Lecture Notes 2025

Optimisation, Control, and Data

Dario Prandi
dario.prandi@centralesupelec.fr

Contents

Part I

Optimisation

Chapter 1

Convex Analysis

This chapter closely follows chapter 5 of the lecture notes [2]. For most of the proof in this chapter we refer to [1] or [3].

1.1 Convex sets

Definition 1.1.1
A set $K \subset \mathbb{R}^d$ is *convex* if

$$tx + (1-t)y \in K \qquad \forall x, y \in K,\ t \in [0,1] \tag{1.1}$$

We have the following fact.

Proposition 1.1.1
The set $K \subset \mathbb{R}^d$ is convex if and only if for any $n \in \mathbb{N}$, and $t_1, \ldots, t_n \geq 0$ such that $\sum_{i=1}^n t_i = 1$, it holds

$$x_1, \ldots, x_n \in K \implies \sum_{i=1}^n t_i x_i \in K. \tag{1.2}$$

Proof: Property (1.2) with $n = 2$ is exactly the definition of K is convex. The statement then follows by induction on n. ■

Definition 1.1.2
The convex hull $\operatorname{conv}(\Omega)$ of $\Omega \subset \mathbb{R}^d$ is the smallest convex set K containing Ω.

By Proposition 1.1.1, it is immediate to observe that

$$\operatorname{conv}(\Omega) = \left\{ \sum_{i=1}^n t_i x_i \mid t_i \geq 0, \sum_{i=1}^n t_i = 1,\ x_i \in \Omega \right\}. \tag{1.3}$$

Example 1.1.1 (Convex sets)

- Unit ball w.r.t. any norm.
- Vector subspaces.
- Hyperplanes, i.e., for any $v \in \mathbb{R}^d$ and $\lambda \in \mathbb{R}$,

$$H_{v,\lambda} := \{x \in \mathbb{R}^d \mid \langle v, x \rangle \geq \lambda\}. \tag{1.4}$$

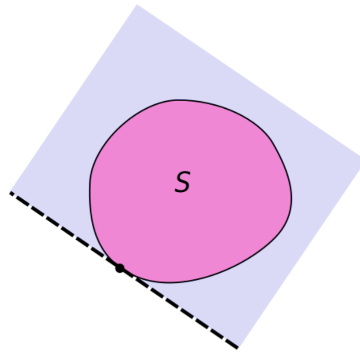

Figure 1.1: Supporting hyperplane for a set S.

An important result (that we will not prove) on convex sets is the following.

Theorem 1.1.1 Separation theorem

Let $K_1, K_2 \subset \mathbb{R}^d$ be two convex sets such with disjoint interior. Then there exists $v \in \mathbb{R}^d$ and $\lambda \in \mathbb{R}$ such that

$$K_1 \subset H_{v,\lambda}, \qquad K_2 \subset \mathbb{R}^d \setminus H_{v,\lambda}. \tag{1.5}$$

Here, $H_{v,\lambda}$ is defined in (1.4).

As a direct consequence, we have the following (see Figure 1.1).

Corollary 1.1.2 Supporting hyperplane theorem

Let $K \subset \mathbb{R}^d$ be a convex set and $x \in \partial K$. Then, there exists a supporting hyperplane of K containing x_0. That is, there exists $v \in \mathbb{R}^d$ and $\lambda \in \mathbb{R}$ such that $K \subset H_{v,\lambda}$ and $x_0 \in \partial H_{v,\lambda}$.

When K is a convex polygon, it is natural to expect it to be determined by its vertices. In order to formalize this intuition we need the following.

Definition 1.1.3

Let $K \subset \mathbb{R}^d$ be a convex set. A point $x \in K$ is an extremum of K if for any $y, z \in K$ and $t \in (0,1)$ we have that

$$x = ty + (1-t)z \implies x = y = z. \tag{1.6}$$

The set of exterma of K is denoted by $\operatorname{extr}(K)$.

In particular, for a convex polygon $\operatorname{extr}(K)$ is the set of its vertices.

Proposition 1.1.2

Let $K \subset \mathbb{R}^d$ be a convex set that is compact. Then,

$$\operatorname{conv}(K) = \operatorname{conv}\left(\operatorname{extr}(K)\right). \tag{1.7}$$

Figure 1.2: Two examples of polar cone.

1.2 Cones

Definition 1.2.1
A set $K \subset \mathbb{R}^d$ is a cone if

$$tx \in K \qquad \forall x \in K,\ t \geq 0. \tag{1.8}$$

Observe that every cone contains the origin.

