Convex Analysis and Optimization

1.1 Motivational example

Consider a rocket at rest on the surface of the Earth at time $t = 0$. The rocket begins by flying straight up until it reaches an altitude of h feet within a given time $t = T$. We are concerned with the gas flow strategies that minimize the gas consumption on the interval of time $[0, T]$

$$\int_0^T |u(t)|dt$$

The equation of motion of the rocket can be modeled via:

$$\ddot{x}(t) = u(t) - g, \quad t \in (0, T)$$

with $x(0) = \dot{x}(0) = 0$ and $x(T) = h$. The natural space to look for minimizers is $L^1([0, T])$, and so the problem can be formulated as follows

$$(\mathcal{P}) \qquad \inf_{u \in L^1([0,T])} \left\{ \int_0^T |u(t)|dt \,\middle|\, \int_0^T (T-t)u(t)dt = h + \frac{1}{2}gT^2 \right\}$$

However $L^1([0, T])$ is neither reflexive nor the (topological) dual of another space. This fact yields to compactness issues that don't allow to guarantee the existence of solutions to $(\mathcal{P})$.

To overcome the lack of appropriate compactness properties of $L^1([0, T])$ we embed this space into a larger one, the space of Radon measures $\mathcal{M}([0, T])$. Recall that the Riesz representation theorem implies that $\mathcal{M}([0, T])$ is isometrically isomorphic to the topological dual to the space of continuous functions $\mathcal{C}([0, T])$. Therefore the relaxed problem is

$$(\widetilde{\mathcal{P}}) \qquad \inf_{\mu \in \mathcal{M}([0,T])} \left\{ \int_0^T |d\mu(t)| \,\middle|\, \int_0^T (T-t)d\mu(t) = h + \frac{1}{2}gT^2 \right\}$$

Problem $(\widetilde{\mathcal{P}})$ is well-posed and it can be studied in the light of Convex Analysis. This theory will allow us to show that:

- $(\widetilde{\mathcal{P}})$ has a unique solution and that

$$\mathrm{val}(\widetilde{\mathcal{P}}) = \max_{y \in \mathbb{R}} \left\{ y(h + \frac{1}{2}gT^2) \,\middle|\, \max_{t \in [0,T]} \{(T-t)y\} \leq 1 \right\} = \frac{1}{T}(h + \frac{1}{2}gT^2)$$

- the solution to $(\widetilde{\mathcal{P}})$ is a Dirac delta at $t = 0$, and in consequence $\mathrm{val}(\mathcal{P}) = \mathrm{val}(\widetilde{\mathcal{P}})$.

1.2 Extended Real-Valued Functions.

In our analysis it will be convenient to consider functions that can take values on the *extended Real line* $\mathbb{R} \cup \{+\infty\} = (-\infty, +\infty]$ and not just on $\mathbb{R} = (-\infty, +\infty)$. For example, les us consider the minimization problem on a space X

$$\inf_C f_0 := \inf\{f_0(x) \mid x \in C\} \tag{1.1}$$

where $f_0 \colon X \to \mathbb{R}$ is the function to be minimized and $C \subseteq X$ is a set of constraints. Let us define $\delta_C \colon X \to \mathbb{R} \cup \{+\infty\}$, the *indicator function* of the set C, via

$$\delta_C(x) := \begin{cases} 0 & x \in C, \\ +\infty & x \notin C. \end{cases}$$

Using the the convention that

$$\alpha + (+\infty) = (+\infty) + \alpha = +\infty, \ \forall \alpha \in \mathbb{R}$$

we check then that

$$\inf_C f_0 = \inf_X (f_0 + \delta_C)$$

In this way, (1.1) can be formulated as an unconstrained problem with an extended Real-valued function

$$\inf\{f(x) \mid x \in X\},$$

where $f \colon X \to \mathbb{R} \cup \{+\infty\}$ is given by $f = f_0 + \delta_C$. This allows to treat problem in an unified way, by hiding the constraints on the definition of the function to be minimized.

1.2.1 Conventions

Given $\lambda \in \mathbb{R}$ and functions $f, g \colon X \to \mathbb{R} \cup \{+\infty\}$, in order to make sense of $f + \alpha g$ we need to introduce some algebraic rules on $\overline{\mathbb{R}} = [-\infty, +\infty]$ that generalize the ones on $\mathbb{R}$. Unless otherwise stated, we assume:

1. $(+\infty) + \alpha = \alpha + (+\infty) = +\infty$, $\forall \alpha \in \mathbb{R} \cup \{+\infty\}$.
2. $(-\infty) + \alpha = \alpha + (-\infty) = -\infty$, $\forall \alpha \in \mathbb{R} \cup \{-\infty\}$.
3. $\alpha \cdot (+\infty) = (+\infty) \cdot \alpha = +\infty$, if $\alpha > 0$.
4. $\alpha \cdot (+\infty) = (+\infty) \cdot \alpha = -\infty$, if $\alpha < 0$.
5. $0 \cdot (+\infty) = (+\infty) \cdot 0 = 0$.
6. $+\infty + (-\infty) = (-\infty) + (+\infty) = +\infty$.

Remark 1.1. *In this setting, the sum and product are not continuous in the sense that if $\alpha_n \xrightarrow[n\to\infty]{} \alpha \in \overline{\mathbb{R}}$ and $\beta_n \xrightarrow[n\to\infty]{} \beta \in \overline{\mathbb{R}}$, then it is not necessarily true that $\alpha_n\beta_n \xrightarrow[n\to\infty]{} \alpha\beta$.*

The rules described above allow us to define then

$$(f + \alpha g)(x) := f(x) + \alpha g(x), \quad \forall x \in X, \ \forall \alpha \in \mathbb{R}.$$

1.2.2 Basic definitions

Given an extended Real-valued function $f\colon X \to \mathbb{R} \cup \{+\infty\}$, its effective domain is the set

$$\operatorname{dom}(f) := \{x \in X \mid f(x) < +\infty\}$$

and its epigraph is the subset of $X \times \mathbb{R}$ defined via

$$\operatorname{epi}(f) := \{(x, z) \in X \times \mathbb{R} \mid f(x) \leq z\}.$$

Most of times we will be interested in a subclass of extended Real-valued function for which minimization problems are not trivial. A function $f\colon X \to \mathbb{R} \cup \{+\infty\}$ is called *proper* if

$$\exists x \in X \text{ so that } f(x) < +\infty.$$

The *infimum* of a proper function f is $\inf_X f := \inf\{f(x) \mid \ x \in \operatorname{dom}(f)\}$. Note that the infimum of a proper function may eventually be $-\infty$. If that is not the case, that is, $\inf_X f > -\infty$, we say that f is *bounded below.*

The set of minimizers of f on X is given by

$$\arg\min(f) := \{x \in X \mid f(x) = \inf_X f\}$$

1.3 Convex functions.

Let X be a Real vectorial space. On the one hand, a subset $C \subseteq X$ is called *c*onvex if

$$\lambda x + (1 - \lambda)y \in C, \quad \forall x, y \in C,\ \forall \lambda \in [0, 1].$$

On the other hand, a extended Real-valued function $f\colon X \to \mathbb{R} \cup \{+\infty\}$ is called *c*onvex if

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y), \quad \forall x, y \in X,\ \forall \lambda \in [0, 1].$$

Example 1.1. *Any affine function or any norm on X is a convex function. Also, δ_C is convex if and only if $C \subseteq X$ is convex.*

Convex functions and convex sets are related through the following relation

$$\operatorname{epi}(f) \text{ is a convex subset of } X \times \mathbb{R} \quad \Longleftrightarrow \quad f\colon X \to \mathbb{R} \cup \{+\infty\} \text{ is convex}$$

Furthermore, the class of convex function is closed under positive linear combinations, that is, if $f, g\colon X \to \mathbb{R} \cup \{+\infty\}$ are convex functions and $\alpha \geq 0$ then $f + \alpha g$ is convex as well.

Example 1.2. *Some relevant examples of convex functions on $\mathbb{R}$ are:*

- $f(x) = ax^2 + bx + c$ *with* $a \geq 0$.
- $f(x) = e^{ax}$ *with* $a \in \mathbb{R}$.
- $f(x) = -\ln(x) + \delta_{(0,+\infty)}(x)$.

Further properties of convex functions are summarized below.

Proposition 1.1. *Let $f\colon X \to \mathbb{R} \cup \{+\infty\}$ be a convex function.*

(i) Suppose Y is a vector space. If $A \times Y \to X$ is affine then $f \circ A$ is convex.

(ii) If $\theta\colon \mathbb{R} \cup \{+\infty\} \to \mathbb{R} \cup \{+\infty\}$ is a nondecreasing convex function then $\theta \circ f$ is convex.

(iii) If $(f_i)_{i\in I}$ is a family of convex functions, then $f = \sup_{i\in I} f_i$ is convex.

(iv) Suppose Y is a vector space and $g : X \times Y \to \mathbb{R} \cup \{+\infty\}$ is convex. Then the function $h\colon X \to \mathbb{R} \cup \{+\infty\}$ given by $h(x) = \inf\limits_{y\in Y} g(x, y)$ is convex.

The following is a fundamental property about the continuity of convex function.

Theorem 1.1. *Let $(X, \|\cdot\|)$ be a normed vector space and $f\colon X \to \mathbb{R} \cup \{+\infty\}$ be a convex function. Suppose that f is bounded above on a neighborhood of* $\mathrm{dom}(f)$*, that is,*

$$\exists x_0 \in \mathrm{dom}(f), r > 0, M \in \mathbb{R}, \quad f(x) \le M,\ \forall x \in \mathbb{B}(x_0, r).$$

Then f is locally Lipschitz continuous on $\mathrm{int}(\mathrm{dom}(f))$.

This result has a trivial consequence when $X = \mathbb{R}^n$.

Corollary 1.1. *If $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is convex and proper, then f is locally Lipschitz continuous on* $\mathrm{int}(\mathrm{dom}(f))$*. In particular, if $f\colon \mathbb{R}^n \to \mathbb{R}$ is convex then it is continuous.*

1.4 Lower semicontinuity and minimization

Most of times we will be working with Banach spaces (a complete normed vector space). However, in order to ensure the existence of solutions to minimization problems, we might require to consider other topologies on those spaces, such as the weak and weak-$\star$ topologies.

1.4.1 Overview on Topology

Recall that a topology $\mathcal{T}$ on X is a collection of subsets of X, called open sets, that contains $\emptyset$ and X, which satisfies in addition

- $\bigcup_{i\in I} O_i \in \mathcal{T}$ for any (arbitrary) collection $\{O_i\}_{i\in I} \subseteq \mathcal{T}$.
- $\bigcap_{i=1}^n O_i \in \mathcal{T}$ for any finite collection $\{O_i\}_{i=1}^n \subseteq \mathcal{T}$.

The pair $(X, \mathcal{T})$ is called a topological space. Moreover, a sequence $\{x_n\}_{n\in\mathbb{N}}$ on a topological space $(X, \mathcal{T})$ is said to converge to some $x \in X$ if

$$\forall O \in \mathcal{T} \text{ with } x \in O,\ \exists N \in \mathbb{N} \text{ such that } \forall n \ge N,\ x_n \in O.$$

Some important examples

Let $(X, \|\cdot\|)$ be a Banach space, let $\mathbb{B}_X$ stands for the closed unit ball of X and let X^* be the topological dual of X, that is, the collection of linear continuous functionals $x^* : X \to \mathbb{R}$ endowed with the norm $\|x^*\|_* = \sup\{\langle x^*, x\rangle \mid x \in \mathbb{B}_X\}$, with $\langle x^*, x\rangle = x^*(x)$

- The *strong topology* on X is the collection of subsets $O \subseteq X$ having the property
$$\forall x \in O,\ \exists \varepsilon > 0,\ \{y \in X \mid \ \|x - y\| < \varepsilon\} \subseteq O.$$
A sequence $\{x_n\}_{n\in\mathbb{N}}$ on X converges strongly to $x \in X$ if
$$x_n \xrightarrow[n\to\infty]{} x \quad \iff \quad \|x_n - x\| \xrightarrow[n\to\infty]{} 0.$$
- The *weak topology* on X is the collection of subsets $O \subseteq X$ having the property
$$\forall x \in O,\ \exists x_1^*, \dots, x_n^* \in X^*,\ \exists \varepsilon > 0,\ \{y \in X \mid |\langle x_i^*, y - x\rangle| < \varepsilon,\ \forall i = 1, \dots, n\} \subseteq O.$$
A sequence $\{x_n\}_{n\in\mathbb{N}}$ on X converges weakly to $x \in X$ if
$$x_n \xrightharpoonup[n\to\infty]{} x \quad \iff \quad \langle x^*, x_n\rangle \xrightarrow[n\to\infty]{} \langle x^*, x\rangle, \quad \forall x^* \in X^*.$$
- The *weak-$\star$ topology* on X^* is the collection of subsets $O \subseteq X^*$ having the property
$$\forall x^* \in O,\ \exists x_1, \dots, x_n \in X,\ \exists \varepsilon > 0,\ \{y^* \in X \mid |\langle y^* - x^*, x_i\rangle| < \varepsilon,\ \forall i = 1, \dots, n\} \subseteq O.$$
A sequence $\{x_n^*\}_{n\in\mathbb{N}}$ on X^* converges in the weak-$\star$ sense to $x^* \in X^*$ if
$$x_n^* \xrightharpoonup[n\to\infty]{\star} x^* \quad \iff \quad \langle x_n^*, x\rangle \xrightarrow[n\to\infty]{} \langle x^*, x\rangle, \quad \forall x \in X.$$

Closed sets and compactness

A subset $C \subseteq X$ is called *closed* if $X \setminus C$ is an open set.

