Convex Analysis

Juan Peypouquet

February 24, 2023

Foreword

This is a draft version of the Lecture Notes for the Convex Analysis course WMMA060-05 from the MSc programme in (Applied) Mathematics at the University of Groningen. The contents have been adapted from [3]. Remarks welcome at `j.g.peypouquet@rug.nl`

Contents

Chapter 1

Functional analysis with a convex flavor

This chapter contains the basic tools from functional analysis that will be needed throughout this course. The concepts are presented from a geometric perspective, when possible. We focus on Hilbert spaces to avoid technicalities, although many results can be extended to more general settings.

1.1 Normed spaces

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

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

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

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

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

Example 1.1. The following expressions define norms in $\mathbf{R}^N$:

1. $\|x\|_\infty = \max\limits_{i=1,\ldots,N} |x_i|$;

2. $\|x\|_p = \left(\sum\limits_{i=1}^N |x_i|^p \right)^{1/p}$, for $p \geq 1$.

Exercise 1.1. Using the norms in the previous example as a model, define the norms $\|x\|_\infty$ and $\|x\|_p$ in the space of real sequences.

Example 1.2. The following expressions define norms in the space vector $\mathscr{C}([a,b];\mathbf{R})$ of all continuous real-valued functions defined on the interval $[a,b]$:

1. $\|f\|_\infty = \max\limits_{t \in [a,b]} |f(t)|$;

2. $\|f\|_p = \left(\int_a^b |f(t)|^p \, dt \right)^{1/p}$, for $p \geq 1$.

Let $x \in X$ and $r > 0$. The *closed ball* of radius r centered at x is the set

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

and the *open ball* is

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

A set is *bounded* if there is a ball that contains it.

A set C is *closed* if, for every $x \notin C$, there is $r > 0$ such that $C \cap B(x,r) = \emptyset$. The complements of closed sets are the *open* sets.

Exercise 1.2. Show that every closed ball is a closed set, and that every open ball is an open set.

We say that a sequence (x_n) in X *converges strongly* (or, simply, *converges*) to $\bar{x} \in X$, and write $x_n \to \bar{x}$, as $n \to \infty$ if $\lim\limits_{n\to\infty} \|x_n - \bar{x}\| = 0$. The point $\bar{x}$ is the *(strong) limit* of the sequence.

A set C is *sequentially closed* if every convergent sequence of elements of C has its limit in C.

Exercise 1.3. Prove that C is closed if, and only if, it is sequentially closed.

A sequence (x_n) has the *Cauchy property*, or it is a *Cauchy sequence*, if $\lim\limits_{n,m\to\infty} \|x_n - x_m\| = 0$. Every convergent sequence has the Cauchy property, and every Cauchy sequence is bounded. A *Banach space* is a normed space in which every Cauchy sequence is convergent, a property called *completeness*.

Example 1.3. From the completeness of $\mathbf{R}$, it follows that $\mathbf{R}^N$ is a Banach space with any of the norms defined in Example 1.1.

Example 1.4. From the norms defined in Example 1.2, only $\|\cdot\|_\infty$ makes $\mathscr{C}([a,b];\mathbf{R})$ a Banach space.

We have the following result:

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

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

We shall find several important consequences of this result, especially the Banach-Steinhaus Uniform Boundedness Principle (Theorem 1.8) and the continuity of convex functions in the interior of their domain (Proposition 2.30).

1.2 Bounded linear operators

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

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

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

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

i) L is continuous in 0;

ii) L is bounded; and

iii) L is uniformly Lipschitz-continuous in X.

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

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

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

We have the following:

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

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

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

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

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

$$\operatorname{ran}(L) = L(X) = \{Lx \,:\, x \in X\}.$$

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

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

Theorem 1.8 (Banach-Steinhaus Uniform Boundedness Principle)**.** *Let $(L_\lambda)_{\lambda\in\Lambda}$ be a family of bounded linear operators from a Banach space $(X, \|\cdot\|_X)$ to a normed space $(Y, \|\cdot\|_Y)$. If*

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

for each $x \in X$, then

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

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

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

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

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

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

1.3 Hilbert spaces

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

1.3.1 Basic concepts, properties and examples

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

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

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

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

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

Proposition 1.9. *For each $x, y \in H$ we have*

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

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

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

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

Therefore,

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

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

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

whence ii) holds. □

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

Example 1.10. The following are Hilbert spaces:

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

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

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

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

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

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

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

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

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

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

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

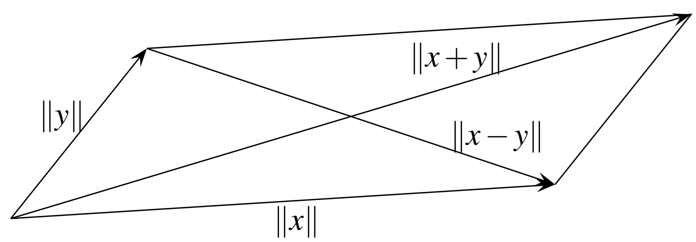
ii) The *colinearity condition*: $x \,\|\, y$ if, and only if, $x = \lambda y$ with $\lambda \in \mathbf{R}$.

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

$$\|x+y\|^2+\|x-y\|^2=2\left(\|x\|^2+\|y\|^2\right) \tag{1.1}$$

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

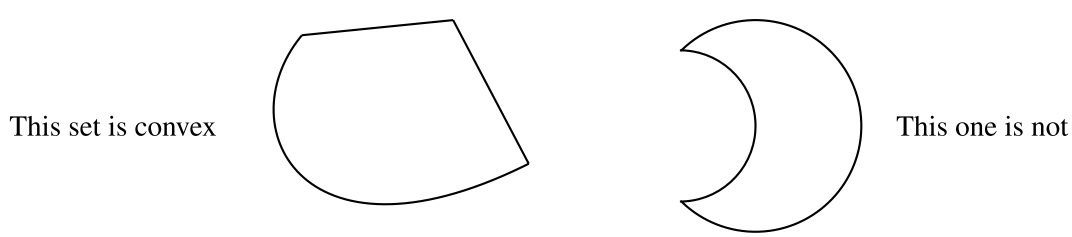
$$\|x+y\|^2=\|x\|^2+\|y\|^2+2\langle x,y\rangle \quad \text{and} \quad \|x-y\|^2=\|x\|^2+\|y\|^2-2\langle x,y\rangle.$$



Example 1.11. The space $X=\mathscr{C}([0,1];\mathbf{R})$ with the norm $\|x\|=\max_{t\in[0,1]}|x(t)|$ is not a Hilbert space. Consider the functions $x,y\in X$, defined by $x(t)=1$ and $y(t)=t$ for $t\in[0,1]$. We have $\|x\|=1$, $\|y\|=1$, $\|x+y\|=2$ and $\|x-y\|=1$. The parallelogram identity (1.1) does not hold. $\square$

1.4 Convex sets

A subset C of a real vector space X is *convex* if $\lambda x+(1-\lambda)y\in C$ whenever $x,y\in C$ and $\lambda\in[0,1]$. In other words, if the segment joining any two points of C also belongs to C.



1.4.1 Some algebraic properties

In this section, X,X_1,X_2 are real vector spaces.

Proposition 1.12. *The intersection of any (possibly infinite) collection of convex sets is convex.*

Proof. Let $C=\cap\{C_j : j\in\mathscr{J}\}$, where C_j is convex for each $j\in\mathscr{J}$. Now, take $x,y\in C$ and $\lambda\in[0,1]$. Then $x,y\in C_j$ for each $j\in\mathscr{J}$. By the convexity of C_j, $\lambda x+(1-\lambda)y\in C_j$. Since this is true for each C_j, we conclude that $\lambda x+(1-\lambda)y\in C$. $\square$

Exercise 1.4. When is the union of convex sets convex?

Proposition 1.13. *Let $C_1, C_2 \subset X$ be convex, and let $\alpha \in \mathbf{R}$. Then, the set*

$$\alpha C_1 + C_2 = \{\, \alpha x_1 + x_2 \ : \ x_1 \in C_1,\ x_2 \in C_2 \,\}$$

is convex.

Proof. Let $C = \alpha C_1 + C_2$, and let $x, y \in C$, and let $\lambda \in [0,1]$. Then, there exist $x_1, y_1 \in C_1$ and $x_2, y_2 \in C_2$ such that $x = \alpha x_1 + x_2$ and $y = \alpha y_1 + y_2$. Therefore,

$$\lambda x + (1-\lambda)y = \lambda(\alpha x_1 + x_2) + (1-\lambda)(\alpha y_1 + y_2) = \alpha\big(\lambda x_1 + (1-\lambda)y_1\big) + \big(\lambda x_2 + (1-\lambda)y_2\big).$$

By convexity, $\lambda x_1 + (1-\lambda)y_1 \in C_1$ and $\lambda x_2 + (1-\lambda)y_2 \in C_2$, and so $\lambda x + (1-\lambda)y \in C$. □

Exercise 1.5. Let B_1 and B_2 be open (respectively, closed) balls, and let $\lambda \in (0,1)$. Prove that the set $B = \lambda B_1 + (1-\lambda)B_2$ is a ball. Compute its center and radius.

Exercise 1.6. Let $C_1 \subset X_1$ and $C_2 \subset X_2$ be convex, and let $L : X_1 \to X_2$ be linear. Show that the sets $L(C_1)$ and $L^{-1}(C_2)$ are convex.

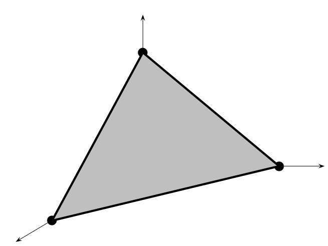
The *convex hull* of a set $S \subset X$ is

$$\operatorname{conv}(S) = \left\{ \sum_{j=1}^{J} \lambda_j x_j \ : \ x_1, \ldots, x_J \in S, \quad \lambda_j \in [0,1] \quad \text{and} \quad \sum_{j=1}^{J} \lambda_j = 1 \right\}.$$

The elements of $\operatorname{conv}(S)$ are the *convex combinations* of the elements of S.

Example 1.14. Given $x, y \in X$, $\operatorname{conv}\big(\{x,y\}\big)$ is the *segment* joining x and y.

Example 1.15. Let $e_1, \ldots, e_N$ denote the canonical vectors of $\mathbf{R}^N$. The set $\operatorname{conv}\big(\{e_1, \ldots, e_N\}\big)$ is usually known as the $(N-1)$-*dimensional simplex*. The picture on the right shows the 2-dimensional simplex in $\mathbf{R}^3$.



Proposition 1.16. *Let $S \subset X$. The set* $\operatorname{conv}(S)$ *is convex and contains S. It is the intersection of all convex sets containing S. In particular, S is convex if, and only if, $S = $* $\operatorname{conv}(S)$.

Proof. It is clear that $S \subset \operatorname{conv}(S)$. Now, let us take $z_1, z_2 \in \operatorname{conv}(S)$ and $\nu \in [0,1]$, and prove that $\mu z_1 + (1-\nu)z_2 \in \operatorname{conv}(S)$. Since $z_1, z_2 \in \operatorname{conv}(S)$, there exist $x_1, y_1, x_2, y_2 \in S$ and $\lambda_1, \lambda_2 \in [0,1]$ such that $z_i = \lambda_i x_i + (1-\lambda_i)y_i$, for $i = 1,2$. Therefore,

$$\mu z_1 + (1-\nu)z_2 = \mu\lambda_1 x_1 + \mu(1-\lambda_1)y_1 + (1-\mu)\lambda_2 x_2 + (1-\mu)(1-\lambda_2)y_2.$$

This point belongs to $\operatorname{conv}(S)$ because

$$\mu\lambda_1 + \mu(1-\lambda_1) + (1-\mu)\lambda_2 + (1-\mu)(1-\lambda_2) = \mu\big(\lambda_1 + 1 - \lambda_1\big) + (1-\mu)\big(\lambda_2 + 1 - \lambda_2\big) = 1,$$

and $x_1, y_1, x_2, y_2 \in S$. □

Theorem 1.17 (Carathéodory's Theorem). *Let $S \subset \mathbf{R}^N$. Every element in* $\mathrm{conv}(S)$ *can be expressed as a convex combination of at most $N+1$ elements of S. That is, for each $x \in \mathrm{conv}(S)$ there exist $x_i \in S$ and $\lambda_i \in [0,1]$, for $i \in \{1,2,\ldots,N+1\}$, such that*

$$x = \sum_{i=1}^{N+1} \lambda_i x_i.$$

Proof. Let $x \in \mathrm{conv}(S)$, and let m be the smallest positive integer such that x can be written as

$$x = \sum_{i=1}^{m} \lambda_i x_i.$$

It follows that $\lambda_i > 0$ for $i = 1,\ldots,m$. We embed $\mathbf{R}^N$ into $\mathbf{R}^{N+1}$ by $z \mapsto (z,1)$. If the vectors $(x_1,1),\ldots,(x_m,1)$ are linearly independent in $\mathbf{R}^{N+1}$, then necessarily $m \le N+1$. Suppose they are not, so there exist $\alpha_1,\ldots,\alpha_m$, with at least one of them positive, such that

$$\sum_{i=1}^{m} \alpha_i x_i = 0.$$

Set

$$\gamma = \min\left\{ \frac{\lambda_i}{\alpha_i} \ : \ \alpha_i > 0 \right\} \qquad \text{and} \qquad \sigma = \sum_{i=1}^{m} (\lambda_i - \gamma\alpha_i),$$

and observe that all the terms in the sum are nonnegative, by the definition of γ. Relabeling the indices if necessary, we may assume that $\gamma = \frac{\lambda_m}{\alpha_m}$. Hence, we can write

$$x = \sum_{i=1}^{m} \left(\frac{\lambda_i - \gamma\alpha_i}{\sigma} \right) x_i = \sum_{i=1}^{m-1} \left(\frac{\lambda_i - \gamma\alpha_i}{\sigma} \right) x_i,$$

because the last coefficient is zero. This contradicts the definition of m and yields the result. □

1.4.2 Basic topological properties, interior, closure

We now discuss some topological properties of convex sets, for which we return to the context of a Hilbert space H.

Proposition 1.18. *If $C \subset H$ is convex, so are* $\mathrm{int}(C)$ *and* $\overline{C}$.

Proof. Let $x_1, x_2 \in \mathrm{int}(C)$ and let $\lambda \in [0,1]$. We shall prove that $x = \lambda x_1 + (1-\lambda)x_2 \in \mathrm{int}(C)$. Since x_1 and x_2 are interior points, there is $\varepsilon > 0$ such that $B(x_i,\varepsilon) \subset C$ for $i = 1,2$. We shall prove that $B(x,\varepsilon) \subset C$, by showing that $x + \varepsilon p \in C$ for every $p \in B(0,1)$. We have

$$x + \varepsilon p = \lambda x_1 + (1-\lambda)x_2 + \lambda\varepsilon p + (1-\lambda)\varepsilon p = \lambda(x_1 + \varepsilon p) + (1-\lambda)(x_2 + \varepsilon p).$$

But $x_i + \varepsilon p \in B(x_i,\varepsilon) \subset C$ for $i = 1,2$. By the convexity of C, $x + \varepsilon p \in C$, and we conclude that $x \in \mathrm{int}(C)$. For the closure, if $x,y \in \overline{C}$, there exist sequences (x_n) and (y_n) in C such that $x_n \to x$ and $y_n \to y$. For each n, we have $\lambda x_n + (1-\lambda)y_n \in C$, so that, passing to the limit, we conclude that $\lambda x + (1-\lambda)y \in \overline{C}$. □

Proposition 1.19. *Let $C \subset H$ be convex, let $x \in \text{int}(C)$, and let $y \in \overline{C}$. Then, $\lambda x + (1-\lambda)y \in \text{int}(C)$ for every $\lambda \in (0,1]$.*

Proof. Take $\lambda \in (0,1]$ and write $z = \lambda x + (1-\lambda)y$. Since $x \in \text{int}(C)$, there is $\rho > 0$ such that $B(x,\rho) \subset C$. On the other hand, since $y \in \overline{C}$, for every $\varepsilon > 0$, there is $p_\varepsilon \in H$ such that $\|p_\varepsilon\| < \varepsilon$ and $y + p_\varepsilon \in C$. We shall pick an appropriate value for ε in a moment. Our purpose is to find $\delta > 0$ such that $z + \delta p \in C$ for all $p \in B(0,1)$. We have

$$z + \delta p = \lambda x + (1-\lambda)y + \delta p = \lambda \left[x + \frac{1-\lambda}{\lambda} p_\varepsilon + \frac{\delta}{\lambda} p \right] + (1-\lambda)(y + p_\varepsilon).$$

Since $y + p_\varepsilon \in C$ and C is convex, it suffices to select ε and δ so that

$$x + \frac{1-\lambda}{\lambda} p_\varepsilon + \frac{\delta}{\lambda} p \in B(x,\rho) \subset C$$

(hence, $z + \delta p$ will be a convex combination of elements of C). In other words, it suffices that

$$\left\| x + \frac{1-\lambda}{\lambda} p_\varepsilon + \frac{\delta}{\lambda} p - x \right\| < \frac{\varepsilon(1-\lambda)}{\lambda} + \frac{\delta}{\lambda} \leq \rho.$$

A suitable choice is $\delta = \varepsilon \leq \frac{\lambda\rho}{2-\lambda}$. $\square$

Exercise 1.7. Let C be a convex subset of H with nonempty interior. Show that $\overline{C} = \overline{\text{int}(C)}$ and $\text{int}(C) = \text{int}(\overline{C})$.

1.4.3 Projection and orthogonality

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

Proposition 1.20. *Let K be a nonempty, closed and convex subset of H and let $x \in H$. Then, there exists a unique point $y^* \in K$ such that*

$$\|x - y^*\| = \min_{y \in K} \|x - y\|. \tag{1.2}$$

Moreover, it is the only element of K such that

$$\langle x - y^*, y - y^* \rangle \leq 0 \quad \textit{for all} \quad y \in K. \tag{1.3}$$

Proof. We shall prove Proposition 1.20 in three steps: first, we verify that (1.2) has a solution; next, we establish the equivalence between (1.2) and (1.3); and finally, we check that (1.3) cannot have more than one solution.