Example 1.2.1 (Cones)

- The second order cone
$$C = \{x = (x', x_n) \in \mathbb{R}^d \times \mathbb{R} \mid \|x'\|_2 \leq x_n\}.$$
- Positive orthant $\mathbb{R}^d_+ = \{x \in \mathbb{R}^d \mid x_i \geq 0, \quad \forall i \in [\![1, d]\!]\}$.
- The set of positive semidefinite matrices $\mathrm{Sym}_+(\mathbb{R}^d)$.

Definition 1.2.2
The conic hull $\mathrm{cone}(\Omega)$ of a set $\Omega \subset \mathbb{R}^d$ is the smallest cone containing Ω. Namely,

$$\mathrm{cone}(\Omega) = \left\{ \sum_{i=1}^{n} t_i x_i \mid t_i \geq 0 \text{ and } x_i \in \Omega \text{ for any } i \in [\![1, n]\!] \right\}. \tag{1.9}$$

Definition 1.2.3
The polar cone K^* of a cone $K \subset \mathbb{R}^d$ is the set

$$K^* := \left\{ y \in \mathbb{R}^d \mid \rangle x, y \langle \geq 0 \quad \forall x \in K \right\}. \tag{1.10}$$

We have the following properties for the polar cone.

Proposition 1.2.1
The polar cone K^* is a closed, convex cone. If, moreover, the cone K is closed, then $K^* * = K$.

1.3 Convex functions

We will work with extended functions $F : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$. The domain of an extended function is

$$\mathrm{dom}(F) = \{x \in \mathbb{R}^d \mid F(x) < +\infty\}. \tag{1.11}$$

An extended function such that $\mathrm{dom}(F) \neq \varnothing$ is called *proper*.

Given a standard function $F : \Omega \longrightarrow \mathbb{R}$, we can identify it with the extended function $\bar{F} : \mathbb{R}^d \longrightarrow \mathbb{R} \cup \{+\infty\}$ defined by

$$\bar{F}(x) = \begin{cases} F(x) & \text{if } x \in \Omega, \\ +\infty & \text{if } x \in \mathbb{R}^d \setminus \Omega. \end{cases} \tag{1.12}$$

Definition 1.3.1 Convex functions

Let $F : \mathbb{R}^d \longrightarrow \mathbb{R} \cup \{+\infty\}$ be an extended function. Then,

- F is convex if
$$F(tx + (1-t)y) \leq tF(x) + (1-t)F(y) \qquad \forall x, y \in \mathbb{R}^d,\ t \in [0,1]. \tag{1.13}$$
- F is strictly convex if
$$F(tx + (1-t)y) < tF(x) + (1-t)F(y) \qquad \forall x, y \in \mathbb{R}^d,\ x \neq y,\ t \in [0,1]. \tag{1.14}$$
- F is strongly convex if there exists $\gamma > 0$ such that
$$F(tx + (1-t)y) \leq tF(x) + (1-t)F(y) - \frac{\gamma}{2} t(t-1)\|x-y\|_2^2 \qquad \forall x, y \in \mathbb{R}^d,\ t \in [0,1]. \tag{1.15}$$

We say that F is *concave* if $-F$ is convex.

Observe that it holds
$$\text{convex} \Longleftarrow \text{strongly convex} \Longleftarrow \text{strictly convex} \tag{1.16}$$

We say that a standard function $F : K \longrightarrow \mathbb{R}$ is convex, strictly convex, strongly convex, or concave, if the same is true for its extension $\bar{F}$. Observe that this requires K to be convex.

Example 1.3.1

- The prototypical convex function, used in the definition of strongly convex, is the quadratic function
$$F(x) = \frac{\|x\|_2^2}{2} = \frac{1}{2}\sum_{i=1}^{d} |x_i|^2. \tag{1.17}$$
- More generally, every norm is convex.
- The norm ℓ_p is strictly convex if and only if $p \in (1, +\infty)$.
- $F(x) = x^\top A x$ is convex if A is positive semidefinite (i.e., $A \in \mathrm{Sym}_{\geq 0}(\mathbb{R}^d)$), and strongly convex if A is positive definite.

Proposition 1.3.1

A function $F : K \longrightarrow \mathbb{R}$ is convex if and only if its epigraph $\mathrm{epi}(F) \subset \mathbb{R}^{d+1}$ is convex. Here, we let
$$\mathrm{epi}(F) = \{(x, r) \mid r \geq F(x)\}. \tag{1.18}$$

Proof: Assume F is convex and let $(x, r), (y, s) \in \mathrm{epi}(F)$. In particular, $r \geq F(x)$ and $s \geq F(y)$. Let $t \in [0,1]$ and observe that
$$tr + (1-t)s \geq tF(x) + (1-t)F(y) \geq F(tx + (1-t)y). \tag{1.19}$$
Hence, $t(x, r) + (1-t)(y, s) \in \mathrm{epi}(F)$. A similar reasoning proves the opposite implication. ■

Figure 1.3: Epigraph of a function.

