Optimization Basic Results

Just Enough for Stochastic Programming

Michel De Lara
Cermics, École des Pontes ParisTech
France

École des Ponts ParisTech

December 4, 2019

Outline of the presentation

Convex functions, coercivity

Existence and uniqueness of a minimum

First-order optimality conditions (the case of equality constraints)

Duality gap and saddle-points

Elements of Lagrangian duality and Uzawa algorithm

Outline of the presentation

Convex functions, coercivity

Existence and uniqueness of a minimum

First-order optimality conditions (the case of equality constraints)

Duality gap and saddle-points

Elements of Lagrangian duality and Uzawa algorithm

Convex sets

Let $N \in \mathbb{N}^*$. We consider subsets of the Euclidian space $\mathbb{R}^N$

- The subset $C \subset \mathbb{R}^N$ is convex if for any $x_1 \in C$, $x_2 \in C$ and $t \in [0,1]$, we have that $tx_1 + (1-t)x_2 \in C$
- An intersection of convex sets is convex
- A segment is convex
- A hyperplane is convex ($H \subset \mathbb{R}^N$ is a hyperplane if there exists $y \in \mathbb{R}^N \backslash \{0\}$ and $b \in \mathbb{R}$ such that $H = \left\{ x \in \mathbb{R}^N \mid \langle x\,, y \rangle + b = 0 \right\}$)
- An affine subspace (intersection of hyperplanes) is convex

Linear and affine functions

Consider a function $f : \mathbb{R}^N \to \mathbb{R}$

- The function f is linear if, for any $x_1 \in \mathbb{R}^N$, $x_2 \in \mathbb{R}^N$ and $t_1 \in \mathbb{R}$, $t_2 \in \mathbb{R}$,

$$f(t_1 x_1 + t_2 x_2) = t_1 f(x_1) + t_2 f(x_2)$$

- The function f is affine if, for any $x_1 \in \mathbb{R}^N$, $x_2 \in \mathbb{R}^N$ and $t \in \mathbb{R}$,

$$f(tx_1 + (1-t)x_2) = tf(x_1) + (1-t)f(x_2)$$

Exercise. Show that $f : \mathbb{R}^N \to \mathbb{R}$ is affine if and only if $g(x) = f(x) - f(0)$ is linear

Convex functions (definitions)

Let $C \subset \mathbb{R}^N$ be an nonempty convex subset of $\mathbb{R}^N$, where $N \in \mathbb{N}^*$, and $f : C \to \mathbb{R}$ be a function

- The function f is affine if, for any $x_1 \in C$, $x_2 \in C$ and $t \in [0, 1]$,

$$f(tx_1 + (1-t)x_2) = tf(x_1) + (1-t)f(x_2)$$

- The function f is convex if, for any $x_1 \in C$, $x_2 \in C$ and $t \in [0, 1]$,

$$f(tx_1 + (1-t)x_2) \leq tf(x_1) + (1-t)f(x_2)$$

- The function f is strictly convex if,
  for any $x_1 \in C$, $x_2 \in C$, $x_1 \neq x_2$, and $t \in]0, 1[$,

$$f(tx_1 + (1-t)x_2) < tf(x_1) + (1-t)f(x_2)$$

- The function f is strongly convex (of modulus $a > 0$) if,
  for any $x_1 \in C$, $x_2 \in C$ and $t \in [0, 1]$,

$$f(tx_1 + (1-t)x_2) \leq tf(x_1) + (1-t)f(x_2) - \frac{a}{2}t(1-t)\|x_1 - x_2\|^2$$

Exercises

Let $C \subset \mathbb{R}^N$ be an nonempty subset of $\mathbb{R}^N$, where $N \in \mathbb{N}^*$

- Show that both definitions of an affine function coincide when $C = \mathbb{R}^N$
- Show that a function $f : C \to \mathbb{R}$ is convex if and only if its epigraph is a convex set subset of $\mathbb{R}^N \times \mathbb{R}$
- Show that a function $f : C \to \mathbb{R}$ is strongly convex of modulus $a > 0$ if and only if $g(x) = f(x) - \frac{a}{2}\|x\|^2$ is convex
- If $f : C \to \mathbb{R}$ is convex, show that f is *not* strictly convex if and only there exists a nonempty convex subset $C' \subset C$ over which f is affine