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

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

by virtue of the parallelogram identity (1.1). Since K is convex, the midpoint between y_n and y_m belongs to K. Therefore,

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

according to the definition of d. We deduce that

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

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

Next, assume (1.2) holds and let $y\in K$. Since K is convex, for each $\lambda\in(0,1)$ the point $\lambda y+(1-\lambda)y^*$ also belongs to K. Therefore,

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

This implies

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

Letting $\lambda\to 0$ we obtain (1.3). Conversely, if (1.3) holds, then

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

for each $y\in K$ and (1.2) holds.

Finally, if $y_1^*,y_2^*\in K$ satisfy (1.3), then

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

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

The point y^* given by Proposition 1.20 is the *projection* of x onto K and will be denoted by $P_K(x)$. The characterization of $P_K(x)$ given by (1.3) says that for each $x\notin K$, the set K lies in the *closed half-space*

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

Corollary 1.21. *Let K be a nonempty, closed and convex subset of H. Then $K=\cap_{x\notin K}\{y\in H : \langle x-P_K(x),y-P_K(x)\rangle\}$.*

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

For subspaces we recover the idea of *orthogonal projection*:

Proposition 1.22. *Let M be a closed subspace of H. Then,*

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

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

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

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

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

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

Proposition 1.23. *Let K be a nonempty, closed and convex subset of H. The function $x \mapsto P_K(x)$ is nonexpansive.*

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

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

We conclude using the Cauchy-Schwarz inequality. □

Theorem 1.24 (Riesz's Representation Theorem)**.** *For each $\ell \in \mathscr{L}(H;\mathbf{R})$ there is a unique $y_\ell \in H$ such that*

$$\ell(h) = \langle y_\ell, h\rangle$$

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

Proof. Let $M = \ker(\ell)$, which is a closed subspace of H because ℓ is linear and continuous. If $M = H$, then $\ell(h) = 0$ for all $h \in H$ and we can take $y_\ell = 0$. If $M \neq H$, pick any $x \notin M$ and define

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

Notice that $\hat{x} \neq 0$ and $\hat{x} \notin M$. Given any $h \in H$, set $u_h = \ell(\hat{x})h - \ell(h)\hat{x}$, so that $u_h \in M$. By Proposition 1.22, we have $\langle \hat{x}, u_h\rangle = 0$. In other words,

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

The vector

$$y_\ell = \frac{\ell(\hat{x})}{\|\hat{x}\|^2}\,\hat{x}$$

has the desired property. It is straightforward to verify that $\ell \mapsto y_\ell$ is a linear isometry. □

1.5 Hahn-Banach Theorem(s)

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

A *hyperplane* is a subset of H of the form

$$H_{y=\alpha} := \{x \in H : \langle y, x\rangle = \alpha\},$$

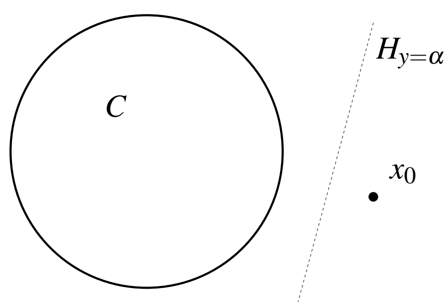
where $y \in H$ and $\alpha \in \mathbf{R}$. Each hyperplane defines open and closed *half-spaces*, which are sets of the form

$$H_{y<\alpha} := \{x \in H : \langle y, x\rangle < \alpha\} \qquad \text{and} \qquad H_{y\leq\alpha} := \{x \in H : \langle y, x\rangle \leq \alpha\},$$

respectively. The halfspaces $H_{y>\alpha}$ and $H_{y\geq\alpha}$ are defined analogously.

1.5.1 Strict separation using the projection

We first show that a closed convex set can be separated from any exterior point by a hyperplane.



Proposition 1.25. *Let C be nonempty, closed and convex. If $x_0 \notin C$, there exist $z \in H \setminus \{0\}$ and $\varepsilon > 0$ such that $\langle z, x\rangle + \varepsilon \leq \langle z, x_0\rangle$ for each $x \in C$.*

Proof. By Proposition 1.20, since C is nonempty, closed and convex, we can project x_0 onto C. The projection $p = P_C(x_0)$ is characterized by $\langle x - p, x_0 - p\rangle \leq 0$ for all $x \in C$. It follows that

$$\langle x - x_0, x_0 - p\rangle + \|x_0 - p\|^2 \leq 0,$$

which we can rewrite as

$$\langle z, x\rangle + \|z\|^2 \leq \langle z, x_0\rangle,$$

by defining $z = x_0 - p$. It suffices to set $\varepsilon = \|z\|^2$, which is positive because $p \in C$ and $x_0 \notin C$. $\square$

Remark 1.26. Proposition 1.25 implies $\sup_{x\in C}\langle z, x\rangle < \langle z, x_0\rangle$. If $x_0 = 0$, this is $\sup_{x\in C}\langle z, x\rangle < 0$

Remark 1.27. Proposition 1.25 also provides the existence of a hyperplane $H_{y=\alpha}$, which is disjoint from C and $\{x_0\}$, and such that $x_0 \in H_{y<\alpha}$ and $C \subset H_{y>\alpha}$.

Exercise 1.8. Let A and B be nonempty, disjoint convex subsets of H. Prove that if A is compact and B is closed, there exist $z \in H \setminus \{0\}$ and $\varepsilon > 0$ such that $\langle y, x\rangle + \varepsilon \leq \langle z, y\rangle$ for each $x \in A$ and $y \in B$. In this setting, comment on Remarks 1.26 and 1.27. Hint: Write $C = B - A$.

1.5.2 The finite dimensional case

In finite dimensional spaces, no topological assumptions are required.

Proposition 1.28. *Given $N \geq 1$, let C be a nonempty and convex subset of $\mathbf{R}^N$ not containing the origin. Then, there exists $v \in \mathbf{R}^N \setminus \{0\}$ such that $v \cdot x \leq 0$ for each $x \in C$. In particular, if $N \geq 2$ and C is open, then*

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

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

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

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

Therefore,

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

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

Exercise 1.9. State a similar result replacing the origin by an arbitrary point $x_0 \notin C$.

Exercise 1.10. In line with Exercise 1.8, what can you say about the separation of two convex sets?

Exercise 1.11. Why is it difficult to extend the proof of Proposition 1.28 to infinite-dimensional spaces?

1.5.3 Non-strict separation

Separation can still be achieved for disjoint convex sets whose closures intersect, although a topological assumption is required.

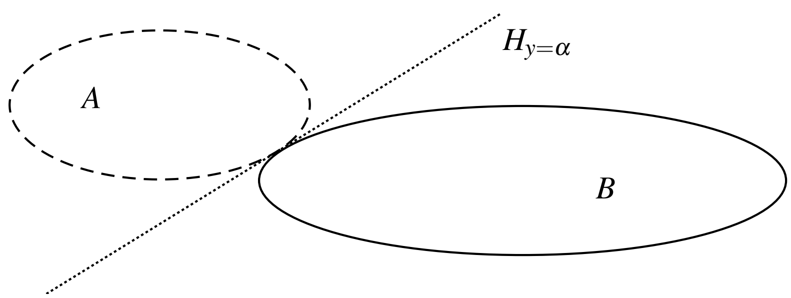
Proposition 1.29. *Let C be nonempty, open and convex. If $0 \notin C$, there is $z \in H \setminus \{0\}$ such that $\langle z, x \rangle < 0$ for each $x \in C$.*

Proof. The proof is achieved in three steps:

First, if the dimension of H is at least 2, there is a nontrivial subspace of H not intersecting C. Take any two-dimensional subspace Y of H. If $Y \cap C = \emptyset$ there is nothing to prove. Otherwise, identify Y with $\mathbf{R}^2$ and use Proposition 1.28 to obtain a subspace of Y disjoint from $Y \cap C$, which clearly gives a subspace of H not intersecting C.

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

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



Exercise 1.12. Let A and B be nonempty, disjoint convex subsets of H. Show that, if A is open, there exists $z \in H \setminus \{0\}$ such that $\langle z, x\rangle < \langle z, y\rangle$ for each $x \in A$ and $y \in B$. Hint: See Exercise 1.8.

Following Exercises 1.8 and 1.12, we can summarize the discussion above as follows:

Theorem 1.30 (Hahn-Banach Separation Theorem)**.** *Let A and B be nonempty, disjoint convex subsets of H.*

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

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

1.6 Weak convergence, sequential closedness and compactness

In this section, we introduce the concept of weak convergence and explore some of its properties.

1.6.1 Motivation: Weierstrass's Theorem

We begin by recalling the following theorem, which is usually discussed in undergraduate calculus.

Theorem 1.31 (Weierstrass's Theorem)**.** *If $f : [a, b] \to \mathbf{R}$ is continuous, then it attains its maximum and its minimum.*

Proof. Let (x_n) be a sequence in $[a,b]$ such that $\lim_{n\to\infty} f(x_n) = \inf_{x\in[a,b]} f(x)$. Such a sequence exists in view of the definition of infimum. Since $[a,b]$ is bounded, so is (x_n). Therefore, it has a subsequence (x_{k_n}), which converges to some $\hat{x}$, which must belong to $[a,b]$, because this set is closed. By the continuity of f, we have

$$f(\hat{x}) = \lim_{n\to\infty} f(x_n) = \inf_{x\in[a,b]} f(x).$$

Therefore $f(\hat{x}) \le f(x)$ for all $x \in [a,b]$. For the maximum, we apply the same procedure to $-f$. □

We used three main arguments in the preceding proof:

1. The fact that $[a,b]$ is bounded allowed us to invoke the Bolzano-Weierstrass property, which states that *every bounded real sequence has a convergent subsequence*.

2. The closedness of $[a,b]$ implies that the limit belongs to $[a,b]$.

3. From the continuity of f, it follows that the value at the limit coincides with the limit of the values.

We shall analyze these properties in closer detail, in order to extend Weierstrass's Theorem to more general spaces. The tools that we will develop will also play an important role in approximation techniques that are at the core of numerical methods used in optimization, (optimal) control, (partial) differential equations and variational analysis.

The Bolzano-Weierstrass property is easily extended to $\mathbf{R}^N$, but not to infinite dimensional spaces.

Example 1.32. In $H = \ell^2(\mathbf{N};\mathbf{R})$, given $n \ge 1$, let (x_n) be the sequence whose i-th component is 1 if $i = n$ and is 0 otherwise. Since $\|x_n\| = 1$ for all n, the sequence (x_n) is bounded. Now, given that

$$\|x_n - x_m\|^2 = \sum_{i\in\mathbf{N}} |x_n(i) - x_m(i)|^2 = 2,$$

whenever $n \neq m$, it cannot have any convergent subsequence.

In the next part, We shall present a different notion of convergence, for which the Bolzano-Weierstrass property is true.

1.6.2 Weakly convergent sequences

We say that a sequence (x_n) in H *converges weakly* to $\bar{x}$, and write $x_n \rightharpoonup \bar{x}$, as $n \to \infty$ if

$$\lim_{n\to\infty} \langle x_n - \bar{x}, y\rangle = 0$$

for all $y \in H$. The point $\bar{x}$ is the *weak limit* of the sequence.

Since $|\langle x_n - \bar{x}, y\rangle| \le \|y\|\,\|x_n - \bar{x}\|$, convergent sequences are weakly convergent and the limits coincide. However, not every weakly convergence sequence is convergent.

Example 1.33. The sequence (x_n) in Example 1.32 is not convergent. However, for each $y \in H$, we have

$$\langle y, x_n \rangle = y(n) \to 0$$

as $n \to \infty$. Therefore, x_n converges weakly to 0 as $n \to \infty$. This example also shows that the origin is a weak sequential limit point of the unit sphere in $\ell^2(\mathbf{N};\mathbf{R})$, which, as a consequence, is not weakly sequentially closed.

The following result contains relevant basic properties of weakly convergent sequences.

Proposition 1.34. *Let (x_n) converge weakly to $\bar{x}$ as $n \to \infty$. Then:*

i) *(x_n) is bounded.*

ii) $\|\bar{x}\| \le \liminf\limits_{n\to\infty} \|x_n\|$.

iii) *If* $\limsup\limits_{n\to\infty} \|x_n\| \le \|\bar{x}\|$, *then (x_n) converges to $\bar{x}$.*

iv) *If (y_n) converges to $\bar{y}$, then* $\lim\limits_{n\to\infty} \langle y_n, x_n \rangle = \langle \bar{y}, \bar{x} \rangle$.

Proof. For i), define the sequence (μ_n) in $\mathscr{L}(H;\mathbf{R})$ by $\mu_n(y) = \langle x_n, y \rangle$, for $y \in H$. Since $\lim\limits_{n\to\infty} \mu_n(y) = \langle y, \bar{x} \rangle$, we have $\sup\limits_{n\in\mathbf{N}} \mu_n(y) < +\infty$ for all $y \in H$. The Banach-Steinhaus Uniform Boundedness Principle (Theorem 1.8) implies $\sup\limits_{n\in\mathbf{N}} \|x_n\| = \sup\limits_{n\in\mathbf{N}} \|\mu_n\|_{\mathscr{L}(H;\mathbf{R})} < +\infty$. Part ii) is trivial if $\bar{x} = 0$. Otherwise, write

$$\|x\|^2 = \langle \bar{x}, x - x_n \rangle + \langle \bar{x}, x_n \rangle \le \langle \bar{x}, x - x_n \rangle + \|\bar{x}\|\,\|x_n\|,$$

and let $n \to \infty$. For iii), notice that

$$0 \le \limsup_{n\to\infty} \|x_n - \bar{x}\|^2 = \limsup_{n\to\infty} \left[\|x_n\|^2 + \|\bar{x}\|^2 - 2\langle x_n, \bar{x} \rangle\right] \le 0.$$

Finally, since (x_n) is bounded,

$$\begin{array}{rcl} |\langle y_n, x_n \rangle - \langle \bar{y}, \bar{x} \rangle| & \le & |\langle y_n - \bar{y}, x_n \rangle| + |\langle \bar{y}, x_n - \bar{x} \rangle| \\ & \le & C\,\|y_n - \bar{y}\| + |\langle \bar{y}, x_n - \bar{x} \rangle| \end{array}$$

for some $C > 0$. As $n \to \infty$, we obtain iv). □

In Hilbert spaces, the Bolzano-Weierstrass property holds for the weak convergence.

Theorem 1.35. *Every bounded sequence in H has a weakly convergent subsequence.*

We will add the proof later. In the meantime, the interested reader may consult [2, Chapters II and V] for full detail, or [1, Chapter 3] for abridged commentaries.

For approximation purposes, it is important to relate properties of sequences with properties of their limits. The concept of closedness plays a fundamental role in this.

1.6.3 Weak closedness

Using the definition of weakly convergent sequences, we can introduce a different notion of closedness. A set $C \subset X$ is *weakly sequentially closed* if every weakly convergent sequence of points in C has its weak limit in C. Since every convergent sequence is weakly convergent, every weakly sequentially closed set is sequentially closed and, therefore, closed. The converse is true for convex sets.

Proposition 1.36. *If $C \subset H$ is closed and convex, it is weakly sequentially closed.*

Proof. Let (x_n) be a sequence in C, such that $x_n \rightharpoonup \bar{x}$, as $n \to \infty$. We shall prove that $\bar{x} \in C$. Suppose, otherwise, that $\bar{x} \notin C$. Since C is closed and convex, by Proposition 1.25 (strict separation), there exist $z \in H$ and $\varepsilon > 0$, such that

$$\langle z, y \rangle + \varepsilon \leq \langle z, \bar{x} \rangle$$

for all $y \in C$. In particular,

$$\langle z, x_n \rangle + \varepsilon \leq \langle z, \bar{x} \rangle$$

for all n. Letting $n \to \infty$, we obtain $\varepsilon \leq 0$, which is a contradiction. □

Proposition 1.36 and Theorem 1.35 together give

Corollary 1.37. *Let $C \subset H$ be closed, convex and bounded. Every sequence in C has a subsequence that converges weakly to a point in C.*

Another useful consequence is the following result, whose proof is left as an exercise.

Corollary 1.38. *Let (y_n) be a bounded sequence in H. If every weakly convergent subsequence has the same weak limit $\hat{y}$, then (y_n) must converge weakly to $\hat{y}$ as $n \to \infty$.*

The *weak topology* is the one generated by the half-spaces. In other words, a set $A \subset H$ is *weakly open* if, for every point $x \in A$ there exist vectors $y_1, \ldots, y_N$ in H, and real numbers $\alpha_1, \ldots, \alpha_N$, such that

$$x \in \bigcap_{i=1}^{N} H_{y_i < \alpha_i} \subset A.$$

Weakly closed sets are the complements of weakly open sets.

Exercise 1.13. Show that if H is finite-dimensional, every weakly open set is open, and viceversa. In other words, the weak topology coincides with the usual topology given by the norm.

Exercise 1.14. Prove that, if H is infinite-dimensional, the open ball $B(0,1)$ is *not* weakly open. Moreover, show that no bounded set can be weakly open.

Exercise 1.15. Prove that the weak topology has the *Hausdorff property*, which means that every two distinct points admit disjoint neighborhoods.

Exercise 1.16. Show that a sequence (y_n) is weakly convergent if, and only if, it is convergent for the weak topology.

Exercise 1.17. Consider the following statements concerning a nonempty set $C \subset X$:

i) C is weakly closed.

ii) C is weakly sequentially closed.

iii) C is sequentially closed.

iv) C is closed.

Show that $i) \Rightarrow ii) \Rightarrow iii) \Leftrightarrow iv) \Leftarrow i)$, and that the four statements are equivalent if C is convex.

