# Convexité, Optimisation et Contrôle Stochastique

Simon Depaule

Novembre 2023

## Table des matières

# 1 Chapitre 1 - Convexité et Optimisation

## 1.1 Ensembles et fonctions convexes

On fixe dans le reste du cours un **espace vectoriel normé** de dimension finie.

**Définition (Ensemble convexe) :** Un sous ensemble $\mathcal{C} \subseteq E$ est dit **convexe** si :

$$\forall \lambda \in [0,1], \forall x, y \in E, \quad \lambda x + (1-\lambda)y \in \mathcal{C}$$

**Définition (Combinaison convexe) :** Pour $x_1, \ldots, x_n \in E$ et $\lambda_1, \ldots, \lambda_n \in \mathbb{R}_+$ tel que $\sum_{i=1}^n \lambda_i = 1$, la quantité $\sum_{i=1}^n \lambda_i x_i = 1$ est appelée **combinaison convexe** des $x_1, \ldots, x_n$

**Propriété :** $\mathcal{C} \subseteq E$ est convexe $\Longleftrightarrow$ toute combinaison convexe d'éléments de $\mathcal{C}$ est dans $\mathcal{C}$.

**Proposition :** Soit $(C_i)_{i\in I}$ une famille de sous ensemble convexes de E. Alors $\bigcap_{i\in I} C_i$ est un **ensemble convexe**.

**Définition (Enveloppe convexe) :** Soit $A \subseteq E$, **l'enveloppe convexe** de A, notée $\text{conv}(A)$ est le plus petit (au sens de l'inclusion) convexe contenant A. Elle est donnée par :

$$\text{conv}(A) = \bigcap_{\substack{A\subset\mathcal{C} \\ \mathcal{C} \text{ convex}}} \mathcal{C}$$

**Proposition :** Pour $A \subseteq E$, conv(A) est l'ensemble des combinaisons convexes de A i.e :

$$\text{conv}(A) = \left\{ \sum_{i=1}^n \lambda_i x_i, \quad n \geq 1, x_1, \ldots, x_n \in A, \lambda_1, \ldots, \lambda_n \in \mathbb{R}_+, \sum_{i=1}^n \lambda_i = 1 \right\}$$

**Définition :** Soit $\mathcal{C}$ convexe de E, $x \in \mathcal{C}$ est dit **point extrémal** si $x \neq \lambda y + (1-\lambda)z$ pour $y, z \in \mathcal{C}$ pour tout $y, z \in \mathcal{C}$ avec $y \neq z$ et $0 < \lambda < 1$.

On note ext(A) l'ensemble des points extrémaux de $\mathcal{C}$.

**Remarque :** Autrement dit, un point x d'un convexe A est extrémal si, pour tout segment de A qui contient x, alors x est une extrémité de ce segment.

- L'ensemble des points extrémaux d'un disque fermé est le cercle qui ferme ce disque.
- L'ensemble des points extrémaux d'un carré est constitué par les 4 sommets du carré.

**Proposition :** Les assertions suivantes sont équivalentes :

i) $x \in \text{ext}(C)$

ii) $x \neq \frac{y+z}{2}$ pour $y, z \in \mathcal{C}$ tel que $y \neq z$

iii) $x = \frac{z+y}{2}$ avec $y, z \in \mathcal{C} \Rightarrow x = y = z$

**Proposition :** $x \in \text{ext}(\mathcal{C}) \iff \mathcal{C} \setminus \{x\}$ est convexe.

**Proposition :** Pour $A \in E$, on a $\text{ext}(\text{conv(A)}) \subset A$

**Théorème (Minkowski en dimension finie - Krein Milman en dimension infinie) :** Si $A \subset E$, A convexe de E, $A = \text{conv}(\text{ext}(A))$

## 1.2 Convex functions

**Definition 1.2.5 (Convex function) :** A function $f : E \to \mathbb{R} \cup \{+\infty\}$ is said to be **convex** if

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

for all $x, y \in E$ with $f(x) < +\infty$, $f(y) < +\infty$ and all $\lambda \in [0, 1]$.

**Remark :**

- We notice that for $f$ convex function, the sets $A_f(a) = \{x \in E : f(x) \leq a\}$ and $B_f(a) = \{x \in E : f(x) < a\}$ are convex sets.
- Unfortunately, the reverse implication is not true. Indeed, for $f : \mathbb{R} \to \mathbb{R}$ monotone function, the sets $A_f(a)$ and $B_f(a)$ are convex sets as they are intervals of $\mathbb{R}$.

**Definition 1.2.6 (Domain of a function) :** The **domain of a function** $f : E \to \mathbb{R} \cup \{+\infty\}$ is the set

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

**Remark 1.2.1 :** The function $f : E \to \mathbb{R} \cup \{+\infty\}$ is convex if and only if

- **1)** its domain $\text{dom}(f)$ is convex and
- **2)** its restriction $f|_{\text{dom}(f)}$ to its domain is convex.

**Definition 1.2.7 (Strictly convex function) :** A function $f : E \to \mathbb{R} \cup \{+\infty\}$ is said to be **strictly convex** if it is convex and

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

for all $x, y \in \text{dom}(f)$ with $x \neq y$ and all $\lambda \in (0, 1)$.