Convex functions on the real line

Proposition

Let $I \subset \mathbb{R}$ be an nonempty interval

- *A C^1 function $f : I \to \mathbb{R}$ is convex if and only if f' is increasing on I*
- *A C^2 function $f : I \to \mathbb{R}$ is convex if and only if $f''(x) \geq 0$, for all $x \in I$*
- *Let $a > 0$. A C^2 function $f : I \to \mathbb{R}$ is a-strongly convex if and only if $f''(x) \geq a$, for all $x \in I$*
- *A C^1 function $f : I \to \mathbb{R}$ is strictly convex if and only if f is convex and the set $\{x \in I \mid f''(x) = 0\}$ is either empty or is a singleton*

<u>Exercise.</u> Study the family of functions $f_\alpha :]0, +\infty[\to \mathbb{R}$ given by $f_\alpha(x) = x^\alpha$. For which values of the parameter α is the function f_α convex? For a given $a > 0$, for which values of the parameter α is the function f_α strongly convex of modulus a? Provide an example of a strictly convex function which is not strongly convex.

Convexity for multivariate functions

The *Hessian* matrix $\mathcal{H}_f(x)$ of a C^2 function $f : \mathbb{R}^N \to \mathbb{R}$ is the $N \times N$ symmetric matrix given by

$$\mathcal{H}_f(x) = \left\{ \frac{\partial^2 f}{\partial x^i \partial x^j}(x) \right\}_{(i,j)\in\{1,\dots,N\}^2}$$

Proposition

Let $C \subset \mathbb{R}^N$ be an nonempty convex subset of $\mathbb{R}^N$, where $N \in \mathbb{N}^*$

- *A C^2 function $f : \mathbb{R}^N \to \mathbb{R}$ is convex on C if and only if the symmetric Hessian matrix $\mathcal{H}_f(x)$ is positive for all $x \in C$*
- *A C^2 function $f : \mathbb{R}^N \to \mathbb{R}$ is strongly convex of modulus $a > 0$ on C if and only if the eigenvalues of the symmetric Hessian matrix $\mathcal{H}_f(x)$ are uniformly bounded below by $a > 0$ on C*

Exercise. Let Q be a $N \times N$ symmetric matrix and $f(x) = 1/2x'Qx$, where x' is the transpose of the vector x. Give conditions on the smallest eigenvalue of Q so that the function f is convex, or strictly convex, or strongly convex of modulus a.

Operations on functions preserving convexity

Proposition

Let $(f_i)_{i\in I}$ be a family of convex functions
Then $\sup_{i\in I} f_i$ is a convex function

Proposition

Let $(f_i)_{i=1,\ldots,n}$ be convex functions
Let $(\alpha_i)_{i=1,\ldots,n}$ be nonnegative numbers
Then $\sum_{i=1}^{m} \alpha_i f_i$ is a convex function

Proposition

Let $f : \mathbb{R}^N \to \mathbb{R}$ be convex
Let A be a $N \times M$ matrix and $b \in \mathbb{R}^N$
Then $y \in \mathbb{R}^M \mapsto f(Ay + b)$ is a convex function

Coercivity

Definition

A function $f : \mathbb{R}^N \to \mathbb{R}$ is coercive if

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

Proposition

A strongly convex function is coercive

Outline of the presentation

Here are the ingredients for a general abstract optimization problem

$$\inf_{u \in \mathbb{U}^{ad}} J(u)$$

- Optimization set $\mathbb{U} (= \mathbb{R}^N)$ containing optimization variables $u \in \mathbb{U}$
- A criterion $J : \mathbb{U} \to \mathbb{R} \cup \{+\infty\}$
- Constraints of the form $u \in \mathbb{U}^{ad} \subset \mathbb{U}$

Examples of classes of optimization problems

$$\inf_{u \in \mathbb{U}^{ad}} J(u)$$

- Linear programming
  - Optimization set $\mathbb{U} = \mathbb{R}^N$
  - Criterion J is linear (affine)
  - Constraints $\mathbb{U}^{ad}$ defined by a finite number of linear (affine) equalities and inequalities
- Convex optimization
  - Criterion J is a convex function
  - Constraints $\mathbb{U}^{ad}$ define a convex set
- Combinatorial optimization
  - Optimization set $\mathbb{U}$ is discrete (binary $\{0, 1\}^N$, integer $\mathbb{Z}^N$, etc.)