Chapter 2

The theory of functions

In all that follows, H is a real Hilbert space. In order to deal with unconstrained and constrained optimization problems in a unified setting, we introduce the *extended real numbers* by adding the *symbol* $+\infty$. The convention $\gamma < +\infty$ for all $\gamma \in \mathbf{R}$ allows us to extend the total order of $\mathbf{R}$ to a total order of $\mathbf{R} \cup \{+\infty\}$. With this convention, we can define functions, defined on a set, with *values* in $\mathbf{R} \cup \{+\infty\}$. The simplest example is the *indicator function* of a set $C \subset H$, defined as

$$\delta_C(x) = \begin{cases} 0 & \text{if } x \in C \\ +\infty & \text{otherwise.} \end{cases}$$

The main interest of introducing these kinds of functions is that, clearly, if $f : H \to \mathbf{R}$, the optimization problems

$$\min\{f(x) : x \in C\} \qquad \text{and} \qquad \min\{f(x) + \delta_C(x) : x \in H\}$$

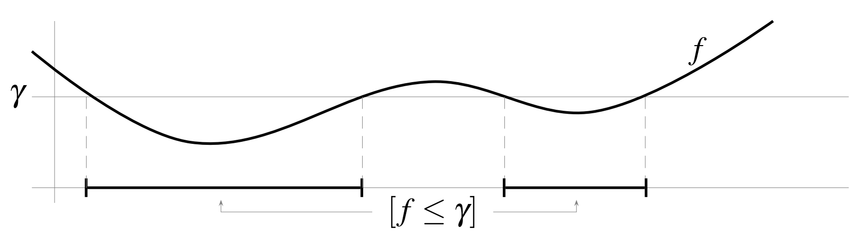
are equivalent. The advantage of the second formulation is that linear, geometric or topological properties of the underlying space H may be exploited.

Let $f : H \to \mathbf{R} \cup \{+\infty\}$ be an extended real-valued function. The *effective domain* (or, simply, the *domain*) of f is the set of points where f is finite. In other words,

$$\mathrm{dom}(f) = \{x \in H \ : \ f(x) < +\infty\}.$$

A function f is *proper* if $\mathrm{dom}(f) \neq \emptyset$. The alternative, namely $f \equiv +\infty$, is not very interesting for the purpose of this course. Given $\gamma \in \mathbf{R}$, the γ-*sublevel set* of f is

$$[f \leq \gamma] = \{x \in H : f(x) \leq \gamma\}.$$



If $x \in \mathrm{dom}(f)$, then $x \in [f \leq f(x)]$. Therefore,

$$\mathrm{dom}(f) = \bigcup_{\gamma \in \mathbf{R}} [f \leq \gamma].$$

On the other hand, recall that

$$\mathrm{argmin}(f) = \{y \in H \ : \ f(y) \leq f(x) \text{ for all } x \in H\}.$$

Observe that

$$\mathrm{argmin}(f) = \bigcap_{\gamma > \inf(f)} [f \leq \gamma]. \tag{2.1}$$

Finally, the *epigraph* of f is the subset of the product space $H \times \mathbf{R}$ defined as

$$\mathrm{epi}(f) = \{(x, \alpha) \in H \times \mathbf{R} \ : \ f(x) \leq \alpha\}.$$

This set includes the graph of f and all the points above it.

2.1 Convex functions

An extended real valued function $f : H \to \mathbf{R} \cup \{+\infty\}$ is *convex* if

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y) \tag{2.2}$$

for each $x, y \in \mathrm{dom}(f)$ and $\lambda \in (0,1)$. Notice that inequality (2.2) holds trivially if $\lambda \in \{0,1\}$ or if either x or y are not in $\mathrm{dom}(f)$.

Exercise 2.1. Prove that $f : H \to \mathbf{R} \cup \{+\infty\}$ is convex if, and only if, $\mathrm{epi}(f)$ is a convex subset of $H \times \mathbf{R}$.

Example 2.1. The indicator function δ_C of a convex set C is a convex function.

Exercise 2.2. Show that if $f : H \to \mathbf{R} \cup \{+\infty\}$ is convex, then each sublevel set $\Gamma_\gamma(f)$ is convex. Provide a counterexample for the converse.

Functions whose level sets are convex are called *quasi-convex*. If the inequality in (2.2) is strict whenever $x \neq y$ and $\lambda \in (0,1)$, we say f is *strictly convex*. Moreover, f is *strongly convex* with parameter $\mu > 0$ (or also μ*-strongly convex*) if

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y) - \frac{\mu}{2}\lambda(1-\lambda)\|x-y\|^2$$

for each $x, y \in \mathrm{dom}(f)$ and $\lambda \in (0,1)$. Clearly, strongly convex functions are strictly convex.

Exercise 2.3. Prove that the graph of a strictly convex function cannot contain a segment. Conclude that a strictly convex function cannot have more than one minimizer.

Exercise 2.4. Show that $f : H \to \mathbf{R} \cup \{+\infty\}$ is μ-strongly monotone if, and only if, the function $g : H \to \mathbf{R} \cup \{+\infty\}$, defined by $g(x) = f(x) - \frac{\mu}{2}\|x\|^2$, is convex.

We shall provide some practical characterizations of convexity for differentiable functions in Subsection 2.3.2.

Example 2.2. Let $B : H \times H \to \mathbf{R}$ be a bilinear function and define $f(x) = B(x,x)$. For each $x, y \in H$ and $\lambda \in (0,1)$, we have

$$f(\lambda x + (1-\lambda)y) = \lambda f(x) + (1-\lambda)f(y) - \lambda(1-\lambda)B(x-y, x-y).$$

Therefore, we have the following:

i) f is convex if, and only if, $B(z,z) \geq 0$ for all $z \in H$;

ii) f is strictly convex if, and only if, $B(z,z) > 0$ for all $z \neq 0$; and

iii) f is μ-strongly convex if, and only if, $B(z,z) \geq \frac{\mu}{2}\|z\|^2$ for all $z \in H$.

In particular, if $A : H \to H$ is linear, the function $f : H \to \mathbf{R}$, defined by $f(x) = \langle Ax, x\rangle$, is convex if, and only if, A is positive semidefinite; strictly convex if, and only if, A is positive definite; and μ-strongly convex if, and only if, A is uniformly elliptic with parameter μ.

Some operations allow us to construct convex functions from others. Here we mention a few examples.

Example 2.3. Suppose $A : X \to Y$ is affine, $f : Y \to \mathbf{R} \cup \{+\infty\}$ is convex and $\theta : \mathbf{R} \to \mathbf{R} \cup \{+\infty\}$ convex and nondecreasing. Then, the function $g = \theta \circ f \circ A : X \to \mathbf{R} \cup \{+\infty\}$ is convex.

Example 2.4. If f and g are convex and if $\alpha \geq 0$, then $f + \alpha g$ is convex. It follows that the set of convex functions is a convex cone.

Example 2.5. If $(f_i)_{i \in I}$ is a family of convex functions, then $\sup(f_i)$ is convex, since $\mathrm{epi}(\sup(f_i)) = \cap \mathrm{epi}(f_i)$ and the intersection of convex sets is convex.

Example 2.6. In general, the infimum of convex functions need not be convex. However, we have the following: Let V be a vector space and let $g : H \times V \to \mathbf{R} \cup \{+\infty\}$ be convex. Then, the function $f : H \to \mathbf{R} \cup \{+\infty\}$ defined by $f(x) = \inf_{v \in V} g(x,v)$ is convex. Here, the facts that g is convex in the product space and the infimum is taken over a whole vector space are crucial.

2.2 Continuity of convex functions

The continuity results presented below reveal how remarkably regular convex functions are.

Lemma 2.7. *Let $f : H \to \mathbf{R} \cup \{+\infty\}$ be convex and fix $x_0 \in H$. If f is bounded from above in a neighborhood of x_0, then $x_0 \in \mathrm{int}(\mathrm{dom}(f))$ and f is Lipschitz-continuous in a neighborhood of x_0. In particular, f is continuous at x_0.*

Proof. There exist $r > 0$ and $K > f(x_0)$ such that $f(z) \leq K$ for every $z \in B(x_0, 2r)$. We shall find $M > 0$ such that $|f(x) - f(y)| \leq M\|x-y\|$ for all $x, y \in B(x_0, r)$. First, consider the point

$$\tilde{y} = y + r\frac{y-x}{\|y-x\|}. \tag{2.3}$$

Since $\|\tilde{y} - x_0\| \leq \|y - x_0\| + r < 2r$, we have $f(\tilde{y}) \leq K$. Solving for y in (2.3), we see that

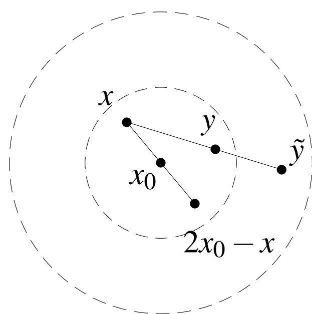
$$y = \lambda \tilde{y} + (1-\lambda)x, \quad \text{with} \quad \lambda = \frac{\|y-x\|}{\|y-x\|+r} \leq \frac{\|y-x\|}{r}.$$

The convexity of f implies

$$f(y) - f(x) \leq \lambda[f(\tilde{y}) - f(x)] \leq \lambda[K - f(x)]. \tag{2.4}$$

Write $x_0 = \frac{1}{2}x + \frac{1}{2}(2x_0 - x)$. We have $f(x_0) \leq \frac{1}{2}f(x) + \frac{1}{2}f(2x_0 - x)$ and

$$-f(x) \leq K - 2f(x_0). \tag{2.5}$$



Combining (2.4) and (2.5), we deduce that

$$f(y) - f(x) \leq 2\lambda[K - f(x_0)] \leq \frac{2(K - f(x_0))}{r}\|x-y\|.$$

Interchanging the roles of x and y we conclude that $|f(x) - f(y)| \leq M\|x-y\|$ with $M = \frac{2(K-f(x_0))}{r}$, which is positive. $\square$

As a consequence, we obtain:

Corollary 2.8. *Let $f : H \to \mathbf{R} \cup \{+\infty\}$ be convex. If f is continuous at $x_0 \in H$, then $x_0 \in$ $\operatorname{int}(\operatorname{dom}(f))$, and f is Lipschitz-continuous in a neighborhood of x_0.*

Exercise 2.5. Let $f : H \to \mathbf{R} \cup \{+\infty\}$ be convex. Show that f is continuous at x_0 if, and only if, $(x_0, \lambda) \in \operatorname{int}(\operatorname{epi}(f))$ for each $\lambda > f(x_0)$.

2.2.1 Transportation of continuity

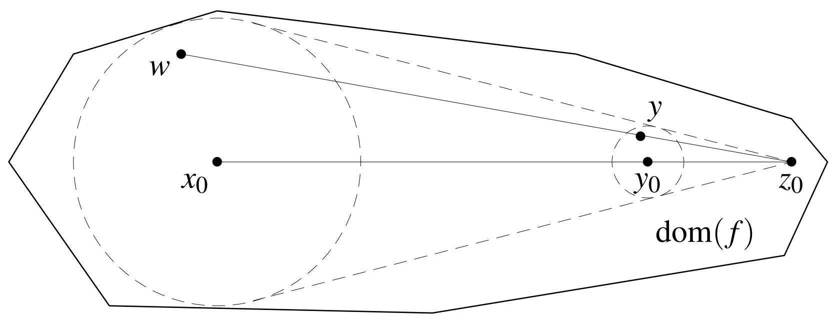
Some local properties turn out to be global in the presence of convexity. For example, we have the following:

Proposition 2.9. *Let $f : H \to \mathbf{R} \cup \{+\infty\}$ be convex. If f is continuous at some $x_0 \in H$, then f is continuous in* $\operatorname{int}(\operatorname{dom}(f))$.

Proof. If f is continuous at x_0, then $x_0 \in \operatorname{int}(\operatorname{dom}(f))$ and there is $r > 0$ such that $f(x) \leq f(x_0) + 1$ for every $x \in B(x_0, r)$. Let $y_0 \in \operatorname{int}(\operatorname{dom}(f))$ and pick $\rho > 0$ such that the point $z_0 = y_0 + \rho(y_0 - x_0)$ belongs to $\operatorname{dom}(f)$. Take $y \in B(y_0, \frac{\rho r}{1+\rho})$ and set $w = x_0 + (\frac{1+\rho}{\rho})(y - y_0)$. Solving for y, we see that

$$y = \left(\frac{\rho}{1+\rho}\right) w + \left(\frac{1}{1+\rho}\right) z_0.$$

On the other hand, $\|w - x_0\| = \left(\frac{1+\rho}{\rho}\right) \|y - y_0\| < r$, and so $w \in B(x_0, r)$ y $f(w) \leq f(x_0) + 1$.



Therefore,

$$f(y) \leq \left(\tfrac{\rho}{1+\rho}\right) f(w) + \left(\tfrac{1}{1+\rho}\right) f(z_0) \leq \max\{f(x_0) + 1, f(z_0)\}.$$

Since this is true for each $y \in B(y_0, \frac{\rho r}{1+\rho})$, we conclude that f is bounded from above in a neighborhood of y_0. Lemma 2.7 implies f is continuous at y_0. □

An immediate consequence is:

Corollary 2.10. *Let $f : H \to \mathbf{R} \cup \{+\infty\}$ be convex. If $[f \leq \gamma]$ has nonempty interior for some $\gamma \in \mathbf{R}$, then f is continuous on* $\operatorname{int}(\operatorname{dom}(f))$.

Proposition 2.9 reveals that a convex function is either continuous in the interior of its domain, or nowhere.

Example 2.11. Let Ω be a bounded domain in $\mathbf{R}^N$, and let $H = L^2(\Omega; \mathbf{R})$. The function $f : H \to \mathbf{R} \cup \{+\infty\}$, defined by

$$f(u) = \int \|\nabla u(\xi)\|^2 \, d\xi,$$

is convex and nowhere continuous.

Proposition 2.9 requires that f be continuous at some point. This hypothesis can be waived in finite-dimensional spaces:

Proposition 2.12. *Let H be finite dimensional and let $f : H \to \mathbf{R} \cup \{+\infty\}$ be convex. Then f is continuous on* $\mathrm{int}(\mathrm{dom}(f))$.

Proof. Let $\{e_1, \ldots, e_N\}$ generate H. Let $x_0 \in \mathrm{int}(\mathrm{dom}(f))$ and take $\rho > 0$ small enough so that $x_0 \pm \rho e_i \in \mathrm{dom}(f)$ for all $i = 1, \ldots, N$. The convex hull C of these points is a neighborhood of x_0 and f is bounded by $\max_{i=1,\ldots,N}\{f(x_0 \pm \rho e_i)\}$ on C. The result follows from Lemma 2.7. □

2.2.2 Convexity and directional derivatives

The *directional derivative* of a function $f : H \to \mathbf{R} \cup \{+\infty\}$ at a point $x \in \mathrm{dom}(f)$ in the direction $h \in H$ is given by

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

whenever this limit exists.

Remark 2.13. If f is convex, a simple computation shows that the quotient $\frac{f(x+th)-f(x)}{t}$ is nondecreasing as a function of t. We deduce that

$$f'(x;h) = \inf_{t>0} \frac{f(x+th) - f(x)}{t},$$

which *exists* in $\mathbf{R} \cup \{-\infty, +\infty\}$.

We have the following:

Proposition 2.14. *Let $f : H \to \mathbf{R} \cup \{+\infty\}$ be proper and convex, and let $x \in$* $\mathrm{dom}(f)$. *Then, the function $\phi_x : H \to [-\infty, +\infty]$, defined by $\phi_x(h) = f'(x;h)$, is convex. If, moreover, f is continuous at x, then ϕ_x is finite and continuous in H.*

Proof. Take $y, z \in X$, $\lambda \in (0,1)$ and $t > 0$. Write $h = \lambda y + (1-\lambda)z$. By the convexity of f, we have

$$\frac{f(x+th) - f(x)}{t} \leq \lambda \frac{f(x+ty) - f(x)}{t} + (1-\lambda) \frac{f(x+tz) - f(x)}{t}.$$

Passing to the limit, $\phi_x(\lambda y + (1-\lambda)z) \leq \lambda \phi_x(y) + (1-\lambda)\phi_x(z)$. Now, if f is continuous in x, it is Lipschitz-continuous in a neighborhood of x, by Corollary 2.8. Then, for all $h \in H$, all sufficiently small $t > 0$ and some $L > 0$, we have

$$-L\|h\| \leq \frac{f(x+th) - f(x)}{t} \leq L\|h\|.$$

It follows that $\mathrm{dom}(\phi_x) = H$ and that ϕ_x is bounded from above in a neighborhood of 0. Using Corollary 2.8 again, we deduce that ϕ_x is continuous in 0, and, by Proposition 2.9, ϕ_x is continuous in $\mathrm{int}(\mathrm{dom}(\phi_x)) = H$. □

2.3 Convexity and gradients

A function $f : H \to \mathbf{R} \cup \{+\infty\}$ is *differentiable (in the sense of Gâteaux)* at $x \in \operatorname{int}\big(\operatorname{dom}(f)\big)$ if the directional derivative

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

exists for all $h \in H$, and there is $g \in H$ such that

$$\langle g, h \rangle = f'(x;h)$$

for all $h \in H$. If this case, the *gradient* of f at x is $\nabla f(x) = g$. As usual, f is *differentiable* on a set A if it is so at every point of A.

This is a weak form of differentiability, which does not imply the continuity of f at x.

Example 2.15. Define $f : \mathbf{R}^2 \to \mathbf{R}$ by

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

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

Exercise 2.6. How is this definition of gradient connected with the partial derivatives studied in undergraduate calculus courses?

Remark 2.16. A stronger notion of differentiability is that of Fréchet, which means that there exists $G \in H$ such that

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

If f is differentiable at x in the sense of Fréchet, then it is continuous there, differentiable in the sense of Gâteaux, and $\nabla f(x) = G$. The reader may want to prove this as an exercise.