Note that if $(X, \|\cdot\|)$ is a Banach space, then any weak open subset of X is also a strong open set. Hence, any weak closed subset is a strong closed subset of X. The converse is true provided that the set in question is also convex.

Lemma 1.1. *Let $(X, \|\cdot\|)$ be a Banach space and $C \subseteq X$ be a convex subset. Then C is closed in the weak topology if and only if it is closed in the strong topology.*

A subset $K \subseteq X$ is called *compact* if any open covering of K admits a finite sub covering, that is, if $\{O_i\}_{i\in I}$ is a collection of open sets of X, then
$$K \subseteq \bigcup_{i\in I} O_i \quad \implies \quad \exists i_1, \dots, i_n \in I \text{ such that } K \subseteq \bigcup_{k=1}^{n} O_{i_k}.$$

In finite dimensional spaces, there is a simple criterion for compactness, the Heine-Borel Theorem that says that
$$K \subseteq \mathbb{R}^n \text{ is compact} \quad \iff \quad K \text{ is bounded and closed.}$$
This criterion holds for infinite dimensional Banach space, however never for the strong topology, but for the weak topologies.

Recall that a topological space $(X, \mathcal{T})$ is said to be *separable* is there is a countable subset $E \subseteq X$ such that E is dense in X, that is, the closure of E agrees with X.

Lemma 1.2 (Banach-Alaoglu Theorem). *The dual closed unit ball*

$$\mathbb{B}_{X^*} := \{x^* \in X^* \mid \|x^*\|_* \leq 1\}$$

is compact in the weak-$\star$ topology on X^*. In particular, X is separable, then every bounded sequence in X^* admits a convergent subsequence in the weak-$\star$ sense.

Let X^{**} be the topological dual of X^* and let us define $J : X \to X^{**}$ via

$$J(x)(x^*) = \langle x^*, x\rangle, \quad \forall x, \in X,\ \forall x^* \in X^*.$$

The mapping J is called the canonical injection from X into X^{**} (it is actually an injective isometry). A Banach space is called *reflexive* is J is surjective on X^{**}, that is, X and X^{**} are isometrically isomorphic.

Lemma 1.3 (Kakutani Theorem). *Let $(X, \|\cdot\|)$ be a Banach space. Then X is reflexive if and only if $\mathbb{B}_X$ is compact in the weak topology on X. In particular, every bounded sequence in X admits a weakly convergent subsequence.*

The fact that in reflexive Banach spaces compactness implies sequentially compactness is just a property, but also a characterization of these spaces.

Lemma 1.4 (Eberlein-Smulian). *Assume that $(X, \|\cdot\|)$ is a Banach space such that every bounded sequence in X admits a weakly convergent subsequence. Then X is reflexive.*

For further details and discussions on weak topologies see [1, Chapter 3].

1.4.2 Lower semicontinuous functions

Definition 1.1. *Let $(X, \mathcal{T})$ be a topological space. A function $f : X \to \mathbb{R} \cup \{+\infty\}$ is called* lower semicontinuous *(lsc) on $(X, \mathcal{T})$ is $\text{epi}(f)$ is closed on $X \times \mathbb{R}$ for the topology $\mathcal{T} \times \mathcal{T}_{\mathbb{R}}$, where $\mathcal{T}_{\mathbb{R}}$ is the usual topology on $\mathbb{R}$.*

Example 1.3. *δ_C is lower semicontinuous on $(X, \mathcal{T})$ if and only if $C \subseteq X$ is closed.*

Proposition 1.2. *Let $f : X \to \mathbb{R} \cup \{+\infty\}$ be a given extended Real-valued function. Then, f is lsc on $(X, \mathcal{T})$ if and only if $\{x \in X \mid f(x) \leq \alpha\}$ is a closed subset on $(X, \mathcal{T})$ for any $\alpha \in \mathbb{R}$. Furthermore, if $x \in dom(f)$, then*

$$\forall \varepsilon > 0,\ \exists O \in \mathcal{T} \text{ with } x \in O \text{ such that } \forall y \in O,\ f(y) \geq f(x) - \varepsilon.$$

Remark 1.2. *Note that f is lsc on $(X, \mathcal{T})$ then*

$$\forall x \in X,\ x_n \xrightarrow[n\to\infty]{} x, \quad \Longrightarrow \quad f(x) \leq \liminf_{n\to+\infty} f(x_n) := \sup_{n\in\mathbb{N}} \inf_{k\geq n} f(x_k).$$

An important result concerning convex functions follows from Lemma 1.1 and Proposition 1.2.

Some basic properties of lsc functions are summarized below.

Proposition 1.3. *Let $\{f_i\}_{i\in I}$ be a family of lsc functions on $(X, \mathcal{T})$. Then $\sup_{i\in I} f_i$ is lsc on $(X, \mathcal{T})$. Furthermore, if I is finite, then $\sum_{i\in I} f_i$ and $\inf_{i\in I} f_i$ are lsc functions on $(X, \mathcal{T})$.*

Example 1.4. *$\|\cdot\|_*$ is lower semicontinuous in the weak-$\star$ topology on X^*.*

Proposition 1.4. *Let $(X, \|\cdot\|)$ be a Banach space. Assume that $f : X \to \mathbb{R} \cup \{+\infty\}$ is convex, then f lsc in the strong topology if and only if f is lsc in the weak topology on X.*

Example 1.5. *$\|\cdot\|$ is lower semicontinuous in the weak topology.*

1.4.3 Inf-compactness and existence of minimizers

Lower semicontinuity is an important issue for the existence of minimizers, but it's not enough. An additional compactness criterion must be evoke.

Definition 1.2. *Let $(X, \mathcal{T})$ be a topological space. A function $f: X \to \mathbb{R} \cup \{+\infty\}$ is said to be* inf-compact *on $(X, \mathcal{T})$ if*

$$\exists x_0 \in X,\ \forall \alpha \leq f(x_0), \quad \{x \in X \mid f(x) \leq \alpha\} \text{ has compact closure on } (X, \mathcal{T})$$

The following is a general theorem for the existence of minimizers.

Theorem 1.2 (Weierstrass-Hilbert-Tonelli Theorem). *Let $(X, \mathcal{T})$ be a topological space and $f: X \to \mathbb{R} \cup \{+\infty\}$ be a proper lsc and inf-compact function on $(X, \mathcal{T})$. Then* $\arg\min(f) \neq \emptyset$.

Assume that $(X, \|\cdot\|)$ is a normed vector space. A function $f: X \to \mathbb{R} \cup \{+\infty\}$ is said to be *coercive* provided that

$$\lim_{\|x\| \to \infty} f(x) = +\infty.$$

The following are corollaries adapted to reflexive Banach spaces and dual spaces.

Corollary 1.2. *Assume that $(X, \|\cdot\|)$ is a reflexive Banach space. Let $f: X \to \mathbb{R} \cup \{+\infty\}$ be a convex coercive function, which in addition is lsc for the strong topology. Then* $\arg\min(f) \neq \emptyset$.

Corollary 1.3. *Assume that $(X, \|\cdot\|)$ is a Banach space. Let $f: X^* \to \mathbb{R} \cup \{+\infty\}$ be a coercive function, which in addition is lsc for the weak-$\star$ topology. Then* $\arg\min(f) \neq \emptyset$.

Example 1.6. *Note that $(\widetilde{\mathcal{P}})$ can be written as*

$$\inf_{\mu \in C([0,T])^*} \{\|\mu\|_* + \delta_C(\mu)\}$$

with $C = \{\mu \in C([0,T])^* \mid \ \langle \mu, \varphi = \alpha\}$ for some $\varphi \in C([0,T])$ and $\alpha \in \mathbb{R}$. Since C is weak-$\star$ closed on $\mathcal{C}([0,T])^*$ and $\|\cdot\|_*$ is lsc for the weak-$\star$ topology on X^*, $(\widetilde{\mathcal{P}})$ has a solution.

1.5 Characterization of convex functions

The convexity of a given function is an algebraic criterion, which may be hard to prove sometimes. Some other criteria for continuously differentiable function are stated below. Recall that if f is differentiable $x \in X$, then its differential at x, denoted by $Df(x)$, is an element of X^* such that

$$\lim_{h \to 0} \frac{|f(x+h) - f(x) - Df(x)(h)|}{\|h\|} = 0$$

If $(X, \langle \cdot, \cdot \rangle)$ is a Hilbert space then it is reflexive and furthermore, by the Riesz representation Theorem, X^* can be identified with X, and so, there is a unique $\nabla f(x) \in X$ such that

$$Df(x)(h) = \langle \nabla f(x), h \rangle, \quad \forall h \in X.$$

See [1, Chapter 5] for further discussions on Hilbert spaces and [2, Chapter 1] for definitions and discussions on Differential Calculus over normed spaces.

Theorem 1.3. *Let $(X, \langle\cdot,\cdot\rangle)$ be a Hilbert space and $f\colon X \to \mathbb{R}\cup\{+\infty\}$ be a continuously differentiable function on* $\mathrm{dom}(f)$*, the latter being a nonempty open subset of X. Then the following are equivalent:*

(i) $f\colon X \to \mathbb{R}\cup\{+\infty\}$ *is convex.*

(ii) $\forall x, y \in \mathrm{dom}(f),\ \langle \nabla f(x) - \nabla f(y), x - y\rangle \geq 0.$

(iii) $\forall x, y \in \mathrm{dom}(f),\ f(y) \geq f(x) + \langle \nabla f(x), y - x\rangle.$

Furthermore, if $f\colon X \to \mathbb{R}\cup\{+\infty\}$ is twice continuously differentiable on $\mathrm{dom}(f)$ *and $X = \mathbb{R}^n$ is finite dimensional, then f is convex is and only if $\nabla^2 f(x)$ is a positive semi-definite matrix of dimension n for any $x \in \mathrm{dom}(f)$.*

Example 1.7. *The following are relevant examples of convex functions on $\mathbb{R}^n$:*

- $f(x) = \frac{1}{2}\langle x, Ax\rangle + \langle b, x\rangle + c$, *with $A \in \mathbb{S}^n_+$, $b \in \mathbb{R}^n$ and $c \in \mathbb{R}$.*
- $f(x) = \|x\|^\alpha$ *with $\alpha \geq 1$.*

1.5.1 Separation of Convex Sets

We now recall some geometric forms of the Hahn-Banach Theorem. From now on $(X, \|\cdot\|)$ is a Banach space and X^* is its topological dual endowed with the norm

$$\|x^*\|_* = \sup\{\langle x^*, x\rangle \mid x \in \mathbb{B}_X\},$$

with $\langle x^*, x\rangle = x^*(x)$ being the usual duality product between X and X^*.

Lemma 1.5 (Hahn-Banach)**.** *Let $A, B \subseteq X$ be two nonempty convex pairwise disjoint subsets.*

(i) If A is open then there exist $x^* \in X^* \setminus \{0\}$ and $\alpha \in \mathbb{R}$ so that

$$\langle x^*, a\rangle < \alpha,\ \forall a \in A \quad \text{and} \quad \langle x^*, b\rangle \geq \alpha,\ \forall b \in B.$$

(ii) If A is closed and B is compact then there exist $x^* \in X^* \setminus \{0\}$, $\alpha \in \mathbb{R}$ and $\varepsilon > 0$ so that

$$\langle x^*, a\rangle \leq \alpha - \varepsilon,\ \forall a \in A \quad \text{and} \quad \langle x^*, b\rangle \geq \alpha + \varepsilon,\ \forall b \in B.$$

The basic idea of the geometric version of the Hahh-Banach Theorem (Lemma 1.5) is that disjoint nonempty convex sets can be separated by an hyperplane. If one of the sets is compact and the other is closed, this can be done in an strict sense. In Figure 1.1 we have sketched a geometric interpretation of this theorem. The picture on the left shows the separation when one of the set is open and the picture on the right shows the strict separation of convex sets. Hahh-Banach Theorem is a consequence of the Zorn's Lemma; see [1, Chapter 1] for details.