Minimum

Definition

We say that $u^\star \in \mathbb{U}$ is a (global) minimum of the optimization problem $\inf_{u \in \mathbb{U}^{ad}} J(u)$ if

$$u^\star \in \mathbb{U}^{ad} \text{ and } J(u^\star) \leq J(u) \; , \;\; \forall u \in \mathbb{U}^{ad}$$

In this case, we write

$$J(u^\star) = \min_{u \in \mathbb{U}^{ad}} J(u)$$

Existence and uniqueness of a minimum

We consider the optimization problem

$$\inf_{u \in \mathbb{U}^{ad}} J(u) \text{ where } \mathbb{U}^{ad} \subset \mathbb{U} = \mathbb{R}^N$$

Proposition

If the criterion J is continuous and the constraint set $\mathbb{U}^{ad}$ is compact (bounded and closed), then there is a minimum

Proposition

If the constraint set $\mathbb{U}^{ad}$ is closed and the criterion J is continuous and coercive, then there is a minimum

Proposition

If the constraint set $\mathbb{U}^{ad}$ is convex, and if the criterion J is strictly convex, a minimum is necessarily unique

Exercises

We consider the optimization problem

$$\inf_{u \in \mathbb{U}^{ad}} J(u)$$

Give an example

- of continuous criterion J and of constraint set $\mathbb{U}^{ad}$ for which there is no minimum
- of criterion J and of compact constraint set $\mathbb{U}^{ad}$ for which there is no minimum
- of continuous criterion J and of unbounded and closed constraint set $\mathbb{U}^{ad}$ for which there is no minimum
- of convex criterion J and of constraint set $\mathbb{U}^{ad}$ for which there is more than one minimum
- of strictly convex criterion J and of constraint set $\mathbb{U}^{ad}$ for which there is more than one minimum

Local minimum

Definition

We say that $u^\star \in \mathbb{U}$ is a local minimum of the optimization problem $\inf_{u \in \mathbb{U}^{ad}} J(u)$ if there exists a neighborhood $\mathcal{V}$ of $u^\star$ in $\mathbb{U}^{ad}$ such that

$$u^\star \in \mathbb{U}^{ad} \text{ and } J(u^\star) \leq J(u) \;, \;\; \forall u \in \mathcal{V}$$

Proposition

If the constraint set $\mathbb{U}^{ad}$ is convex, and if the criterion J is convex, a local minimum is a global minimum

Outline of the presentation

Convex functions, coercivity

Existence and uniqueness of a minimum

First-order optimality conditions (the case of equality constraints)

Duality gap and saddle-points

Elements of Lagrangian duality and Uzawa algorithm

Optimization under equality constraints

We consider the optimization problem

$$\inf_{u \in \mathbb{R}^N} J(u)$$

under the constraint

$$\Theta(u) = 0$$

where Θ is a function with values in $\mathbb{R}^M$

$$\Theta = (\Theta_1, \ldots, \Theta_M) : \mathbb{R}^N \to \mathbb{R}^M$$

whose components are denoted by Θ_j, where j runs from 1 to M

Sufficient condition for qualification in case of equality constraints

Definition

Let $u^\star \in \mathbb{R}^N$. The equality constraints $\Theta(u) = 0$ are said to be *regular* at $u^\star$ if, when $\Theta(u^\star) = 0$, the function Θ is differentiable at $u^\star$ and the vectors $\nabla\Theta_j(u^\star)$, $j \in \{1, \dots, M\}$, are linearly independent

