Stochastic Optimization
Recalls on convex analysis

V. Leclère

November 25 2020

École des Ponts ParisTech

UNIVERSITÉ PARIS-EST

Presentation Outline

Objective of the course

- Uncertainty is present in most optimization problem, sometimes taken into account.
- Two major way of taking uncertainty into account :
  - Robust approach: assuming that uncertainty belongs in some set C, and will be chosen adversarily.
  - Stochastic approach: assuming that uncertainty is a random variable with known law.
- We will take the stochastic approach, considering the multi-stage approach : a first decision is taken, then part of the uncertainty is revealed, before taking a second decision and so on.

Syllabus

- 1st course: Convex toolbox
- 2nd course: Probability toolbox
- 3rd course: two-stage stochastic programm
- 4th course: Bellman operators and Dynamic Programming
- 5th course: Decomposition methods for two stage SP
- 6th course: Stochastic Dual Dynamic Programming

Validation

- The stochastic optimization course is in two part
- Evaluation have 3 components :
  - A paper presentation with P.Carpentier
  - A written exam on my part of the course with theoretical and modelling questions
  - Multiple practical work to send to vincent.leclere@enpc.fr
- Practical work will be done in Julia (www.julialang.com)using jupyter notebook
- Instructions for installing julia / jupyter and using the library can be found at https://github.com/leclere/TP-Saclay
- Practical work will be posted there

Presentation Outline

1. Overview of the course

2. Convex sets and functions
   - Fundamental definitions and results
   - Convex function and minimization
   - Subdifferential and Fenchel-Transform

3. Duality
   - Recall on Lagrangian duality
   - Marginal interpretation of multiplier
   - Fenchel duality

Convex sets

- C is a convex set iff

$$\forall x_1, x_2 \in C, \quad [x_1, x_2] \subset C.$$

- If for all $i \in I$, C_i is convex, then so is $\cap_{i \in I} C_i$
- $C_1 + C_2$, and $C_1 \times C_2$ are convex
- For any set X the convex hull of X is the smallest convex set containing X,

$$\text{conv}(X) := \Big\{ tx_1 + (1-t)x_2 \quad | \quad x_1, x_2 \in C, \quad t \in [0,1] \Big\}.$$

- The closed convex hull of X is the intersection of all half-spaces containing X.

Separation

Let X be a Banach space, and X^* its topological dual (i.e. the set of all continuous linear form on X).

Theorem (Simple separation)

Let A and B be convex non-empty, disjunct subsets of X. Assume that, $int(A) \neq \emptyset$, then there exists a separating hyperplane $(x^*, \alpha) \in X^* \times \mathbb{R}$ such that

$$\langle x^*, a \rangle \leq \alpha \leq \langle x^*, b \rangle \qquad \forall a, b \in A \times B.$$

Theorem (Strong separation)

Let A and B be convex non-empty, disjunct subsets of X. Assume that, A is closed, and B is compact (e.g. a point), then there exists a strict separating hyperplane $(x^*, \alpha) \in X^* \times \mathbb{R}$ such that, there exists $\varepsilon > 0$,

$$\langle x^*, a \rangle + \varepsilon \leq \alpha \leq \langle x^*, b \rangle - \varepsilon \qquad \forall a, b \in A \times B.$$

Convex functions : basic properties

- A function $f : X \to \bar{\mathbb{R}}$ is convex if its epigraph is convex.
- $f : X \to \mathbb{R} \cup \{+\infty\}$ is convex iff

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

- If f, g convex, $\lambda > 0$, then $\lambda f + g$ is convex.
- If f convex non-decreasing, g convex, then $f \circ g$ convex.
- If f convex and a affine, then $f \circ a$ is convex.
- If $(f_i)_{i \in I}$ is a family of convex functions, then $\sup_{i \in I} f_i$ is convex.

Convex functions : further definitions and properties

- The domain of a convex function is $\mathrm{dom}(f) = \{x \in X \mid f(x) < +\infty\}$.
- The level set of a convex function is $\mathrm{lev}_\alpha(f) = \{x \in X \mid f(x) \leq \alpha\}$
- A function is lower semi continuous (lsc) iff for all $\alpha \in \mathbb{R}$, lev_α is closed.
- The domain and the level sets of a convex function are convex.
- A convex function is proper if it never takes $-\infty$, and $\mathrm{dom}(f) \neq \emptyset$.
- A function is coercive if $\lim_{\|x\| \to \infty} f(x) = +\infty$.