A function $h : X \to \mathbb{R}$ is called affine continuous on X if $\exists x^* \in X^*$ and $\alpha \in \mathbb{R}$ so that

$$h(x) = \langle x^*, x\rangle + \alpha, \quad \forall x \in X$$

Let us define

$$\Gamma_0(X) = \left\{ f\colon X \to \mathbb{R}\cup\{+\infty\} \,\middle|\, f \text{ is proper and } f = \sup_{i\in I} h_i \text{ with } h_i \text{ affine continuous on } X \right\}$$

Note that is $f \in \Gamma_0(X)$ then f is convex and lsc on X. The converse is true, and it is a consequence of the Hahn-Banach Theorem.

Theorem 1.4. *$f \in \Gamma_0(X)$ if and only is f is proper convex lsc on X.*

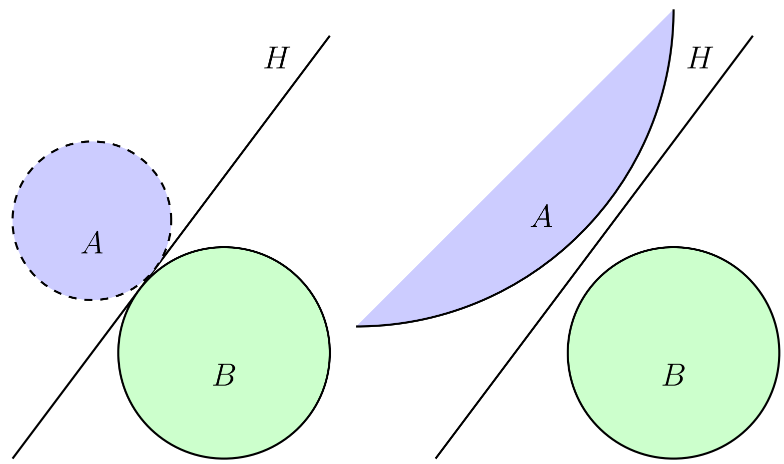

Figure 1.1: Hanh-Banach Theorem

1.6 The Fenchel conjugate

The fact that any proper convex lsc function can be written as the supremum of affine continuous function on X implies that there is at least some $x_0^* \in X^*$ such that

$$\langle x_0^*, x\rangle - f(x) < +\infty, \quad \forall x \in X$$

Hence, the supremum of $\langle x_0^*, x\rangle - f(x)$ over $x \in X$ is a Real number, this means that the function $f^* : X^* \to \mathbb{R} \cup \{+\infty\}$ defined via

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

is proper. Furthermore, since $X \subseteq X^{**}$ via the canonical injection, then $f^* \in \Gamma_0(X^*)$. This function is very important in convex analysis and it is called the *Fenchel conjugate* of f.

The following is an inequality that also plays a central role on the theory and it is called the *Young-Fenchel inequality*

$$\forall x \in X,\ \forall x^* \in X^*,\ f(x) + f^*(x^*) \geq \langle x^*, x\rangle.$$

We will study later on some criteria to turn the Young-Fenchel inequality into an equation.

1.6.1 Some properties and examples

Proposition 1.5. *Let $f \in \Gamma_0(X)$, then:*

1. $(\alpha f + \beta)^*(y) = \alpha f^*(\frac{y}{\alpha}) - \beta$, *for any $\alpha > 0$ and $\beta \in \mathbb{R}$.*

2. *given $x_0 \in X$, if $g(x) = f(x - x_0)$, we have $g^*(x^*) = f^*(x^*) + \langle x^*, x_0\rangle$.*

3. *given $x_0^* \in X^*$, if $g(x) = f(x) + \langle x_0^*, x\rangle$ then $g^*(x^*) = f^*(x^* - x_0^*)$.*

Example 1.8. *Let $p \in (1, +\infty)$ and consider the function $f : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ given by*

$$f(x) = \frac{1}{p}|x|^p, \quad \forall x \in \mathbb{R}.$$

Letting $q \in (1, +\infty)$ be such that $1 = \frac{1}{p} + \frac{1}{q}$, then the Fenchel conjugate of f is given by

$$f^*(x^*) = \frac{1}{q}|x^*|^q, \quad \forall x^* \in \mathbb{R}$$

Proposition 1.6. *Let $f : X \to \mathbb{R} \cup \{+\infty\}$ be a radially defined function, that is, there is $\varphi : [0, +\infty) \to \mathbb{R} \cup \{+\infty\}$ with $[0, +\infty) \cap \operatorname{dom}(\varphi) \neq \emptyset$ such that*

$$f(x) = \varphi(\|x\|), \quad \forall x \in X$$

Then then Fenchel conjugate of f is given by

$$f^*(x^*) = \varphi^*(\|x^*\|_*), \quad \forall x^* \in X^*.$$

Example 1.9. *Let $p \in (1, +\infty)$ and $[a, b] \subseteq \mathbb{R}$, recall that if $X = L^p([a, b])$ then $X^* = L^q([a, b])$ with $q \in (1, +\infty)$ being such that $1 = \frac{1}{p} + \frac{1}{q}$. Consequently, if*

$$f(u) = \frac{1}{p}\|u\|_p^p, \ \forall u \in L^p([a, b]) \quad \Longrightarrow \quad f^*(v) = \frac{1}{q}\|v\|_q^q, \ \forall v \in L^q([a, b]).$$

Example 1.10. *Suppose X is a Hilbert space and make the identification $X = X^*$. Then $f(x) = \frac{1}{2}\|x\|^2 = \langle x, x\rangle$ is the unique $f \in \Gamma_0(X)$ that satisfies $f = f^*$*

Example 1.11 (support function)**.** *Let $C \subseteq X$ be a nonempty convex closed subset. Then δ_C^* is given by*

$$\sigma_C(x^*) = \sup_{x \in C}\langle x^*, x\rangle, \quad \forall x^* \in X^*$$

This mapping is called the support function of C. Furthermore, consider the following cases:

1. *If $C = \{x_0\}$ then $\sigma_{\{x_0\}}(x^*) = \langle x^*, x_0\rangle$.*

2. *If $C = \mathbb{B}_X$ then $\sigma_{\mathbb{B}_X}(x^*) = \|x^*\|_*$.*

3. *If C is a cone, that is, $\alpha C \subseteq C$ for any $\alpha > 0$, then $\sigma_C(x^*) = \delta_{C^0}(x^*)$ where*

   $$C^0 = \{x^* \in X^* \mid \ \forall x \in C, \ \langle x^*, x\rangle \leq 0\}$$

   is the polar cone to C.

4. *If C is a subvector space then $\sigma_C(y) = \delta_{C^\perp}(x^*)$, where*

   $$C^\perp = \{x^* \in X^* \mid \forall x \in C, \ \langle x^*, x\rangle = 0\}$$

   is the orthogonal subspace to C.

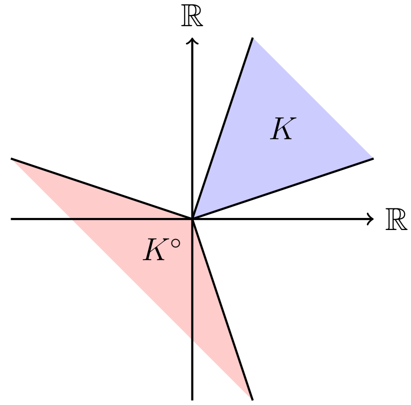

Figure 1.2: polar cone on $\mathbb{R}^2$

1.6.2 Biconjugate

Let us assume for a moment that $(X, \|\cdot\|)$ is a reflexive Banach space, and let us identify X^{**} with X by means of the canonical injection.

Given $f \in \Gamma_0(X)$, we define the *biconjugate* of f, $f^{**} : X \to \mathbb{R} \cup \{+\infty\}$ as the Fenchel conjugate of $f^* \in \Gamma_0(X^*)$, that is,

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

By the Young-Fenchel inequality we have that $f^{**} \leq f$. Hence, since by definition f^{**} is convex and lsc on X^*, we have that $f^{**} \in \Gamma_0(X)$ provided that $f \in \Gamma_0(X)$.

Proposition 1.7. *If $(X, \|\cdot\|)$ is a reflexive Banach space and $f \in \Gamma_0(X)$, then $f = f^{**}$.*

If X is not reflexive (as for instance if $X = C([a.b])$), the duality product need to be restraint to a weaker topology. Recall that:

- The weak topology on X is the minimal collection of open sets of X satisfying the definition of a topology and that makes the maps $x \mapsto \langle x^*, x \rangle$ continuous for any $x^* \in X^*$.
- The weak-$\star$ topology on X^* is the minimal collection of open sets of X^* satisfying the definition of a topology and that makes the maps $x^* \mapsto \langle x^*, x \rangle$ continuous for any $x \in X$.

Hence, by endowing X with the weak topology and the dual space X^* with the weak-$\star$ topology, we have that the usual duality product is enough to represent all the linear continuous functions on X and X^*. Indeed, on the one hand, by definition of the weak topology on X the maps $x \mapsto \langle x^*, x \rangle$ are continuous for any $x^* \in X^*$. On the other hand, it can be proved (see [1, Proposition 3.14]) that for any linear weak-$\star$ continuous map $\ell : X^* \to \mathbb{R}$ there is a unique $x \in X$ such that

$$\ell(x^*) = \langle x^*, x \rangle, \quad \forall x^* \in X^*.$$

We say then that X and X^* are in **duality** and then X can be identified with the dual space of X^* (endowed with the weak-$\star$), and so the biconjugate is well defined. If we are not in the reflexive case we might assume that this property holds in whatever it follows.

1.7 The subdifferential

We now introduce a concept that generalizes the idea of gradient of a function. The definition suits well to convex function, however the definition doesn't require the convexity of the function at hand.

Definition 1.3. *Let $f\colon X \to \mathbb{R} \cup \{+\infty\}$ be an extended Real-valued function. A subgradient of f at $x \in X$ is any $x^* \in X^*$ that satisfies*

$$f(x) + \langle x^*, y - x\rangle \leq f(y), \quad \forall y \in X.$$

The set of all subgradients of f at x, denoted $\partial f(x)$, is called the subdifferential of f at x.

Remark 1.3. *Let us point out that $\partial f(x)$ is a (possibly empty) convex and closed subset of X^* for any $x \in X$. Furthermore, if $f(x) = +\infty$ then $\partial f(x) = \emptyset$.*

Example 1.12. *Let us see some examples:*

- *If $f(x) = |x|$, for any $x \in \mathbb{R}$, then $\partial f(0) = [-1, 1]$.*
- *If $f(x) = \sqrt{x} + \delta_{(-\infty,o)}(x)$, for any $x \in \mathbb{R}$, then $\partial f(0) = \emptyset$.*
- *If $f(x) = \delta_C(x)$, then $\partial\delta_C(x)$ is the normal cone to C at $x \in X$, that is,*

$$N_C(x) := \partial\delta_C(x) = \{x^* \in X^* \mid \langle x^*, y - x\rangle \leq 0,\ \forall y \in C\}.$$

1.7.1 Properties of the subdifferential

Proposition 1.8. *Let $f\colon X \to \mathbb{R} \cup \{+\infty\}$ proper and $x \in \operatorname{dom}(f)$. Then*

$$x^* \in \partial f(x) \quad \iff \quad f(x) + f^*(x^*) = \langle x^*, x\rangle$$

Moreover, if $f \in \Gamma_0(X)$ with $(X, \|\cdot\|)$ being a Banach space in duality with X^*, then

$$x^* \in \partial f(x) \quad \iff \quad x \in \partial f^*(x^*).$$

The subdifferential of a function enjoy further topological properties when the function is convex proper and lsc on X.

Theorem 1.5. *Let $f\colon X \to \mathbb{R} \cup \{+\infty\}$ be a convex function and $x \in X$ such that $f(x) \in \mathbb{R}$ and f is continuous at x. Then $\partial f(x) \neq \emptyset$ and it is a weak-$\star$ compact subset of X^*.*

1.7.2 Directional derivatives

Recall that the directional derivatives of a function is

$$f'(x; d) := \lim_{t \to 0^+} \frac{f(x + td) - f(x)}{t}, \quad \forall d \in X.$$

This derivative enjoys further structural properties when it is convex.

Proposition 1.9. *Let $f : X \to \mathbb{R} \cup \{+\infty\}$ and suppose that f is finite at $x \in X$. Then*

$$f'(x; d) := \inf_{t>0} \frac{f(x + td) - f(x)}{t}, \quad \forall d \in X.$$

Moreover, $d \mapsto f'(x; d)$ is sublinear, and

$$\partial f(x) = \{x^* \in X^* \mid \ \langle x^*, d\rangle \leq f'(x; d), \ \forall d \in X\}.$$

We can characterize the directional derivatives in term of the subdifferential.