Let $u^\star \in \mathbb{R}^N$. In case

- either the equality constraints $\Theta(u) = 0$ are regular at $u^\star$
- or the function Θ is affine

we say that the equality constraints $\Theta(u) = 0$ are qualified at $u^\star$

Lagrangian

Definition

The Lagrangian $\mathcal{L} : \mathbb{R}^N \times \mathbb{R}^M \to \mathbb{R}$ is defined by

$$\mathcal{L}(u, \lambda) = J(u) + \langle \lambda, \Theta(u) \rangle = J(u) + \sum_{j=1}^{M} \lambda_j \Theta_j(u)$$

The variables λ are called (Lagrange) multipliers

First-order optimality conditions (necessary)
KKT optimality conditions

Proposition

We suppose that the criterion J is differentiable. Let $u^\star \in \mathbb{R}^N$.
If the equality constraints $\Theta(u) = 0$ are qualified at $u^\star$,
then a *necessary condition* for $u^\star$ to be a *local minimum* of J,
among the u such that $\Theta(u) = 0$,
is that there exists a vector $\lambda^\star$ of $\mathbb{R}^M$ such that

$$\frac{\partial \mathcal{L}}{\partial u}(u^\star, \lambda^\star) = 0 \text{ and } \frac{\partial \mathcal{L}}{\partial \lambda}(u^\star, \lambda^\star) = 0$$

expressing the first-order optimality conditions
(KKT optimality conditions)

First-order optimality conditions (sufficient)

Proposition

Let $u^\star \in \mathbb{R}^N$. We suppose that

- the criterion J is differentiable and convex
- in the equality constraints $\Theta(u) = 0$, the function Θ is affine

Then a *sufficient condition* for $u^\star$ to be a *minimum* of J,
among the u such that $\Theta(u) = 0$,
is that there exists a vector $\lambda^\star$ of $\mathbb{R}^M$ such that

$$\frac{\partial \mathcal{L}}{\partial u}(u^\star, \lambda^\star) = 0 \text{ and } \frac{\partial \mathcal{L}}{\partial \lambda}(u^\star, \lambda^\star) = 0$$

Outline of the presentation

Convex functions, coercivity

Existence and uniqueness of a minimum

First-order optimality conditions (the case of equality constraints)

Duality gap and saddle-points

Elements of Lagrangian duality and Uzawa algorithm

Duality gap

Consider a function

$$\phi : \mathbb{X} \times \mathbb{Y} \to \mathbb{R}$$

Proposition

We have the inequality

$$\inf_x \sup_y \phi(x, y) \geq \sup_y \inf_x \phi(x, y)$$

Notice that we minimize in the first variable x (primal variable) and maximize in the second variable y (dual variable)

Definition

The duality gap is

$$\inf_x \sup_y \phi(x, y) - \sup_y \inf_x \phi(x, y) \geq 0$$

Saddle-point

Definition

We say that $(\overline{x}, \overline{y}) \in \mathbb{X} \times \mathbb{Y}$ is a saddle-point if

- $y \mapsto \phi(\overline{x}, y)$ achieves a *maximum* at $\overline{y}$
- $x \mapsto \phi(x, \overline{y})$ achieves a *minimum* at $\overline{x}$

or, equivalently

$$\phi(x, \overline{y}) \geq \phi(\overline{x}, \overline{y}) \geq \phi(\overline{x}, y)$$

Proposition

When there exists a saddle-point, there is no duality gap
(that is, the duality gap is zero)

Existence of a saddle point

Proposition

Suppose that $\phi : \mathbb{X} \times \mathbb{Y} \to \mathbb{R}$

- is continuous,
- convex-concave (convex in the variable x, concave in the variable y),
- there exists two convex closed sets $X \subset \mathbb{X}$ and $Y \subset \mathbb{Y}$ such that
  - there exists a $\hat{x} \in X$ such that $\lim_{\|y\| \to +\infty} \phi(\hat{x}, y) = -\infty$, or the set Y is bounded,
  - there exists a $\hat{y} \in Y$ such that $\lim_{\|x\| \to +\infty} \phi(x, \hat{y}) = +\infty$, or the set X is bounded.