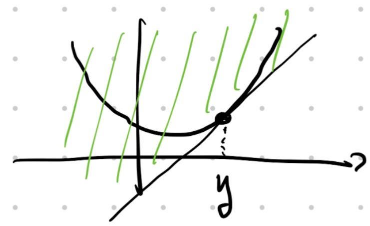

Figure 1.4: Graphical representation of Proposition 1.3.2.

Proposition 1.3.2 Differential characterisations of convexity

Let $F : \mathbb{R}^d \longrightarrow \mathbb{R}$ be an everywhere differentiable function. Then,

- F is convex if and only if
$$F(y) \geq F(x) + \langle \nabla F(x), y - x \rangle, \qquad \forall x, y \in \mathbb{R}^d. \tag{1.20}$$
- F is strongly convex with parameter $\gamma > 0$ if and only if
$$F(y) \geq F(x) + \langle \nabla F(x), y - x \rangle + \frac{\gamma}{2} \|x - y\|_2^2, \qquad \forall x, y \in \mathbb{R}^d. \tag{1.21}$$
- If F is everywhere twice differentiable, then it is convex if and only if
$$\operatorname{Hess} F(x) \geq 0 \qquad \forall x \in \mathbb{R}^d. \tag{1.22}$$
Here, we denoted by $\operatorname{Hess} F(x)$ the Hessian of F at x.

Proposition 1.3.3

Let $F : K \longrightarrow \mathbb{R}$ be convex. Then, F is continuous on the interior of K.

Proof: Let $x_0 \in \operatorname{int}(K)$ and consider $r > 0$ such that $B(x_0, r) \subset K$. Without loss of generality, we assume $x_0 = 0$ (otherwise, replace the function F by its translation $G(x) = F(x) - F(x_0)$).

Convexity will allow to bound the difference $F(y) - F(0)$ with the values of F on the sphere $\partial B(0, r)$. However, without continuity, the function F need not be bounded on the compact set $\partial B(0, r)$, and hence we need some additional care.

Pick $d + 1$ linearly independent points $v_0, \ldots v_{d+1} \in \partial B(0, r)$, and consider the corresponding symplex

$$\Delta = \operatorname{conv}\left(\{v_0, \ldots, v_{d+1}\}\right) = \left\{ \sum_{i=1}^{d+1} t_i v_i \mid t_i \geq 0, \sum_i t_i = 1 \right\} \subset B(0, r). \tag{1.23}$$

Then, letting $M = \max_{i \in [\![1, d+1]\!]} F(v_i)$, the fact that F is convex yields that for any $x = \sum_{i=1}^{d+1} t_i v_i \in \Delta$ it holds

$$F(x) \leq \sum_{i=1}^{d+1} t_i F(v_i) \leq M. \tag{1.24}$$

In particular, we can fix a radius $r' < r$ such that $B(0, r') \subset \Delta$ where F is bounded.

We now proceed to bound the difference $F(x) - F(0)$. Let $x \in U \subset B(0, r')$ and set $t = \|x\|/r'$. In particular, $t \in [0, 1]$ and the ray $\{sx \mid s \geq 0$ meets the sphere $\partial B(x_0, r')$ at the point

$$y = \frac{r'}{\|x\|}(x). \tag{1.25}$$

In particular, $x = (1 - t)0 + ty$. By convexity and (1.24), we have

$$F(x) \leq (1 - t)F(0) + tF(y) \leq .(1 - t)F(0) + tM \implies F(x) - F(0) \leq t(M - F(0)). \tag{1.26}$$

To derive a bound from below, we proceed similarly, considering

$$z = \frac{r'}{\|x\| - r'} x. \tag{1.27}$$

Indeed, we then have $0 = (1-t)x + tx$, where $t = \|x\|/r'$ as above. Then, convexity and the fact that $z \in B(0, r')$ yield

$$F(0) \leq (1-t)F(x) + tM \implies F(x) - F(0) \geq -\frac{t}{1-t}(M - F(0)). \tag{1.28}$$

Combining (1.26) and (1.28), we obtain

$$-\frac{t}{1-t}(M - F(0)) \leq F(x) - F(0) \leq t(M - F(0)), \qquad t = \frac{\|x\|}{r'} \tag{1.29}$$

Since x was arbitrary in $B(0, r')$ we can take the limit as $x \to 0$, which implies $t \to 0$ and thus that

$$\lim_{x \to 0} |F(x) - F(0)| = 0, \tag{1.30}$$

concluding the proof. ■

The following result is at the core of the relation between optimisation and convexity.