Corollary 1.4. *Let $f \in \Gamma_0(X)$ and suppose that f is finite and continuous at $x \in X$. Then*

$$f'(x; d) = \max_{x^* \in \partial f(x)} \langle x^*, d\rangle = \sigma_{\partial f(x)}(d), \quad \forall d \in X$$

Moreover, if $X = \mathbb{R}^n$, then f is differentiable at x if and only if $\partial f(x) = \{x^*\}$.

1.7.3 Application to optimization

Theorem 1.6 (Fermat's rule I)**.** *Let $f \in \Gamma_0(X)$. Then*

$$x \in \arg\min(f) \iff 0 \in \partial f(x)$$

Furthermore,

$$\arg\min_X f = \partial f^*(0)$$

which is convex and weak-$\star$ compact if f^* is finite and continuous at $x^* = 0$.

Often we are interested in optimization problems with explicit constraint

$$(\mathcal{P}) \quad \inf_{x \in C} f(x),$$

The Fermat's rule can be refined in this context by means of a sum rule for subdifferentials.

Proposition 1.10 (Moreau-Rockafellar Theorem)**.** *Let $f_1, f_2 \in \Gamma_0(X)$. Suppose that f_1 is continuous at some $\bar{x} \in \mathrm{dom}(f_1) \cap \mathrm{dom}(f_2)$. Then*

$$\forall x \in X, \ \partial f_1(x) + \partial f_2(x) = \partial(f_1 + f_2)(x).$$

Recall that $(\mathcal{P})$ can be written as

$$\inf_{x \in X} \{f(x) + \delta_C(x)\}.$$

Theorem 1.7 (Fermat's rule II)**.** *Let $f \in \Gamma_0(X)$ and $C \subseteq X$ be a nonempty convex closed subset. Suppose there exists $\bar{x} \in \mathrm{int}(C)$ such that f is finite at $\bar{x}$. Then, $x \in C$ is a solution to $(\mathcal{P})$ if and only if*

$$0 \in \partial f(x) + N_C(x)$$

or equivalently

$$\exists x^* \in \partial f(x) \quad \text{such that } \langle x^*, y - x\rangle \geq 0, \quad \forall y \in C.$$

Besides of the sum rule, the subdifferential satisfies some composition rules. The following result is of particular interest for the duality theory we want to develop later.

Proposition 1.11. *Let $A \colon X \to Y$ be a continuous linear function and $f \in \Gamma_0(Y)$, where $(Y, |\cdot\|)$ is another Banach space. Suppose that f is continuous at some $y \in \mathrm{dom}(f)$, then we have that*

$$\partial(f \circ A)(x) = A^* \partial f(Ax), \quad \forall x \in X.$$

1.8 Duality in convex optimization

We start now the most relevant part of this introduction to convex analysis, the theory of duality. Let us start with a familiar case studied in linear optimization.

Example 1.13. *Consider the optimization problem*

$$(\mathrm{P}_L) \qquad \min_{x\in\mathbb{R}^n}\{\langle c,x\rangle \mid \ Ax\le b,\ x\ge 0\}$$

where $c\in\mathbb{R}^n$, $b\in\mathbb{R}^m$ *y* A *is a matrix of dimension* $m\times n$. *Then, the dual problem of* (P_L) *is*

$$(\mathrm{D}_L) \qquad \min_{x\in\mathbb{R}^n}\{\langle b,y\rangle \mid \ A^Ty\ge -c, y\ge 0\}$$

The duality theory in linear programming says:

- *duality is symmetric, that is, the dual of* (D_L) *is* (P_L).
- $\mathrm{val}\,(\mathrm{P}_L)+\mathrm{val}\,(\mathrm{D}_L)\ge 0$ *(weak duality); recall that that by convention* $+\infty+(-\infty)=+\infty$.
- *If* $\mathrm{val}\,(\mathrm{P}_L)$ *is finite, then* $\mathrm{val}\,(\mathrm{D}_L)$ *is also finite and* $\mathrm{val}\,(\mathrm{P}_L)+\mathrm{val}\,(\mathrm{D}_L)=0$ *(strong duality).*

Something similar can be said for general convex optimization problem, that is for

$$(\mathrm{P}) \qquad \inf\{f(x)\mid x\in X\}$$

where $(X,\|\cdot\|)$ is a given Banach space in duality with its dual X^* and $f\in\Gamma_0(X)$.

Definition 1.4. *Consider another Banach space* $(Y,\|\cdot\|)$ *also in duality with its dual* Y^*. *A function* $\varphi\in\Gamma_0(X\times Y^*)$ *is called a perturbation of* f *if*

$$\varphi(x,0)=f(x),\quad \forall x\in X.$$

By the Young-Fenchel inequality, for any perturbation function of f we have that

$$\varphi(x,0)+\varphi^*(0,y)\ge 0,\quad \forall x\in X,\ y\in Y.$$

It follows then that

$$\mathrm{val}\,(\mathrm{P})+\inf_{y\in Y}\varphi^*(0,y)\ge 0$$

Thus, a natural way to define a dual problem to (P) using a perturbation of f is via

$$(\mathrm{D}) \qquad \inf\{\varphi^*(0,y)\mid y\in Y\}$$

Problem (D) is called the dual problem to (P) associated with the perturbation φ. Therefore, weak duality always holds true

$$\mathrm{val}\,(\mathrm{P})+\mathrm{val}\,(\mathrm{D})\ge 0$$

Example 1.14. *The perturbation*

$$\varphi(x,y^*)=\langle c,x\rangle+\delta_{\mathbb{R}^m_-}(Ax-b-y^*)+\delta_{\mathbb{R}^n_+}(x),\quad \forall x\in\mathbb{R}^n,\ y^*\in\mathbb{R}^m.$$

provides the dual problem (D_L) *related to* (P_L).

Remark 1.4. *The fact that* $\varphi\in\Gamma_0(X\times Y^*)$ *and the spaces* X *and* Y *are both in duality with their respective dual spaces, imply that the biconjugate* φ^{**} *is well defined and agrees with* φ. *Consequently, duality defined in this way is symmetric, that is, the dual problem of* (D)

$$(\mathrm{DD}) \qquad \inf\{\varphi^{**}(x^{**},0)\mid x^{**}\in X^{**}\}$$

can be identified with (P) *via the canonical injection* $x\mapsto\langle\cdot,x\rangle$.

1.8.1 Strong duality

We have already seen that weak duality always holds. Let us now focus on the strong duality, that is, we seek for criteria in order to get

$$\text{val}\,(\text{P}) + \text{val}\,(\text{D}) = 0 \tag{1.2}$$

Let us introduce the *primal value function* associated with the perturbation φ via

$$\vartheta(y^*) := \inf_{x\in X} \varphi(x, y^*), \quad \forall y^* \in Y^*.$$

In a similar way we define the *dual value function* via

$$\omega(x^*) = \inf_{y\in Y} \varphi^*(x^*, y), \quad \forall x^* \in X^*.$$

It's clear then that $\vartheta(0) = \text{val}\,(\text{P})$ and $\omega(0) = \text{val}\,(\text{D})$. Moreover, these two functions are convex, but not necessarily proper or lower semicontinuous. Furthermore, we have that

$$\vartheta^*(y) = \varphi(0, y), \quad \text{and} \quad \omega^*(x) = \varphi(x, 0), \qquad \forall x \in X,\ y \in Y.$$

Theorem 1.8 (Duality theorem)**.** *Let $\varphi \in \Gamma_0(X \times Y^*)$ be a perturbation function for* (P)*. Suppose there is $x \in X$ such that $y^* \mapsto \varphi(x, y^*)$ is finite and (strongly) continuous at $y^* = 0$.*

- *If $\vartheta(0) = -\infty$ then dual problem* (D) *is infeasible.*
- *If $\vartheta(0) > -\infty$ then* (1.2) *holds and $\partial v(0) \neq \emptyset$ is the set of minimizer of* (D)*.*

Remark 1.5. *By the symmetry of duality, an analogous theorem can be stated for the dual problem, characterizing in particular the solution(s) to the primal problem. Note that the solution to the dual problem is unique if the primal value function is differentiable. The solution in this case is $\nabla\vartheta(0)$.*

Remark 1.6. *In the preceding theorem, since $y^* \mapsto \varphi(x, y^*)$ is continuous with respect to the norm topology on Y^*, then the set of minimizer of* (D) *is a nonempty bounded closed and convex subset of Y.*

Optimal solutions to primal and dual problems can be characterized as follows.

Theorem 1.9 (optimality conditions)**.** *Let $\varphi \in \Gamma_0(X \times Y^*)$ be a perturbation function for* (P)*. Let $x \in X$ and $y \in Y$. Then the following statements are equivalent*

(i) *x solves* (P)*, y solves* (P) *and* (1.2) *holds.*

(ii) $(0, y) \in \partial\varphi(x, 0)$ *or equivalently* $(x, 0) \in \partial\varphi^*(0, y)$.

(ii) $\varphi^*(0, y) + \varphi(x, 0) = 0$.

1.8.2 Fenchel-Rockafellar's duality Theorem

In this final part we study a duality theorem that will allow us to solve completely the example at the introduction. We look in particular a way to provide qualification condition for strong duality and existence of solutions in terms of the original data, and not the perturbation function.

Let X and Y be two Banach spaces in duality with their dual spaces. Let $A : X \to Y^*$ be a linear continuous mapping, $f \in \Gamma_0(X)$ and $g \in \Gamma_0(Y^*)$. We are concerned with the optimization problem

$$(P_0) \qquad \inf\{f(x) + g(Ax) \mid x \in X\}$$

We can associate to (P_0) the dual problem

$$(D_0) \qquad \inf\{f^*(-A^*y) + g^*(y) \mid y \in Y\}$$

via the perturbation function $\varphi(x, y^*) = f(x) + g(Ax + y^*)$.

Theorem 1.10 (Fenchel-Rockafellar's duality theorem)**.** *Suppose that (P_0) and (D_0) are given as above, where X and Y are two Banach spaces in duality with their dual spaces, $A : X \to Y^*$ is a linear continuous mapping, $f \in \Gamma_0(X)$ and $g \in \Gamma_0(Y^*)$.*

- *If $\operatorname{val}(P_0) \in \mathbb{R}$ and $0 \in \operatorname{int}(\operatorname{dom}(g) - A\operatorname{dom}(f))$, then (D_0) has a solution.*
- *If $\operatorname{val}(D_0) \in \mathbb{R}$ and $0 \in \operatorname{int}(\operatorname{dom}(f^*) - A^*\operatorname{dom}(g^*))$, then (P_0) has a solution.*

In either case we have strong duality, that is, $\operatorname{val}(P_0) + \operatorname{val}(D_0) = 0$. Moreover, the following are equivalent:

1. *$\bar{x}$ solves (P_0), $\bar{y}$ solves (D_0) and $\operatorname{val}(P_0) + \operatorname{val}(D_0) = 0$.*
2. *Extremality relations: $f(\bar{x}) + f^*(-A^*\bar{y}) = \langle \bar{x}, -A^*\bar{y}\rangle$ and $g(A\bar{x}) + g^*(\bar{y}) = \langle A\bar{x}, \bar{y}\rangle$.*
3. *Euler-Lagrange equations: $-A^*\bar{y} \in \partial f(\bar{x})$ and $\bar{y} \in \partial g(A\bar{y})$.*

A consequence of the Fenchel-Rockafellar's duality theorem is the following result.

Corollary 1.5. *Let X be a Banach space in duality with X^*. Let $M \subseteq X$ be a given vector subspace and $M^\perp$ be its orthogonal space, that is,*

$$M^\perp = \{x^* \in X^* \mid \langle x^*, x\rangle = 0, \quad \forall x \in M\}.$$

Hence, given $z^* \in X^*$

$$\inf_{x^* \in M^\perp} \|x^* - z^*\|_* = \sup\{\langle z^*, x\rangle \mid x \in M, \ \|x\| \leq 1\}.$$

Furthermore, the infimum is attained at $\bar{x}^* \in M^\perp$. Also, $x_0 \in M$ attains the maximum if and only if $\langle z^* - x^*, \bar{x}\rangle = \|\bar{z}^* - x^*\|_*\|\bar{x}\|$

Appendix

A.1 Sketch of Proposition 1.11's proof

The inclusion $A^*\partial f(Ax) \subseteq \partial(f\circ A)(x)$ is rather simple, so we skip it and focus on other one.

