c\}.$$

1. Prove that $P = \inf_{(\phi,\psi)\in\mathcal{C}^0(X)\times\mathcal{C}^0(Y)} f(\Lambda(\phi,\psi)) + g(\phi,\psi)$, where

$$f : \sigma \in \mathcal{C}^0(X\times Y) \mapsto = \begin{cases} 0 & \text{if } \sigma \leq c \\ +\infty & \text{otherwise,} \end{cases}$$
$$g : (\phi,\psi) \in E \times F \mapsto -(\langle\mu|\phi\rangle + \langle\nu|\psi\rangle).$$

2. Prove that f^* and g^* are given by:

$$f^* : \gamma \in \mathcal{M}(X\times Y) : \gamma \mapsto = \begin{cases} \langle\gamma|c\rangle & \text{if } \gamma \in \mathcal{M}_+(X\times Y) \\ +\infty & \text{otherwise} \end{cases}$$

$$g^* : (\kappa,\lambda) \in \mathcal{M}(X)\times\mathcal{M}(Y) \mapsto \begin{cases} 0 & \text{if } (\kappa,\lambda) = -(\mu,\nu) \\ +\infty & \text{otherwise} \end{cases}$$

3. Let $\gamma \in \mathcal{M}(X\times Y)$. The *marginal of* γ *on* X is the measure $\Pi_X\gamma \in \mathcal{M}(X)$ defined by $\Pi_X\gamma : \phi \in \mathcal{C}^0(X) = \gamma(\phi\oplus 0)$. The marginal on Y is defined similarly. Prove that the adjoint Λ^* of γ is given by

$$\Lambda^*(\gamma) = (\Pi_X\gamma, \Pi_Y\gamma) \in \mathcal{M}(X)\times\mathcal{M}(Y).$$

4. Deduce from Fenchel-Rockafellar that $P = D$ where

$$\begin{aligned} D &:= \max_{\gamma\in\mathcal{C}^0(X\times Y)^*} -g^*(-\Lambda^*\gamma) - f^*(\gamma) \\ &= \max\{-\langle\gamma|c\rangle \mid \gamma \in \mathcal{M}_+(X\times Y) \text{ s.t. } \Pi_X\gamma = \nu \text{ and } \Pi_Y\gamma = \nu\}. \end{aligned}$$

This duality formula is due to Leonid Kantorovich, and is one of the most important results in the theory of optimal transport.

References

[AAC92] Giovanni Alberti, Luigi Ambrosio, and Piermarco Cannarsa. On the singularities of convex functions. *Manuscripta Math*, 76(3-4):421–435, 1992.

[AB86] Hédy Attouch and Haïm Brezis. Duality for the sum of convex functions in general banach spaces. In *North-Holland Mathematical Library*, volume 34, pages 125–133. Elsevier, 1986.

[BC+11] Heinz H Bauschke, Patrick L Combettes, et al. *Convex analysis and monotone operator theory in Hilbert spaces*, volume 408. Springer, 2011.

[Bre10] Haim Brezis. *Functional analysis, Sobolev spaces and partial differential equations*. Springer Science & Business Media, 2010.

[BV+10] Jonathan M Borwein, Jon D Vanderwerff, et al. *Convex functions: constructions, characterizations and counterexamples*, volume 172. Cambridge University Press Cambridge, 2010.

[CS04] Piermarco Cannarsa and Carlo Sinestrari. *Semiconcave functions, Hamilton-Jacobi equations, and optimal control*, volume 58. Springer Science & Business Media, 2004.

[ET99] Ivar Ekeland and Roger Temam. *Convex analysis and variational problems.* SIAM, 1999.

[Hör55] Lars Hörmander. Sur la fonction d'appui des ensembles convexes dans un espace localement convexe. *Arkiv för matematik*, 3(2):181–186, 1955.

[Kap11] Samuel Kaplan. *The Bidual of $C(X)$.* Elsevier, 2011.

[M+95] Robert J McCann et al. Existence and uniqueness of monotone measure-preserving maps. *Duke Mathematical Journal*, 80(2):309–324, 1995.

[Mar70] B Martinet. Régularisation d'inéquations variationnelles par approximations successives. rev. française informat. *Recherche Opérationnelle*, 4:154–158, 1970.

[Mar72] Bernard Martinet. Détermination approchée d'un point fixe d'une application pseudo-contractante. *CR Acad. Sci. Paris*, 274(2):163–165, 1972.

[Roc70] R Tyrrell Rockafellar. *Convex analysis.* Number 28. Princeton university press, 1970.

[Roc76] R Tyrrell Rockafellar. Monotone operators and the proximal point algorithm. *SIAM journal on control and optimization*, 14(5):877–898, 1976.

[Zal02] Constantin Zalinescu. *Convex analysis in general vector spaces.* World scientific, 2002.