**Proposition 1.2.6 :** Let $C$ be a **convex subset** of $E$ and $f : E \to \mathbb{R} \cup \{+\infty\}$ be a **strictly convex function** such that $C \subset \mathrm{dom}(f)$. Let $x_0 \in C$ such that

$$f(x_0) = \max_{x \in C} f(x).$$

Then $x_0 \in \mathrm{ext}(C)$.

**Proposition 1.2.7 (Convex extension) :** If $f : C \to \mathbb{R}$ is a **convex function defined on the convex set** $C$, it can be **extended to the convex function** $\tilde{f} : E \to \mathbb{R} \cup \{+\infty\}$ by setting

$$\tilde{f}(x) = \begin{cases} f(x) & \text{if } x \in C, \\ +\infty & \text{else.} \end{cases}$$

**Definition 1.2.8 (Epigraph of a function) :** Let $f : E \to \mathbb{R} \cup \{+\infty\}$.

- Its **epigraph** is the subset of $E \times \mathbb{R}$ defined by

$$\mathrm{epi}(f) = \{(x, t) \in E \times \mathbb{R} : f(x) \leq t\}.$$

- The **strict epigraph** of $f$ is defined by

$$\mathrm{stepi}(f) = \{(x, t) \in E \times \mathbb{R} : f(x) < t\}.$$

**Proposition 1.2.8 :**

i) The function $f : E \to \mathbb{R} \cup \{+\infty\}$ is **convex** if and only if $\mathrm{epi}(f)$ is convex.

ii) If the function $f : E \to \mathbb{R} \cup \{+\infty\}$ is **strictly convex**, then $\mathrm{stepi}(f)$ is convex.

**Proposition 1.2.9 :**

(i) Let $(f_i)_{i \in I}$ be a family of **convex functions**. Its **upper envelope** $f$ defined by $f = \sup_{i \in I} f_i$ is a **convex function**.

(ii) If $g$ is the pointwise limit of a sequence $(g_n)_{n \geq 0}$ of convex functions, then $g$ is convex.

We may also need to consider convex functions valued in $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\} \cup \{-\infty\}$. We then take the convexity property of the epigraph as the definition.

**Definition 1.2.9 (Convex function valued in $\overline{\mathbb{R}}$).** The function $f : E \to \overline{\mathbb{R}}$ is said to be **convex** if its epigraph

$$\mathrm{epi}(f) = \{(x, t) \in E \times \mathbb{R} : f(x) \leq t\}.$$

is convex.

**Remark :** In the case where $f$ is valued in $\mathbb{R} \cup \{+\infty\}$, the definition of convexity coincides with the initial one. For the case where $f$ is valued in $\overline{\mathbb{R}}$, we have the following result.

**Proposition 1.2.10 :** A function $f : E \to \overline{\mathbb{R}}$ is convex if and only if

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

for all $x, y \in E$ with $f(x) < +\infty$, $f(y) < +\infty$ and all $\lambda \in [0,1]$, with the convention $-\infty + t = -\infty$, for $t < +\infty$, $a \cdot (-\infty) = -\infty$, for $a > 0$, and $0 \cdot (-\infty) = 0$.

**Definition 1.2.10 (Proper convex function) :** A **convex function** is said to be **proper** if it does not take the value $-\infty$, and is not identically equal to $+\infty$. If not, the function is said to be improper.

# 2 Chapter 2 : Structure of convex sets

## 2.1 Topological properties

**Proposition 2.1.11 :** Let $C$ be a convex set. The adherence $\mathrm{adh}(C)$ of $C$ is convex.

**Definition 2.1.11 (Closed convex hull) :** The closed convex hull of a subset $A$ of $E$ is the adherence of its convex hull $\mathrm{adh}(\mathrm{conv}(A))$.

**Proposition 2.1.12 :**

- (i) $\mathrm{adh}(\mathrm{conv}(A))$ is the smallest closed convex set containing $A$.
- (ii) If $A_1, \dots, A_p$ are convex compact subsets of $E$, then $\mathrm{conv}(A_1, \dots, A_p)$ is compact.
- (iii) If $A$ is a compact subset of $E$, $\mathrm{conv}(A)$ is compact.

**Proposition 2.1.13 :** Let $C$ be a convex subset of $E$ such that $\mathrm{int}(C) \neq \emptyset$. Then $\mathrm{int}(C)$ is **convex**. Moreover, we have $\mathrm{adh}(\mathrm{int}(C)) = \mathrm{adh}(C)$ and $\mathrm{int}(\mathrm{adh}(C)) = \mathrm{int}(C)$.

## 2.2 Separation of convex sets

**Theorem 2.2.2 (Projection on convex sets) :** Let $C$ be a **closed convex subset** of $E$ and $x_0 \notin C$. There exists a unique $y_0 \in C$, called the **projection of $x_0$ on $C$**, such that

$$|x_0 - y_0| = \inf_{y \in C} |x_0 - y|.$$

The projection of $x_0$ on $C$ is characterized by the following inequality

$$\langle x_0 - y_0, y - y_0 \rangle \leq 0 \tag{2.2.1}$$

for all $y \in C$.

**Definition 2.2.12 (Affine Hyperplan) :** An **affine hyperplane** $H$ is a subset of $E$ such that the set

$$H - x = \{y - x : y \in H\}$$

is a vector hyperplane of $E$ for some $x \in E$.