Given that f is continuous at some $\bar{x} \in \mathrm{dom}(f)$, we have that $\mathrm{int}(\mathrm{dom}(f)) \neq \emptyset$. Furthermore, we have that

$$\mathrm{int}(\mathrm{dom}(f)) \subseteq \{(y,\lambda) \in Y\times\mathbb{R} \mid\ f(y) < \lambda\}.$$

Therefore, given $x \in X$ and $x^* \in \partial(f\circ A)(x)$, the set

$$S = \{(Az, f(Ax) + \langle x^*, z-x\rangle) \in Y\times\mathbb{R} \mid\ z \in X\}$$

can be separated from $\mathrm{int}(\mathrm{dom}(f))$. Indeed, both sets are convex and nonempty, and moreover if $(y,\lambda) \in S\cap \mathrm{int}(\mathrm{dom}(f))$, then for some $z \in X$ we must have that $y = Az$, $\lambda = f(Ax) + \langle x^*, z-x\rangle$ and

$$f(Az) = f(y) < \lambda = f(Ax) + \langle x^*, z-x\rangle$$

But, this inequality is not possible because $x^* \in \partial(f\circ A)(x)$. Hence, $S\cap\mathrm{int}(\mathrm{dom}(f)) = \emptyset$ and by Hahn-Banach, there are $y^* \in Y^*$ and $\alpha \in \mathbb{R}$ (no both zero at the same time) so that

$$\langle y^*, y\rangle + \alpha\lambda < \langle y^*, \tilde{y}\rangle + \alpha\tilde{\lambda}, \qquad \forall (y,\lambda) \in \mathrm{int}(\mathrm{dom}(f)),\ (\tilde{y},\tilde{\lambda}) \in S.$$

Evaluating at $y = \tilde{y} = Ax$, $\lambda > f(Ax) = \tilde{\lambda}$, we get that $\alpha < 0$ and so by normalizing, we can assume that $\alpha = -1$ and $y^* \neq 0$. Thus, letting $\lambda \to f(y)$ we get

$$\langle y^*, y\rangle - f(y) \leq \langle y^*, Az\rangle - f(Ax) - \langle x^*, z-x\rangle, \qquad \forall y \in \mathrm{dom}(f),\ z \in X. \tag{A.3}$$

On the one hand, evaluating (A.3) at $y = Ax$ and $z = x \pm d$ for some $d \in X\setminus\{0\}$, we get that

$$\langle A^*y^* - x^*, d\rangle = 0,$$

from where we get that $x^* = A^*y^*$. On the other hand, evaluating (A.3) at $z = x$ we get that

$$f(Ax) + \langle y^*, y - Ax\rangle \leq f(y), \quad \forall y \in \mathrm{dom}(f),$$

from where we get that $y^* \in \partial f(Ax)$ and so $x^* \in A^*\partial f(Ax)$, which completes the proof.

Bibliography

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

[2] H. Cartan *Differential calculus.* Hermann; 1971.

Optimal control theory

1.1 Introduction

In this part of the course we are concerned with optimal control problems, which are in simple words, Calculus of Variation problems with an additional dynamical constraints of the type

(1.1) $$\dot{x}(t) \in F(x(t)), \text{ a.e. } t \in [0,T].$$

Here $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is a given set-valued map or multifunction, that is, $F(x) \subseteq \mathbb{R}^n$ with $F(x)$ possibly being the empty set. This set-valued maps is called *dynamics*.

We recall that a Calculus of Variation problems consists in minimizing a certain functional

$$J(x) := \int_0^T L(t,x(t),\dot{x}(t))dt + \ell(x(0),x(T)).$$

over $AC[0,T]$, the space of all the absolutely continuous arcs $x : [0,T] \to \mathbb{R}^n$. The functional in question is composed of two parts, an accumulative cost given by a function $L : [0,T] \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ called *Lagrangian* and an *end-points cost* $\ell : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$.

Essentially, an optimal control problem consists in minimizing the functional $J(x)$ over all $x \in AC[0,T]$ with the additional dynamical constraint (1.1).

1.1.1 Dynamical systems

It's common to find in the literature (1.1) written as a controlled ordinary differential equation

(1.2) $$\dot{x}(t) = f(x(t),u(t)), \quad u(t) \in U \text{ a.e. } t \in [0,T],$$

where $U \subseteq \mathbb{R}^m$ is a given nonempty set known as the *control space*, $u : [0,T] \to U$ is called a *control* or *input* and $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ is also referred to as *dynamics function*. The arc $x : [0,T] \to \mathbb{R}^n$ is also usually called *state* of the control system.

Notable examples of controlled vector field are listed below:

- **Linear systems:**
$$\dot{x}(t) = Ax(t) + Bu(t), \quad \text{for a.e. } t \in [0,T].$$
where A and B are given Real-valued matrices of dimension $n \times n$ and $n \times m$, respectively

- **Control-affine systems:**
$$\dot{x}(t) = f_0(x(t)) + \sum_{i=1}^{m} u_i(t) f_i(x(t)), \quad u(t) = (u_1(t),\dots,u_m(t)), \quad \text{for a.e. } t \in [0,T].$$

Differential inclusions vs. Differential equations

A relation between (1.1) and (1.2) can be established via

$$F(x) = f(x,U) := \{v \in \mathbb{R}^n \mid \exists u \in U,\ v = f(x,u)\}.$$

Indeed, if $x \in AC[0,T]$ is a solution to (1.2) associated with $u : [0,T] \to U$, and since

$$f(x(t),u(t)) \in f(x,U), \qquad \text{a.e. } t \in [0,T]$$

then $x \in AC[0,T]$ is also a solution to (1.1). The converse is not straightforward and requires further developments and some assumptions on the dynamics. The following result is known as the Filippov Measurable Selection Theorem (for a proof see [2, Theorem 2.3].

Lemma 1.1 (Filippov)**.** *Suppose $\phi : \mathbb{R} \times \mathbb{R}^m \to \mathbb{R}^n$ is continuous and $v : \mathbb{R} \to \mathbb{R}^n$ is a measurable function. Suppose $U \subseteq \mathbb{R}^m$ is a compact set such that $v(t) \in \phi(t,U)$ for a.e. $t \in [0,T]$. Then, there exists a measurable function $u : [0,T] \to U$ such that*

$$\dot{x}(t) = \phi(t,u(t)) \qquad a.e.\ t \in [0,T].$$

The following result implies that, essentially, the formulation as differential inclusion and the formulation as controlled ordinary differential equation of a dynamical system are equivalent.

Theorem 1.1. *Let $U \subseteq \mathbb{R}^m$ be a compact set and $f : \mathbb{R}^n \times U \to \mathbb{R}^n$ be a continuous function. Let $x \in AC[0,T]$ be a solution to (1.1) with $F(x) = f(x,U)$. Then there exists a measurable function $u : [0,T] \to U$ such that (1.2) holds.*

Proof. It's enough to use the Filippov Measurable Selection Theorem with

$$v(t) = \dot{x}(t) \qquad \text{and} \qquad \phi(t,u) = f(x(t),u).$$

□

The Gronwall's Lemma

Another important property of trajectories of dynamical systems is that the growth of their norm can be estimated if the dynamics has *linear growth*, that is, there exists $c_F > 0$ such that

$$\sup\{|v| \mid v \in F(x)\} \le c_F(1+|x|), \quad \forall x \in \mathbb{R}^n.$$

Proposition 1.1. *If $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ has linear growth, then any solution of (1.1) satisfies*

$$|x(t) - x(s)| \le (e^{c_F(t-s)} - 1)\,(|x(s)|+1), \quad \text{for any} \quad 0 \le s < t \le T.$$

Furthermore,

$$|\dot{x}(t)| \le c_F e^{C_F t}(|x(0)|+1), \quad \text{for a.e.} \quad t \in [0,T].$$

1.1.2 Particular optimal Control problems

In optimal control, the velocity of the trajectory has an indirect impact on the the cost to be minimized. It's the control parameter that plays the main role. We now introduce several types of optimal control problem. Note that in any case a Calculus of variation problem is recovered by the trivial dynamical constraint

$$\dot{x}(t) = u(t), \quad \text{for a.e. } t \in [0,T].$$

- **Lagrange problem:** For $T \in (0,+\infty)$, minimize

$$\int_0^T L(x(t),u(t))dt$$

  over all measurable function $u : [0,T] \to U$ and $x \in AC[0,T]$ such that (1.2) is satisfied and for $x_0, x_T \in \mathbb{R}^n$ given we also have $x(0) = x_0$ and $x(T) = x_T$.

- **Bolza problem:** For $T \in (0,+\infty)$, minimize

$$\int_0^T L(x(t),u(t))dt + g(x(T))$$

  over all measurable function $u : [0,T] \to U$ and $x \in AC[0,T]$ such that (1.2) is satisfied, with $x(0) = x_0$ for some $x_0 \in \mathbb{R}^n$ given.

- **Mayer problem:** For $T \in (0,+\infty)$, minimize $g(x(T))$ over all $x \in AC[0,T]$ such that (1.1) is satisfied, with $x(0) = x_0$ for some $x_0 \in \mathbb{R}^n$ given.

- **Infinite horizon problem:** For a *discount factor* $\lambda > 0$, minimize

$$\int_0^{+\infty} e^{-\lambda t} L(x(t),u(t))dt$$

  over all measurable function $u : [0,+\infty) \to U$ and $x \in AC[0,+\infty)$ such that (1.2) is satisfied, with $x(0) = x_0$ for some $x_0 \in \mathbb{R}^n$ given.

- **Minimum time problem:** Given target $S \subseteq \mathbb{R}^N$, minimize $T > 0$ such that $x(T) \in S$ over all $x \in AC[0,T]$ such that (1.1) is satisfied, with $x(0) = x_0$ for some $x_0 \in \mathbb{R}^n$ given.

1.1.3 Existence of solutions

Let us now turn our attention into some conditions that ensure the existence of minimizers of an optimal control problem. Let us focus on the Bolza problem. A fundamental assumption required is the convexity of a certain augmented dynamics

$$(H_0) \qquad \forall x \in \mathbb{R}^n, \quad \{(v,r) \in \mathbb{R}^n \times \mathbb{R} \mid \exists u \in U,\ v = f(x,u),\ L(x,u) \le r\} \text{ is convex}$$

This assumption is satisfied in some recognizable cases. For example this is the case if U is a convex set of $\mathbb{R}^m$, the dynamical system is control-affine and the Lagrangian is a convex with respect to the control ($u \mapsto L(x,u)$ is a convex function).

In order to prove the existence of solution of an optimal control problem, we need to use some classical compactness theorem of Functional Analysis, which we recall by the sake of completeness.

Lemma 1.2 (Ascoli-Arzelà). *Let I be an interval of $\mathbb{R}$ and $C(I)$ denote the space of continuous functions from I to $\mathbb{R}^n$ supplied with the topology of uniform convergence on compact subintervals of I. Then a bounded sequence $\{x_n\} \subseteq C(I)$ has a subsequence that converges uniformly on compact subintervals of I provided it is equicontinuous, that is,*

$$\forall \varepsilon > 0, \exists \delta > 0,\ \forall t,s \in I,\ |t-s| < \delta \quad \Longrightarrow \quad |x_n(t) - x_n(s)| < \varepsilon, \quad \forall n \in \mathbb{N}$$

Lemma 1.3 (Dunford-Pettis). *Let I be an interval of $\mathbb{R}$, μ be a measure on I and $L^1(I,d\mu)$ denote the equivalence class of $d\mu$-integrable functions from I to $\mathbb{R}^n$. A bounded sequence $\{x_n\} \subseteq L^1(I,d\mu)$ has compact weak closure on $L^1(I,d\mu)$ if and only if it is equi-integrable:*

$$\forall \varepsilon > 0, \exists \delta > 0,\ \forall A \subseteq I \text{ measurable},\ \mu(A) < \delta \quad \Longrightarrow \quad \int_A |x_n(t)| d\mu(t) < \varepsilon, \quad \forall n \in \mathbb{N}$$

and for any $\varepsilon > 0$ there is $A \subseteq I$ measurable with $\mu(A) < +\infty$ such that

$$\int_{I \setminus A} |x_n(t)| d\mu(t) < \varepsilon, \quad \forall n \in \mathbb{N}$$

Theorem 1.2. *Suppose that (H_0) is satisfied. Assume in addition that*

- *$U \subseteq \mathbb{R}^m$ is nonempty and compact.*
- *$f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ is continuous, with $x \mapsto f(x,U)$ having linear growth.*
- *$L : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is continuous.*
- *$g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is lower semicontinuous.*

If the Bolza problem is feasible, then there is an optimal control that solves the Bolza problem.

Proof. Let us sketch the main steps in order to prove Theorem 1.2.