Convex functions : polyhedral functions

- A polyhedra is a finite intersection of half-spaces, thus convex.
- A polyhedral function is a function whose epigraph is a polyhedra.
- Finite intersection, cartesian product and sum of polyhedra is polyhedra.
- In particular a polyhedral function is convex lsc, with polyhedral domain and level sets.
- If $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ is polyhedral, then it can be written as

$$\begin{aligned} f(x) = \min_{\theta} \quad & \theta \\ s.t. \quad & \alpha_\kappa^\top x + \beta_\kappa \leq \theta & \forall \kappa \leq k \\ & \gamma_\kappa{}^\top x + \delta_\kappa \leq 0 & \forall \kappa \leq k' \end{aligned}$$

Convex functions : polyhedral approximations

- f is convex iff it is above all its tangeant.
- Let $\{x_\kappa, g_\kappa\}_{\kappa \leq k}$ be a collection of (sub-)gradient, that is such that $f \geq \langle g_\kappa, \cdot - x_\kappa \rangle + f(x_\kappa)$, then

$$\underline{f}_k : x \mapsto \max_{\kappa \leq k} \langle g_\kappa, x - x_\kappa \rangle + f(x_\kappa)$$

  is a polyhedral outer-approximation of f.
- Let $\{x_\kappa\}_{\kappa \leq k}$ be a collection of point in $dom(f)$. Then,

$$\bar{f}_k : x \mapsto \min_{\sigma \in \Delta_k} \left\{ \sum_{\kappa=1}^{k} \sigma_\kappa f(x_\kappa) \;\middle|\; \sum_{\kappa=1}^{k} \sigma_\kappa x_\kappa = x \right\}$$

  is a polyhedral inner-approximation of f.

Convex functions : polyhedral approximations

- f is convex iff it is above all its tangeant.
- Let $\{x_\kappa, g_\kappa\}_{\kappa \leq k}$ be a collection of (sub-)gradient, that is such that $f \geq \langle g_\kappa, \cdot - x_\kappa \rangle + f(x_\kappa)$, then

$$\underline{f}_k : x \mapsto \max_{\kappa \leq k} \langle g_\kappa, x - x_\kappa \rangle + f(x_\kappa)$$

is a polyhedral outer-approximation of f.

- Let $\{x_\kappa\}_{\kappa \leq k}$ be a collection of point in $dom(f)$. Then,

$$\bar{f}_k : x \mapsto \min_{\sigma \in \Delta_k} \left\{ \sum_{\kappa=1}^{k} \sigma_\kappa f(x_\kappa) \;\middle|\; \sum_{\kappa=1}^{k} \sigma_\kappa x_\kappa = x \right\}$$

is a polyhedral inner-approximation of f.

Convex functions : polyhedral approximations

- f is convex iff it is above all its tangeant.
- Let $\{x_\kappa, g_\kappa\}_{\kappa \leq k}$ be a collection of (sub-)gradient, that is such that $f \geq \langle g_\kappa, \cdot - x_\kappa \rangle + f(x_\kappa)$, then

$$\underline{f}_k : x \mapsto \max_{\kappa \leq k} \langle g_\kappa, x - x_\kappa \rangle + f(x_\kappa)$$

  is a polyhedral outer-approximation of f.
- Let $\{x_\kappa\}_{\kappa \leq k}$ be a collection of point in $dom(f)$. Then,

$$\bar{f}_k : x \mapsto \min_{\sigma \in \Delta_k} \Big\{ \sum_{\kappa=1}^{k} \sigma_\kappa f(x_\kappa) \;\Big|\; \sum_{\kappa=1}^{k} \sigma_\kappa x_\kappa = x \Big\}$$

  is a polyhedral inner-approximation of f.

Convex functions : strict and strong convexity

- $f : X \to \mathbb{R} \cup \{+\infty\}$ is strictly convex iff
  $$\forall t \in]0,1[, \quad \forall x, y \in X, \qquad f(tx + (1-t)y) < tf(x) + (1-t)f(y).$$
- $f : X \to \mathbb{R} \cup \{+\infty\}$ is α-convex iff $\forall x, y \in X$
  $$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\alpha}{2}\|y - x\|^2.$$
- If $f \in C^1(\mathbb{R}^n)$
  - $\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq 0$ iff f convex
  - if strict inequality holds, then f strictly convex
- If $f \in C^2(\mathbb{R}^n)$,
  - $\nabla^2 f \succeq 0$ iff f convex
  - if $\nabla^2 f \succ 0$ then f strictly convex
  - if $\nabla^2 f \succeq \alpha I$ then f is α-convex

Convex functions : strict and strong convexity

- $f : X \to \mathbb{R} \cup \{+\infty\}$ is strictly convex iff

$$\forall t \in]0,1[, \quad \forall x, y \in X, \qquad f(tx + (1-t)y) < tf(x) + (1-t)f(y).$$

- $f : X \to \mathbb{R} \cup \{+\infty\}$ is α-convex iff $\forall x, y \in X$

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

- If $f \in C^1(\mathbb{R}^n)$
  - $\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq 0$ iff f convex
  - if strict inequality holds, then f strictly convex
- If $f \in C^2(\mathbb{R}^n)$,
  - $\nabla^2 f \succeq 0$ iff f convex
  - if $\nabla^2 f \succ 0$ then f strictly convex
  - if $\nabla^2 f \succeq \alpha I$ then f is α-convex

Convex functions : strict and strong convexity

- $f : X \to \mathbb{R} \cup \{+\infty\}$ is strictly convex iff

$$\forall t \in]0,1[, \quad \forall x, y \in X, \qquad f(tx + (1-t)y) < tf(x) + (1-t)f(y).$$

- $f : X \to \mathbb{R} \cup \{+\infty\}$ is α-convex iff $\forall x, y \in X$

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

- If $f \in C^1(\mathbb{R}^n)$
  - $\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq 0$ iff f convex
  - if strict inequality holds, then f strictly convex
- If $f \in C^2(\mathbb{R}^n)$,
  - $\nabla^2 f \succeq 0$ iff f convex
  - if $\nabla^2 f \succ 0$ then f strictly convex
  - if $\nabla^2 f \succeq \alpha I$ then f is α-convex

Presentation Outline

1. Overview of the course
2. Convex sets and functions
   - Fundamental definitions and results
   - Convex function and minimization
   - Subdifferential and Fenchel-Transform
3. Duality
   - Recall on Lagrangian duality
   - Marginal interpretation of multiplier
   - Fenchel duality

Convex optimization problem

$$\min_{x \in C} \quad f(x)$$

Where C is closed convex and f convex finite valued, is a convex optimization problem.