Theorem 1.3.1

Let $F : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ be a convex extended function. Then,

- Any local minimum of F is global.
- The set of minima of F is convex.
- If F is strictly convex and admits a minimum, this minimum is unique.
- If F is real-valued and strongly convex, then it has a unique minimum.

Proof: Assume that $x^\star$ is a local minimum, i.e., there exists $r > 0$ such that $F(x^\star) \leq F(x)$ for any $x \in B(0, r)$. Let $y \in \mathbb{R}^d$ and consider a point on the ray starting at $x^\star$ and passing through y:

$$z = x^\star + s(y - x^\star) = (1-s)x^\star + sy \qquad s \geq . \tag{1.31}$$

Taking $s < \min\{1, r'/\|x^\star - y\|\}$ we have that $z \in B(0, r)$. Hence, by local minimality of $x^\star$ and convexity of F we have

$$F(x^\star) \leq F(z) \leq (1-s)F(x^\star) + sF(y) \implies F(x^\star) \leq F(y). \tag{1.32}$$

This concludes the proof of the first point.

Assume now that x_1, x_2 are minima for F. This clearly implies that $F(x_1) = F(x_2) =: m$, and thus, by convexity of F, for any $t \in [0, 1]$ we have

$$m \leq F(tx_1 + (1-t)x_2) \leq tF(x_1) + (1-t)F(x_2) = m \implies F(tx_1 + (1-t)x_2). \tag{1.33}$$

This implies that $tx_1 + (1-t)x_2$ is a minimum for any $t \in [0, 1]$, thus proving the second point.

The same argument as above in the case of a strictly convex function yields to

$$m \leq F(tx_1 + (1-t)x_2) < m \qquad \text{if } x_1 \neq x_2. \tag{1.34}$$

This implies immediately that the minimum is unique.

Assume, finally, that F is strongly convex. Since it is strictly convex, we just need to prove the existence of a minimum. By Proposition 1.3.3 we have that F is continuos, and thus it suffices to prove its coercivity: $F(x) \to +\infty$ if $\|x\| \to +\infty$. We provide a proof of this fact in the case where F is differentiable (the general case can be obtained similarly using Proposition 1.4.2, proven later on). In this case, by Proposition 1.3.2 we have that

$$F(y) \geq F(0) + \langle \nabla F(0), y \rangle + \frac{\gamma}{2}\|y\|_2^2 \qquad \forall y \in \mathbb{R}^d. \tag{1.35}$$

Since $\langle \nabla F(0), y \rangle \leq \|y\|_2$, the quadratic term on the right-hand side of the above equation, implies that the limit as $\|y\|_2 \to +\infty$ is $+\infty$. ■

Remark: Strict convexity is not enough to ensure the existence of a minimum. Consider, for example, $F(x) = e^x$.

1.4 Convex conjugate and sub-differential

Definition 1.4.1

The convex conjugate (of Fenchel dual) of an extended function $F : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is the function $F^* : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ defined by

$$F^*(y) = \sup_{x \in \mathbb{R}^d} \left[\langle x, y\rangle - F(x)\right]. \tag{1.36}$$

Recall the following.

Definition 1.4.2

A function $F : \mathbb{R}^d \to \mathbb{R}$ is *lower semicontinuous* (l.s.c.) if

$$\liminf_{x \to x_0} F(x) \geq F(x_0), \qquad \forall x_0 \in \mathbb{R}^d. \tag{1.37}$$

Equivalently, F is l.s.c. if its epigraph is closed.

Example 1.4.1

- Every continuous function is lower semicontinuous.
- For any set $\Omega \subset \mathbb{R}^d$, the $0 - \infty$ characteristic function

$$\chi_K = \begin{cases} 0 & \text{if } x \in \Omega, \\ +\infty & \text{otherwise,} \end{cases} \tag{1.38}$$

  is lower semicontinuous, but not continuous.

Proposition 1.4.1

Let $F : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$. Then,

1. The convex conjugate F^* is a lower semicontinuous convex function.
2. We have the Fenchel (or Young, or Fenchel-Young) inequality

$$\langle x, y\rangle \leq F(x) + F^*(y) \tag{1.39}$$

Proof: For any $y_1, y_2 \in \mathbb{R}^d$ and $t \in [0, 1]$ we have

$$\langle x, ty_1 + (1-t)y_2\rangle - F(x) = t\left(\langle x, y_1\rangle - F(x)\right) + (1-t)\left(\langle x, y_2\rangle - F(x)\right). \tag{1.40}$$

Taking the supremum for $x \in \mathbb{R}^d$ of the above, and recalling that $\sup(g(x) + h(x)) \leq \sup g(x) + \sup h(x)$ proves convexity of F^*.