1. Using Gronwall's Lemma and the fact that the Bolza problem is feasible, we have that the value of the Bolza problem

$$V = \inf_{x \in AC[0,T]} \left\{ \int_0^T L(x(t),u(t))dt + g(x(T)) \,\middle|\, \exists\, u : [0,T] \to U \text{ s.t. (1.2) holds and } x(0) = x_0 \right\}$$

   is finite, and so, we there is a minimizing sequence $\{x_n, u_n\}$ such that each $x_n \in AC[0,T]$ with $x_n(0) = x_0$, each $u_n : [0,T] \to U$ is measurable, $\dot{x}_n(t) = f(x_n(t), u_n(t))$ for a.e. $t \in [0,T]$ and

$$\int_0^T L(x_n(t), u_n(t))dt + g(x_n(T)) \to V \quad \text{as } n \to +\infty$$

2. Using Gronwall's Lemma and Ascoli-Arzelà theorem, we prove that $\{x_n\}$ is relatively compact on $C[0,T]$.

3. Then, using Gronwall's Lemma and Dunford-Pettis theorem, we prove that $\{\dot{x}_n\}$ is relatively weakly compact on $L^1([0,T],dt)$.

4. We combine step 2 and 3 to show that there is $x \in AC[0,T]$ such that, up to a subsequence, $x_n \to x$ uniformly on $[0,T]$ and $\dot{x}_n \rightharpoonup \dot{x}$ weakly on $L^1([0,T],dt)$.

5. We use Gronwall's Lemma, Dunford-Pettis theorem and the fact that L is uniformly continuous on $\mathbb{B}(x_0,(e^{c_F T}-1)(|x_0|+1)) \times U$ in order to prove that $\{L(x_n,u_n)\}$ weakly converges, passing into a subsequence if necessary, to some function $r \in L^1([0,T],dt)$.

6. The next step consists in showing that there are $v_n \in \overline{\text{co}\{x_n\}}$ and $r_n \in \overline{\text{co}\{L(x_n,u_n)\}}$ such that $v_n(t) \to \dot{x}(t)$ and $r_n(t) \to r(t)$ for a.e. $t \in [0,T]$.

7. We now define $F(x) = \{(v,r) \in \mathbb{R}^n \times \mathbb{R} \mid \exists u \in U,\ v = f(x,u),\ L(x,u) \leq r\}$, and show that
$$(x(t),r(t)) \in F(x(t)), \quad \text{for a.e. } t \in [0,T]$$
To do this, we use (H_0), the previous step and the fact that
$$(x_n(t),L(x_n(t),u_n(t))) \in F(x_n(t)), \quad \text{for a.e. } t \in [0,T]$$

8. Using the Filippov's Measurable selection theorem, we get that there are measurable functions $u : [0,T] \to U$ and $\xi : [0,T] \to [0,+\infty)$ such that
$$\dot{x}(t) = f(x(t),u(t)) \quad \text{and} \quad r(t) - \xi(t) = L(x(t),u(t)), \quad \text{for a.e. } t \in [0,T]$$

9. Finally, the conclusion is reached by using the lower semiconitnuity of g and the fact that the function constantly equal to 1 belongs to $L^\infty([0,T],dt)$, and so
$$\int_0^T L(x_n(t),u_n(t)))dt \to \int r(t)dt \quad \text{as} \quad n \to +\infty.$$

$\square$

1.1.4 Value functions and Hamilton-Jacobi equations

The Value Function is the mapping that associates the initial data of the problem with the minimal cost-to-go of the optimal control problem. For the optimal control problems we have introduced, it takes the following forms:

- **Bolza problem:**
$$\vartheta(\tau,\xi) := \inf_{x \in AC[\tau,T]} \left\{ \int_\tau^T L(x(t),u(t))dt + g(x(T)) \,\middle|\, \exists\, u : [\tau,T] \to U \text{ s.t. (1.2) holds and } x(\tau) = \xi \right\}$$

- **Mayer problem:**
$$\vartheta(\tau,\xi) := \inf_{x \in AC[\tau,T]} \{g(x(T)) |\ x \text{ satisfies (1.1) and } x(\tau) = \xi\}$$

- **Infinite horizon problem:** For a *discount factor* $\lambda > 0$,
$$\vartheta(\xi) := \inf_{x \in AC[0,+\infty)} \left\{ \int_0^{+\infty} L(x(t),u(t))dt \,\middle|\, \exists\, u : [0,+\infty) \to U \text{ s.t. (1.2) holds and } x(0) = \xi \right\}$$

- **Minimum time problem:** For a target $S \subseteq \mathbb{R}^N$,
$$\mathcal{T}^S(\xi) := \inf_{x \in AC[0,T]} \{T \mid x \text{ satisfies (1.1)}, x(T) \in S \text{ and } x(0) = \xi\}$$

Bellman's Dynamic programming principle

The interest in studying the value function of an optimal control problem lies in the potentiality of computing this mapping before solving the optimization problem. The most powerful tool for doing so is the Hamilton-Jacobi (HJ) approach, which is a technique based on a functional equation known as the *Dynamic Programming Principle (DPP)*. This methodology dates from the 1950's and was first studied by Bellman and his coauthors.

This equation has different forms based on the issue at hand:

- **Bolza problem:** for any $\tau \in [0,T]$ and $h \in [0,T-\tau]$

$$V(\tau,\xi) = \inf_{x\in AC[\tau,\tau+h]} \left\{ \int_{\tau}^{\tau+h} L(x(t),u(t))dt + V(\tau+h,x(\tau+h)) \,\middle|\, \begin{array}{c} u:[\tau,\tau+h]\to U \text{ s.t. (1.2)} \\ \text{holds and } x(\tau)=\xi \end{array} \right\}$$

- **Mayer problem:** for any $\tau \in [0,T]$ and $h \in [0,T-\tau]$

$$V(\tau,\xi) = \inf_{x\in AC[\tau,\tau+h]} \{V(\tau+h,x(\tau+h)) \mid x \text{ satisfies (1.1) and } x(\tau)=\xi\}$$

- **Infinite horizon problem:** for any $\tau \in [0,+\infty)$

$$V(x) = \inf_{x\in AC[0,\tau]} \left\{ \int_0^{\tau} e^{-\lambda s} L(x(t),u(t))dt + e^{-\lambda\tau}V(x(\tau)) \,\middle|\, \begin{array}{c} u:[0,\tau]\to U \text{ s.t. (1.2)} \\ \text{holds and } x(0)=\xi \end{array} \right\}$$

- **Minimum time problem:** for any $\tau \in [0,T^S(\xi)]$

$$T^S(\xi) = \inf_{x\in AC[0,\tau]} \left\{ \tau + T^S(x(\tau)) \mid x \text{ satisfies (1.1) and } x(0)=\xi \right\}$$

The main advantage of this method is that essentially, the Value Function is the unique mapping that verifies the DPP and therefore, the idea is to find an equivalent formulation of this optimality principle in terms of a partial differential equation called the *Hamilton-Jacobi-Bellman (HJB) equation*. The corresponding HJB equations for the value functions given above are as follows:

- **Bolza problem:**

$$-\partial_t V(t,x) + H(x,\nabla_x V(t,x)) = 0, \quad \forall t \in (0,T), \qquad V(T,x) = g(x)$$

  where the Hamiltonian is given by

$$H(x,y) := \sup_{u\in U}\{-\langle f(x,u),y\rangle - L(x,u)\}.$$

- **Mayer problem:**

$$-\partial_t V(t,x) + H(x,\nabla_x V(t,x)) = 0, \quad \forall t \in (0,T), \qquad V(T,x) = g(x)$$

  where the Hamiltonian is given by

$$H(x,y) := \sup_{v\in F(x)}\{-\langle v,y\rangle\}.$$

- **Infinite horizon problem:**
$$\lambda V(x) + H(x, \nabla V(x)) = 0$$
where the Hamiltonian is the same as for the Bolza problem.

- **Minimum problem:**
$$-1 + H(x, \nabla T(x)) = 0, \quad x \in \operatorname{int}(\operatorname{dom} T) \setminus S, \qquad T(x) = 0, \quad \forall x \in S.$$
where the Hamiltonian is the same as for the Mayer problem.

Let us point out that value functions are rarely differentiable, and consequently, solutions to the HJB equations need to be understood in a weak sense. The most suitable framework to deal with these equations is the *Viscosity Solutions Theory* introduced by Crandall-Lions in 1983. This methodology is based on two semisolution concepts, namely the *viscosity supersolution* and *subsolution*, respectively. The theory provides existence and uniqueness for a much more general class of fully nonlinear Hamilton-Jacobi equations, not necessarily related to an optimal control problem).

In order to introduce the notion of viscosity solution, we introduce the viscosity subdifferential:
$$\partial_V \omega(x) := \left\{ y \in \mathbb{R}^n \,\middle|\, \liminf_{\tilde{x} \to x} \frac{\omega(\tilde{x}) - \omega(x) - \langle y, \tilde{x} - x \rangle}{|\tilde{x} - x|} \geq 0 \right\}$$
and the viscosity superdifferential:
$$\partial^V \omega(x) := \left\{ y \in \mathbb{R}^n \,\middle|\, \limsup_{\tilde{x} \to x} \frac{\omega(\tilde{x}) - \omega(x) - \langle y, \tilde{x} - x \rangle}{|\tilde{x} - x|} \leq 0 \right\}.$$

Definition 1.1. *A continuous function $\omega : \Omega \to \mathbb{R}$ is said to be a viscosity solution of a Hamilton-Jacobi equation*
$$H(x, \nabla \omega) = 0, \quad x \in \Omega$$
if it is a viscosity supersolution, that is,
$$H(x, y) \geq 0, \quad x \in \Omega, \ \forall y \in \partial_V \omega(x)$$
and if it is a viscosity subsolution, that is,
$$H(x, y) \leq 0, \quad x \in \Omega, \ \forall y \in \partial^V \omega(x)$$

Proposition 1.2. *Suppose that*

- *$U \subseteq \mathbb{R}^m$ is nonempty and compact.*
- *$f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ is continuous, with $x \mapsto f(x, u)$ Lipschitz continuous, uniformly on $u \in U$.*
- *$L : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ and $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ are continuous.*

Then the value function of the Bolza problem is a viscosity solution of the corresponding HJ equation.

Let us point out that the definition of a viscosity solution to a HJ equation is sometimes stated in terms of test functions. This is due to the following proposition.

Proposition 1.3. *Let $y \in \mathbb{R}^n$. Then $y \in \partial_V \omega(x)$ at $x \in \operatorname{dom} \omega$ if and only if there exists a continuous function $\varphi : \mathbb{R}^n \to \mathbb{R}$ differentiable at x such that $\nabla \varphi(x) = y$ and $\omega - \varphi$ attains a local minimum at x. Similarly, $y \in \partial^V \omega(x)$ at $x \in \operatorname{dom} \omega$ if and only if there exists a continuous function $\varphi : \mathbb{R}^n \to \mathbb{R}$ differentiable at x such that $\nabla \varphi(x) = y$ and $\omega - \varphi$ attains a local maximum at x.*

1.2 Fully convex Bolza problem problems

The discussion now is centered about the fully convex case, that is, the Lagrangian end-point-cost are assumed to be convex. This means in particular that the functional to be minimized is convex. It is worthy to mention that we are now adopting the same formulation as in Calculus of variations, that is, $L(x,v)$ instead of a control formulation. For this reason, we assume $L:\mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$ and $\ell:\mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$ belong to the class Γ_0, that is, they are convex proper and lower semicontinuous functions.

1.2.1 Primal problem and implicit constraints

Given a Lagrangian $L:\mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$ and an end-points cost $\ell:\mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$, we are concerned with the Bolza problem:

$$(P_0)\qquad \text{Minimize } \int_0^T L(x(t),\dot{x}(t))dt+\ell(x(0),x(T)),\quad \text{over all } x\in AC[0,T].$$

By allowing L to take infinite values, we are handling implicitly constraints over the state of system. Indeed, let us define

$$(1.3)\qquad \mathcal{X}:=\{x\in\mathbb{R}^n \mid \exists v\in\mathbb{R}^n,\ L(x,v)<+\infty\}\quad\text{and}\quad F(x):=\{v\in\mathbb{R}^n \mid L(x,v)<+\infty\}.$$

In a similar way, by allowing ℓ to take infinite values, we are handling implicitly constraints over end-points of the system. In this case we define define

$$(1.4)\qquad A:=\{(a,b)\in\mathbb{R}^n\times\mathbb{R}^n \mid \ell(a,b)<+\infty\}.$$

Note the for any feasible arc for (P_0), we must have that

$$(1.5)\qquad x(t)\in X,\quad \forall t\in[0,T],\quad \dot{x}(t)\in F(x(t)),\quad \text{for a.e. } t\in[0,T]\quad\text{and}\quad (x(0),x(T))\in A$$

The implicit constraints of the Fully convex Bolza problem at hand satisfied some basic structural conditions, which we summarize below.