Then, there exists a saddle point for the function ϕ on $X \times Y$

Outline of the presentation

Convex functions, coercivity

Existence and uniqueness of a minimum

First-order optimality conditions (the case of equality constraints)

Duality gap and saddle-points

Elements of Lagrangian duality and Uzawa algorithm

Optimization under equality constraints

- We consider the optimization problem

$$\inf_{u \in \mathbb{R}^N} J(u)$$

under the constraint

$$\Theta(u) = 0$$

where $\Theta = (\Theta_1, \ldots, \Theta_M) : \mathbb{R}^N \to \mathbb{R}^M$

- The Lagrangian $\mathcal{L} : \mathbb{R}^N \times \mathbb{R}^M \to \mathbb{R}$ is defined by

$$\mathcal{L}(u, \lambda) = J(u) + \langle \lambda, \Theta(u) \rangle = J(u) + \sum_{j=1}^{M} \lambda_j \Theta_j(u)$$

Primal problem

Definition

The primal optimization problem is

$$\inf_{u \in \mathbb{R}^N} \sup_{\lambda \in \mathbb{R}^M} \mathcal{L}(u, \lambda)$$

Proposition

The original and the primal optimization problems have the same solutions (in $u \in \mathbb{R}^N$)

Dual problem

Definition

The dual optimization problem is

$$\sup_{\lambda \in \mathbb{R}^M} \inf_{u \in \mathbb{R}^N} \mathcal{L}(u, \lambda)$$

Definition

The dual function is $D : \mathbb{R}^M \to \mathbb{R} \cup \{-\infty\}$ given by

$$D(\lambda) = \inf_{u \in \mathbb{R}^N} \mathcal{L}(u, \lambda)$$

Proposition

The dual function is concave

Proposition

When there exists a saddle-point for the Lagrangian, primal and dual problems are equivalent

First-order optimality conditions and saddle point

Proposition

We suppose that

- the criterion J is differentiable and convex
- in the equality constraints $\Theta(u) = 0$, the function Θ is affine

Let $u^\star \in \mathbb{R}^N$ be a *minimum* of J, among the u such that $\Theta(u) = 0$.
Then, there exists a vector $\lambda^\star$ of $\mathbb{R}^M$ such that
$(u^\star, \lambda^\star)$ is a saddle point of the Lagrangian $\mathcal{L}$, that is,

$$u \mapsto \mathcal{L}(u, \lambda^\star)$$

achieves a minimum at $u^\star$, and

$$\lambda \mapsto \mathcal{L}(u^\star, \lambda)$$

achieves a maximum at $\lambda^\star$

Existence of a minimum and of a saddle point

Proposition

We suppose that

- the criterion J is differentiable and strongly convex
- in the equality constraints $\Theta(u) = 0$, the function Θ is affine

Then

- there exists a unique *minimum* $u^\star \in \mathbb{R}^N$ of J among the u such that $\Theta(u) = 0$
- there exists a vector $\lambda^\star$ of $\mathbb{R}^M$ such that $(u^\star, \lambda^\star)$ is a saddle point of the Lagrangian $\mathcal{L}$

The Uzawa algorithm or dual gradient algorithm

We suppose that

- the criterion J is differentiable and a-strongly convex
- in the equality constraints $\Theta(u) = 0$,
  the function Θ is affine, with norm κ

Then, when $0 < \rho < 2a/\kappa^2$, the following algorithm converges towards the (unique) minimum of

$$\inf_{u \in \mathbb{R}^N} J(u) \ , \ \ \Theta(u) = 0$$

Data: Initial multiplier $\lambda^{(0)}$, step ρ
Result: minimum and multiplier;
repeat
  $u^{(k)} = \arg\min_{u \in \mathbb{R}^N} \mathcal{L}(u, \lambda^{(k)})$ (minimization w.r.t. the first variable) ;
  $\lambda^{(k+1)} = \lambda^{(k)} + \rho \Theta(u^{(k)})$ (gradient step for the second variable) ;
until $\Theta(u^{(k)}) = 0$;

Algorithm 1: Dual Gradient Algorithm