- If C is compact and f proper lsc, then there exists an optimal solution.
- If f proper lsc and coercive, then there exists an optimal solution.
- The set of optimal solutions is convex.
- If f is strictly convex the minimum (if it exists) is unique.
- If f is α-convex the minimum exists and is unique.

Constraints and infinite values

A very standard trick in optimization consists in replacing constraints by infinite value of the cost function.

$$\min_{x \in C \subset X} f(x) = \min_{x \in X} f(x) + \mathbb{I}_C(x).$$

where

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

- If f is lsc and C is closed, then $f + \mathbb{I}_C$ is lsc.
- If f is proper and C is bounded, then $f + \mathbb{I}_C$ is coercive.
- Thus, from a theoretical point of view, we do not need to explicitely write constraint in a problem.

Presentation Outline

Subdifferential of convex function

Let X be a Banach space, $f : X \to \bar{\mathbb{R}}$.

- X^* is the topological dual of X, that is the set of continuous linear form on X.
- The subdifferential of f at $x \in \mathrm{dom}(f)$ is the set of slopes of all affine minorants of f exact at x:

$$\partial f(x) := \Big\{ x^* \in X^* \quad | \quad f(\cdot) \geq \langle x^*, \cdot - x \rangle + f(x) \Big\}.$$

- If f is convex and derivable at x then

$$\partial f(x) = \big\{ \nabla f(x) \big\}.$$

Partial infimum

Let $f : X \times Y \to \bar{\mathbb{R}}$ be a jointly convex and proper function, and define

$$v(x) = \inf_{y \in Y} f(x, y)$$

then v is convex.

If v is proper, and $v(x) = f(x, y^\sharp(x))$ then

$$\partial v(x) = \{ g \in X^* \quad | \quad (g, 0) \in \partial f(x, y^\sharp(x)) \}$$

proof:

$$\begin{aligned} g \in \partial v(x) \quad & \Leftrightarrow \quad \forall x', \qquad v(x') \geq v(x) + \langle g, x' - x \rangle \\ & \Leftrightarrow \quad \forall x', y' \quad f(x', y') \geq f(x, y^\sharp(x)) + \left\langle \begin{pmatrix} g \\ 0 \end{pmatrix}, \begin{pmatrix} x' \\ y' \end{pmatrix} - \begin{pmatrix} x \\ y^\sharp(x) \end{pmatrix} \right\rangle \\ & \Leftrightarrow \quad \begin{pmatrix} g \\ 0 \end{pmatrix} \in \partial f(x, y^\sharp(x)) \end{aligned}$$

Partial infimum

Let $f : X \times Y \to \bar{\mathbb{R}}$ be a jointly convex and proper function, and define

$$v(x) = \inf_{y \in Y} f(x, y)$$

then v is convex.
If v is proper, and $v(x) = f(x, y^\sharp(x))$ then

$$\partial v(x) = \{g \in X^* \quad | \quad (g, 0) \in \partial f(x, y^\sharp(x))\}$$

proof:

$$\begin{aligned} g \in \partial v(x) \quad &\Leftrightarrow \quad \forall x', \qquad v(x') \geq v(x) + \langle g, x' - x \rangle \\ &\Leftrightarrow \quad \forall x', y' \quad f(x', y') \geq f(x, y^\sharp(x)) + \left\langle \begin{pmatrix} g \\ 0 \end{pmatrix}, \begin{pmatrix} x' \\ y' \end{pmatrix} - \begin{pmatrix} x \\ y^\sharp(x) \end{pmatrix} \right\rangle \\ &\Leftrightarrow \quad \begin{pmatrix} g \\ 0 \end{pmatrix} \in \partial f(x, y^\sharp(x)) \end{aligned}$$

Partial infimum

Let $f : X \times Y \to \bar{\mathbb{R}}$ be a jointly convex and proper function, and define

$$v(x) = \inf_{y \in Y} f(x, y)$$

then v is convex.
If v is proper, and $v(x) = f(x, y^\sharp(x))$ then

$$\partial v(x) = \{g \in X^* \quad | \quad (g, 0) \in \partial f(x, y^\sharp(x))\}$$

proof:

$$\begin{aligned} g \in \partial v(x) \quad &\Leftrightarrow \quad \forall x', \qquad v(x') \geq v(x) + \langle g, x' - x\rangle \\ &\Leftrightarrow \quad \forall x', y' \quad f(x', y') \geq f(x, y^\sharp(x)) + \left\langle \begin{pmatrix} g \\ 0 \end{pmatrix}, \begin{pmatrix} x' \\ y' \end{pmatrix} - \begin{pmatrix} x \\ y^\sharp(x) \end{pmatrix} \right\rangle \\ &\Leftrightarrow \quad \begin{pmatrix} g \\ 0 \end{pmatrix} \in \partial f(x, y^\sharp(x)) \end{aligned}$$