From Riesz representation Theorem, we deduce that a subset $H$ of $E$ is an **affine hyperplane** if and only if there exist a **linear form $f$** and a **constant $c$** such that

$$H = \{x \in E : f(x) = c\}.$$

**Theorem 2.2.3 (Separation of a point and a closed convex set) :** Let $C$ be a **closed convex subset** of $E$ and $x_0 \in E$ such that $x_0 \notin C$. There exists an **affine hyperplane** $H$ strictly separating $x_0$ from $C$. That is, there exist a linear form $f$ and a constant $c$ such that

$$f(x_0) > c \quad \text{and} \quad f(y) < c \quad \forall y \in C.$$

**Theorem 2.2.4 (Separation of a point from an open convex set) :** Let $C$ be an **open convex subset** of $E$ and $x_0 \notin C$. There exists an affine hyperplane $H$ such that $x_0 \in H$ and $H \cap C = \emptyset$.

**Corollary 2.2.1 (Hahn-Banach Theorem) :** Let $C_1$ and $C_2$ be two nonempty convex subsets of $E$ such that $C_1 \cap C_2 = \emptyset$. Suppose that $C_1$ is open. There exists an affine hyperplane $H$ strictly separating $C_1$ and $C_2$. More precisely, there exists a linear form $f$ on $E$ and a constant $c$ such that

$$f(x_1) < c \leq f(x_2)$$

for all $x_1 \in C_1$ and $x_2 \in C_2$.

**Corollary 2.2.2 :** Let $C_1$ and $C_2$ be two nonempty convex subsets of $E$ such that $C_1 \cap C_2 = \emptyset$. There exists an affine hyperplane $H$ separating $C_1$ and $C_2$. More precisely, there exists a nonzero linear form $f$ on $E$ and a constant $c$ such that

$$f(x_1) \leq c \leq f(x_2)$$

for all $x_1 \in C_1$ and $x_2 \in C_2$.

**Corollary 2.2.3 :** Let $f : E \to \mathbb{R} \cup \{+\infty\}$ be a convex function and $g : E \to \mathbb{R} \cup \{-\infty\}$ a concave function, i.e., $-g$ convex, such that $f \geq g$ on $E$. Suppose that there exists $x_0 \in E$ such that $f(x_0)$ and $g(x_0)$ are finite and $g$ is continuous at $x_0$. Then, there exists an affine form $h$ such that $f \geq h \geq g$ on $E$.

# 3 Chapter 3 : Regularity of convex functions

## 3.1 Convexity and continuity

**Définition :** Pour $f : E \to \mathbb{R} \cup \{+\infty\}$ et $x \in \text{int}(\text{dom}(f))$. On dit que f est **localement Lipschitz** en x s'il existe $r > 0$ et $L > 0$ tels que $B(x, r) \subset \text{dom}(f)$ et :

$$\forall y, z \in B(x, r), \quad |f(z) - f(y)| \leq L|z - y|$$

**Proposition 3.1.15 (Continuity of convex functions) :** Let $f : E \to \mathbb{R} \cup \{+\infty\}$ be a **convex function** and $x_0 \in \text{int}(\text{dom}(f))$. If $f$ is **upper bounded in the neighborhood of $x_0$**, Then f is **continuous**, and also **locally Lipschitz** in the neighborhood of $x_0$.

**Definition 3.1.13 (Locally Lipschitz functions) :** A function $f$ is said to be **locally Lipschitz** on $\text{int}(\text{dom}(f))$ if it is Lipschitz in the neighborhood of all points of $\text{int}(\text{dom}(f))$.

**Corollary 3.1.5 :** Let $f : E \to \mathbb{R} \cup \{+\infty\}$ be a **convex function**. The following assertions are equivalent.

(i) There exists $x_0 \in \text{dom}(f)$ such that $f$ is **upper bounded** at the neighborhood of $x_0$.

(ii) There exists $x_0 \in \text{dom}(f)$ such that $f$ is **continuous** at $x_0$.

(iii) $\text{int}(\text{dom}(f)) \neq \emptyset$ and $f$ is **continuous** on $\text{dom}(f)$.

(iv) $\text{int}(\text{dom}(f)) \neq \emptyset$ and $f$ is **locally Lipschitz** on $\text{dom}(f)$.

**Corollary 3.1.6 :** Let $f : E \to \mathbb{R} \cup \{+\infty\}$ be a **convex function**. Then, $f$ is **continuous** and even **locally Lipschitz** on $\text{int}(\text{dom}(f))$.

## 3.2 Convexity and differentiability

**Definition 3.2.14 (Directional derivative) :** Let $f : E \to \mathbb{R} \cup \{+\infty\}$, $x_0 \in \text{dom}(f)$ and $h \in E \setminus \{0\}$. The function $f$ is said to be **right-differentiable at $x_0$ in the direction $h$** if

$$\lim_{t \to 0^+} \frac{f(x_0 + th) - f(x_0)}{t}$$

exists in $\mathbb{R} \cup \{+\infty\}$. When it exists, this limit is denoted by $f'_d(x_0, h)$ and is called the **directional derivative of $f$ in the direction $h$ at $x_0$.**

**Proposition 3.2.16 :** A **convex function** $f : E \to \mathbb{R} \cup \{+\infty\}$ is **right-differentiable** at any $x_0 \in \text{dom}(f)$ in any direction $h$. Moreover, we have the inequality

$$f(x) - f(x_0) \geq f'_d(x_0, x - x_0) \tag{3.2.2}$$

for all $x \in E$.