Lower semicontinuity of F^* follows since it is the supremum for $x \in \mathbb{R}^d$ of $g_x(y) := \langle x, y\rangle - F(x)$, which is affine and in particular lower semicontinuous. Indeed, the supremum of a family of l.s.c. functions is l.s.c..

The second point (Fenchel inequality) is a direct consequence of the definition of F^*. ■

Example 1.4.2

- Let $F(x) = \frac{1}{2}\|x\|_2^2$. Then, $F^*(y) = \frac{1}{2}\|y\|_2^2 = F(y)$. This is the only function with this property.

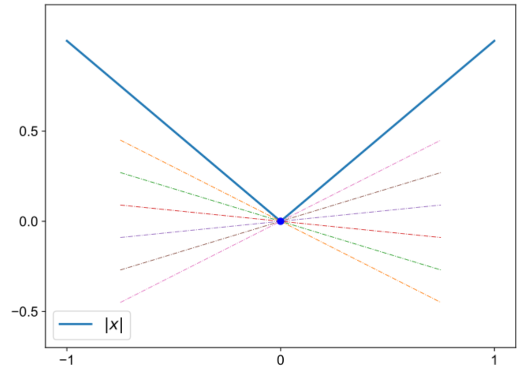

Figure 1.5: Visualization of the subgradients of $F(x) = |x|$ at $x = 0$. Image from this website.

- Let $F = \chi_K$ be the $0 - \infty$ characteristic function of a convex set $K \subset \mathbb{R}^d$ defined in (1.38). Then,

$$F^*(y) = \sup_{x \in K} \langle x, y \rangle. \tag{1.41}$$

Definition 1.4.3

The *subdifferential* of a convex extended function $F : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ at $x \in \mathbb{R}^d$ is the set

$$\partial F(x) = \left\{ v \in \mathbb{R}^d \mid F(y) \geq F(x) + \langle v, y - x \rangle, \qquad \forall y \in \mathbb{R}^d \right\}. \tag{1.42}$$

A vector $v \in \partial F(x)$ is called a *subgradient* for F at x.

Example 1.4.3

Consider $F(x) = |x|$. Then,

$$\partial F(x) = \begin{cases} \{\mathrm{sgn}(x)\} & \text{if } x \neq 0, \\ [-1, 1] & \text{if } x = 0. \end{cases} \tag{1.43}$$

Here, $\mathrm{sgn}(x) = x/|x|$ is the sign function. See Figure 1.5.

Theorem 1.4.1

Let $F : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ be a convex function. Then, $x \in \mathbb{R}^d$ is a minimum for F if and only if $0 \in \partial F(x)$

Proof: The fact that x is a minimum means that $F(x) \leq F(y)$ for any $y \in \mathbb{R}^d$, which is the definition of $0 \in \partial F(x)$. ■

We have the following.

Proposition 1.4.2

Let $F : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ be a convex function. Then,

- For any $x \in \mathbb{R}^d$ the subdifferential $\partial F(x)$ is non-empty.
- It holds that
$$\partial F(x) = \left\{ v \in \mathbb{R}^d \mid F^*(v) + F(x) = \langle x, v \rangle \right\}. \tag{1.44}$$
- If F is differentiable at $x \in \mathbb{R}^d$, then $\partial F(x) = \{\nabla F(x)\}$.

Proof: The first part of the theorem is a consequence of the Supporting Hyperplane Theorem (see Corollary 1.1.2) and Proposition 1.3.1. Indeed, the latter implies that the epigraph epi(F) is convex and hence, by the former, any of its boundary point admits a supporting hyperplane. Using the fact that $\partial\,\mathrm{epi}(F) = \{(x, F(x)) \mid x \in \mathbb{R}^d\}$ allows to conclude.

To prove the second statement, observe that $v \in \partial F(x)$ is equivalent to

$$\langle y, v\rangle - F(y) \leq \langle x, v\rangle - F(x), \qquad \forall y \in \mathbb{R}^d. \tag{1.45}$$

Taking the sup for $y \in \mathbb{R}^d$ yields that $F^*(v) \leq \langle x, v\rangle - F(x)$. The opposite inequality follows from Fenchel inequality (see Proposition 1.4.1).