Convex function : regularity

- Assume f convex, then f is continuous on the relative interior of its domain, and Lipschtiz on any compact contained in the relative interior of its domain.
- A proper convex function is subdifferentiable on the relative interior of its domain
- If f is convex, it is L-Lipschitz iff $\partial f(x) \subset B(0, L), \quad \forall x \in \mathrm{dom}(f)$

Fenchel transform

Let X be a Banach space, $f : X \to \bar{\mathbb{R}}$ convex proper.

- The Fenchel transform of f, is $f^* : X^* \to \bar{\mathbb{R}}$ with

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

- f^* is convex lsc as the supremum of affine functions.
- $f \leq g$ implies that $f^* \geq g^*$.
- If f is proper convex lsc, then $f^{**} = f$, otherwise $f^{**} \leq f$.

Fenchel transform

Let X be a Banach space, $f : X \to \bar{\mathbb{R}}$ convex proper.

- The Fenchel transform of f, is $f^* : X^* \to \bar{\mathbb{R}}$ with

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

- f^* is convex lsc as the supremum of affine functions.
- $f \leq g$ implies that $f^* \geq g^*$.
- If f is proper convex lsc, then $f^{**} = f$, otherwise $f^{**} \leq f$.

Fenchel transform

Let X be a Banach space, $f : X \to \bar{\mathbb{R}}$ convex proper.

- The Fenchel transform of f, is $f^* : X^* \to \bar{\mathbb{R}}$ with

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

- f^* is convex lsc as the supremum of affine functions.
- $f \leq g$ implies that $f^* \geq g^*$.
- If f is proper convex lsc, then $f^{**} = f$, otherwise $f^{**} \leq f$.

Fenchel transform

Let X be a Banach space, $f : X \to \bar{\mathbb{R}}$ convex proper.

- The Fenchel transform of f, is $f^* : X^* \to \bar{\mathbb{R}}$ with

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

- f^* is convex lsc as the supremum of affine functions.
- $f \leq g$ implies that $f^* \geq g^*$.
- If f is proper convex lsc, then $f^{**} = f$, otherwise $f^{**} \leq f$.

Fenchel transform and subdifferential

- By definition $f^*(x^*) \geq \langle x^*, x \rangle - f(x)$ for all x,
- thus we always have (Fenchel-Young) $f(x) + f^*(x^*) \geq \langle x^*, x \rangle$.
- Recall that $x^* \in \partial f(x)$ iff for all x', $f(x') \geq f(x) + \langle x^*, x' - x \rangle$ iff

$$\langle x^*, x \rangle - f(x) \geq \langle x^*, x' \rangle - f(x') \qquad \forall x'$$

that is

$$x^* \in \partial f(x) \Leftrightarrow x \in \arg\max_{x' \in X} \left\{ \langle x^*, x' \rangle - f(x') \right\} \Leftrightarrow f(x) + f^*(x^*) = \langle x^*, x \rangle$$

- From Fenchel-Young equality we have

$$\partial v^{**}(x) \neq \emptyset \implies \partial v^{**}(x) = \partial v(x) \text{ and } v^{**}(x) = v(x).$$

- If f proper convex lsc

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

Fenchel transform and subdifferential

- By definition $f^*(x^*) \geq \langle x^*, x\rangle - f(x)$ for all x,
- thus we always have (Fenchel-Young) $f(x) + f^*(x^*) \geq \langle x^*, x\rangle$.
- Recall that $x^* \in \partial f(x)$ iff for all x', $f(x') \geq f(x) + \langle x^*, x' - x\rangle$ iff

$$\langle x^*, x\rangle - f(x) \geq \langle x^*, x'\rangle - f(x') \qquad \forall x'$$

that is

$$x^* \in \partial f(x) \Leftrightarrow x \in \arg\max_{x' \in X} \left\{\langle x^*, x'\rangle - f(x')\right\} \Leftrightarrow f(x) + f^*(x^*) = \langle x^*, x\rangle$$

- From Fenchel-Young equality we have

$$\partial v^{**}(x) \neq \emptyset \quad \Longrightarrow \quad \partial v^{**}(x) = \partial v(x) \text{ and } v^{**}(x) = v(x).$$

- If f proper convex lsc

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

Fenchel transform and subdifferential

- By definition $f^*(x^*) \geq \langle x^*, x\rangle - f(x)$ for all x,
- thus we always have (Fenchel-Young) $f(x) + f^*(x^*) \geq \langle x^*, x\rangle$.
- Recall that $x^* \in \partial f(x)$ iff for all x', $f(x') \geq f(x) + \langle x^*, x' - x\rangle$ iff

$$\langle x^*, x\rangle - f(x) \geq \langle x^*, x'\rangle - f(x') \qquad \forall x'$$

  that is

$$x^* \in \partial f(x) \Leftrightarrow x \in \arg\max_{x' \in X} \big\{\langle x^*, x'\rangle - f(x')\big\} \Leftrightarrow f(x) + f^*(x^*) = \langle x^*, x\rangle$$