Example 2.17. Let $B : H \times H \to \mathbf{R}$ be a bilinear function:

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

for all $x,y,z \in H$ and $\alpha \in \mathbf{R}$. Suppose also that B is continuous: $|B(x,y)| \leq \beta \|x\| \|y\|$ for some $\beta \geq 0$ and all $x,y \in H$. The function $f : H \to \mathbf{R}$, defined by $f(x) = B(x,x)$, is differentiable (even in the sense of Fréchet) and $\nabla f(x)h = B(x,h) + B(h,x)$. If B is symmetric, which means that $B(x,y) = B(y,x)$ for all $x,y \in H$, then $\nabla f(x)h = 2B(x,h)$.

In this section, we discuss three aspects of the relationship between convexity and derivatives, namely: optimality conditions, characterizations of convexity, and smoothness.

2.3.1 Optimality conditions

We know from undergraduate calculus courses that minimizers of differentiable functions are critical. For convex functions, global minimizers are the only possible critical points.

Theorem 2.18. *Let $f : H \to \mathbf{R} \cup \{+\infty\}$ be differentiable at $\hat{x} \in \operatorname{int}\big(\operatorname{dom}(f)\big)$.*

i) If $f(\hat{x}) \leq f(y)$ for all $y \in B(\hat{x}, r)$, then $\nabla f(\hat{x}) = 0$.

ii) If $\nabla f(\hat{x}) = 0$ and f is convex, then $f(\hat{x}) \leq f(y)$ for all $y \in H$.

Proof. Given $d \in H$, we have $f(\hat{x} + td) \geq f(\hat{x})$, for all $t > 0$ sufficiently small. Rearranging the terms and passing to the limit, we obtain

$$\langle \nabla f(\hat{x}), d \rangle = f'(\hat{x}; d) = \lim_{t \to 0} \frac{f(\hat{x} + td) - f(\hat{x})}{t} \geq 0.$$

Since this is true for every $d \neq 0$, it implies that $\nabla f(\hat{x}) = 0$. Conversely, if there is $z \in H$ such that $f(z) < f(\hat{x})$, convexity gives

$$f\big(\hat{x} + t(z - \hat{x})\big) - f(\hat{x}) = f\big(tz + (1-t)\hat{x}\big) - f(\hat{x}) \leq t\big(f(z) - f(\hat{x})\big)$$

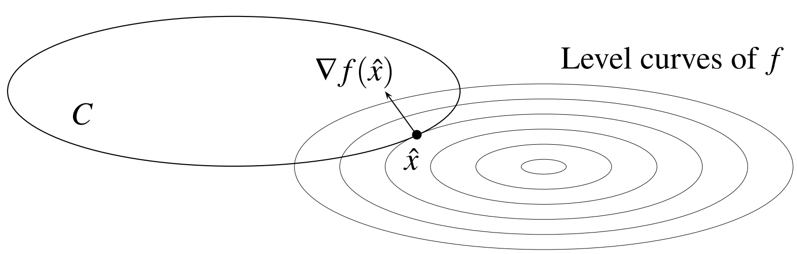
for all $t \in (0,1)$. It follows that $\langle \nabla f(\hat{x}), z - \hat{x} \rangle \leq f(z) - f(\hat{x}) < 0$, and so $\nabla f(\hat{x}) \neq 0$. □

Exercise 2.7. Let $C \subset H$ be convex, let $\hat{x} \in C$, and let $f : H \to \mathbf{R} \cup \{+\infty\}$ be differentiable at $\hat{x}$. Prove the following:

i) If $f(\hat{x}) \leq f(y)$ for all $y \in C$, then $\langle \nabla f(\hat{x}), y - \hat{x} \rangle \geq 0$ for all $y \in C$.

ii) If f is convex and $\langle \nabla f(\hat{x}), y - \hat{x} \rangle \geq 0$ for all $y \in C$, then $f(\hat{x}) \leq f(y)$ for all $y \in C$.

To fix the ideas, consider a differentiable function on $X = \mathbf{R}^2$. Theorem 2.18 states that the gradient of f at $\hat{x}$ must point *inwards*, with respect to C. In other words, f can only decrease by leaving the set C. This situation is depicted below:



2.3.2 Characterizations of convexity for differentiable functions

The results presented in this section and the next are useful for two reasons: they provide properties of convex functions, as well as a tool to determine whether a function is convex or not.

Proposition 2.19. *Let $A \subset H$ be open and convex, and let $f : A \to \mathbf{R}$ be differentiable. The following are equivalent:*

i) *f is convex.*

ii) *$f(y) \geq f(x) + \langle \nabla f(x), y - x\rangle$ for every $x, y \in A$.*

iii) *$\langle \nabla f(x) - \nabla f(y), x - y\rangle \geq 0$ for every $x, y \in A$.*

Proof. By convexity,

$$f(\lambda y + (1-\lambda)x) \leq \lambda f(y) + (1-\lambda) f(x)$$

for all $y \in X$ and all $\lambda \in (0,1)$. Rearranging the terms we get

$$\frac{f(x + \lambda(y-x)) - f(x)}{\lambda} \leq f(y) - f(x).$$

As $\lambda \to 0$ we obtain *ii*). From *ii*), we immediately deduce *iii*). To prove that *iii*) $\Rightarrow$ *i*), define $\phi : [0,1] \to \mathbf{R}$ by

$$\phi(\lambda) = f\big(\lambda x + (1-\lambda)y\big) - \lambda f(x) - (1-\lambda) f(y).$$

Then $\phi(0) = \phi(1) = 0$ and

$$\phi'(\lambda) = \langle \nabla f\left(\lambda x + (1-\lambda)y\right), x - y\rangle - f(x) + f(y)$$

for $\lambda \in (0,1)$. Take $0 < \lambda_1 < \lambda_2 < 1$ and write $x_i = \lambda_i x + (1-\lambda_i)y$ for $i = 1,2$. A simple computation shows that

$$\phi'(\lambda_1) - \phi'(\lambda_2) = \frac{1}{\lambda_1 - \lambda_2}\langle \nabla f(x_1) - \nabla f(x_2), x_1 - x_2\rangle \leq 0.$$

In other words, ϕ' is nondecreasing. Since $\phi(0) = \phi(1) = 0$, there is $\bar{\lambda} \in (0,1)$ such that $\phi'(\bar{\lambda}) = 0$. Since ϕ' is nondecreasing, we must have $\phi' \leq 0$ (and so ϕ is nonincreasing) on $[0, \bar{\lambda}]$ and next $\phi' \geq 0$ (whence ϕ is nondecreasing) on $[\bar{\lambda}, 1]$. It follows that $\phi(\lambda) \leq 0$ on $[0,1]$, and so, f is convex. □

It is possible to obtain characterizations for strict and strong convexity as well.

Exercise 2.8. Let $A \subset X$ be open and convex, and let $f : A \to \mathbf{R}$ be differentiable. Prove that the following statements are equivalent:

i) f is strictly convex.

ii) $f(y) > f(x) + \langle \nabla f(x), y - x\rangle$ for any distinct $x, y \in A$.

iii) $\langle \nabla f(x) - \nabla f(y), x - y\rangle > 0$ for any distinct $x, y \in A$.

Exercise 2.9. Let $A \subset X$ be open and convex, and let $f : A \to \mathbf{R}$ be differentiable. Show that the following statements are equivalent:

i) f is strongly convex with constant $\alpha > 0$.

ii) $f(y) \geq f(x) + \langle \nabla f(x), y - x\rangle + \frac{\alpha}{2}\|x - y\|^2$ for every $x, y \in A$.

iii) $\langle \nabla f(x) - \nabla f(y), x - y\rangle \geq \alpha\|x - y\|^2$ for every $x, y \in A$.

2.3.3 Second order characterizations

Let $f : H \to \mathbf{R} \cup \{+\infty\}$, and let $A \subset \operatorname{int}\big(\operatorname{dom}(f)\big)$ be open, and let f be differentiable in A. We say f is *twice differentiable* at $x \in A$ if there exists $M \in \mathscr{L}(H;H)$ such that

$$\langle Mk, h\rangle = \lim_{t\to 0} \frac{\langle \nabla f(x+tk) - \nabla f(x), h\rangle}{t}$$

for all $h, k \in H$. The *Hessian* of f at x is $\nabla^2 f(x) = M$. The operator $\nabla^2 f(x)$ is symmetric for each $x \in A$ where $\nabla^2 f$ is continuous.

With this notion, we can obtain second order characterizations of convexity.

Proposition 2.20. *Let $A \subset X$ be open and convex, and let $f : A \to \mathbf{R}$ be twice differentiable on A.*

a) f is convex if, and only if, $\langle \nabla^2 f(x)d, d\rangle \geq 0$ for every $x \in A$ and $d \in H$.

b) f is strictly convex if, and only if, $\langle \nabla^2 f(x)d, d\rangle > 0$ for every $x \in A$ and $d \neq 0$.

c) f is strongly convex with constant α if, and only if, $\langle \nabla^2 f(x)d, d\rangle \geq \frac{\alpha}{2}\|d\|^2$ for every $x \in A$ and $d \in H$.

Proof. We shall discuss part $a)$ and leave the rest to the reader. To this end, let us prove that condition $iii)$ in Proposition 2.19 implies that $\langle \nabla^2 f(x)d, d\rangle \geq 0$ for every $x \in A$ and $d \in H$, which in turn implies that f is convex.

For $t > 0$ and $h \in H$, we have $\langle \nabla f(x+th) - \nabla f(x), th\rangle \geq 0$. Dividing by t^2 and passing to the limit as $t \to 0$, we obtain $\langle \nabla^2 f(x)h, h\rangle \geq 0$. Finally, defining ϕ as in Proposition 2.19, we see that

$$\phi''(\lambda) = \langle \nabla^2 f\big(\lambda x + (1-\lambda)y\big)(x-y), x-y\rangle \geq 0.$$

It follows that ϕ' is nondecreasing and we conclude as before. □

2.3.4 Smoothness

A function $f : H \to \mathbf{R}$ is *L-smooth* if it is differentiable, and ∇f is Lipschitz continuous with constant $L \geq 0$. For smooth functions, we can quantify the precision of the first-order approximation, as follows:

Lemma 2.21. *If $f : H \to \mathbf{R}$ is L-smooth, then*

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

for each $x, y \in H$.

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

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

Therefore,

$$f(y)-f(x)=\int_0^1\langle\nabla f(x),h\rangle\,dt+\int_0^1\langle\nabla f(x+th)-\nabla f(x),h\rangle\,dt.$$

It follows that

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

as claimed. □

Exercise 2.10. Prove that if f is smooth, it is differentiable in the sense of Fréchet (see Remark 2.16.

Theorem 2.22 (Baillon-Haddad Theorem)**.** *Let $f:H\to\mathbf{R}$ be convex and differentiable, and let $L>0$. The following statements are equivalent:*

i) *f is L-smooth;*

ii) $f(z)\leq f(x)+\langle\nabla f(x),z-x\rangle+\frac{L}{2}\|z-x\|^2$*, for all* $x,z\in H$*;*

iii) $f(x)\geq f(y)+\langle\nabla f(y),x-y\rangle+\frac{1}{2L}\|\nabla f(x)-\nabla f(y)\|^2$*, for all* $x,y\in H$*; and*

iv) $\langle\nabla f(x)-\nabla f(y),x-y\rangle\geq\frac{1}{L}\|\nabla f(x)-\nabla f(y)\|^2$*, for all* $x,y\in H$*.*

Proof. We shall prove that $i)\Rightarrow ii)\Rightarrow iii)\Rightarrow iv)\Rightarrow i)$. The first implication is true, even if f is not convex (see Lemma 2.21). Suppose *ii*) holds. Subtract $\langle\nabla f(y),z\rangle$ to both sides, write $h_y(z)=f(z)-\langle\nabla f(y),z\rangle$ and rearrange the terms to obtain

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

The function h_y is convex, differentiable and $\nabla h_y(y)=0$. We deduce that $h_y(y)\leq h_y(z)$ for all $z\in X$, and so

$$h_y(y)\leq h_x(x)+\langle\nabla f(x)-\nabla f(y),z\rangle+\frac{L}{2}\|z-x\|^2. \tag{2.6}$$

Replace

$$z=x-\frac{1}{L}\big(\nabla f(x)-\nabla f(y)\big)$$

in (2.6) to obtain

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

which is precisely *iii*). Interchanging the roles of x and y and adding the resulting inequality, we obtain *iv*). The last implication is straightforward. □

2.3.5 Functions which are both strongly convex and smooth

Proposition 2.23. *If $f : H \to \mathbf{R}$ is L-smooth and μ-strongly convex, then*

$$\langle \nabla f(x) - \nabla f(y), x-y\rangle \geq \frac{\mu L}{L+\mu}\|x-y\|^2 + \frac{1}{L+\mu}\|\nabla f(x) - \nabla f(y)\|^2 \tag{2.7}$$

for all $x,y \in H$.

Proof. Since f is μ-strongly convex, the function g, defined by $g(x) = f(x) - \frac{\mu}{2}\|x\|^2$ is convex (see Exercise 2.4). We also know that $\nabla g(x) = \nabla f(x) - \mu x$. Next, we have

$$\begin{aligned}
g(y) &= f(y) - \frac{\mu}{2}\|y\|^2 \\
&\leq f(x) + \langle \nabla f(x), y-x\rangle + \frac{L}{2}\|y-x\|^2 - \frac{\mu}{2}\left(\|x\|^2 + \|y-x\|^2 + 2\langle x, y-x\rangle\right) \\
&= f(x) - \frac{\mu}{2}\|x\|^2 + \langle \nabla f(x) - \mu x, y-x\rangle + \frac{L-\mu}{2}\|y-x\|^2 \\
&= g(x) + \langle \nabla g(x), y-x\rangle + \frac{L-\mu}{2}\|y-x\|^2.
\end{aligned}$$

Since g is convex, we may use the fact that $ii) \Rightarrow iv)$ in Theorem 2.22, to conclude that

$$\langle \nabla g(x) - \nabla g(y), x-y\rangle \geq \frac{1}{L-\mu}\|\nabla g(x) - \nabla g(y)\|^2.$$

Therefore,

$$\begin{aligned}
\langle \nabla f(x) - \nabla f(y), x-y\rangle &= \langle \nabla g(x) - \nabla g(y), x-y\rangle + \mu\|y-x\|^2 \\
&\geq \frac{1}{L-\mu}\|\nabla g(x) - \nabla g(y)\|^2 + \mu\|y-x\|^2 \\
&= \frac{1}{L-\mu}\|\nabla f(x) - \nabla f(y) - \mu(x-y)\|^2 + \mu\|y-x\|^2 \\
&= \frac{1}{L-\mu}\|\nabla f(x) - \nabla f(y)\|^2 + \frac{\mu L}{L-\mu}\|y-x\|^2 \\
&\quad - \frac{2\mu}{L-\mu}\langle \nabla f(x) - \nabla f(y), x-y\rangle.
\end{aligned}$$

Multiplying by $L-\mu$ and rearranging the terms, we obtain

$$(L+\mu)\langle \nabla f(x) - \nabla f(y), x-y\rangle \geq \|\nabla f(x) - \nabla f(y)\|^2 + \mu L\|y-x\|^2,$$

which is equivalent to (2.7) □

2.4 Lower-semicontinuity and existence of minimizers

A function $f : H \to \mathbf{R} \cup \{+\infty\}$ is *lower-semicontinuous* at a point $x \in H$ if for each $\alpha < f(x)$ there is $r > 0$ such that $f(y) > \alpha$ for all $y \in B(x,r)$.

Lower-semiconitnuity can also be characterized in terms of convergent sequences:

Exercise 2.11. Show that $f : H \to \mathbf{R} \cup \{+\infty\}$ is lower-semicontinuous at $x \in \mathrm{dom}(f)$ if, and only if,

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

for every sequence (x_n) that converges to x.

Example 2.24. Let $f : \mathbf{R}^M \to \mathbf{R} \cup \{+\infty\}$ be proper, lower-semicontinuous and bounded from below. Define $F : L^2(0,T;\mathbf{R}^M) \to \mathbf{R} \cup \{+\infty\}$ by

$$F(u) = \int_0^T f(u(t))\,dt.$$

We shall see that F is lower-semicontinuous. Let (u_n) be a sequence in the domain of F converging strongly to some $\bar{u} \in L^2(0,T;\mathbf{R}^M)$. Extract a subsequence (u_{k_n}) such that

$$\lim_{n\to\infty} F(u_{k_n}) = \liminf_{n\to\infty} F(u_n).$$

Next, since (u_{k_n}) converges in $L^2(0,T;\mathbf{R}^M)$, we may extract yet another subsequence (u_{j_n}) such that $(u_{j_n}(t))$ converges to $\bar{u}(t)$ for almost every $t \in [0,T]$ (see, for instance, [1, Theorem 4.9]). Since f is lower-semicontinuous, we must have $\liminf_{n\to\infty} f(u_{j_n}(t)) \geq f(\bar{u}(t))$ for almost every t. By Fatou's Lemma (see, for instance, [1, Lemma 4.1]), we obtain

$$F(\bar{u}) \leq \liminf_{n\to\infty} F(u_{j_n}) = \lim_{n\to\infty} F(u_{k_n}) = \liminf_{n\to\infty} F(u_n),$$

and so F is lower-semicontinuous.

If f is lower-semicontinuous at every point of its domain, we just say f is lower-semicontinuous.

Example 2.25. The indicator function δ_C of a closed set C is lower-semicontinuous.

Example 2.26. If f and g are lower-semicontinuous and if $\alpha \geq 0$, then $f + \alpha g$ is lower-semicontinuous. In other words, the set of lower-semicontinuous functions is a convex cone.

Lower-semicontinuity can also be characterized in terms of the sublevel sets and the epigraph:

Proposition 2.27. *Let $f : H \to \mathbf{R} \cup \{+\infty\}$. The following statements are equivalent:*

i) f is lower-semicontinuous;

ii) for each $\gamma \in \mathbf{R}$, the sublevel set $[f \leq \gamma]$ is closed; and

iii) the set $\mathrm{epi}(f)$ *is closed in $H \times \mathbf{R}$.*

Example 2.28. If $(f_i)_{i\in I}$ is a family of lower-semicontinuous functions, then $\sup(f_i)$ is lower-semicontinuous, since $\mathrm{epi}(\sup(f_i)) = \cap\,\mathrm{epi}(f_i)$ and the intersection of closed sets is closed.