Concerning the proof of the last statement, the fact that $\nabla F(x) \in \partial F(x)$ follows from the characterisation of convexity for differentiable functions given in Proposition 1.3.2. To prove the opposite implication, let $v \in \partial F(x)$ and observe that by definition of subgradient the directional derivative $\partial_h F(x)$ of f in the direction $h \in \mathbb{R}^d$ at x satisfies

$$\partial_h F(x) = \lim_{t\to 0} \frac{F(x+th) - F(x)}{t} \geq \langle v, h\rangle. \tag{1.46}$$

Since we know that $\partial_h F(x) = \langle \nabla F(x), h\rangle$, we have that

$$\langle \nabla F(x) - v, h\rangle \geq 0, \qquad \forall h \in \mathbb{R}^d. \tag{1.47}$$

But this implies that $\nabla F(x) = v$, concluding the proof. ■

Thanks to the previous result, we are in a position to prove the following property of the convex biconjugate.

Theorem 1.4.2 Fenchel-Moreau Theorem

The biconjugate F^** is the largest convex lower semicontinuous function satisfying $F^* * (x) \leq F(x)$ for any $x \in \mathbb{R}^d$. In particular, $F^** = F$ if F is convex and proper.

Proof: We have that $-F^*(y) = \inf_{x\in\mathbb{R}^d} (F(x) - \langle x, y\rangle)$, which implies that for any $y, z \in \mathbb{R}^d$ it holds

$$\langle z, y\rangle - F^*(y) \leq \langle z - x, y\rangle + F(x), \qquad \forall x \in \mathbb{R}^d. \tag{1.48}$$

In particular, considering $z = x$ we have

$$F^* * (x) = \sup_{y\in\mathbb{R}^d} (\langle x, y\rangle - F^*(y)) \leq F(x), \tag{1.49}$$

proving the first part of the statement.

Since F^{**} is convex and l.s.c. by Proposition 1.4.1, in order to complete the proof it suffices to show that if F is convex, then

$$F^* * (x) \geq F(x), \qquad \forall x \in \mathbb{R}^d. \tag{1.50}$$

Let $v \in \partial F(x)$, which exists thanks to Proposition 1.4.2. For such a v, using the characterisation of the subdifferential in Proposition 1.4.2, we have

$$F^*(v) = \langle x, v\rangle - F(x), \tag{1.51}$$

so that $F^* * (z) \geq \langle v, z - x\rangle + F(x)$ for any $z \in \mathbb{R}^d$. Picking $z = x$ allows to conclude. ■

Proposition 1.4.3

Let $F : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ be a convex function and $x, y \in \mathbb{R}^d$. Then, the following are equivalent:

i. $y \in \partial F(x)$.

ii. $F(x) + F^*(y) = \langle x, y\rangle$.

If, additionally, F is l.s.c. , then the above are also equivalent to

iii. $x \in \partial F^*(y)$.

Proof: To show that i is equivalent to ii, we just need to show that $y \in \partial F(x)$ is equivalent to

$$F(x) + F^*(y) \leq \langle x, y \rangle. \tag{1.52}$$

Indeed, the opposite inequality is always true due to Fenchel's inequality (see Proposition 1.4.1).

Observe that the fact that $y \in \partial F(x)$ means that

$$\langle x, y \rangle F(x) \geq \langle z, y \rangle F(z), \qquad \forall z \in \mathbb{R}^d. \tag{1.53}$$

That is, the function $z \mapsto \langle z, y \rangle F(z)$ attains its maximum at $z = x$. But, by definition of F^*, this is equivalent to (1.52), thus proving that i is equivalent to ii.

To complete the proof, observe that by Theorem 1.4.2 the lower semicontinuity of F yield that $F^* * = F$, so that ii is equivalent to $F^* * (x) + F^*(y) = \langle x, y \rangle$. Using the fact that $i \iff ii$ with F replaced by F^* completes the proof. ∎

1.5 Convex optimization problems

Definition 1.5.1

An optimization problem is a minimization problem of the form

$$\min_{x \in \mathbb{R}^d} F_0(x) \qquad \text{subject to} \qquad Ax = y \qquad \text{and} \qquad F_j(x) \leq 0, \qquad j \in [\![1, M]\!]. \tag{OP}$$

Here,

1. $F_0 : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is the *objective function*;
2. $F_1, \ldots, F_M : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ are the *constraing functions*;
3. $A \in \mathbb{R}^{m \times n}$ and $y \in \mathbb{R}^m$ provide the *equality constraints*;

The optimization problem is *convex* (resp. *linear*) if $F_0, \ldots, F_M$ are convex (resp. linear) functions.