- From Fenchel-Young equality we have

$$\partial v^{**}(x) \neq \emptyset \quad \Longrightarrow \quad \partial v^{**}(x) = \partial v(x) \ \text{ and } \ v^{**}(x) = v(x).$$

- If f proper convex lsc

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

Fenchel transform and subdifferential

- By definition $f^*(x^*) \geq \langle x^*, x\rangle - f(x)$ for all x,
- thus we always have (Fenchel-Young) $f(x) + f^*(x^*) \geq \langle x^*, x\rangle$.
- Recall that $x^* \in \partial f(x)$ iff for all x', $f(x') \geq f(x) + \langle x^*, x' - x\rangle$ iff

  $$\langle x^*, x\rangle - f(x) \geq \langle x^*, x'\rangle - f(x') \qquad \forall x'$$

  that is

  $$x^* \in \partial f(x) \Leftrightarrow x \in \arg\max_{x' \in X} \left\{\langle x^*, x'\rangle - f(x')\right\} \Leftrightarrow f(x) + f^*(x^*) = \langle x^*, x\rangle$$

- From Fenchel-Young equality we have

  $$\partial v^{**}(x) \neq \emptyset \quad \Longrightarrow \quad \partial v^{**}(x) = \partial v(x) \text{ and } v^{**}(x) = v(x).$$

- If f proper convex lsc

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

Presentation Outline

Weak duality

The problem

$$(P) \quad \min_{x \in \mathbb{R}^n} \quad f(x)$$
$$s.t. \quad c_i(x) = 0 \qquad \forall i \in [\![1, n_E]\!]$$
$$c_j(x) \leq 0 \qquad \forall j \in [\![n_E + 1, n_E + n_I]\!]$$

can be written

$$\min_{x \in \mathbb{R}^n} \quad \max_{\lambda \in \mathbb{R}^{n_E}, \mu \in \mathbb{R}_+^{n_I}} \quad \mathcal{L}(x, \lambda, \mu)$$

where

$$\mathcal{L}(x, \lambda, \mu) := f(x) + \sum_{i=1}^{n_E + n_I} \lambda_i c_i(x)$$

The dual problem is

$$(D) \qquad \max_{\lambda \in \mathbb{R}^{n_E} \times \mathbb{R}_+^{n_I}} \quad \min_{x \in \mathbb{R}^n} \quad \mathcal{L}(x, \lambda, \mu)$$

and we have, without assumption

$$v_D \leq v_P.$$

Weak duality

The problem

$$(P) \quad \min_{x\in\mathbb{R}^n} \quad f(x)$$
$$s.t. \quad c_i(x) = 0 \qquad \forall i \in [\![1, n_E]\!]$$
$$c_j(x) \leq 0 \qquad \forall j \in [\![n_E+1, n_E+n_I]\!]$$

can be written

$$\min_{x\in\mathbb{R}^n} \quad \max_{\lambda\in\mathbb{R}^{n_E}, \mu\in\mathbb{R}_+^{n_I}} \quad \mathcal{L}(x,\lambda,\mu)$$

where

$$\mathcal{L}(x,\lambda,\mu) := f(x) + \sum_{i=1}^{n_E+n_I} \lambda_i c_i(x)$$

The dual problem is

$$(D) \qquad \max_{\lambda\in\mathbb{R}^{n_E}\times\mathbb{R}_+^{n_I}} \quad \min_{x\in\mathbb{R}^n} \quad \mathcal{L}(x,\lambda,\mu)$$

and we have, without assumption

$$v_D \leq v_P.$$

Linear Programming duality

$$\min_{x \geq 0} \quad c^\top x$$
$$s.t. \quad Ax = b$$

is equivalent to

$$\min_{x \geq 0} \max_{\lambda} (c - A^\top \lambda)^\top x + b^\top \lambda$$

and the dual problem is

$$\max_{\lambda} \quad b^\top \lambda$$
$$s.t. \quad A^\top \lambda \leq c$$

with equality between both problem except if there is neither primal nor dual admissible solution.

Strong duality

The duality gap is the difference between the primal value and dual value of a problem.
Consider problem

$$
\begin{aligned}
(P) \quad & \min_{x \in \mathbb{R}^n} \quad f(x) & \\
& s.t. \quad c_i(x) = 0 & \forall i \in [\![1, n_E]\!] \\
& \qquad\quad c_j(x) \leq 0 & \forall j \in [\![n_E + 1, n_E + n_I]\!]
\end{aligned}
$$

with (P) convex in the sense that f is convex, c_I is convex lsc, c_I is affine.
If further the constraints are qualified, then there is no duality gap.