2.4.1 Lower-semicontinuous convex functions

Pretty much as closed convex sets are intersections of closed half-spaces, any lower-semicontinuous convex function can be represented as a supremum of continuous affine functions:

Proposition 2.29. *Let $f : H \to \mathbf{R} \cup \{+\infty\}$ be proper. Then, f is convex and lower-semicontinuous if, and only if, there exists a family $(f_i)_{i\in I}$ of continuous affine functions on H such that $f = \sup(f_i)$.*

Proof. Suppose f is convex and lower-semicontinuous and let $x_0 \in H$. We shall prove that, for every $\lambda_0 < f(x_0)$, there exists a continuous affine function α such that $\alpha(x) \leq f(x)$ for all $x \in \mathrm{dom}(f)$ and $\lambda_0 < \alpha(x_0) < f(x_0)$. Since $\mathrm{epi}(f)$ is nonempty, closed and convex, and $(x_0, \lambda_0) \notin \mathrm{epi}(f)$, by Proposition 1.25, there exist $(z,s) \in H \times \mathbf{R} \setminus \{(0,0)\}$ and $\varepsilon > 0$ such that

$$\langle z, x_0 \rangle + s\lambda_0 + \varepsilon \leq \langle z, x \rangle + s\lambda \tag{2.8}$$

for all $(x,\lambda) \in \mathrm{epi}(f)$. Clearly, $s \geq 0$. Otherwise, we may take $x \in \mathrm{dom}(f)$ and λ sufficiently large to contradict (2.8). We distinguish two cases:

<u>$s > 0$:</u> We may assume, without loss of generality, that $s = 1$. Then, we set

$$\alpha(x) = \langle -z, x \rangle + [\langle z, x_0 \rangle + \lambda_0 + \varepsilon],$$

and take $\lambda = f(x)$ to deduce that $f(x) \geq \alpha(x)$ for all $x \in \mathrm{dom}(f)$ and $\alpha(x_0) > \lambda_0$. Observe that this is valid for each $x_0 \in \mathrm{dom}(f)$.

<u>$s = 0$:</u> Necessarily, $x_0 \notin \mathrm{dom}(f)$. Therefore, $f(x_0) = +\infty$. Set

$$\alpha_0(x) = \langle -z, x \rangle + [\langle z, x_0 \rangle + \varepsilon],$$

and observe that $\alpha_0(x) \leq 0$ for all $x \in \mathrm{dom}(f)$ and $\alpha_0(x_0) = \varepsilon > 0$. Now take $\hat{x} \in \mathrm{dom}(f)$ and use the argument of the case $s > 0$ to obtain a continuous affine function $\hat{\alpha}$ such that $f(x) \geq \hat{\alpha}(x)$ for all $x \in \mathrm{dom}(f)$. Given $n \in \mathbf{N}$ set $\alpha_n = \hat{\alpha} + n\alpha_0$. We conclude that $f(x) \geq \alpha_n(x)$ for all $x \in \mathrm{dom}(f)$ and $\lim\limits_{n\to\infty} \alpha_n(x_0) = \lim\limits_{n\to\infty} (\hat{\alpha}(x_0) + n\varepsilon) = +\infty = f(x_0)$.

The converse is straightforward, since $\mathrm{epi}(\sup(f_i)) = \cap\, \mathrm{epi}(f_i)$. □

Lower-semicontinuous convex functions are continuous in the interior of their domains.

Proposition 2.30. *Let $f : H \to \mathbf{R} \cup \{+\infty\}$ be lower-semicontinuous and convex. Then f is continuous on* $\mathrm{int}(\mathrm{dom}(f))$.

Proof. Fix $x_0 \in \mathrm{int}(\mathrm{dom}(f))$. Without any loss of generality we may assume that $x_0 = 0$. Take $\gamma > f(0)$. Given $x \in H$, define $g_x : \mathbf{R} \to \mathbf{R} \cup \{+\infty\}$ by $g_x(t) = f(tx)$. Since $0 \in \mathrm{int}(\mathrm{dom}(g_x))$, we deduce that g_x is continuous at 0 by Proposition 2.12. Therefore, there is $t_x > 0$ such that $t_x x \in [f \leq \gamma]$. Repeating this argument for each $x \in H$, we see that $\bigcup_{n\geq 1} n[f \leq \gamma] = H$. Baire's Theorem 1.5 shows that $[f \leq \gamma]$ has nonempty interior, and we conclude by Corollary 2.10. □

2.4.2 Existence of minimizers

By analogy with the characterization of lower-semicontinuity given in Exercise 2.11, we say that a function $f : H \to \mathbf{R} \cup \{+\infty\}$ is *weakly sequentially lower-semicontinuous* at $x \in \mathrm{dom}(f)$ if

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

for every sequence (x_n) converging weakly to x. From Propositions 1.36 and 2.27, we obtain

Proposition 2.31. *If $f : H \to \mathbf{R} \cup \{+\infty\}$ is quasi-convex and lower-semicontinuous, it is weakly sequentially lower-semicontinuous.*

Example 2.32. Let $f : \mathbf{R}^M \to \mathbf{R} \cup \{+\infty\}$ be proper, convex, lower-semicontinuous and bounded from below. As in Example 2.24, define $F : L^2(0,T;\mathbf{R}^M) \to \mathbf{R} \cup \{+\infty\}$ by

$$F(u) = \int_0^T f(u(t))\,dt.$$

Clearly, F is proper and convex, and we already proved (in Example 2.24) that F is lower-semicontinuous. Proposition 2.31 shows that F is weakly sequentially lower-semicontinuous.

We shall say (x_n) is a *minimizing sequence* for $f : H \to \mathbf{R} \cup \{+\infty\}$ if $\lim_{n\to\infty} f(x_n) = \inf(f)$. An important property of sequentially lower-semicontinuous functions is that the limits of convergent minimizing sequences are minimizers.

Proposition 2.33. *Let (x_n) be a minimizing sequence for $f : H \to \mathbf{R} \cup \{+\infty\}$.*

i) If (x_n) converges to $\bar{x}$ and f is lower-semicontinuous there, then $\bar{x} \in \mathrm{argmin}(f)$.

ii) If (x_n) converges weakly to $\bar{x}$ and f is weakly sequentially lower-semicontinuous there, then $\bar{x} \in \mathrm{argmin}(f)$.

A function $f : H \to \mathbf{R} \cup \{+\infty\}$ is *coercive* if $\lim_{\|x\|\to\infty} f(x) = \infty$.

Exercise 2.12. Prove that $f : H \to \mathbf{R} \cup \{+\infty\}$ is coercive if, and only if, $[f \leq \gamma]$ is bounded for each $\gamma \in \mathbf{R}$.

We are now in a position to state and prove the following result, concerning the existence of minimizers:

Theorem 2.34. *If $f : H \to \mathbf{R} \cup \{+\infty\}$ is proper, quasi-convex, coercive and lower-semicontinuous, then $\mathrm{argmin}(f)$ is nonempty, convex, closed and bounded. If, moreover, f is strictly convex, then $\mathrm{argmin}(f)$ is a singleton.*

Proof. Let (x_n) be a minimizing sequence for f. Since $f(x_n) \to \inf(f)$, there is $\gamma \in \mathbf{R}$ such that $x_n \in [f \leq \gamma]$ for all n. The set $[f \leq \gamma]$ is bounded, because f is coercive. It follows that the sequence (x_n) is bounded. By Theorem 1.35, it has a subsequence (x_{k_n}), which converges weakly to some $\bar{x}$. By Proposition 2.31, f is weakly sequentially lower-semicontinuous. Therefore, Proposition 2.33 implies that $\bar{x} \in \mathrm{argmin}(f) \neq \emptyset$. Moreover, since $\mathrm{argmin}(f) = [f \leq \min(f)]$, this set is convex, closed and bounded, because f is quasi-convex, lower-semicontinuous and coercive, respectively. If f is strictly convex, it cannot have more than one minimizer. $\square$

The reader will find several similarities and differences with the discussion following Weierstrass's Theorem in Section 1.6.

If f is strongly convex, it is strictly convex and coercive. Therefore, we deduce:

Corollary 2.35. *If $f : H \to \mathbf{R} \cup \{+\infty\}$ is proper, strongly convex and lower-semicontinuous, then* $\mathrm{argmin}(f)$ *is a singleton.*

A direct application of Corollary 2.35 is the Lax-Milgram Theorem, of great relevance in the theory of partial differential equations.

Corollary 2.36 (Lax-Milgram Theorem)**.** *Let $B : H \times H \to \mathbf{R}$ be a symmetric bilinear form such that*

$$B(x,x) \geq \mu\|x\|^2 \qquad \text{and} \qquad B(x,y) \leq L\|x\|\,\|y\|$$

for all $x,y \in H$ and some $\mu, L > 0$. For every $z \in H$, there is a unique $\hat{x}_z \in H$ such that $B(\hat{x}_z, y) = \langle z, y\rangle$ for all $y \in H$. Moreover, the function $z \mapsto \hat{x}_z$ is continuous.

Proof. Apply Corollary 2.35 to the continuous and strongly convex function $f : H \to \mathbf{R} \cup \{+\infty\}$, defined by

$$f(x) = \frac{1}{2}B(x,x) - \langle z, x\rangle,$$

to deduce that it has a unique minimizer $\hat{x}_z$. By Theorem 2.18, we have

$$0 = \langle \nabla f(\hat{x}_z), y\rangle = B(\hat{x}_z, y) - \langle z, y\rangle,$$

for all $y \in H$. For the continuity, since $\mu\|\hat{x}_z\|^2 \leq B(\hat{x}_z, \hat{x}_z) = \langle z, \hat{x}_z\rangle \leq \|z\|\,\|\hat{x}_z\|$, the Cauchy-Schwarz inequality yields $\|\hat{x}_z\| \leq \frac{1}{\mu}\|z\|$. $\square$

Remark 2.37. The Lax-Milgram Theorem can be restated as follows: if we replace $z \in H$ by $\ell \in \mathscr{L}(H;\mathbf{R})$, we will obtain $B(\hat{x}_z, y) = \ell(y)$ instead of $B(\hat{x}_z, y) = \langle z, y\rangle$. It amounts to the same, in view of Riesz's Representation Theorem 1.24. Can you provide an alternative (and straightforward) proof of the latter, based on the arguments given above?

Chapter 3

Subdifferential calculus

3.1 Subgradients and subdifferentials

Let $f : H \to \mathbf{R}$ be convex and assume it is differentiable at a point $x \in H$. By convexity, Proposition 2.19 gives

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle \tag{3.1}$$

for each $y \in H$. This shows that the hyperplane[1]

$$V = \{(y,z) \in H \times \mathbf{R} : f(x) + \langle \nabla f(x), y - x \rangle = z\}$$

lies below the set $\mathrm{epi}(f)$ and touches it at the point $(x, f(x))$.

This idea can be generalized to nondifferentiable functions. Let $f : H \to \mathbf{R} \cup \{+\infty\}$ be proper and convex. A point $x^* \in H$ is a *subgradient* of f at x if

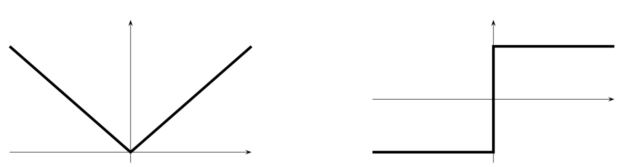
$$f(y) \geq f(x) + \langle x^*, y - x \rangle \tag{3.2}$$

for all $y \in H$, or, equivalently, if (3.2) holds for all y in a neighborhood of x. The set of all subgradients of f at x is the *subdifferential* of f at x and is denoted by $\partial f(x)$. If $\partial f(x) \neq \emptyset$, we say f is *subdifferentiable* at x. The *domain* of the subdifferential is

$$\mathrm{dom}(\partial f) = \{x \in H \ : \ \partial f(x) \neq \emptyset\}.$$

By definition, $\mathrm{dom}(\partial f) \subset \mathrm{dom}(f)$. The inclusion may be strict though, as in Example 3.3 below. Let us see some examples:

Example 3.1. For $f : \mathbf{R} \to \mathbf{R}$, given by $f(x) = |x|$, we have $\partial f(x) = \{-1\}$ if $x < 0$, $\partial f(0) = [-1, 1]$, and $\partial f(x) = \{1\}$ for $x > 0$.



Graph of f Graph of ∂f

[1]In the terminology of Section 1.5, we have $V = H_{z=\alpha}$, where $z = (\nabla f(x), -1)$ and $\alpha = \langle \nabla f(x), x \rangle - f(x)$.

Example 3.2. More generally, if $f : H \to \mathbf{R}$ is given by $f(x) = \|x - x_0\|$, with $x_0 \in H$, then

$$\partial f(x) = \begin{cases} \overline{B}(0,1) & \text{if} \quad x = x_0 \\ \frac{x-x_0}{\|x-x_0\|} & \text{if} \quad x \neq x_0. \end{cases}$$

Example 3.3. Define $g : \mathbf{R} \to \mathbf{R} \cup \{+\infty\}$ by $g(x) = +\infty$ if $x < 0$, and $g(x) = -\sqrt{x}$ if $x \geq 0$. Then

$$\partial g(x) = \begin{cases} \emptyset & \text{if} \quad x \leq 0 \\ \left\{-\frac{1}{2\sqrt{x}}\right\} & \text{if} \quad x > 0. \end{cases}$$

Notice that $0 \in \mathrm{dom}(g)$ but $\partial g(0) = \emptyset$, thus $\mathrm{dom}(\partial g) \subsetneq \mathrm{dom}(g)$.

Example 3.4. Let $h : \mathbf{R} \to \mathbf{R} \cup \{+\infty\}$ be given by $h(x) = +\infty$ if $x \neq 0$, and $h(0) = 0$. Then

$$\partial h(x) = \begin{cases} \emptyset & \text{if} \quad x \neq 0 \\ \mathbf{R} & \text{if} \quad x = 0. \end{cases}$$

Observe that h is subdifferentiable but not continuous at 0. □

Example 3.5. More generally, let C be a nonempty, closed and convex subset of H and let $\delta_C : H \to \mathbf{R} \cup \{+\infty\}$ be the indicator function of C:

$$\delta_C(x) = \begin{cases} 0 & \text{if } x \in C \\ +\infty & \text{otherwise.} \end{cases}$$

Given $x \in H$, the set $N_C(x) = \partial \delta_C(x)$ is the *normal cone* to C at x. It is given by

$$N_C(x) = \{x^* \in H : \langle x^*, y - x\rangle \leq 0 \text{ for all } y \in C\}$$

if $x \in C$, and $N_C(x) = \emptyset$ if $x \notin C$. Intuitively, the normal cone contains the directions that point *outwards* with respect to C. If C is a closed affine subspace, that is $C = \{x_0\} + V$, where $x_0 \in H$ and V is a closed subspace of H, then $N_C(x) = V^\perp$ for all $x \in C$.

The subdifferential is indeed an extension of the notion of derivative:

Proposition 3.6. *Let $f : H \to \mathbf{R} \cup \{+\infty\}$ be convex. If f is differentiable at x, then $x \in \mathrm{dom}(\partial f)$ and $\partial f(x) = \{\nabla f(x)\}$.*

Proof. First, the gradient inequality (3.1) and the definition of the subdifferential together imply $\nabla f(x) \in \partial f(x)$. Now take any $x^* \in \partial f(x)$. By definition,

$$f(y) \geq f(x) + \langle x^*, y - x\rangle$$

for all $y \in H$. Take any $h \in H$ and $t > 0$, and write $y = x + th$ to deduce that

$$\frac{f(x+th) - f(x)}{t} \geq \langle x^*, h\rangle.$$

Passing to the limit as $t \to 0$, we deduce that $\langle \nabla f(x) - x^*, h\rangle \geq 0$. Since this must be true for each $h \in H$, necessarily $x^* = \nabla f(x)$. □

Proposition 3.36 provides a converse result.

A convex function may have more than one subgradient. However, we have:

Proposition 3.7. *For each $x \in H$, the set $\partial f(x)$ is closed and convex.*

Proof. Let $x_1^*, x_2^* \in \partial f(x)$ and $\lambda \in (0,1)$. For each $y \in H$ we have

$$\begin{array}{rcl} f(y) & \geq & f(x) + \langle x_1^*, y-x \rangle \\ f(y) & \geq & f(x) + \langle x_2^*, y-x \rangle \end{array}$$

Add λ times the first inequality and $1-\lambda$ times the second one to obtain

$$f(y) \geq f(x) + \langle \lambda x_1^* + (1-\lambda) x_2^*, y-x \rangle.$$

Since this holds for each $y \in X$, we have $\lambda x_1^* + (1-\lambda) x_2^* \in \partial f(x)$. To see that $\partial f(x)$ is closed, take a sequence (x_n^*) in $\partial f(x)$, converging to some x^*. We have

$$f(y) \geq f(x) + \langle x_n^*, y-x \rangle$$

for each $y \in H$ and $n \in \mathbf{N}$. As $n \to \infty$ we obtain

$$f(y) \geq f(x) + \langle x^*, y-x \rangle.$$

It follows that $x^* \in \partial f(x)$. □

Recall that if $f : \mathbf{R} \to \mathbf{R}$ is convex and differentiable, then its derivative is nondecreasing. The following result generalizes this fact:

Proposition 3.8. *Let $f : H \to \mathbf{R} \cup \{+\infty\}$ be convex. If $x^* \in \partial f(x)$ and $y^* \in \partial f(y)$, then*

$$\langle x^* - y^*, x-y \rangle \geq 0.$$

Proof. Note that $f(y) \geq f(x) + \langle x^*, y-x \rangle$ and $f(x) \geq f(y) + \langle y^*, x-y \rangle$. It suffices to add these inequalities and rearrange the terms. □

This property of the subdifferential operator ∂f is known as *monotonicity*.