**Lemma 3.2.2 :** Let $f : E \to \mathbb{R} \cup \{+\infty\}$ be a **convex function**, $x_0 \in \mathrm{dom}(f)$, and $h \in E \setminus \{0\}$. The function $\Delta_{x_0,h} : (0, +\infty) \to \mathbb{R} \cup \{+\infty\}$ defined by

$$\Delta_{x_0,h}(t) = \frac{f(x_0 + th) - f(x_0)}{t}, \quad t \in (0, +\infty)$$

is **nondecreasing.**

**Corollary 3.2.8 (Minimum of a convex function) :** A **convex proper function** $f$ reaches its minimum at $x_0 \iff f'_d(x_0, h) \geq 0 \quad \forall h \in E \setminus \{0\}$

**Proposition 3.2.17 :** Let $f : E \to \mathbb{R} \cup \{+\infty\}$ be a function differentiable on $\mathrm{dom}(f)$ that is supposed to be convex. Then

$$f \ \textbf{convex} \iff f(y) - f(x) \geq \langle \nabla f(x), y - x \rangle \quad \forall x \in \mathrm{dom}(f), \forall y \in E$$

**Proposition 3.2.18 :** Let $f : E \to \mathbb{R} \cup \{+\infty\}$ be a function differentiable on $\mathrm{dom}(f)$ that is supposed to be convex. Then

$$f \ \textbf{convex} \iff \langle \nabla f(y) - \nabla f(x), y - x \rangle \geq 0 \quad \forall x, y \in \mathrm{dom}(f)$$

**Proposition 3.2.19 (Coercivity of the gradient) :** Let $f$ be a **convex and differentiable** function from $E$ to $\mathbb{R}$. We suppose that **$\nabla f$ is $L$-Lipschitz continuous** on $E$. Then

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \frac{1}{L} \|\nabla f(x) - \nabla f(y)\|^2 \quad \forall x, y \in E$$

**Proposition 3.2.20 :** Let $f : E \to \mathbb{R} \cup \{+\infty\}$ be a function differentiable on $\mathrm{dom}(f)$. Suppose that the function $t \in \mathbb{R}_+ \mapsto \nabla f(x + th)$ is differentiable and

$$\left\langle \frac{\nabla f(x + th) - \nabla f(x)}{t}, h \right\rangle \xrightarrow[t \to 0^+]{} Q(x, h)$$

with $Q(x, h) \geq 0$ for all $h \in E$ and all $x \in \mathrm{dom}(f)$. **Then $f$ is convex.**

# 4 Chapter 4 : Optimization of differentiable functions

## 4.1 Optimality coniditions

**Framework :** In this chapter we are interested in solving the problem

$$\min_{x \in K} f(x) \tag{4.1.1}$$

where $f : E \to \mathbb{R}$ is a **differentiable convex function** such that

$$\lim_{|x| \to +\infty} f(x) = +\infty \tag{4.1.2}$$

and $K$ is a **closed convex subset** of $E$.

**Theorem 4.1.5 :** Under the previous assumptions, the optimisation problem **(4.1.1)** has **at least one solution**. If moreover $f$ is **strictly convex** then **(4.1.1)** has a **unique solution.**

**Theorem 4.1.6 :** Suppose that $f$ is **convex and differentiable** and $K$ is **closed and convex**. Then, $x^*$ is a solution of **(4.1.1)** if and only if

$$\langle \nabla f(x^*), v - x^* \rangle \geq 0 \quad \forall v \in K$$

**Lagrange multipliers :** We consider the case where the set $K$ is defined by $m \geq 1$ functions $F_1, \ldots, F_m$ from $E$ to $\mathbb{R}$ as follows

$$K = \{x \in E : F_i(x) = 0 \text{ for } i = 1, \ldots, m\}.$$

**Theorem 4.1.7 :** Suppose that $f$ is **differentiable** and that the functions $F_1, \ldots, F_m$ are $C^1$. Let $x^* \in E$ be a solution to **(4.1.1)** such that $(\nabla F_i(x^*))_{i=1}^m$ is a family of **linearly independent** vectors of $E$. Then there exist $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$ such that

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla F_i(x^*) = 0.$$

The coefficients $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$ are called **Lagrange multipliers.**

We next turn to the case where the set $K$ is defined by inequalities instead of equalities. More precisely, we suppose that $K$ is given by

$$K = \{x \in E : F_i(x) \leq 0 \text{ for } i = 1, \ldots, m\}. \tag{4.1.3}$$

Before stating the main result, we need to introduce the following definition.

**Definition 4.1.15 (Qualified constraints) :** The constraints defining $K$ in **(4.1.3)** are said to be **qualified** at $x^* \in K$ if there exists $w \in E$ such that

$$\langle \nabla F_i(x^*), w \rangle < 0,$$

or

$$\langle \nabla F_i(x^*), w \rangle = 0 \quad \text{and } F_i \text{ is affine}$$

for any $i = 1, \ldots, m$ such that $F_i(x^*) = 0$.