Recall KKT

Assume that f, g_i and h_j are differentiable. Assume that $x^\sharp$ is an optimal solution of (P), and that the constraints are qualified in $x^\sharp$. Then we have

$$\begin{cases} \nabla_x \mathcal{L}(x^\sharp, \lambda^\sharp) = \nabla f(x^\sharp) + \sum_{i=1}^{n_E+n_i} \lambda_i^\sharp \nabla c_i(x^\sharp) = 0 \\ c_E(x^\sharp) = 0 \\ 0 \leq \lambda_I \perp c_I(x^\sharp) \leq 0 \end{cases}$$

Presentation Outline

Perturbed problem

Consider the perturbed problem

$$
\begin{aligned}
(P_p) \quad & \min_{x\in\mathbb{R}^n} \quad f(x) \\
& s.t. \quad c_i(x) + p_i = 0 \qquad \forall i \in [\![1, n_E]\!] \\
& \qquad\;\; c_j(x) + p_j \leq 0 \qquad \forall j \in [\![n_E+1, n_I+n_E]\!]
\end{aligned}
$$

with value $v(p)$, and optimal multiplier (for $p = 0$) λ_0.

Linear programming case

$$\begin{aligned} v(p) := \min_{x \geq 0} \quad & c^\top x \\ s.t. \quad & Ax + p = b \end{aligned}$$

by LP duality (assuming at least one admissible primal solution) we have

$$\begin{aligned} v(p) = \max_{\lambda} \quad & -b^\top \lambda + p^\top \lambda \\ s.t. \quad & A^\top \lambda \leq c \end{aligned}$$

Note λ_0 the optimal multiplier for (P_0), note that it is admissible for (D_p), hence $v(p) \geq -b^\top \lambda_0 + p^\top \lambda_0$. By strong duality we have $v(0) = -b^\top \lambda_0$, hence

$$v(p) \geq v(0) + \lambda_0^\top p$$

or

$$\lambda_0 \in \partial v(0).$$

Linear programming case

$$v(p) := \min_{x \geq 0} \quad c^\top x$$
$$s.t. \quad Ax + p = b$$

by LP duality (assuming at least one admissible primal solution) we have

$$v(p) = \max_{\lambda} \quad -b^\top \lambda + p^\top \lambda$$
$$s.t. \quad A^\top \lambda \leq c$$

Note λ_0 the optimal multiplier for (P_0), note that it is admissible for (D_p), hence $v(p) \geq -b^\top \lambda_0 + p^\top \lambda_0$. By strong duality we have $v(0) = -b^\top \lambda_0$, hence

$$v(p) \geq v(0) + \lambda_0^\top p$$

or

$$\lambda_0 \in \partial v(0).$$

Linear programming case

$$v(p) := \min_{x \geq 0} \quad c^\top x$$
$$s.t. \quad Ax + p = b$$

by LP duality (assuming at least one admissible primal solution) we have

$$v(p) = \max_{\lambda} \quad -b^\top \lambda + p^\top \lambda$$
$$s.t. \quad A^\top \lambda \leq c$$

Note λ_0 the optimal multiplier for (P_0), note that it is admissible for (D_p), hence $v(p) \geq -b^\top \lambda_0 + p^\top \lambda_0$. By strong duality we have $v(0) = -b^\top \lambda_0$, hence

$$v(p) \geq v(0) + \lambda_0^\top p$$

or

$$\lambda_0 \in \partial v(0).$$

Optimality condition by saddle point

Let $\Lambda := \mathbb{R}^{n_E} \times \mathbb{R}_+^{n_I}$. $(x^\sharp, \lambda^\sharp)$ is a saddle-point of $\mathcal{L}$ on $\mathbb{R}^n \times \Lambda$ iff

$$\forall \lambda \in \Lambda, \quad \mathcal{L}(x^\sharp, \lambda) \leq \mathcal{L}(x^\sharp, \lambda^\sharp) \leq \mathcal{L}(x, \lambda^\sharp), \quad \forall x \in \mathbb{R}^n$$

Consider $(\bar{x}, \bar{\lambda}) \in \mathbb{R}^n \times \Lambda$. Then $\bar{\lambda} \in \arg\max_{\lambda \in \Lambda} \mathcal{L}(\bar{x}, \lambda)$ iff $c_E(\bar{x}) = 0$ and $0 \leq \bar{\lambda}_I \perp c_I(\bar{x}) \leq 0$.

Theorem

If $(x^\sharp, \lambda^\sharp)$ is a saddle-point of $\mathcal{L}$ on $\mathbb{R}^n \times \Lambda$, then $x^\sharp$ is an optimal solution of (P).

Note that we need no assumption for this result.

Optimality condition by saddle point

Let $\Lambda := \mathbb{R}^{n_E} \times \mathbb{R}_+^{n_I}$. $(x^\sharp, \lambda^\sharp)$ is a saddle-point of $\mathcal{L}$ on $\mathbb{R}^n \times \Lambda$ iff

$$\forall \lambda \in \Lambda, \quad \mathcal{L}(x^\sharp, \lambda) \leq \mathcal{L}(x^\sharp, \lambda^\sharp) \leq \mathcal{L}(x, \lambda^\sharp), \quad \forall x \in \mathbb{R}^n$$

Consider $(\bar{x}, \bar{\lambda}) \in \mathbb{R}^n \times \Lambda$. Then $\bar{\lambda} \in \arg\max_{\lambda \in \Lambda} \mathcal{L}(\bar{x}, \lambda)$ iff $c_E(\bar{x}) = 0$ and $0 \leq \bar{\lambda}_I \perp c_I(\bar{x}) \leq 0$.