For strongly convex functions, we have the following result:

Proposition 3.9. *Let $f : H \to \mathbf{R} \cup \{+\infty\}$ be strongly convex with parameter α. Then, for each $x^* \in \partial f(x)$ and $y \in H$, we have*

$$f(y) \geq f(x) + \langle x^*, y-x \rangle + \frac{\alpha}{2} \|x-y\|^2.$$

Moreover, for each $y^* \in \partial f(y)$, we have $\langle x^* - y^*, x-y \rangle \geq \alpha \|x-y\|^2$.

The definition of the subdifferential has a straightforward yet remarkable consequence, namely:

Theorem 3.10 (Fermat's Rule)**.** *Let $f : H \to \mathbf{R} \cup \{+\infty\}$ be proper and convex. Then $\hat{x}$ is a global minimizer of f if, and only if, $0 \in \partial f(\hat{x})$.*

Observe the relationship with Theorem 2.18.

3.2 Subdifferentiablility and calculus rules

Although subdifferentiability does not imply continuity (see Example 3.4), the converse is true.

Proposition 3.11. *Let $f : H \to \mathbf{R} \cup \{+\infty\}$ be convex. If f is continuous at x, then $\partial f(x)$ is nonempty and bounded.*

Proof. By Corollary 2.8, $x \in \operatorname{int}\big(\operatorname{dom}(f)\big)$, from which it follows that $\operatorname{int}(\operatorname{epi}(f)) \neq \emptyset$. Part *i)* of the Hahn-Banach Separation Theorem 1.30, with $A = \operatorname{int}(\operatorname{epi}(f))$ and $B = \{(x, f(x))\}$, gives $(z,s) \in H \times \mathbf{R} \setminus \{(0,0)\}$ such that

$$\langle z, y\rangle + s\lambda \leq \langle z, x\rangle + sf(x)$$

for every $(y,\lambda) \in \operatorname{int}(\operatorname{epi}(f))$. Taking $y = x$, we deduce that $s \leq 0$. If $s = 0$ then $\langle z, y-x\rangle \leq 0$ for every y in a neighborhood of x. Hence $z = 0$, which is a contradiction. We conclude that $s < 0$. Therefore,

$$\lambda \geq f(x) + \langle z^*, y-x\rangle,$$

with $z^* = -z/s$, for every $\lambda > f(y)$. Letting λ tend to $f(y)$, we see that

$$f(y) \geq f(x) + \langle z^*, y-x\rangle.$$

This implies $z^* \in \partial f(x) \neq \emptyset$. On the other hand, since f is continuous at x, it is Lipschitz-continuous on a neighborhood of x, by Corollary 2.8. If $x^* \in \partial f(x)$, then

$$f(x) + \langle x^*, y-x\rangle \leq f(y) \leq f(x) + M\|y-x\|,$$

and so, $\langle x^*, y-x\rangle \leq M\|y-x\|$ for every y in a neighborhood of x. We conclude that $\|x^*\| \leq M$. In other words, $\partial f(x) \subset \bar{B}(0,M)$. □

As a consequence, we have the following:

Corollary 3.12. *Let $f : H \to \mathbf{R} \cup \{+\infty\}$ be convex and lower-semicontinuous. Then,*

$$\operatorname{int}\big(\operatorname{dom}(f)\big) \subset \operatorname{dom}(\partial f) \subset \operatorname{dom}(f).$$

In particular, if $\operatorname{int}\big(\operatorname{dom}(f)\big) \neq \emptyset$, *then* $\overline{\operatorname{dom}(\partial f)} = \overline{\operatorname{dom}(f)}$

3.2.1 Approximate subdifferentiability

As seen in Example 3.3, it may happen that $\operatorname{dom}(\partial f) \subsetneq \operatorname{dom}(f)$. This can be an inconvenient when trying to implement optimization algorithms. Given $\varepsilon > 0$, a point $x^* \in H$ is an *ε-approximate subgradient* of f at $x \in \operatorname{dom}(f)$ if

$$f(y) \geq f(x) + \langle x^*, y-x\rangle - \varepsilon \tag{3.3}$$

for all $y \in H$ (compare with the subdifferential inequality (3.2)). The set $\partial_\varepsilon f(x)$ of such vectors is the *ε-approximate subdifferential* of f at x. Clearly, $\partial f(x) \subset \partial_\varepsilon f(x)$ for each x, and so

$$\operatorname{dom}(\partial f) \subset \operatorname{dom}(\partial_\varepsilon f) \subset \operatorname{dom}(f)$$

for each $\varepsilon > 0$. It turns out that the last inclusion is, in fact, an equality.

Proposition 3.13. *Let $f : H \to \mathbf{R} \cup \{+\infty\}$ be proper, convex and lower-semicontinuous. Then* $\mathrm{dom}(f) = \mathrm{dom}(\partial_\varepsilon f)$ *for all* $\varepsilon > 0$.

Proof. Let $\varepsilon > 0$ and let $x \in \mathrm{dom}(f)$. By Proposition 2.29, f can be represented as the pointwise supremum of continuous affine functions. Therefore, there exist $x^* \in H$ and $\mu \in \mathbf{R}$, such that

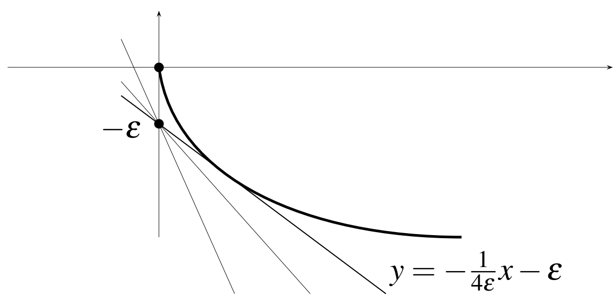
$$f(y) \geq \langle x^*, y\rangle + \mu$$

for all $y \in X$ and

$$\langle x^*, x\rangle + \mu \geq f(x) - \varepsilon.$$

Adding these two inequalities, we obtain precisely (3.3). $\square$

Example 3.14. Let us recall from Example 3.3 that the function $g : \mathbf{R} \to \mathbf{R} \cup \{+\infty\}$ given by $g(x) = +\infty$ if $x < 0$, and $g(x) = -\sqrt{x}$ if $x \geq 0$; satisfies $0 \in \mathrm{dom}(g)$ but $\partial g(0) = \emptyset$. A simple computation shows that $\partial_\varepsilon g(0) = \left(-\infty, \frac{1}{4\varepsilon}\right]$ for $\varepsilon > 0$. Observe that $\mathrm{dom}(\partial g) \subsetneq \mathrm{dom}(\partial_\varepsilon g) = \mathrm{dom}(g)$.



3.2.2 Subdifferential of the sum of convex functions and a chain rule

A natural question is whether the subdifferential of the sum of two functions, is the sum of their subdifferentials. We begin by showing that this is not always the case.

Example 3.15. Let $f, g : \mathbf{R} \to \mathbf{R} \cup \{+\infty\}$ be given by

$$f(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ +\infty & \text{if } x > 0 \end{cases} \qquad \text{and} \qquad g(x) = \begin{cases} +\infty & \text{if } x < 0 \\ -\sqrt{x} & \text{if } x \geq 0. \end{cases}$$

We have

$$\partial f(x) = \begin{cases} \{0\} & \text{if } x < 0 \\ [0, +\infty) & \text{if } x = 0 \\ \emptyset & \text{if } x > 0 \end{cases} \qquad \text{and} \qquad \partial g(x) = \begin{cases} \emptyset & \text{if } x \leq 0 \\ \left\{-\frac{1}{2\sqrt{x}}\right\} & \text{if } x > 0. \end{cases}$$

Therefore, $\partial f(x) + \partial g(x) = \emptyset$ for every $x \in \mathbf{R}$. On the other hand, $f + g = \delta_{\{0\}}$, which implies $\partial(f+g)(x) = \emptyset$ if $x \neq 0$, but $\partial(f+g)(0) = \mathbf{R}$. We see that $\partial(f+g)(x)$ may differ from $\partial f(x) + \partial g(x)$.

Roughly speaking, the problem with the function in the example above is that the intersection of their domains is too small. We have the following:

Theorem 3.16 (Moreau-Rockafellar Theorem)**.** *Let $f,g : H \to \mathbf{R} \cup \{+\infty\}$ be proper, convex and lower-semicontinuous. For each $x \in H$, we have*

$$\partial f(x) + \partial g(x) \subset \partial(f+g)(x). \tag{3.4}$$

Equality holds for every $x \in H$ if f is continuous at some $x_0 \in \mathrm{dom}(g)$.

Proof. If $x^* \in \partial f(x)$ and $z^* \in \partial g(x)$, then

$$f(y) \geq f(x) + \langle x^*, y-x\rangle \quad \text{and} \quad g(y) \geq g(x) + \langle z^*, y-x\rangle$$

for each $y \in H$. Adding both inequalities, we conclude that

$$f(y)+g(y) \geq f(x)+g(x)+\langle x^*+z^*, y-x\rangle$$

for each $y \in H$ and so, $x^*+z^* \in \partial(f+g)(x)$. Suppose now that $u^* \in \partial(f+g)(x)$. We have

$$f(y)+g(y) \geq f(x)+g(x)+\langle u^*, y-x\rangle \tag{3.5}$$

for every $y \in H$. We shall find $x^* \in \partial f(x)$ and $z^* \in \partial g(x)$ such that $x^*+z^* = u^*$. To this end, consider the following nonempty convex sets:

$$\begin{aligned} B &= \{(z,\mu) \in H \times \mathbf{R} \,:\, g(z)-g(x) \leq -\mu\} \\ C &= \{(y,\lambda) \in H \times \mathbf{R} \,:\, f(y)-f(x)-\langle u^*, y-x\rangle \leq \lambda\}. \end{aligned}$$

Define $h : H \to \mathbf{R} \cup \{+\infty\}$ as $h(y) = f(y)-f(x)-\langle u^*, y-x\rangle$. Since h is continuous in x_0 and $C = \mathrm{epi}(h)$, the open convex set $A = \mathrm{int}(C)$ is nonempty, by Corollary 2.8. Moreover, $A \cap B = \emptyset$ by inequality (3.5). Using part $i)$ of Hahn-Banach Separation Theorem 1.30, we obtain $(w,s) \in H \times \mathbf{R} \setminus \{(0,0)\}$ such that

$$\langle w,y\rangle + s\lambda \leq \langle w,z\rangle + s\mu$$

for each $(y,\lambda) \in A$ and each $(z,\mu) \in B$. Taking $(y,\lambda) = (x,1) \in A$ and $(z,\mu) = (x,0) \in B$, we deduce that $s \leq 0$. On the other hand, if $s = 0$, taking $z = x_0$ we see that $\langle w, x_0 - y\rangle \geq 0$ for every y in a neighborhood of x_0. This implies $w = 0$ and contradicts the fact that $(w,s) \neq (0,0)$. Therefore, $s < 0$, and we may write

$$\langle z^*, y\rangle + \lambda \geq \langle z^*, z\rangle + \mu \tag{3.6}$$

with $z^* = w/s$. Since the point $\big(z, g(x)-g(z)\big)$ belongs to B, inequality (3.6) yields

$$g(z)-g(x) \geq \langle z^*, z-y\rangle - \lambda$$

for every $(y,\lambda) \in \mathrm{int}(C)$. By Proposition 1.19, we can let $(y,\lambda) \to (x,0) \in C$, to deduce that $z^* \in \partial g(x)$. In a similar fashion, taking $(z,\mu) = (x,0) \in B$, we deduce from (3.6) that

$$\lambda \geq \langle -z^*, y-x\rangle$$

for all $(y,\lambda) \in \mathrm{int}(C)$. Letting $\lambda \to f(y)-f(x)-\langle u^*, y-x\rangle$, it follows that

$$f(y) \geq f(x) + \langle u^*-z^*, y-x\rangle,$$

whence $u^*-z^* \in \partial f(x)$. □

Observe that, if $\partial(f+g)(x) = \partial f(x) + \partial g(x)$ for all $x \in X$, then

$$\operatorname{dom}\big(\partial(f+g)\big) = \operatorname{dom}(\partial f) \cap \operatorname{dom}(\partial g).$$

Exercise 3.1 (Stampacchia's Theorem)**.** Let $B : H \times H \to \mathbf{R}$ be a symmetric bilinear form such that

$$B(x,x) \geq \mu \|x\|^2 \qquad \text{and} \qquad B(x,y) \leq L\|x\|\,\|y\|$$

for all $x, y \in H$ and some $\mu, L > 0$, and let $K \subset H$ be nonempty, closed and convex. Prove that, for every $z \in H$, there is a unique $\hat{x}_z \in K$ such that $B(\hat{x}_z, y - \hat{x}_z) = \langle z, y - \hat{x}_z \rangle$ for all $y \in K$.

Using the Hahn-Banach Separation Theorem 1.30 in a similar way, we can obtain a rule for the subdifferential of the composition with a linear function:

Exercise 3.2 (Chain Rule)**.** Let X, Y be real Hilbert spaces, let $A \in \mathscr{L}(X;Y)$ and let $f : Y \to \mathbf{R} \cup \{+\infty\}$ be proper, convex and lower-semicontinuous. Show that, for each $x \in X$, we have

$$A^* \partial f(Ax) \subset \partial(f \circ A)(x),$$

and that equality holds for every $x \in X$ if f is continuous at some $y_0 \in \operatorname{ran}(A)$. Hint: Find inspiration in the proof of Theorem 3.16.

Combining the Exercise 3.2 and Theorem 3.16, we obtain the following:

Corollary 3.17. *Let X, Y be real Hilbert spaces, let $A \in \mathscr{L}(X;Y)$, and let $f : Y \to \mathbf{R} \cup \{+\infty\}$ and $g : X \to \mathbf{R} \cup \{+\infty\}$ be proper, convex and lower-semicontinuous. For each $x \in X$, we have*

$$A^* \partial f(Ax) + \partial g(x) \subset \partial(f \circ A + g)(x).$$

Equality holds for every $x \in X$ if there is $x_0 \in \operatorname{dom}(g)$ such that f is continuous at Ax_0.

This result will be useful in the context of Fenchel-Rockafellar duality (see Theorem 3.31).

3.3 Moreau envelope

In this subsection, we present a regularizing and smoothing technique for convex functions defined on a Hilbert space. By adding a quadratic term, one is able to force the existence and uniqueness of a minimizer. This fact, in turn, has three major consequences: first, it is the core of an important minimization method, known as the proximal point algorithm, which we shall study later; second, it allows us to construct a smooth version of the function having the same minimizers; and third, it allows us to approximate solutions of the *differential inclusion* $-\dot{x}(t) \in \partial f\big(x(t)\big)$.

3.3.1 The proximity operator

Given $\lambda > 0$ and $x \in H$, consider the function $f_{(\lambda,x)} : H \to \mathbf{R} \cup \{+\infty\}$, defined by

$$f_{(\lambda,x)}(z) = f(z) + \frac{1}{2\lambda}\|z - x\|^2.$$

The following property will be useful later on, especially when we study the *proximal point algorithm* (actually, it is the core of that method):

Proposition 3.18. *For each $\lambda > 0$ and $x \in H$, the function $f_{(\lambda,x)}$ has a unique minimizer $\bar{x}$. Moreover, $\bar{x}$ is characterized by the inclusion*

$$-\frac{\bar{x} - x}{\lambda} \in \partial f(\bar{x}). \tag{3.7}$$

Proof. Since f is proper, convex and lower-semicontinuous, $f_{(\lambda,x)}$ is proper, strictly convex, lower-semicontinuous and coercive. Existence and uniqueness of a minimizer $\bar{x}$ is given by Theorem 2.34. Finally, Fermat's Rule (Theorem 3.10) and the Moreau-Rockafellar Theorem 3.16 imply that $\bar{x}$ must satisfy

$$0 \in \partial f_{(\lambda,x)}(\bar{x}) = \partial f(\bar{x}) + \frac{\bar{x} - x}{\lambda},$$

which gives the result. □

Solving for $\bar{x}$ in (3.7), we can write

$$\bar{x} = (I + \lambda \partial f)^{-1} x, \tag{3.8}$$

where $I : H \to H$ denotes the identity function. In view of Proposition 3.18, the expression $J_\lambda = (I + \lambda \partial f)^{-1}$ defines a function $J_\lambda : H \to H$, called the *proximity operator* of f, with parameter λ. It is also known as the *resolvent* of the operator ∂f, with parameter λ, due to its analogy with the resolvent of a linear operator. We have the following:

Proposition 3.19. *If $f : H \to \mathbf{R} \cup \{+\infty\}$ is proper, lower-semicontinuous and convex, then $J_\lambda : H \to H$ is (everywhere defined and) nonexpansive.*

Proof. Let $\bar{x} = J_\lambda(x)$ and $\bar{y} = J_\lambda(y)$, so that

$$-\frac{\bar{x} - x}{\lambda} \in \partial f(\bar{x}) \qquad \text{y} \qquad -\frac{\bar{y} - y}{\lambda} \in \partial f(\bar{y}).$$

In view of the monotonicity of ∂f (Proposition 3.8), we have

$$\langle (\bar{x} - x) - (\bar{y} - y), \bar{x} - \bar{y} \rangle \leq 0.$$

Therefore,

$$0 \leq \|\bar{x} - \bar{y}\|^2 \leq \langle x - y, \bar{x} - \bar{y} \rangle \leq \|x - y\| \, \|\bar{x} - \bar{y}\|, \tag{3.9}$$

and we conclude that $\|\bar{x} - \bar{y}\| \leq \|x - y\|$. □

Example 3.20. The indicator function δ_C of a nonempty, closed and convex subset of H is proper, convex and lower-semicontinuous. Given $\lambda > 0$ and $x \in H$, $J_\lambda(x)$ is the unique solution of

$$\min\left\{\delta_C(z) + \frac{1}{2\lambda}\|z-x\|^2 : z \in H\right\} = \min\{\|z-x\| : z \in C\}.$$

Independently of λ, $J_\lambda(x)$ is the point in C which is closest to x. In other words, it is the projection of x onto C, which we have denoted by $P_C(x)$. From Proposition 3.19, we recover the known fact that the function $P_C : H \to H$, defined by $P_C(x) = J_\lambda(x)$, is nonexpansive (see Proposition 1.23).