**Theorem 4.1.8 :** Suppose that $K$ is given by **(4.1.3)**. Let $x^* \in K$ such that the functions $f$ and $(F_i)_{i=1}^m$ are differentiable at $x^*$ and the **constraints are qualified** at $x^*$. Then, if $x^*$ is a **local minimizer** of $f$ over $K$, there exist $\lambda_1, \ldots, \lambda_m \in \mathbb{R}_+$ called Lagrange multipliers, such that

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla F_i(x^*) = 0,$$

and $\lambda_i = 0$ if $F_i(u) < 0$ for $i = 1, \ldots, m$.

**Theorem 4.1.9** Suppose that the functions $f$ and $(F_i)_{i=1}^m$ are **convex, continuous on $E$ and differentiable on $K$** defined by **(4.1.3)**. Let $x^* \in K$ be such that the constraints are **qualified** at $x^*$. Then, $x^*$ is a global minimum of $f$ over $K$ if and only if there exist $\lambda_1, \ldots, \lambda_m \in \mathbb{R}_+$ called Lagrange multipliers, such that

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla F_i(x^*) = 0,$$

and $\lambda_i = 0$ if $F_i(x^*) < 0$ for $i = 1, \ldots, m$.

## 4.2 Gradient descent

**Framework :** For a convex and differentiable function $f$, there exist several gradient descent methods allowing to construct minimizing sequences. In this section, we shall present a classical method using the gradient $\nabla f$ of $f$ to approximate solutions of the minimization problem

$$\min_{x \in E} f(x). \tag{4.2.4}$$

- We consider $\gamma \in (0, +\infty)$ called step size, $x_0 \in E$ and we define the sequence $(x_n)_{n \geq 0}$ by

$$x_{n+1} = x_n - \gamma \nabla f(x_n), \quad n \geq 0. \tag{4.2.5}$$

- Denote by $T$ the map from $E$ to $E$ defined by

$$T(x) = x - \gamma \nabla f(x) = (I - \gamma \nabla f)(x) \quad \text{for} \quad x \in E$$

- The sequence $(x_n)_{n \in \mathbb{N}}$ is given by

$$x_{n+1} = T(x_n), \quad n \geq 0.$$

Without any assumption on $\nabla f$, this algorithm may diverge, especially if $\gamma$ is taken too large. However, if we suppose that $\nabla f$ is Lipschitz continuous, we are able to make this algorithm converge to a minimizer by choosing correctly $\gamma$ as the following result shows.

**Theorem 4.2.10 :** Suppose that $f$ is **convex differentiable** with **gradient $L$-Lipschitz**. Suppose also that $f$ has a global minimum at $x^* \in E$. Then for any $x_0 \in E$, $t$, and any $\gamma \in \left(0, \frac{2}{L}\right)$ the sequence $(x_n)_{n \geq 0}$ defined by (4.2.5) converges to $x^*$.

**Lemma 4.2.3 :** Let $T$ be a map from $E$ to $E$, 1-Lipschitz continuous and admitting at least one fixed point. Let $x_0 \in E$ and $(x_n)_{n \geq 0}$ the sequence defined by (4.2.5). If $\lim_{n \to +\infty} |x_{n+1} - x_n| = 0$, $(x_n)_{n \geq 0}$ converges to a fixed point of $T$.

**Lemma 4.2.4 :** Let $f$ be a **convex differentiable function** whose gradient $\nabla f$ is $L$-Lipschitz continuous. For $\gamma \in \left(0, \frac{2}{L}\right]$, the map $T = \mathrm{Id} - \gamma \nabla f$ is 1-Lipschitz continuous.

# 5 Chapter 5 : Duality of convex functions

## 5.1 Upper envelopes of affine functions

**Definition 5.1.16 :** An affine function $h : E \to \mathbb{R}$ is a function such that

$$h(\lambda x + (1-\lambda)y) = \lambda h(x) + (1-\lambda)h(y)$$

for all $\lambda \in \mathbb{R}$ and $x, y \in E$. Equivalently, $h$ is affine if and only if there exists $g : E \to \mathbb{R}$ linear form such that $h = g + h(0)$.

**Theorem 5.1.13 :** A function $f : E \to \mathbb{R} \cup \{+\infty\}$ is convex and lsc if and only if it is the upper envelope of affine functions.

## 5.2 Fenchel-Moreau conjugate

**Definition 5.2.17 (Conjugate function) :** Let $f : E \to \mathbb{R}$. Its conjugate function (in the sense of Fenchel-Moreau) is the function $f^* : E \to \mathbb{R}$ defined by

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

**Definition 5.2.18 (Convex lsc regularization) :** The convex lsc regularization of a function $f : E \to \mathbb{R}$ is the greatest lsc convex function $\hat{f} : E \to \mathbb{R}$ dominated by $f$.

**Proposition 5.2.21 :** The conjugate function $f^*$ of a function $f$ is either a convex lsc function from $E$ to $\mathbb{R} \cup \{+\infty\}$, or identically equal to $-\infty$.

## 5.3 Bi-conjugate

**Definition 5.3.19 :** The biconjugate function of $f : E \to \mathbb{R}$ is defined as the conjugate of the conjugate $f^*$ of $f$ and is denoted by $(f^*)^*$.

**Theorem 5.3.14 (Bi-conjugate function theorem) :** The bi-conjugate $(f^*)^*$ is equal to the convex lsc regularization of $f$ if $f$ is lower bounded by an affine function. If not, $(f^*)^*$ is identically equal to $-\infty$.