Theorem

If $(x^\sharp, \lambda^\sharp)$ is a saddle-point of $\mathcal{L}$ on $\mathbb{R}^n \times \Lambda$, then $x^\sharp$ is an optimal solution of (P).

Note that we need no assumption for this result.

Optimality condition by saddle point

Let $\Lambda := \mathbb{R}^{n_E} \times \mathbb{R}_+^{n_I}$. $(x^\sharp, \lambda^\sharp)$ is a saddle-point of $\mathcal{L}$ on $\mathbb{R}^n \times \Lambda$ iff

$$\forall \lambda \in \Lambda, \quad \mathcal{L}(x^\sharp, \lambda) \leq \mathcal{L}(x^\sharp, \lambda^\sharp) \leq \mathcal{L}(x, \lambda^\sharp), \quad \forall x \in \mathbb{R}^n$$

Consider $(\bar{x}, \bar{\lambda}) \in \mathbb{R}^n \times \Lambda$. Then $\bar{\lambda} \in \arg\max_{\lambda \in \Lambda} \mathcal{L}(\bar{x}, \lambda)$ iff $c_E(\bar{x}) = 0$ and $0 \leq \bar{\lambda}_I \perp c_I(\bar{x}) \leq 0$.

Theorem

If $(x^\sharp, \lambda^\sharp)$ is a saddle-point of $\mathcal{L}$ on $\mathbb{R}^n \times \Lambda$, then $x^\sharp$ is an optimal solution of (P).

Note that we need no assumption for this result.

Convex case

If (P) is convex in the sense that f is convex, c_I is convex and c_E is affine, then v is convex.

Theorem

Assume that v is convex, then

$$\partial v(0) = \left\{\lambda \in \Lambda \quad | \quad (x,\lambda) \text{ is a saddle point of } \mathcal{L}\right\}$$

In particular, $\partial v(0) \neq \emptyset$ iff there exists a saddle point of $\mathcal{L}$.

Theorem (Slater's qualification condition)

Consider a convex optimisation problem. Assume that c_E' is onto, and there exists $x \in rint(\mathrm{dom}(f))$ with $c_I(x) < 0$, and c_I continuous at x, then if x^* is an optimal solution, there exists λ^* such that (x^*,λ^*) is a saddle-point of the Lagrangian. Further, v is locally Lipschitz around 0.

Convex case

If (P) is convex in the sense that f is convex, c_I is convex and c_E is affine, then v is convex.

Theorem

Assume that v is convex, then

$$\partial v(0) = \{\lambda \in \Lambda \quad | \quad (x, \lambda) \text{ is a saddle point of } \mathcal{L}\}$$

In particular, $\partial v(0) \neq \emptyset$ iff there exists a saddle point of $\mathcal{L}$.

Theorem (Slater's qualification condition)

Consider a convex optimisation problem. Assume that c_E' is onto, and there exists $x \in rint(\mathrm{dom}(f))$ with $c_I(x) < 0$, and c_I continuous at x, then if x^* is an optimal solution, there exists λ^* such that (x^*, λ^*) is a saddle-point of the Lagrangian. Further, v is locally Lipschitz around 0.

Presentation Outline

Duality by abstract perturbation

Let $\mathbb{X}$ and $\mathbb{Y}$ be Banach spaces. There exists an abstract duality framework for $\min_{x\in\mathbb{X}} f(x)$ by considering a perturbation function $\Phi : \mathbb{X} \times \mathbb{Y} \to \mathbb{R} \cup \{+\infty\}$ (with $\Phi(\cdot, 0) = f$).

$$(\mathcal{P}_y) \qquad v(y) := \inf_{x\in\mathbb{X}} \Phi(x, y).$$

We have

$$\begin{aligned} v^*(y^*) &= \sup_{y\in\mathbb{Y}} \langle y^*, y\rangle - v(y) \\ &= \sup_{x,y} \langle y^*, y\rangle - \Phi(x, y) = \Phi^*(0, y^*) \end{aligned}$$

Thus we have

$$(\mathcal{D}_y) \qquad v^{**}(y) = \sup_{y^*\in\mathbb{Y}^*} \langle y^*, y\rangle - \Phi^*(0, y^*)$$

Generically

$$\mathrm{val}(\mathcal{D}_y) = v^{**}(y) \le v(y) = \mathrm{val}(\mathcal{P}_y)$$

Duality by abstract perturbation

Let $\mathbb{X}$ and $\mathbb{Y}$ be Banach spaces. There exists an abstract duality framework for $\min_{x\in\mathbb{X}} f(x)$ by considering a perturbation function $\Phi : \mathbb{X} \times \mathbb{Y} \to \mathbb{R} \cup \{+\infty\}$ (with $\Phi(\cdot, 0) = f$).

$$(\mathcal{P}_y) \qquad v(y) := \inf_{x\in\mathbb{X}} \Phi(x, y).$$

We have

$$\begin{aligned} v^*(y^*) &= \sup_{y\in\mathbb{Y}} \langle y^*, y\rangle - v(y) \\ &= \sup_{x,y} \langle y^*, y\rangle - \Phi(x, y) = \Phi^*(0, y^*) \end{aligned}$$