Definition 1.5.2

Consider an optimization problem (OP). Then,

- The set $\Phi \subset \mathbb{R}^d$ of points $x \in \mathbb{R}^d$ satisfying the constraints is the set of *feasible points*. That is,

$$\Phi = \left\{ x \in \mathbb{R}^d \mid Ax = y, \qquad F_j(x) \leq 0 \qquad \forall j \in [\![1, M]\!] \right\}. \tag{1.54}$$

  In particular, Φ is convex if (OP) is convex.
- Problem (OP) is *feasible* if it admits at least a feasible point (i.e., $\Phi \neq \emptyset$).
- The *optimal value* is $p^\star = \min_{x \in \Phi} F(x_0)$.
- A *minimizer* is a feasible point $x^\star$ such that $F_0(x^\star) \leq F_0(x)$ for all feasible $x \in \Phi$. That is, $F_0(x^\star) = p^\star$.

Observe that the constrained optimization problem (OP) is equivalent to the uncostrained optimization problem

$$\min_{x \in \mathbb{R}^d} F_0(x) + \chi_\Phi, \tag{1.55}$$

where χ_Φ is the $0 - \infty$ characteristic function defined in (1.38).

Let us introduce the notation

$$\mathbb{R}^M = \{ \nu \in \mathbb{R}^M \mid \nu_j \geq 0 \quad \forall j \in [\![1, M]\!] \}. \tag{1.56}$$

Definition 1.5.3 Lagrange and Lagrange dual functions

The *Lagrange function* of the optimization problem (OP) is the function $F : \mathbb{R}^d \times \mathbb{R}^m \times \mathbb{R}^M_+ \to \mathbb{R} \cup \{+\infty\}$ defined by

$$L(x, \xi, \nu) = F_0(x) + \langle \xi, Ax - y \rangle + \sum_{j=1}^{m} \nu_j F_j(x). \tag{1.57}$$

The *Lagrange dual function* is the function $H : \mathbb{R}^m \times \mathbb{R}^M_+ \to \mathbb{R} \cup \{-\infty\} \cup \{+\infty\}$, defined by

$$H(\xi, \nu) = \inf_{x \in \mathbb{R}^d} L(x, \xi, \nu). \tag{1.58}$$

Proposition 1.5.1

The dual function is always concave. Moreover, if $x^\star$ is a minimizer of (OP), we have

$$H(\xi, \nu) \le F(x^\star), \qquad \forall \xi \in \mathbb{R}^m,\ \nu \in \mathbb{R}^M_+. \tag{1.59}$$

Proof: Observe that $-H$ is the supremum w.r.t. $x \in \mathbb{R}^d$ of the functions $g_x(\xi, \nu) = -L(x, \xi, \nu)$. The function g_x is affine, and thus convex. Hence, $-H$ is the pointwise supremum of the family $\{g_x\}_{x \in \mathbb{R}^d}$ of convex function. It is immediate to check that it is convex, and thus that H is concave.

On the other hand, for any feasible point $x \in \Phi$, since $\nu_j \ge 0$ for any $j \in [\![1, M]\!]$, we have

$$\langle \xi, Ax - y \rangle + \sum_{j=1}^{m} \nu_j F_j(x) \le 0. \tag{1.60}$$

Then, $L(x, \xi, \nu) \le F_0(x) \le F_0(x^\star)$ and, as a consequence,

$$H(\xi, \nu) \le \inf_{x \in \Phi} L(x, \xi, \nu) \le F_0(x^\star). \tag{1.61}$$

This completes the proof of the statement. ∎

The previous result suggests to introduce the following.

Definition 1.5.4 Primal and dual problem

The *dual problem* to (OP), which is called the *primal problem*, it the optimization problem

$$\max_{\xi \in \mathbb{R}^m,\ \nu \in \mathbb{R}^M} H(\xi, \nu) \qquad \text{subject to} \qquad \nu_j \ge 0 \quad \forall j \in [\![1, M]\!]. \tag{DP}$$

- A pair $(\xi, \nu) \in \mathbb{R}^m \times \mathbb{R}^M_+$ is called *dual feasible*.
- The *dual optimal value* is the solution $d^\star$ of (DP).
- A *dual optimal* or *optimal Lagrange multiplier* is a feasible maximizer $(\xi^\star \nu^\star) \in \mathbb{R}^m \times \mathbb{R}^M_+$.
- A *primal-dual optimal* is a triple $(x^\star, \xi^\star, \nu^\star)$ where $x^\star$ is a minimizer for (OP) and $(\xi^\star, \nu^\star)$ is a dual optimal.