**Corollary 5.3.9 :** Let $f : E \to \mathbb{R}$ be a convex proper lsc function. We then have $(f^*)^* = f$. The equality also holds if $f$ is identically equal to $-\infty$.

# 6 Chapter 6 : Diffusion processes

## 6.1 SDEs with random coefficients

**Framework :** We recall in this section the basic results for stochastic differential equations

$$dX_t = b_t(X_t)dt + \sigma_t(X_t)dW_t, \quad t \in [0, T]. \tag{6.1.1}$$

Here, $b$ and $\sigma$ are $\mathcal{F} \otimes \mathcal{B}(\mathbb{R}^n)$-progressively measurable functions from $[0, T] \times \Omega \times \mathbb{R}^n$ to $\mathbb{R}^n$ and $\mathbb{R}^{n \times d}$ respectively. In particular, for every fixed $x \in \mathbb{R}^n$, the processes $\{b_t(x), \sigma_t(x), t \in [0, T]\}$ are $\mathcal{F}$-progressively measurable.

**Definition 6.1.1 :** A strong solution of **(6.1.1)** is an $\mathcal{F}$-progressively measurable process $X$ such that

$$\int_0^T \left(|b_t(X_t)| + |\sigma_t(X_t)|^2\right) dt < \infty \quad \text{a.s.}$$

and

$$X_t = X_0 + \int_0^t b_s(X_s)ds + \int_0^t \sigma_s(X_s)dW_s, \quad t \in [0, T].$$

**Definition :**

- $\mathbb{H}^2 = \{\text{stochastic process } U \text{ such that } \mathbb{E}\left[\int_0^T |U_s|^2\, ds\right] < \infty\}$
- Let $c$ be a positive constant. We define the norm $\|\cdot\|_{\mathbb{H}^2_c}$ by

$$\|\phi\|_{\mathbb{H}^2_c} = \mathbb{E}\left[\left(\int_0^T e^{-ct}|\phi_s|^2 ds\right)^{1/2}\right].$$

**Theorem 6.1.1 :** Let $X_0$ be a **square integrable** random variable **independent** of $W$. Assume that the processes $b(\cdot)$ and $\sigma(\cdot)$ are in $\mathbb{H}^2$, and there exists a constant $L > 0$ such that

$$|b_t(x) - b_t(y)| + |\sigma_t(x) - \sigma_t(y)| \leq L|x - y| \quad \forall t \in [0, T] \text{ and } x, y \in \mathbb{R}^n. \tag{6.1.2}$$

Then there exists a unique **strong solution** of (6.1.1) in $\mathbb{H}^2$. Moreover, we have

$$\mathbb{E}\left[\sup_{t \in [0,T]} |X_t|^2\right] \leq C\left(1 + \mathbb{E}[|X_0|^2]\right) \tag{6.1.3}$$

for some constant $C = C(T, L)$ depending only on $T$ and $L$.

**Remark :** The previous result can be easily extended to any initial time $t \in [0, T]$ instead of 0. In the sequel, we shall denote by $X^{t,x} = \{X_s^{t,x},\, t \leq s \leq T\}$ the process solution to

$$X_s = x + \int_t^s b_u(X_u)du + \int_t^s \sigma_u(X_u)dW_u, \quad u \in [t, T],$$

for $t \in [0, T]$ and $x \in \mathbb{R}^n$.

**Theorem 6.1.2 :** Suppose that assumptions of Theorem 6.1.1 hold.

(i) There exists a constant $C > 0$ such that

$$\mathbb{E}\left[\sup_{t\leq s\leq t'}|X_s^{t,x} - X_s^{t,x'}|^2\right] \leq Ce^{Ct'}|x - x'|^2$$

for $t, t' \in [0, T]$ such that $t \leq t'$ and $x, x' \in \mathbb{R}^n$.

(ii) Moreover, if we have

$$B := \sup_{0\leq s<s'\leq T}(s' - s)^{-1}\mathbb{E}\int_s^{s'}(|b_r(0)|^2 + |\sigma_r(0)|^2)dr < +\infty$$

then, there exists a constant $C > 0$ such that

$$\mathbb{E}\left[\sup_{t'\leq s\leq T}|X_s^{t,x} - X_s^{t',x}|^2\right] \leq Ce^{CT}(B + |x|^2)(t' - t)$$

for $t, t' \in [0, T]$ such that $t < t'$ and $x \in \mathbb{R}^n$.

## 6.2 Markov SDEs

**Proposition 6.2.1 (Markov property) :** We have

$$\mathbb{E}\left[\Phi\left(X_r^{t,x}, u \leq r \leq s\right)\middle|\mathcal{F}_u\right] = \mathbb{E}\left[\Phi\left(X_r^{t,x}, u \leq r \leq s\right)\middle|X_u^{t,x}\right]$$

for any $u \in [t, s]$ and any measurable bounded (or nonnegative) function

$$\Phi : C([u, s]) \to \mathbb{R}.$$

## 6.3 Connection with PDEs

Let $X^{t,x}$ be the unique solution to the SDE

$$X_s = x + \int_t^s b_u(X_u)du + \int_t^s \sigma_u(X_u)dW_u, \quad s \in [t, T], \tag{1}$$

for $t \in [0, T]$ and $x \in \mathbb{R}^n$.