Lemma 1.4. *X and A are nonempty convex sets, $F:\mathbb{R}^n\rightrightarrows\mathbb{R}^n$ is a proper convex set-valued map, that is,*

$$\lambda F(x)+(1-\lambda)F(y)\subseteq F(\lambda x+(1-\lambda)y),\quad \forall x,y\in\mathbb{R}^n,\ \lambda\in[0,1],$$

its images are convex subsets of $\mathbb{R}^n$ and $\mathrm{dom}(F)=\mathcal{X}$.

1.2.2 Examples

The convexity assumptions may seem rather restrictive. However, they are enough to treat a large range of problems. Let us now present some examples to illustrate the scope of the theory we want to develop.

1) **Linear-Quadratic problems:** The main feature of these problems is that the functional to be minimized is an arbitrary quadratic function and the dynamical constraint is a linear system.

Let $U \subseteq \mathbb{R}^m$ be a convex closed nonempty set and consider some matrices $A, P, Q, S \in \mathbb{M}_{n\times n}$, $R \in \mathbb{M}_{m\times m}$ and $B \in \mathbb{M}_{n\times m}$, with P, Q, R and S being positive semi-definite. The Linear-Quadratic problem is

$$(\mathrm{P_{LQ}}) \qquad \begin{cases} \text{Minimize over all } x \in AC[0,T] \text{ the functional:} \\ \dfrac{1}{2}\displaystyle\int_0^T \left(\langle Px(t), x(t)\rangle + \langle Ru(t), u(t)\rangle\right) dt + \dfrac{1}{2}\left(\langle Qx(0), x(0)\rangle + \langle Sx(T), x(T)\rangle\right), \\ \text{ subject to the dynamical constraint:} \\ \dot{x}(t) = A(t)x(t) + B(t)u(t), \quad u(t) \in U, \quad \text{ for a.e. } t \in [0,T]. \end{cases}$$

Under these circumstances, the accumulative cost is

$$L(x,v) = \frac{1}{2}\langle Px, x\rangle + \inf_{u\in U}\{\langle Ru, u\rangle \mid \ v = Ax + Bu\}$$

and the dynamics is

$$F(x) = Ax + BU.$$

2) **Optimal problem with mixed-convex cosntraints:** In a control problem, the dynamical constraint may also be written in an implicit form. For example, suppose we are concerned with

$$(\mathrm{P_{mix}}) \qquad \begin{cases} \text{Minimize } \displaystyle\int_0^T L_0(x(t), \dot{x}(t))dt + \ell_0(x(0), x(T)), \\ \text{over all } x \in AC[0,T] \text{ that satisfies} \\ L_i(x(t), \dot{x}(t)) \leq 0, \quad \forall i = 1, \ldots, k, \ \text{ for a.e. } t \in [0,T] \\ \ell_j(x(0), x(T)) \leq 0, \quad \forall j = 1, \ldots, l \end{cases}$$

We recover the formulation as a Calculus of Variation problem by setting the dynamics as

$$F(x) := \{v \in \mathbb{R}^n \mid \ L_i(x,v) \leq 0, \quad \forall i = 1, \ldots, k, \},$$

and the Lagrangian as:

$$L(x,v) := \begin{cases} L_0(x,v) & \text{if } v \in F(x) \\ +\infty & \text{otherwise.} \end{cases}$$

Furthermore, the end-points cost is given explicitly by

$$\ell = \{(a,b) \in \mathbb{R}^n \times \mathbb{R}^n \mid \ \ell_j(a,b) \leq 0, \quad \forall j = 1, \ldots, l\}.$$

1.2.3 Dual problem

This fully convex framework we are concerned with yields naturally to a duality theory. We can associated a dual problem with (P_0) by conjugate operation. The dual problem is

$$(D_0)\qquad \text{Minimize } \int_0^T M(y(t),\dot{y}(t))dt + m(y(0),y(T)), \quad \text{over all } y \in AC[0,T],$$

where the Lagrangian $M:\mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$ and an end-points cost $m:\mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$ of the dual problem are given by

$$M(y,w) := L^*(w,y) \quad \text{and} \quad m(q,b) := \ell^*(q,-b) \tag{1.6}$$

Given that the Lagrangian and end-points cost of the primal problem are assumed to be proper convex lower semicontinuous function, it is not difficult to see that it also holds M and m are proper convex lower semicontinuous functions.

By symmetry, this formulation considers implicitly constraints over the state of system. Indeed, the state constraints and dynamics of the dual problem are

$$Y := \{y\in\mathbb{R}^n \mid \exists w\in\mathbb{R}^n,\, M(y,w)<+\infty\} \quad \text{and} \quad G(y) := \{w\in\mathbb{R}^n \mid M(y,w)<+\infty\}. \tag{1.7}$$

and the end-points constraint is

$$B := \{(a,b)\in\mathbb{R}^n\times\mathbb{R}^n \mid m(a,b)<+\infty\}. \tag{1.8}$$

Likewise for the primal problem, any feasible arc for (D_0) satisfies as well

$$y(t)\in Y, \quad \forall t\in[0,T], \quad \dot{y}(t)\in G(y(t)), \quad \text{for a.e. } t\in[0,T] \quad \text{and} \quad (y(0),y(T))\in B \tag{1.9}$$

Moreover, the symmetry between primal and dual problem can be take further to provide analogous structural properties for the implicit constraints. Lemma 1.4 can be reformulated for the dual problem in the following terms.

Lemma 1.5. *Y and B are nonempty convex sets, $G:\mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is a proper convex set-valued map with convex images and* $\mathrm{dom}(G)=Y$.

On the other hand, thanks to the Young-Fenchel inequality, for any $x,y\in AC[0,T]$ feasible for (P_0) and (D_0), respectively, we have

$$\int_0^T L(x(t),\dot{x}(t))dt + \ell(x(0),x(T)) + \int_0^T M(y(t),\dot{y}(t))dt + m(y(0),y(T)) \geq 0.$$

Thus, with the convention $+\infty-\infty=-\infty+\infty=+\infty$, the primal and dual problem satisfy the weak duality property:

$$\mathrm{val}(P_0)+\mathrm{val}(D_0)\geq 0 \tag{1.10}$$

1.2.4 Lower semicontinuity properties of the functional

Consider $\Phi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ and $\varphi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$. We denote by $J_{\Phi,\phi} : AC[0,T] \to \mathbb{R} \cup \{+\infty\}$ the functional defined via

$$J_{\Phi,\phi}(z) := \int_0^T \Phi(z(t), \dot{z}(t))dt + \varphi(z(0), z(T)), \quad \forall z \in AC[0,T]$$

Note that $J_{L,\ell}$ and $J_{M,m}$ are the functionals to be minimized in (P_0) and (D_0), respectively.

Note that we can make the following identification: $AC[0,T] \cong \mathbb{R}^n \times L^1[0,T]$, this is because an arc $x : [0,T] \to \mathbb{R}^n$ belongs to $AC[0,T]$ if there is $v \in L^1$ such that

$$x(t) = x(0) + \int_0^t v(s)ds, \quad \forall t \in [0,T].$$

Proposition 1.4. *If $\Phi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ and $\varphi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ are proper convex and lower semicontinuous functions. Then the functional $J_{\Phi,\phi} : AC[0,T] \to \mathbb{R} \cup \{+\infty\}$ is convex and lower semicontinuous with respect to the strong and weak topologies on $AC[0,T]$.*

Note that the presence of possible state and end-point constraints may preclude the properness of the functionals. This is the case, if one of the two problems (primal or dual) is not feasible.

1.2.5 Duality of the infimum

Let us now turn our attention into condition in order to ensure the strong duality property between (P_0) and (D_0), that is

$$\text{val}(P_0) + \text{val}(D_0) = 0 \tag{1.11}$$

Since $AC[0,T]$ can be identified with $\mathbb{R}^n \times L^1[0,T]$, the topological dual of $AC[0,T]$ can be identified with $\mathbb{R}^n \times L^\infty[0,T]$ with the duality product:

$$\ll x, (a,p) \gg = \langle x(0), a \rangle + \int_0^T \langle \dot{x}(t), p(t) \rangle dt, \quad \forall x \in AC[0,T],\ (a,p) \in \mathbb{R}^n \times L^\infty[0,T]$$

In order to define a perturbation to study the relation between the primal and dual problem, we introduce further notation: Consider $\Phi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ and $\varphi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, we denote by $\Psi_{\Phi,\phi} : AC[0,T] \times \mathbb{R}^n \times L^\infty[0,T] \to \mathbb{R} \cup \{+\infty\}$ the functional defined by

$$\Psi_{\Phi,\phi}(z,a,p) := \int_0^T \Phi(z(t)+p(t), \dot{z}(t))dt + \varphi(z(0)+a, z(T)), \quad \forall z \in AC[0,T],\ (a,p) \in \mathbb{R}^n \times L^\infty[0,T]$$

and the corresponding value function

$$\vartheta_{\Phi,\varphi}(a,p) = \inf_{x \in AC[0,T]} \Psi_{\Phi,\varphi}(x,a,p), \quad \forall (a,p) \in \mathbb{R}^n \times L^\infty[0,T]$$

Hence, $\Psi_{L,\ell}$ and $\Psi_{M,m}$ are perturbation for the primal and dual problems because

$$\text{val}(P_0) = \vartheta_{L,\ell}(0,0) \quad \text{and} \quad \text{val}(D_0) = \vartheta_{M,m}(0,0)$$

Proposition 1.5. *We have that*

$$\vartheta^*_{M,m}(x(0),\dot{x}) = J_{L,\ell}(x) \quad \text{and} \quad \vartheta^*_{L,\ell}(y(0),\dot{y}) = J_{M,m}(y) \quad \forall x,y \in AC[0,T].$$

Moreover, if $\vartheta_{L,\ell}$ and $\vartheta_{M,m}$ are proper and lower semicontinuous on $\mathbb{R}^n \times L^\infty[0,T]$ for the weak-$\star$ topology, then

$$\vartheta_{M,m}(a,p) = J^*_{L,\ell}(a,p) \quad \text{and} \quad \vartheta_{L,\ell}(a,p) = J^*_{M,m}(a,p) \quad \forall (a,p) \in \mathbb{R}^n \times L^\infty.$$

The proof of the preceding result relies on a much general property. It's worthy to mention that here we are assuming that L^1 and L^∞ are in duality, that is, L^1 is endowed with the weak topology and L^∞ with the weak-$\star$ topology.

Lemma 1.6. *Let $f \in \Gamma_0(\mathbb{R}^n \times \mathbb{R}^n)$ and $1 \le p_1, p_2 \le +\infty$. Then, the functional $I_f : L^{p_1}[0,T] \times L^{p_2}[0,T] \to \mathbb{R} \cup \{+\infty\}$ defined via*

$$I_f(\alpha,\beta) = \int_0^t f(\alpha(t),\beta(t))dt, \quad \forall (\alpha,\beta) \in L^{p_1}[0,T] \times L^{p_2}[0,T]$$

is convex proper and lower semicontinuous, and its Fenchel conjugate is given by

$$(I_f)^*(u,v) = I_{f^*}(u,v) := \int_0^t f^*(u(t),v(t))dt, \quad \forall (u,v) \in L^{q_1}[0,T] \times L^{q_2}[0,T]$$

where $\frac{1}{p_1} + \frac{1}{q_1} = 1$ and $\frac{1}{p_2} + \frac{1}{q_2} = 1$.

Remark 1.1. *Note that* $\mathrm{val}(P_0) = J^*_{L,\ell}(0,0)$ *and* $\mathrm{val}(D_0) = J^*_{M,m}(0,0)$. *Hence, if any of the value functions is proper and lower semicontinuous on $\mathbb{R}^n \times L^\infty[0,T]$ for the weak-$\star$ topology, then (1.11) holds. In general, if neither value function satisfies this property, we only have that*

$$\overline{\vartheta_{M,m}}(0,0) = \mathrm{val}(P_0) \quad \text{and} \quad \overline{\vartheta_{L,\ell}}(0,0) = \mathrm{val}(D_0),$$

where $\overline{f}$ is the weak-$\star$ lower semicontinuous envelop of a function $f : \mathbb{R}^n \times L^\infty[0,T] \to \mathbb{R} \cup \{+\infty\}$, that is,

$$\overline{f}(a,p) = \inf\left\{ \liminf_{\alpha\in\Lambda} f(a_\alpha,p_\alpha) \,\middle|\, \{(a_\alpha,p_\alpha)\}_{\alpha\in\Lambda} \text{ is a net such that } (a_\alpha,p_\alpha) \xrightarrow{\star} (a,p) \right\}$$

The proof of the preceding result also relies on some classical theorem for integrals, which we recall by sake of completeness.