Definition 1.5.5 Duality

The primal-dual problems always satisfy *weak duality*, that is $d^\star \le p^\star$ where $d^\star$ is the dual optimal value and $p^\star$ is the primal optimal value.
We say that the problems enjoy *strong duality* if it holds

$$p^\star = d^\star. \tag{1.62}$$

The above shows the interest of the dual problem: when strong duality holds, in order to solve the minimization problem (OP) it suffices to solve the dual problem (DP).

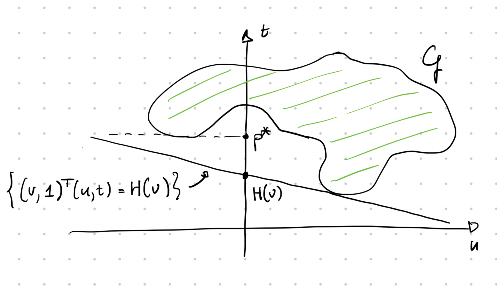

Figure 1.6: Geometric interpretation. The value of the dual function $H(v)$ identifies a supporting hyperplane for the set $\mathcal{G}$.

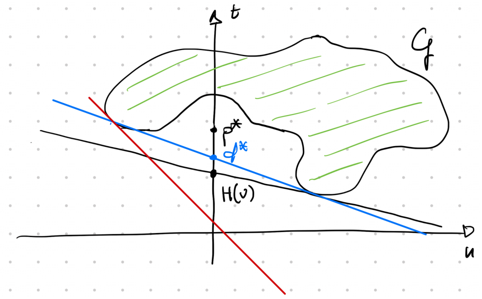

Figure 1.7: Geometric interpretation. Solving the dual problem yields the blue hyperplane. In this case $p^\star > d^\star$ and strong duality does not hold.

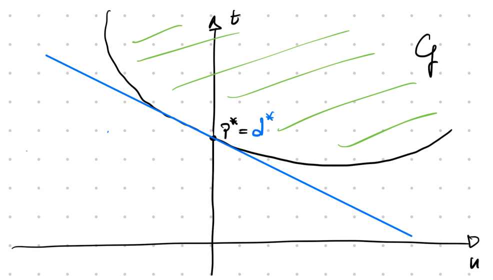

Figure 1.8: Geometric interpretation of Slater's condition. When the set $\mathcal{G}$ is convex and has interior that intersects the left-hand plane, the best supporting hyperplane yields the optimal value $p^\star$.

The following is the most used criterion for strong duality.

Theorem 1.5.1 Slater's constraint quantification

Assume that $F_0, \ldots, F_M$ are convex functions with domain $\mathrm{dom}(\mathbb{R}^d)$ and that $F_0(x) \geq -c$ for some $c \geq 0$. Then, strong duality holds if there exists $x \in \Phi \subset \mathbb{R}^d$ such that $F_j(x) < 0$ for any $j \in [\![1, M]\!]$.

For a proof of the above result, we refer to [1, Section 5.3.2].

Geometric interpretation

Let us follow [1, Section 5.3] and present a geometric interpretation of the previous discussion. Assume that there are no equality constraints and a single inequality constraint, and define

$$\mathcal{G} = \{(F_1(x), F_0(x)) \mid x \in \mathbb{R}^d\}. \tag{1.63}$$

By construction, the problem is feasible if and only if $\mathcal{G}$ intersects the left-half plane. Furthermore, we have

$$p^\star = \min\{t \mid (u,t) \in \mathcal{G},\, u \leq 0\}. \tag{1.64}$$

Since $L(x, v) = (v, 1)^\top (F_1(x), F_0(x))$, we also have

$$H(v) = \inf\{(v,1)^\top (u,t) \mid (u,t) \in \mathcal{G}\}. \tag{1.65}$$

Hence, if this infimum is finite, the inequality $(v,1)^\top (u,t) \geq H(v)$ defines a supporting hyperplane for $\mathcal{G}$.

If the problem is convex, then $\mathcal{G}$ is convex and under Slater's condition its interior intersects the left-hand plane. This insures that strong duality holds.

Bibliography

[1] Stephen P. Boyd and Lieven Vandenberghe. *Convex Optimization*. Version 29. Cambridge New York Melbourne New Delhi Singapore: Cambridge University Press, 2023. 716 pp. ISBN: 978-0-521-83378-3.

[2] Massimo Fornasier. “Foundations of Data Analysis”.

[3] Ralph Tyrell Rockafellar. *Convex Analysis*. Princeton Landmarks in Mathematics and Physics. Princeton: Princeton University Press, 2015. 470 pp. ISBN: 978-0-691-01586-6.