**Definition :** We next define the operator $\mathcal{L}$ by

$$\mathcal{L}\varphi(t, x) = \lim_{h\to 0}\frac{\mathbb{E}[\varphi(X_{t+h}^{t,x})] - \varphi(x)}{h}$$

for any function $\varphi : \mathbb{R}^n \to \mathbb{R}$ such that $\mathcal{L}\varphi$ is well defined. $\mathcal{L}$ is called the generator of the diffusion.

**Remark :** From Itô's formula, $\mathcal{L}\varphi$ is well defined for any $\varphi$ bounded $C^2$ with bounded derivatives and we have

$$\mathcal{L}\varphi(t, x) = b_t(x) \cdot \nabla\varphi(x) + \frac{1}{2}\operatorname{Tr}(\sigma_t(x)\sigma_t(x)^\top\nabla^2\varphi(x))$$

for $t \in [0, T]$ and $x \in \mathbb{R}^n$. The generator provides a connection between diffusion processes and linear partial differential equations.

**Proposition 6.3.2 :** Suppose that the function $v$ defined by

$$v(t,x) = \mathbb{E}[g(X_T^{t,x})], \quad (t,x) \in [0,T] \times \mathbb{R}^n,$$

is in $C^{1,2}([0,T] \times \mathbb{R}^n)$. Then, $v$ solves the partial differential equation:

$$\partial_t v + \mathcal{L}v = 0 \quad \text{on } [0,T) \times \mathbb{R}^n$$

with terminal condition

$$v(T,\cdot) = g \text{ on } \mathbb{R}^n.$$

**Feynman-Kac representation of Cauchy problem :** We consider the following linear partial differential equation called Cauchy problem

$$\begin{cases} \partial_t v + \mathcal{L}v - kv + f = 0, & \text{on } [0,T] \times \mathbb{R}^n, \\ v(T,\cdot) = g, & \text{on } \mathbb{R}^n. \end{cases} \tag{6.3.6}$$

with $k$ and $f$ two functions from $[0,T] \times \mathbb{R}^n$ to $\mathbb{R}$. The next result provides a representation of this purely deterministic problem by means of stochastic differential equations.

**Theorem 6.3.3 :** Assume that the coefficients $b$ and $\sigma$ satisfy the assumptions of Theorem 6.1.1. Assume further that the function $k$ is uniformly lower bounded, and $f$ has quadratic growth in $x$ uniformly in $t$. Let $v \in C^{1,2}([0,T] \times \mathbb{R}^n) \cap C^0([0,T] \times \mathbb{R}^n)$ be a solution of (6.3.6) with quadratic growth in $x$ uniformly in $t$. Then

$$v(t,x) = \mathbb{E}\left[\int_t^T \beta_s^{t,x} f(s, X_s^{t,x})\, ds + \beta_T^{t,x} g(X_T^{t,x})\right]$$

where $X^{t,x}$ is the unique solution to (6.3.5) and

$$\beta_s^{t,x} = e^{-\int_t^s k(u, X_u^{t,x})\, du}, \quad s \in [t,T],$$

for $(t,x) \in [0,T] \times \mathbb{R}^d$.

# 7 Chapter 7 : Optimal control of diffusion processes and dynamic programming

## 7.1 Stochastic control problem

**Definition (Set of controls) :** We fix a subset $A$ of some $\mathbb{R}^p$ for $p \geq 1$. We suppose that $A$ is bounded and we denote by $\mathcal{A}$ the set of $\mathcal{F}$-progressive processes $(\alpha_t)_{t\in[0,T]}$ valued in $A$.

**Controlled diffusion process :** We fix two functions $b, \sigma : \mathbb{R}^n \times A \to \mathbb{R}^n, \mathbb{R}^{d\times n}$. We suppose that $b$ and $\sigma$ are continuous and there exists a constant $L$ such that

$$|b(x,a) - b(x',a')| + |\sigma(x,a) - \sigma(x',a')| \leq L(|x - x'| + |a - a'|),$$

for all $(x,a),(x',a') \in \mathbb{R}^n \times A$. From Theorem 6.1.1, we have existence and uniqueness of the controlled process $X_s^{t,x,\alpha}$ defined by

$$X_s^{t,x,\alpha} = x + \int_t^s b(X_u^{t,x,\alpha}, \alpha_u)du + \int_t^s \sigma(X_u^{t,x,\alpha}, \alpha_u)dW_u, \quad (7.1.1)$$

for any $(t,x) \in [0,T] \times \mathbb{R}^n$ and any control process $\alpha \in \mathcal{A}$.

**Proposition 7.1.3 :** For $p \geq 1$, there exists a constant $C_p$ such that:

$$\mathbb{E}\left[\sup_{s\in[t,T]} |X_s^{t,x,\alpha}|^p\right] \leq C_p(1 + |x|^p),$$

$$\mathbb{E}\left[\sup_{s\in[t,T]} |X_s^{t,x,\alpha} - X_s^{t,x',\alpha}|^p\right] \leq C_p|x - x'|^p,$$

$$\mathbb{E}\left[\sup_{t\leq s\leq T} |X_s^{t,x,\alpha} - X_{s\vee(t+h),x,\alpha}^{t+h}|^p\right] \leq C_p h^{p/2}(1 + |x|^p),$$

$$\mathbb{E}\left[\sup_{s\in[t,T]} |X_s^{t,x,\alpha} - X_s^{t,x,\alpha'}|^p\right] \leq C_p \mathbb{E}\left[\int_t^T |\alpha_s - \alpha'_s|^p ds\right],$$

for all $t \in [0,T]$, $h \in [0, T-t]$, $x, x' \in \mathbb{R}^n$, and $\alpha, \alpha' \in \mathcal{A}$.