Thus we have

$$(\mathcal{D}_y) \qquad v^{**}(y) = \sup_{y^*\in\mathbb{Y}^*} \langle y^*, y\rangle - \Phi^*(0, y^*)$$

Generically

$$\mathrm{val}(\mathcal{D}_y) = v^{**}(y) \leq v(y) = \mathrm{val}(\mathcal{P}_y)$$

Duality by abstract perturbation

Let $\mathbb{X}$ and $\mathbb{Y}$ be Banach spaces. There exists an abstract duality framework for $\min_{x\in\mathbb{X}} f(x)$ by considering a perturbation function $\Phi : \mathbb{X} \times \mathbb{Y} \to \mathbb{R} \cup \{+\infty\}$ (with $\Phi(\cdot, 0) = f$).

$$(\mathcal{P}_y) \qquad v(y) := \inf_{x\in\mathbb{X}} \Phi(x, y).$$

We have

$$\begin{aligned} v^*(y^*) &= \sup_{y\in\mathbb{Y}} \langle y^*, y\rangle - v(y) \\ &= \sup_{x,y} \langle y^*, y\rangle - \Phi(x, y) = \Phi^*(0, y^*) \end{aligned}$$

Thus we have

$$(\mathcal{D}_y) \qquad v^{**}(y) = \sup_{y^*\in\mathbb{Y}^*} \langle y^*, y\rangle - \Phi^*(0, y^*)$$

Generically

$$\mathrm{val}(\mathcal{D}_y) = v^{**}(y) \le v(y) = \mathrm{val}(\mathcal{P}_y)$$

Solution of the dual as subgradient

Note that the set of solution of the dual is $S(\mathcal{D}_y) = \partial v^{**}(y)$.
Recall that, for v proper convex,

$$\partial v^{**}(x) \neq \emptyset \quad \Longrightarrow \quad \partial v^{**}(x) = \partial v(x) \text{ and } v^{**}(x) = v(x)$$

Thus, if v is proper convex and subdifferentiable at y (or equivalently if $S(\mathcal{D}_y) \neq \emptyset$), then,

$$\mathrm{val}(\mathcal{D}_y) = \mathrm{val}(\mathcal{P}_y)$$
$$S(\mathcal{D}_y) = \partial v(y)$$

Finally, as a convex function is subdifferentiable on the relative interior of its domain, a sufficient qualification condition (to have a zero dual gap and existence of multipliers), is that

$$0 \in rint(\mathrm{dom}(v)).$$

Solution of the dual as subgradient

Note that the set of solution of the dual is $S(\mathcal{D}_y) = \partial v^{**}(y)$.
Recall that, for v proper convex,

$$\partial v^{**}(x) \neq \emptyset \quad \Longrightarrow \quad \partial v^{**}(x) = \partial v(x) \text{ and } v^{**}(x) = v(x)$$

Thus, if v is proper convex and subdifferentiable at y (or equivalently if $S(\mathcal{D}_y) \neq \emptyset$), then,

$$\mathrm{val}(\mathcal{D}_y) = \mathrm{val}(\mathcal{P}_y)$$
$$S(\mathcal{D}_y) = \partial v(y)$$

Finally, as a convex function is subdifferentiable on the relative interior of its domain, a sufficient qualification condition (to have a zero dual gap and existence of multipliers), is that

$$0 \in rint(\mathrm{dom}(v)).$$

Solution of the dual as subgradient

Note that the set of solution of the dual is $S(\mathcal{D}_y) = \partial v^{**}(y)$.
Recall that, for v proper convex,

$$\partial v^{**}(x) \neq \emptyset \quad \Longrightarrow \quad \partial v^{**}(x) = \partial v(x) \text{ and } v^{**}(x) = v(x)$$

Thus, if v is proper convex and subdifferentiable at y (or equivalently if $S(\mathcal{D}_y) \neq \emptyset$), then,

$$\mathrm{val}(\mathcal{D}_y) = \mathrm{val}(\mathcal{P}_y)$$
$$S(\mathcal{D}_y) = \partial v(y)$$

Finally, as a convex function is subdifferentiable on the relative interior of its domain, a sufficient qualification condition (to have a zero dual gap and existence of multipliers), is that

$$0 \in rint(\mathrm{dom}(v)).$$

Recovering the Lagrangian dual

Problem $(\mathcal{P}_y)$ can be written

$$\begin{aligned} \min_{x,z} \quad & \Phi(x,z) \\ s.t. \quad & z = y \end{aligned}$$

with Lagrangian dual

$$\max_{y^* \in Y^*} \inf_{x,z \in X \times Y} \Phi(x,z) + \langle y^*, y - z \rangle = \max_{y^* \in Y^*} \langle y^*, y \rangle - \underbrace{\sup_{x,z \in X \times Y} \Big\{ \langle y^*, z \rangle - \Phi(x,z) \Big\}}_{\Phi^*(0,y^*)}$$

Hence, we recover the Fenchel dual from the Lagrangian dual.

For next week

- Install Julia / Jupyter / JuMP (see instructions https://github.com/leclere/TP-Saclay)
- Run the CrashCourse notebook to get used with those tools (there are other resources available on the web as well)
- Contact me vincent.leclere@enpc.fr in case of trouble