3.3.2 The Moreau envelope

By Proposition 3.18, for each $\lambda > 0$ and $x \in H$, the function $f_{(\lambda,x)}$ of f has a unique minimizer $\bar{x}$. The *Moreau envelope* of f, with parameter $\lambda > 0$, is the function $f_\lambda : H \to \mathbf{R}$ defined by

$$f_\lambda(x) = \min_{z\in H}\left\{f_{(\lambda,x)}(z)\right\} = \min_{z\in H}\left\{f(z) + \frac{1}{2\lambda}\|z-x\|^2\right\} = f(\bar{x}) + \frac{1}{2\lambda}\|\bar{x}-x\|^2.$$

An immediate consequence of the definition is that

$$\inf_{z\in H} f(z) \le f_\lambda(x) \le f(x)$$

for all $\lambda > 0$ and $x \in H$. Therefore,

$$\inf_{z\in H} f(z) = \inf_{z\in H} f_\lambda(z),$$

and $\hat{x}$ minimizes f on H if, and only if, it minimizes f_λ on H for all $\lambda > 0$.

Example 3.21. Let C be a nonempty, closed and convex subset of H, and let $f = \delta_C$. Then, for each $\lambda > 0$ and $x \in H$, we have

$$f_\lambda(x) = \min_{z\in C}\left\{\frac{1}{2\lambda}\|z-x\|^2\right\} = \frac{1}{2\lambda}\mathrm{dist}(x,C)^2,$$

where $\mathrm{dist}(x,C)$ denotes the distance from x to C.

A remarkable property of the Moreau envelope is given by the following:

Proposition 3.22. *Let $\lambda > 0$. The function f_λ is differentiable (even in the sense of Fréchet), and*

$$Df_\lambda(x) = \frac{1}{\lambda}(x-\bar{x}),$$

for all $x \in H$. Moreover, f_λ is convex and Df_λ is Lipschitz-continuous with constant $1/\lambda$.

Proof. First observe that

$$\begin{aligned} f_\lambda(y) - f_\lambda(x) &= f(\bar{y}) - f(\bar{x}) + \frac{1}{2\lambda}\left[\|\bar{y}-y\|^2 - \|\bar{x}-x\|^2\right] \\ &\ge -\frac{1}{\lambda}\langle \bar{x}-x, \bar{y}-\bar{x}\rangle + \frac{1}{2\lambda}\left[\|\bar{y}-y\|^2 - \|\bar{x}-x\|^2\right], \end{aligned}$$

by the subdifferential inequality and inclusion (3.7). Straightforward algebraic manipulations yield

$$f_\lambda(y) - f_\lambda(x) - \frac{1}{\lambda}\langle x-\bar{x}, y-x\rangle \geq \frac{1}{2\lambda}\|(\bar{y}-\bar{x})-(y-x)\|^2 \geq 0. \tag{3.10}$$

Interchanging the roles of x and y, we obtain

$$f_\lambda(x) - f_\lambda(y) - \frac{1}{\lambda}\langle y-\bar{y}, x-y\rangle \geq 0.$$

We deduce that

$$0 \leq f_\lambda(y) - f_\lambda(x) - \frac{1}{\lambda}\langle x-\bar{x}, y-x\rangle \leq \frac{1}{\lambda}\left[\|y-x\|^2 - \langle \bar{y}-\bar{x}, y-x\rangle\right] \leq \frac{1}{\lambda}\|y-x\|^2,$$

in view of (3.9). Writing $y = x+h$, we conclude that

$$\lim_{\|h\|\to 0} \frac{1}{\|h\|}\left| f_\lambda(x+h) - f_\lambda(x) - \frac{1}{\lambda}\langle x-\bar{x}, h\rangle \right| = 0,$$

and f_λ is Fréchet-differentiable. The convexity follows from the fact that

$$\langle Df_\lambda(x) - Df_\lambda(y), x-y\rangle \geq 0,$$

and the characterization given in Proposition 2.19. Finally, from (3.9), we deduce that

$$\|(x-\bar{x})-(y-\bar{y})\|^2 = \|x-y\|^2 + \|\bar{x}-\bar{y}\|^2 - 2\langle x-y, \bar{x}-\bar{y}\rangle \leq \|x-y\|^2.$$

Therefore,

$$\|Df_\lambda(x) - Df_\lambda(y)\| \leq \frac{1}{\lambda}\|x-y\|,$$

and so, Df_λ is Lipschitz-continuous with constant $1/\lambda$. □

Remark 3.23. Observe that $Df_\lambda(x) \in \partial f(\bar{x})$.

3.4 Fenchel conjugate

The *Fenchel conjugate* of a proper function $f : H \to \mathbf{R} \cup \{+\infty\}$ is the function $f^* : H \to \mathbf{R} \cup \{+\infty\}$ defined by

$$f^*(x^*) = \sup_{x\in H}\{\langle x^*, x\rangle - f(x)\}. \tag{3.11}$$

Since f^* is a supremum of continuous affine functions, it is convex and lower-semicontinuous (see Proposition 2.29). Moreover, if f is bounded from below by a continuous affine function (for instance if f is proper and lower-semicontinuous), then f^* is proper. As a consequence of the definition, we deduce the following:

Proposition 3.24 (Fenchel-Young Inequality)**.** *Let $f : H \to \mathbf{R} \cup \{+\infty\}$. For all $x, x^* \in H$, we have*

$$f(x) + f^*(x^*) \geq \langle x^*, x\rangle. \tag{3.12}$$

If f is convex, then equality in (3.12) *holds if, and only if, $x^* \in \partial f(x)$.*

In particular, if f is convex, then $\partial f(H) \subset \mathrm{dom}(f^*)$.

Another direct consequence of the Fenchel-Young Inequality (3.24) is:

Corollary 3.25. *If $f : H \to \mathbf{R}$ is convex and differentiable, then $\nabla f(H) \subset \mathrm{dom}(f^*)$ and*

$$f^*(\nabla f(x)) = \langle \nabla f(x), x \rangle - f(x)$$

for all $x \in H$.

Let us see some examples:

Example 3.26. Let $f : \mathbf{R} \to \mathbf{R}$ be defined by $f(x) = e^x$. Then

$$f^*(x^*) = \begin{cases} x^* \ln(x^*) - x^* & \text{if} \quad x^* > 0 \\ 0 & \text{if} \quad x^* = 0 \\ +\infty & \text{if} \quad x^* < 0. \end{cases}$$

This function is called the *Bolzmann-Shannon Entropy*.

Example 3.27. Let $f : H \to \mathbf{R}$ be defined by $f(x) = \langle x_0^*, x \rangle + \alpha$ for some $x_0^* \in H$ and $\alpha \in \mathbf{R}$. Then,

$$f^*(x^*) = \sup_{x \in X} \{ \langle x^* - x_0^*, x \rangle - \alpha \} = \begin{cases} +\infty & \text{if} \quad x^* \neq x_0^* \\ -\alpha & \text{if} \quad x^* = x_0^*. \end{cases}$$

In other words, $f^* = \delta_{\{x_0^*\}} - \alpha$.

Example 3.28. Let δ_C be the indicator function of a nonempty, closed convex subset C of H. Then,

$$\delta_C^*(x^*) = \sup_{x \in C} \{ \langle x^*, x \rangle \}.$$

This is the *support function* of the set C, and is denoted by $\sigma_C(x^*)$.

Example 3.29. The Fenchel conjugate of a radial function is radial: Let $\phi : [0,\infty) \to \mathbf{R} \cup \{+\infty\}$ and define $f : H \to \mathbf{R} \cup \{+\infty\}$ by $f(x) = \phi(\|x\|)$. Extend ϕ to all of $\mathbf{R}$ by $\phi(t) = +\infty$ for $t < 0$. Then

$$\begin{aligned} f^*(x^*) &= \sup_{x \in H} \{ \langle x^*, x \rangle - \phi(\|x\|) \} \\ &= \sup_{t \geq 0} \sup_{\|h\|=1} \{ t \langle x^*, h \rangle - \phi(t) \} \\ &= \sup_{t \in \mathbf{R}} \{ t \|x^*\| - \phi(t) \} \\ &= \phi^*(\|x^*\|). \end{aligned}$$

For instance, for $\phi(t) = \frac{1}{2} t^2$, we obtain $f^*(x^*) = \frac{1}{2} \|x^*\|^2$.

Example 3.30. Let $f : H \to \mathbf{R} \cup \{+\infty\}$ be proper and convex, and let $x \in \mathrm{dom}(f)$. For $h \in H$, set

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

(see Remark 2.13). Let us compute ϕ_x^*:

$$\begin{aligned} \phi_x^*(x^*) &= \sup_{h\in H} \{ \langle x^*, h\rangle - \phi_x(h) \} \\ &= \sup_{h\in H} \sup_{t>0} \left\{ \langle x^*, h\rangle - \frac{f(x+th)-f(x)}{t} \right\} \\ &= \sup_{t>0} \sup_{z\in H} \left\{ \frac{f(x) + \langle x^*, z\rangle - f(z) - \langle x^*, x\rangle}{t} \right\} \\ &= \sup_{t>0} \left\{ \frac{f(x) + f^*(x^*) - \langle x^*, x\rangle}{t} \right\}. \end{aligned}$$

By the Fenchel-Young Inequality (Proposition 3.24), we conclude that $\phi_x^*(x^*) = 0$ if $x^* \in \partial f(x)$, and $\phi_x^*(x^*) = +\infty$ otherwise. In other words, $\phi_x^*(x^*) = \delta_{\partial f(x)}(x^*)$.

Another easy consequence of the definition is:

Exercise 3.3. If $f \leq g$, then $f^* \geq g^*$. In particular,

$$\left(\sup_{i\in I}(f_i) \right)^* \leq \inf_{i\in I}(f_i^*) \qquad \text{and} \qquad \left(\inf_{i\in I}(f_i) \right)^* = \sup_{i\in I}(f_i^*)$$

for any family $(f_i)_{i\in I}$ of functions on H with values in $\mathbf{R} \cup \{+\infty\}$.

3.5 Fenchel-Rockafellar duality

Let X and Y be real Hilbert spaces, and let $A \in \mathscr{L}(X;Y)$. Consider two proper, lower-semicontinuous and convex functions $f : X \to \mathbf{R} \cup \{+\infty\}$ and $g : Y \to \mathbf{R} \cup \{+\infty\}$. The *primal problem* in Fenchel-Rockafellar duality is given by:

$$(P) \qquad \inf_{x\in X} f(x) + g(Ax).$$

Its optimal value is denoted by α and the set of *primal solutions* is S. Consider also the *dual problem*:

$$(D) \qquad \inf_{y\in Y} f^*(-A^*y) + g^*(y),$$

with optimal value α^*. The set of *dual solutions* is denoted by S^*.

By the Fenchel-Young Inequality (Proposition 3.24), for each $x \in X$ and $y \in Y$, we have $f(x) + f^*(-A^*y) \geq \langle -A^*y, x\rangle$ and $g(Ax) + g^*(y) \geq \langle y, Ax\rangle$. Thus,

$$f(x) + g(Ax) + f^*(-A^*y) + g^*(y) \geq \langle -A^*y, x\rangle + \langle y, Ax\rangle = 0 \tag{3.13}$$

for all $x \in X$ and $y \in Y$, and so $\alpha + \alpha^* \geq 0$. The *duality gap* is $\alpha + \alpha^*$.

Let us characterize the primal-dual solutions:

Theorem 3.31. *The following statements concerning points $\hat{x} \in X$ and $\hat{y} \in Y$ are equivalent:*

i) $-A^*\hat{y} \in \partial f(\hat{x})$ *and* $\hat{y} \in \partial g(A\hat{x})$;

ii) $f(\hat{x}) + f^*(-A^*\hat{y}) = \langle -A^*\hat{y}, \hat{x} \rangle$ *and* $g(A\hat{x}) + g^*(\hat{y}) = \langle \hat{y}, A\hat{x} \rangle$;

iii) $f(\hat{x}) + g(A\hat{x}) + f^*(-A^*\hat{y}) + g^*(\hat{y}) = 0$; *and*

iv) $\hat{x} \in S$ *and* $\hat{y} \in S^*$ *and* $\alpha + \alpha^* = 0$.

Moreover, if $\hat{x} \in S$ *and there is* $x \in \mathrm{dom}(f)$ *such that* g *is continuous in* Ax*, then there exists* $\hat{y} \in Y$ *such that all four statements hold.*

Proof. Statements $i)$ and $ii)$ are equivalent by the Fenchel-Young Inequality (Proposition 3.24) and, clearly, they imply $iii)$. But $iii)$ implies $ii)$ since

$$\left[f(x) + f^*(-A^*y) - \langle -A^*y, x \rangle \right] + \left[g(Ax) + g^*(y) - \langle y, Ax \rangle \right] = 0$$

and each term in brackets is nonnegative. Next, $iii)$ and $iv)$ are equivalent in view of (3.13). Finally, if $\hat{x} \in S$ and there is $x \in \mathrm{dom}(f)$ such that g is continuous in Ax, then

$$0 \in \partial(f + g \circ A)(\hat{x}) = \partial f(\hat{x}) + A^* \partial g(A\hat{x})$$

(see Exercise 3.17). Hence, there is $\hat{y} \in \partial g(A\hat{x})$ such that $-A^*\hat{y} \in \partial f(\hat{x})$, which is $i)$. □

For the *linear programming problem*, the preceding argument gives:

Example 3.32. Consider the *linear programming problem*:

$$(LP) \qquad \min_{x \in \mathbf{R}^N} \{ c \cdot x \, : \, Ax \leq b \},$$

where $c \in \mathbf{R}^N$, A is a matrix of size $M \times N$, and $b \in \mathbf{R}^M$. This problem can be recast in the form of (P) by setting $f(x) = c \cdot x$ and $g(y) = \delta_{R^M_+}(b - y)$. The Fenchel conjugates are easily computed (Examples 3.27 and 3.28) and the dual problem is

$$(DLP) \qquad \min_{y \in \mathbf{R}^M} \{ b \cdot y \, : \, A^*y + c = 0, \text{ and } y \geq 0 \}.$$

If (LP) has a solution and there is $x \in \mathbf{R}^N$ such that $Ax < b$, then the dual problem has a solution, there is no duality gap and the primal-dual solutions $(\hat{x}, \hat{y})$ are characterized by Theorem 3.31. Observe that the number of variables and constraints in (LP) and (DLP) are inverted.

3.6 The biconjugate

If f^* is proper, the *biconjugate* of f is the function $f^{**} : H \to \mathbf{R} \cup \{+\infty\}$ defined by

$$f^{**}(x) = \sup_{x^* \in H} \{ \langle x^*, x \rangle - f^*(x^*) \}.$$

Being a supremum of continuous affine functions, the biconjugate f^{**} is lower-semicontinuous and convex.

Remark 3.33. From the Fenchel-Young Inequality (Proposition 3.24), we deduce that $f^{**}(x) \le f(x)$ for all $x \in H$. In particular, f^{**} is proper. A stronger conclusion for lower-semicontinuous convex functions is given in Proposition 3.35. □

Example 3.34. Let $f : H \to \mathbf{R}$ be defined by $f(x) = \langle x_0^*, x\rangle + \alpha$ for some $x_0^* \in H$ and $\alpha \in \mathbf{R}$. We already computed $f^* = \delta_{\{x_0^*\}} - \alpha$ in Example 3.27. We deduce that

$$f^{**}(x) = \sup_{x^* \in X^*} \{\langle x^*, x\rangle - f^*(x^*)\} = \langle x_0^*, x\rangle + \alpha.$$

Therefore, $f^{**} = f$. □

The situation discussed in Example 3.34 for continuous affine functions is not exceptional. On the contrary, we have the following:

Proposition 3.35. *Let $f : H \to \mathbf{R} \cup \{+\infty\}$ be proper. Then, f is convex and lower-semicontinuous if, and only if, $f^{**} = f$.*

Proof. We already mentioned in Remark 3.33 that always $f^{**} \le f$. On the other hand, since f is convex and lower-semicontinuous, there exists a family $(f_i)_{i\in I}$ of continuous affine functions on H such that $f = \sup_{i\in I}(f_i)$, by Proposition 2.29. As in Example 3.3, we see that $f \le g$ implies $f^{**} \le g^{**}$. Therefore,

$$f^{**} \ge \sup_{i\in I}(f_i^{**}) = \sup_{i\in I}(f_i) = f,$$

because $f_i^{**} = f_i$ for continuous affine functions (see Example 3.34). The converse is straightforward, since $f = f^{**}$ is a supremum of continuous affine functions. □

Exercise 3.4. Let $f : H \to \mathbf{R} \cup \{+\infty\}$ be proper. Show that f^{**} is the greatest lower-semicontinuous and convex function below f.