**Reward Functions and Gain :**

- We fix two **reward functions** $f : \mathbb{R}^n \times A \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$.
- We assume that $f$ and $g$ are **locally Lipschitz continuous**, that is, for any $N > 0$, there exists a constant $L_N$ such that

$$|f(x,a) - f(x',a')| + |g(x) - g(x')| \leq L_N(|x - x'| + |a - a'|),$$

  for any $(x,a),(x',a') \in \mathbb{R}^n \times A$ such that $|x| \leq N$ and $|x'| \leq N$.

- We also assume that $f$ and $g$ have **polynomial growth**, that is, there exist a constant $C$ and an integer $p$ such that

$$|f(x,a)| + |g(x)| \leq C(1 + |x|^p),$$

  for all $(x,a) \in \mathbb{R}^n \times A$.

- We next define the **gain functional** $J : [0,T] \times \mathbb{R}^n \times \mathcal{A} \to \mathbb{R}$ by

$$J(t,x,\alpha) = \mathbb{E}\left[g(X_T^{t,x,\alpha}) + \int_t^T f(X_s^{t,x,\alpha}, \alpha_s)ds\right],$$

  for $(t,x,\alpha) \in [0,T] \times \mathbb{R}^n \times \mathcal{A}$.

- We observe that $J(t,x,\alpha)$ is well-defined under the polynomial growth assumption from Proposition 7.1.3.
- We now define the **value function** $v$ of the considered stochastic control problem by

$$v(t,x) = \sup_{\alpha \in \mathcal{A}_t} J(t,x,\alpha). \tag{7.1.2}$$

  for $(t,x) \in [0,T] \times \mathbb{R}^n$, where $\mathcal{A}_t = \{\alpha \in \mathcal{A} : \alpha \text{ independent of } \mathcal{F}_t\}, \quad t \in [0,T]$.

## 7.2 Dynamic programming principle

**Proposition 7.2.4 :** For a compact set $\Theta \subset [0,T] \times \mathbb{R}^n$, there exists a real map $\lambda_\Theta : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\lambda_\Theta(r) \to 0$ as $r \to 0$ and

$$|J(t,x,\alpha) - J(s,y,\alpha)| \leq \lambda_\Theta(|t-s| + |x-y|) \tag{7.2.3}$$

for all $(t,x),(s,y) \in \Theta$ and $\alpha \in \mathcal{A}$. The value function $v$ is locally uniformly continuous and has polynomial growth.

**Theorem 7.2.4 (Dynamic programming principle) :** The value function $v$ satisfies

$$v(t,x) = \sup_{\alpha \in \mathcal{A}_t} \mathbb{E}\left[v\big(\theta^\alpha, X_{\theta^\alpha}^{t,x,\alpha}\big) + \int_t^{\theta^\alpha} f\big(X_s^{t,x,\alpha}, \alpha_s\big)ds\right]$$

for any $(t,x) \in [0,T] \times \mathbb{R}^n$ and any family of stopping times $\{\theta^\alpha, \alpha \in \mathcal{A}\}$ valued in $[t,T]$.

## 7.3 Dynamic programming equation

We prove in this section that, if $v$ is smooth enough, it solves a PDE called the Hamilton-Jacobi-Bellman (HJB) equation.

- More precisely, define the second-order local operator $\mathcal{L}^a$, for $a \in A$, by

$$\mathcal{L}^a \varphi(t,x) = \partial_t \varphi(t,x) + b(x,a) \cdot \nabla \varphi(t,x) + \frac{1}{2} \mathrm{Tr}\big(\sigma(x,a)\sigma^\top(x,a)\nabla^2 \varphi(t,x)\big)$$

  for $(t,x) \in [0,T] \times \mathbb{R}^n$ and $\varphi \in C^{1,2}([0,T] \times \mathbb{R}^n)$.

- Then, the HJB equation takes the following form:

$$\sup_{a \in A} \Big\{ \mathcal{L}^a v(t,x) + f(x,a) \Big\} = 0, \quad (t,x) \in [0,T] \times \mathbb{R}^n \tag{7.3.6}$$

  together with the terminal condition:

$$v(T,x) = g(x), \quad x \in \mathbb{R}^n. \tag{7.3.7}$$

- To simplify notations, we denote by $\mathcal{H}$ the operator defined by

$$\mathcal{H}\varphi(t,x) = \sup_{a \in A} \Big\{ \mathcal{L}^a \varphi(t,x) + f(x,a) \Big\}$$

  for $(t,x) \in [0,T] \times \mathbb{R}^n$ and $\varphi \in C^{1,2}([0,T] \times \mathbb{R}^n)$.

**Theorem 7.3.5 :** Suppose that $v \in C^0([0,T] \times \mathbb{R}^n) \cap C^{1,2}([0,T] \times \mathbb{R}^n)$. Then, $v$ is a solution to (7.3.6)-(7.3.7).