Lemma 1.7 (Dominated convergence theorem)**.** *Let $f : [0,T] \to \mathbb{R}$ and $\{f_n\}$ be a sequence of functions in $L^1[0,T]$ such that*

- *$f_n(t) \to f(t)$ for a.e. on $[0,T]$.*
- *$|f_n(t)| \le g(t)$ for a.e. on $[0,T]$ for some $g \in L^1[0,T]$.*

Then $f \in L^1[0,T]$ and $f_n \to f$ in $L^1[0,T]$.

Lemma 1.8 (Fatou's Lemma)**.** *Let $\{f_n\}$ be a sequence of functions in $L^1[0,T]$ such that $f_n(t) \ge 0$ for a.e. on $[0,T]$. Then*

$$\int_0^T \liminf_{n\to+\infty} f_n(t)dt \le \liminf_{n\to+\infty} \int_0^T f_n(t)dt.$$

1.2.6 Duality gap

Let us now provide a criterion in order to make at least one of the value functions of the perturbed problems to be lower semicontinuous on $\mathbb{R}^n \times L^\infty[0,T]$ for the weak-$\star$ topology.

Suppose that $\vartheta_{\Phi,\varphi}$ is finite and bounded above on the set

$$\{(a,p) \in \mathbb{R}^n \times L^\infty[0,T] \mid \ |a-\bar{a}| + \|p-\bar{p}\|_{L^1} < \varepsilon\}$$

for some $\varepsilon > 0$ and $(\bar{a},\bar{p}) \in \mathbb{R}^n \times L^\infty[0,T]$. This is for example the case if Φ is locally Lipschitz continuous with respect to the first variable. Then, given that $\vartheta_{\Phi,\varphi}$ is convex, the function is continuous everywhere on $\mathbb{R}^n \times L^\infty[0,T]$ for the norm

$$\|(a,p)\| = |a| + \|p\|_{L^1}$$

The weak topology on the normed space $(L^\infty[0,T], \|\cdot\|_{L^1})$ can be identified with the weak-$\star$ topology on $L^\infty[0,T]$ (as topological dual of $L^1[0,T]$). Now, let us recall if $(X,\|\cdot\|)$ is a normed space, then a convex function on X is lower semicontinuous for the strong topology if and only if it is for the weak topology. This in consequence means that the $\vartheta_{\Phi,\varphi}$ is weak-$\star$ lower semicontinuous on $L^\infty[0,T]$.

The following case is an example of an optimal control problem, where there is duality gap, that is,

$$\mathrm{val}(P_0) + \mathrm{val}(D_0) > 0$$

Example 1.1. *If $L(x,v) = \delta_{[0,+\infty)}(x)$ and $\ell(a,b) = a$, then* $\mathrm{val}(P_0) = 0$. *However, in this case $M(y,w) = \delta_{\{0\}\times(-\infty,0]}(y,w)$ and $m(a,b) = \delta_{\{(1,0)\}}(a,b)$, then (D_0) is not feasible and* $\mathrm{val}(D_0) = +\infty$.

1.2.7 Optimality conditions

The duality condition can be expressed in terms of the a generalized version of the Euler-Lagrange equation and a transversality condition. We recall that the classical Euler-Lagrange equation found in mechanics and in classical Calculus of Variations is

$$\frac{\partial}{\partial t}\partial_v\Phi(x(t),\dot{x}(t)) - \partial_x\Phi(x(t),\dot{x}(t)) = 0$$

If we set $y(t) = \partial_v\Phi(x(t),\dot{x}(t))$, then the Euler-Lagrange equation takes the form

$$(\dot{y}(t), y(t)) \in \nabla\Phi(x(t),\dot{x}(t))$$

In the fully convex setting we expect to work with subdifferentiable Lagrangian rather than with smooth ones. For this reason the Euler-Lagrange equation need to be understood in a weaker sense.

Definition 1.2. *An arc $\bar{z} \in AC[0,T]$ is called an extremal of problem*

$$\inf_{z\in AC[0,T]} J_{\Phi,\varphi}(z)$$

if there is another arc $p \in AC[0,T]$ (called co-extremal of $J_{\Phi,\varphi}$ associated with $\bar{z}$) such that the generalized Euler-Lagrange equation *is satisfied:*

$$(\dot{p}(t), p(t)) \in \partial\Phi(\bar{z}(t),\dot{\bar{z}}(t)), \quad a.e.\ t \in [0,T] \tag{1.12}$$

as well as the transversality condition

$$(p(0), -p(T)) \in \partial\varphi(\bar{z}(0),\bar{z}(T)) \tag{1.13}$$

Similarly as done for abstract convex optimization problem, the optimality conditions provided by the duality theorem say that the co-extremals are in indeed optimal solution to the dual problem.

Theorem 1.3. *Consider two arcs $\bar{x}, \bar{y} \in AC[0,T]$. Then, the following are equivalent:*

1. *$\bar{x}$ is an extremal for problem (P_0) with $\bar{y}$ being a corresponding co-extremal.*
2. *$\bar{y}$ is an extremal for problem (D_0) with $\bar{x}$ being a corresponding co-extremal.*
3. *$\bar{x}$ solves (P_0), $\bar{y}$ solves (D_0) and the strong duality property (1.11) holds.*

1.2.8 Hamiltonian system

The optimality conditions can also be written in terms of a generalized Hamiltonian systems, derived from the generalized Euler Lagrange equation. In classical Calculus of Variations

$$(-\dot{y}(t), \dot{x}(t)) \in \nabla H(x(t), y(t))$$

The Hamiltonian of the problems at hand have the form

$$H(x,y) := \sup_{v \in \mathbb{R}^n} \{\langle y, v\rangle - L(x,v)\},$$

This function is seldom differentiable, so we need as well to understand the Hamiltonian systems in a weaker sense. Note that this function is convex in y for any x fixed and concave in x for any y fixed, we might say that the Hamiltonian $H : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}$ is a *concave-convex function*.

In order to extend the notion of a Hamiltonian system to this nonsmooth setting, we introduce the following partial subdifferential for concave-convex Hamiltonian H.

$$\begin{aligned}
\partial_x H(x,y) &:= \{w \in \mathbb{R}^n \mid H(z,y) \leq H(x,y) + \langle w, z-x\rangle,\ \forall z \in \mathbb{R}^n\} \\
\partial_y H(x,y) &:= \{v \in \mathbb{R}^n \mid H(x,y) + \langle v, z-y\rangle \leq H(x,z),\ \forall z \in \mathbb{R}^n\}
\end{aligned}$$

That is, $\partial_x H$ is the negative of subdifferential of the convex function $x \mapsto -H(x,y)$ and $\partial_y H$ is the subdifferential of the convex function $y \mapsto H(x,y)$.

It turns out that the optimality conditions for an arc $x \in ac$ to be a solution of (P_0) involves the existence of an adjoint arc $y \in ac$, which turns out to be an optimal solution of (D_0), and in such case, both arcs are characterized as a solution of the generalized Hamiltonian system

$$-\dot{y}(t) \in \partial_x H(x(t), y(t)) \quad \text{and} \quad \dot{x}(t) \in \partial_y H(x(t), y(t)), \quad \text{a.e. } t \in [0,T] \tag{1.14}$$

Proposition 1.6. *Consider two arcs $\bar{x}, \bar{y} \in AC[0,T]$. Then, the following are equivalent:*

1. *The pair $(\bar{x}, \bar{y})$ satisfies the generalized Hamiltonian systems (1.14).*
2. *$\bar{x}$ satisfies the generalized Euler-Lagrange equation (1.12) for L with $\bar{y}$ as co-extremal.*
3. *$\bar{y}$ satisfies the generalized Euler-Lagrange equation (1.12) for M with $\bar{x}$ as co-extremal.*

The Hamiltonian formulation of the optimality conditions are useful because it turns the quest of optimal trajectories into the study of the existence of solution to a differential inclusion:

Let us consider the set-valued map $\Gamma : \mathbb{R}^n \times \mathbb{R}^n \rightrightarrows \mathbb{R}^n \times \mathbb{R}^n$ defined via:

$$\Gamma(x,y) := \{(v,-w) \in \mathbb{R}^n \times \mathbb{R}^n \mid w \in \partial_x H(x,y),\ v \in \partial_y H(x,y)\}$$

It's not difficult to see that a pair (x,y) satisfies the generalized Hamiltonian systems (1.14) if and only if it's a solution to the differential inclusion

$$(\dot{x}(t),\dot{y}(t)) \in \Gamma(x(t),y(t)) \tag{1.15}$$

We see also that $\Gamma(x,y)$ is a convex set for any $(x,y) \in \mathbb{R}^n \times \mathbb{R}^n$. It is also compact and nonempty whenever H is continuous; because it's essentially made of two subdifferentials. Besides, Γ also satisfies (on the interior of the domain of H) a continuity property for set-valued maps called *upper semicontinuity*.

Definition 1.3. *A set-valued map $\Gamma : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ is called upper semicontinuous at $z \in \mathbb{R}^m$ if for any open set $O \subseteq \mathbb{R}^m$ that contains $\Gamma(z)$, there is $\delta > 0$ such that $\Gamma(\tilde{z}) \subseteq O$ for each $\tilde{z} \in \mathbb{B}(z,\delta)$.*

The fact that Γ is upper semicontinuous on the interior of the domain of H is a direct consequence of the following result for subdifferential of convex functions.

Lemma 1.9. *Let $f \in \Gamma_0(\mathbb{R}^n)$, then for any $x \in \operatorname{int}(\operatorname{dom}(f))$ and any $\varepsilon > 0$ there is $\delta > 0$ such that*

$$\partial f(y) \subseteq \partial f(x) + \mathbb{B}(0,\varepsilon), \qquad \forall y \in \mathbb{B}(x,\delta).$$

Under these circumstances, a general theorem for existence of local solution to the differential inclusion (1.15) can be invoked.

Proposition 1.7 ([1, Theorem 2.1.3])**.** *Suppose $\Gamma : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ is upper semicontinuous on a neighborhood of $z_0 \in \mathbb{R}^m$ with nonempty, convex and compact images. Then there exists $T > 0$ such that*

$$\dot{z}(t) \in \Gamma(z(t))$$

has a solution $z : [0,T] \to \mathbb{R}^m$ with $z(0) = z_0$.

1.2.9 Maximum principle

Let us now illustrate the relation with the *Pontryagin Maximum principle* and the fully convex approach we have taken. For this purpose, let us pick up the notation introduced originally for optimal control problems, with explicit dependence on the control. For simplicity we consider the Lagrange problem with fixed end-points, that is:

$$\text{Minimize} \quad \int_0^T \Lambda(x(t),u(t))dt$$

over all measurable function $u : [0,T] \to U$ and $x \in AC[0,T]$ such that

$$\dot{x}(t) = Ax(t) + Bu(t), \quad u(t) \in U \text{ a.e. } t \in [0,T],$$

is satisfied, with $x(0) = x_0$ and $x(T) = x_T$ be given.

The Hamiltonian of the problem is the same that appeared when we discussed the Hamilton-Jacobi equations, that is,

$$H(x,y) = \max_{u \in U} \{\langle Ax + Bu, y\rangle - \Lambda(x,u)\}$$

Theorem 1.4. *Consider two arcs $\bar{x},\bar{y} \in AC[0,T]$ and a measurable function $\bar{u} : [0,T] \to U$. Assume that $U \subseteq \mathbb{R}^m$ is nonempty convex and compact. Furthermore, assume that $\bar{x}(0) = x_0$, $\bar{x}(T) = x_T$ and*

$$\dot{\bar{x}}(t) = A\bar{x}(t) + B\bar{u}(t), \quad \textit{for a.e. } t \in [0,T].$$

Then the following are equivalent:

- *The pair $(\bar{x},\bar{y})$ satisfies the generalized Hamiltonian systems (1.14)*
- *The maximum principle is satisfied:*

$$H(\bar{x}(t),\bar{y}(t)) = \langle A\bar{x}(t) + B\bar{u}(t), \bar{y}(t)\rangle - \Lambda(\bar{x}(t),\bar{u}(t))$$

Moreover, if Λ is differentiable with respecto to x, then

$$\dot{y}(t) = -A^* y(t) + \nabla_x \Lambda(\bar{x}(t),\bar{u}(t)), \quad \textit{for a.e. } t \in [0,T].$$

Bibliography

[1] J.-P. Aubin and A. Cellina. *Differential inclusions: set-valued maps and viability theory*. Springer-Verlag, 1984.

[2] G. Smirnov. *Introduction to the theory of differential inclusions*. American Mathematical Society, 2002.