Proof. If $g : H \to \mathbf{R} \cup \{+\infty\}$ is a lower-semicontinuous and convex function such that $g \le f$, then $g = g^{**} \le f^{**}$. □

Proposition 3.35 also gives the following converse to Proposition 3.6:

Proposition 3.36. *Let $f : H \to \mathbf{R} \cup \{+\infty\}$ be convex. Suppose f is continuous in x_0 and $\partial f(x_0) = \{x_0^*\}$. Then f is differentiable in x_0 and $\nabla f(x_0) = x_0^*$.*

Proof. As we saw in Propostion 2.14, the function $\phi_{x_0} : H \to \mathbf{R}$ defined by $\phi_{x_0}(h) = f'(x_0; h)$ is convex and continuous in H. We shall see that actually $\phi_{x_0}(h) = \langle x_0^*, h\rangle$ for all $h \in H$. First observe that

$$\phi_{x_0}(h) = \phi_{x_0}^{**}(h) = \sup_{x^* \in H} \{\langle x^*, h\rangle - \phi_{x_0}^*(x^*)\}.$$

Next, as we computed in Example 3.30, $\phi_{x_0}^*(x^*) = \delta_{\partial f(x_0)}(x^*)$. Hence,

$$\phi_{x_0}(h) = \sup_{x^* \in H} \{\langle x^*, h\rangle - \delta_{\{x_0^*\}}(x^*)\} = \langle x_0^*, h\rangle.$$

We conclude that f is differentiable in x_0 and $\nabla f(x_0) = x_0^*$. □

Combining the Fenchel-Young Inequality (Proposition 3.24) and Proposition 3.35, we obtain the following:

Proposition 3.37 (Legendre-Fenchel Reciprocity Formula)**.** *Let $f : H \to \mathbf{R} \cup \{+\infty\}$ be proper, lower-semicontinuous and convex. Then*

$$x^* \in \partial f(x) \qquad \textit{if, and only if,} \qquad x \in \partial f^*(x^*).$$

3.7 Constrained problems and optimality conditions

Let $C \subset H$ be closed and convex, and let $f : H \to \mathbf{R} \cup \{+\infty\}$ be proper, lower-semicontinuous and convex. The following result characterizes the solutions of the optimization problem

$$\min\{ f(x) : x \in C \}.$$

Proposition 3.38. *Let $f : H \to \mathbf{R} \cup \{+\infty\}$ be proper, lower-semicontinuous and convex and let $C \subset H$ be closed and convex. Assume either that f is continuous at some point of C, or that there is an interior point of C where f is finite. Then $\hat{x}$ minimizes f on C if, and only if, there is $p \in \partial f(\hat{x})$ such that $-p \in N_C(\hat{x})$.*

Proof. By Fermat's Rule (Theorem 3.10) and the Moreau-Rockafellar Theorem 3.16, $\hat{x}$ minimizes f on C if, and only if,

$$0 \in \partial(f + \delta_C)(\hat{x}) = \partial f(\hat{x}) + N_C(\hat{x}).$$

This is equivalent to the existence of $p \in \partial f(\hat{x})$ such that $-p \in N_C(\hat{x})$. □

Let C be a closed affine subspace of H, namely, $C = \{x_0\} + V$, where $x_0 \in H$ and V is a closed subspace of H. Then $N_C(\hat{x}) = V^\perp$. We obtain:

Corollary 3.39. *Let $C = \{x_0\} + V$, where $x_0 \in X$ and V is a closed subspace of H, and let $f : H \to \mathbf{R} \cup \{+\infty\}$ be a proper, lower-semicontinuous and convex function and assume that f is continuous at some point of C. Then $\hat{x}$ minimizes f on C if, and only if, $\partial f(\hat{x}) \cap V^\perp$.*

3.7.1 Affine constraints

Let X, Y be real Hilbert spaces, let $A \in \mathscr{L}(X;Y)$ and let $b \in Y$. We shall derive optimality conditions for the problem of minimizing a function $f : H \to \mathbf{R} \cup \{+\infty\}$ over a set C of the form:

$$C = \{ x \in X : Ax = b \}, \tag{3.14}$$

which we assume to be nonempty. Before doing so, let us recall that, given $A \in \mathscr{L}(X;Y)$, the adjoint of A is the operator $A^* : Y \to X$ defined by the identity

$$\langle A^* y, x \rangle_X = \langle y, Ax \rangle_Y,$$

for $x \in X$ and $y \in Y$. It is possible to prove (see, for instance, [1, Theorem 2.19]) that, if A has closed range, then $\ker(A)^\perp = \operatorname{ran}(A^*)$.

We have the following:

Theorem 3.40. *Let C be defined by* (3.14)*, where A has closed range. Let $f : X \to \mathbf{R} \cup \{+\infty\}$ be a proper, lower-semicontinuous and convex function and assume that f is continuous at some point of C. Then $\hat{x}$ minimizes f on C if, and only if, $A\hat{x} = b$ and there is $\hat{y} \in Y$ such that $-A^*\hat{y} \in \partial f(\hat{x})$.*

Proof. First observe that $N_C(\hat{x}) = \ker(A)^\perp$. Since A has closed range, $\ker(A)^\perp = \operatorname{ran}(A^*)$. We deduce that $-\hat{p} \in N_C(\hat{x}) = \operatorname{ran}(A^*)$ if, and only if, $\hat{p} = A^*\hat{y}$ for some $\hat{y} \in Y$, and conclude using Proposition 3.38. □

Exercise 3.5. Explain how this is related to (is a particular case of) Theorem 3.31.

3.7.2 Nonlinear constraints and Lagrange multipliers

Consider proper, lower-semicontinuous and convex functions $f, g_1, \ldots, g_m : H \to \mathbf{R} \cup \{+\infty\}$, along with continuous affine functions $h_1, \ldots, h_p : H \to \mathbf{R}$. We shall derive optimality conditions for the problem of minimizing f over the set C defined by

$$C = \{x \in H : g_i(x) \le 0 \text{ for all } i, \text{ and } h_j(x) = 0 \text{ for all } j\}, \tag{3.15}$$

assuming this set is nonempty. Since each g_i is convex and each h_j is affine, the set C is convex. To simplify the notation, write

$$S = \operatorname{argmin}\{ f(x) : x \in C\} \qquad \text{and} \qquad \alpha = \inf\{ f(x) : x \in C\}.$$

We begin by establishing an auxiliary result, which is a consequence of the Hahn-Banach Separation Theorem 1.30.

Lemma 3.41. *Let $\phi_0, \ldots, \phi_N : H \to \mathbf{R} \cup \{+\infty\}$ be convex functions and suppose*

$$\inf_{x \in H} \left[\max_{i=0,\ldots,N} \phi_i(x) \right] \ge 0. \tag{3.16}$$

Then, there exists $\lambda_0, \ldots, \lambda_N \ge 0$ (and not all zero) such that

$$\inf_{x \in H} \left[\lambda_0 \phi_0(x) + \cdots + \lambda_N \phi_N(x) \right] \ge 0. \tag{3.17}$$

Moreover, if $\max\limits_{i=0,\ldots,N} \phi_i(\bar{x}) = 0$, then $\lambda_i \phi_i(\bar{x}) = 0$ for $i = 0, \ldots, N$.

Proof. Let $D = \bigcap\limits_{i=0}^{N} \operatorname{dom}(\phi_i)$. If $D = \emptyset$, there is nothing to prove. If $D \ne \emptyset$, define

$$A = \bigcup_{x \in D} \left[\prod_{i=0}^{N} (\phi_i(x), +\infty) \right].$$

The set $A \subset \mathbf{R}^{N+1}$ is nonempty ($D \ne \emptyset$), open (union of open sets), convex (because each ϕ_i is convex), and does not contain the origin (by (3.16)). In view of the Hahn-Banach Separation Theorem 1.30, there exist $\lambda_0, \ldots, \lambda_N \in \mathbf{R}$ (and not all zero) such that

$$\lambda_0 z_0 + \cdots + \lambda_N z_N \ge 0 \tag{3.18}$$

for all $(z_0, \ldots, z_N) \in A$. If some $\lambda_i < 0$, the left-hand side of (3.18) can be made negative by taking z_i large enough. Therefore, $\lambda_i \ge 0$ for each i. Passing to the limit in (3.18), we obtain (3.17). Finally, if $\max\limits_{i=0,\ldots,N} \phi_i(\bar{x}) = 0$, then $\phi_i(\bar{x}) \le 0$ for all $i = 0, \ldots, N$ and so, $\lambda_i \phi_i(\bar{x}) = 0$, in view of (3.17). □

We begin by showing the following intermediate result, which is interesting in its own right:

Proposition 3.42. *There exist* $\lambda_0,\ldots,\lambda_m \geq 0$*, and* $\mu_1,\ldots\mu_p \in \mathbf{R}$ *(not all zero) such that*

$$\lambda_0\,\alpha \leq \lambda_0 f(x) + \sum_{i=1}^{m}\lambda_i g_i(x) + \sum_{j=1}^{p}\mu_j h_j(x)$$

for all $x \in X$.

Proof. If $\alpha = -\infty$, the result is trivial. Otherwise, set $N = m+2p$, $\phi_0 = f - \alpha$, $\phi_i = g_i$ for $i = 1,\ldots,m$, $\phi_{m+j} = h_j$ and $\phi_{m+p+j} = -h_j$ for $j = 1,\ldots,p$. Then, inequality (3.16) holds, in view of the definition of α. We conclude by applying Lemma 3.41 and regrouping the terms containing the affine parts. □

A more precise and useful result can be obtained under a *qualification condition*:

Slater's condition: There exists $x_0 \in \mathrm{dom}(f)$ such that $g_i(x_0) < 0$ for $i = 1,\ldots m$, and $h_j(x_0) = 0$ for $j = 1,\ldots p$.

Roughly speaking, this means that the constraint given by the system of inequalities is *thick* in the subspace determined by the linear equality constraints.

Corollary 3.43. *If Slater's condition holds, there exist* $\hat{\lambda}_1,\ldots,\hat{\lambda}_m \geq 0$*, and* $\hat{\mu}_1,\ldots,\hat{\mu}_p \in \mathbf{R}$*, such that*

$$\alpha \leq f(x) + \sum_{i=1}^{m}\hat{\lambda}_i g_i(x) + \sum_{j=1}^{p}\hat{\mu}_j h_j(x)$$

for all $x \in H$.

Proof. If $\lambda_0 = 0$ in Proposition 3.42, Slater's condition cannot hold. It suffices to divide the whole expression by $\lambda_0 > 0$ and rename the other variables. □

As a consequence of the preceding discussion we obtain the first-order optimality condition for the constrained problem, namely:

Theorem 3.44. *If* $\hat{x} \in S$ *and Slater's condition holds, there exist* $\hat{\lambda}_1,\ldots,\hat{\lambda}_m \geq 0$*, and* $\hat{\mu}_1,\ldots,\hat{\mu}_p \in \mathbf{R}$*, such that* $\hat{\lambda}_i g_i(\hat{x}) = 0$ *for all* $i = 1,\ldots,m$*, and*

$$0 \in \partial\left(f + \sum_{i=1}^{m}\hat{\lambda}_i g_i\right)(\hat{x}) + \sum_{j=1}^{p}\hat{\mu}_j \nabla h_j(\hat{x}). \tag{3.19}$$

Conversely, if $\hat{x} \in C$ *and there exist* $\hat{\lambda}_1,\ldots,\hat{\lambda}_m \geq 0$*, and* $\hat{\mu}_1,\ldots,\hat{\mu}_p \in \mathbf{R}$*, such that* $\hat{\lambda}_i g_i(\hat{x}) = 0$ *for all* $i = 1,\ldots,m$ *and* (3.19) *holds, then* $\hat{x} \in S$*.*

Proof. If $\hat{x} \in S$, the inequality in Corollary 3.43 is in fact an equality. Proposition 3.42 implies $\hat{\lambda}_i g_i(\hat{x}) = 0$ for all $i = 1, \ldots, m$. Fermat's Rule (Theorem 3.10) gives (3.19). Conversely, we have

$$\begin{aligned} f(\hat{x}) &\leq f(x) + \sum_{i=1}^{m} \hat{\lambda}_i g_i(x) + \sum_{j=1}^{p} \hat{\mu}_j h_j(x) \\ &\leq \sup_{\lambda_i \geq 0} \sup_{\mu_j \in \mathbf{R}} \left[f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \mu_j h_j(x) \right] \\ &= f(x) + \delta_C(x) \end{aligned}$$

for all $x \in H$. It follows that $\hat{x}$ minimizes f on C. □

Remark 3.45. By the Moreau-Rockafellar Theorem 3.16, the inclusion

$$0 \in \partial f(\hat{x}) + \sum_{i=1}^{m} \hat{\lambda}_i \partial g_i(\hat{x}) + \sum_{j=1}^{p} \hat{\mu}_j \nabla h_j(\hat{x})$$

implies (3.19). Moreover, they are equivalent if there is a point where $f, g_1, \ldots, g_m$ are all finite, and at most one of them is not continuous.

3.7.3 Lagrangian duality

For $(x, \lambda, \mu) \in H \times \mathbf{R}_+^m \times \mathbf{R}^p$, the *Lagrangian* for the constrained optimization problem studied before is

$$L(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \mu_j h_j(x).$$

Observe that

$$\sup_{(\lambda,\mu) \in \mathbf{R}_+^m \times \mathbf{R}^p} L(x, \lambda, \mu) = f(x) + \delta_C(x).$$

We shall refer to the problem of minimizing f over C as the *primal problem*. Its value is

$$\alpha = \inf_{x \in H} \left[\sup_{(\lambda,\mu) \in \mathbf{R}_+^m \times \mathbf{R}^p} L(x, \lambda, \mu) \right],$$

and S is the set of *primal solutions*. By inverting the order in which the supremum and infimum are taken, we obtain the *dual problem* (now in the sense of Lagrangian duality), whose value is

$$\alpha^* = \sup_{(\lambda,\mu) \in \mathbf{R}_+^m \times \mathbf{R}^p} \left[\inf_{x \in H} L(x, \lambda, \mu) \right].$$

The set S^* of points at which the supremum is attained is the set of *dual solutions* or *Lagrange multipliers*.

Clearly, $\alpha^* \leq \alpha$. The difference $\alpha - \alpha^*$ is the *duality gap* between the primal and dual problems.

We say $(\hat{x},\hat{\lambda},\hat{\mu}) \in H \times \mathbf{R}^m_+ \times \mathbf{R}^p$ is a saddle point of L if

$$L(\hat{x},\lambda,\mu) \leq L(\hat{x},\hat{\lambda},\hat{\mu}) \leq L(x,\hat{\lambda},\hat{\mu}) \tag{3.20}$$

for all $(x,\lambda,\mu) \in H \times \mathbf{R}^m_+ \times \mathbf{R}^p$.

Primal and dual solutions are further characterized by:

Theorem 3.46. *Let $(\hat{x},\hat{\lambda},\hat{\mu}) \in H \times \mathbf{R}^m_+ \times \mathbf{R}^p$. The following are equivalent:*

i) $\hat{x} \in S$, $(\hat{\lambda},\hat{\mu}) \in S^*$ and $\alpha = \alpha^*$;

ii) $(\hat{x},\hat{\lambda},\hat{\mu})$ is a saddle point of L; and

iii) $\hat{x} \in C$, $\hat{\lambda}_i g_i(\hat{x}) = 0$ for all $i = 1,\ldots,m$ and (3.19) holds.

Moreover, if $S \neq \emptyset$ and Slater's condition holds, then $S^* \neq \emptyset$ and $\alpha = \alpha^*$.

Proof. Let i) hold. Since $\hat{x} \in S$, we have

$$L(\hat{x},\lambda,\mu) \leq \sup_{\lambda,\mu} L(\hat{x},\lambda,\mu) = \alpha$$

for all $(\lambda,\mu) \in \mathbf{R}^m_+ \times \mathbf{R}^p$. Next,

$$\alpha^* = \inf_x L(x,\hat{\lambda},\hat{\mu}) \leq L(x,\hat{\lambda},\hat{\mu}).$$

for all $x \in H$. Finally,

$$\alpha^* = \inf_x L(x,\hat{\lambda},\hat{\mu}) \leq L(\hat{x},\hat{\lambda},\hat{\mu}) \leq \sup_{\lambda,\mu} L(\hat{x},\lambda,\mu) = \alpha.$$

Since $\alpha^* = \alpha$, we obtain (3.20), which gives ii). Now, suppose ii) holds. By (3.20), we have

$$f(\hat{x}) + \delta_C(\hat{x}) = \sup_{\lambda,\mu} L(\hat{x},\lambda,\mu) \leq \inf_x L(x,\hat{\lambda},\hat{\mu}) = \alpha^* < +\infty,$$

and so, $\hat{x} \in C$. Moreover,

$$f(\hat{x}) = f(\hat{x}) + \delta_C(\hat{x}) \leq L(\hat{x},\hat{\lambda},\hat{\mu}) = f(\hat{x}) + \sum_{i=1}^{m} \hat{\lambda}_i g_i(\hat{x}) \leq f(\hat{x}).$$

It follows that $\sum_{i=1}^{m} \hat{\lambda}_i g_i(\hat{x}) = 0$ and we conclude that each term is zero, since they all have the same sign. The second inequality in (3.20) gives (3.19). Next, assume iii) holds. By Theorem 3.44, $\hat{x} \in S$. On the other hand,

$$L(\hat{x},\hat{\lambda},\hat{\mu}) = f(\hat{x}) = \alpha \geq \alpha^* = \sup_{\lambda,\mu} \left[\inf_x L(x,\lambda,\mu) \right],$$

and $(\hat{\lambda},\hat{\mu}) \in S^*$. We conclude that $\alpha^* = \inf_x L(x,\hat{\lambda},\hat{\mu}) = L(\hat{x},\hat{\lambda},\hat{\mu}) = \alpha$. Finally, if $S \neq \emptyset$ and Slater's condition holds, by Theorem 3.44, there is $(\hat{\lambda},\hat{\mu})$ such that condition iii) in Theorem 3.46 holds. Then, $(\hat{\lambda},\hat{\mu}) \in S^*$ and $\alpha = \alpha^*$. $\square$

Remark 3.47. According to Theorem 3.46, if a Lagrange multiplier $(\hat{\lambda}, \hat{\mu})$ is known, we can recover the solution $\hat{x}$ for the original (primal) constrained problem as a solution of

$$\min_{x \in H} L(x, \hat{\lambda}, \hat{\mu}),$$

which is an unconstrained problem.

Exercise 3.6. Explain the relationship with Theorem 3.40.

Bibliography

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

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

[3] Peypouquet J. Convex Optimization in Normed Spaces: Theory, Methods and Examples. Springer; 2015.