IMCA Instituto de Matemática y Ciencias Afines

Smart Energy and Stochastic Optimization

◇

SHORT COURSES

Review of convexity and optimization

P. Carpentier — J. P. Chancelier — M. De Lara

Cermics, ENPC, IP Paris, France

November 26, 2024

Outline of the presentation

Optimization problems, convex functions, local and global minima

First-order optimality conditions

Lagrangian duality and Uzawa algorithm

Outline of the presentation

Optimization problems, convex functions, local and global minima

First-order optimality conditions

Lagrangian duality and Uzawa algorithm

Outline of the presentation

Ingredients for a general optimization problem

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

- Optimization space $\mathcal{U}$, optimization variables $u \in \mathcal{U}$
- Constraints $u \in U^{ad} \subset \mathcal{U}$ (admissible set)
- Criterion $J : \mathcal{U} \to \mathbb{R} \cup \{+\infty\}$

As a remark, we have

$$\inf_{u \in U^{ad}} J(u) = \inf_{u \in \mathcal{U}} \Big(J(u) + \iota_{U^{ad}}(u) \Big)$$

$\iota_{U^{ad}}$ being the indicator function of the set U^{ad}

$$\iota_{U^{ad}}(u) = \begin{cases} 0 & \text{if } u \in U^{ad} \\ +\infty & \text{if } u \notin U^{ad} \end{cases}$$

Some classes of optimization problems in finite dimension

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

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

Some useful formulas in optimization

- Linearity formula. For any function $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ we have

$$\inf_{x \in \mathcal{X}} \Big(a + f(x) \Big) = a + \inf_{x \in \mathcal{X}} f(x) \; , \quad \forall a \in \mathbb{R}$$

$$\inf_{x \in \mathcal{X}} \alpha f(x) = \alpha \inf_{x \in \mathcal{X}} f(x) \; , \qquad \forall \alpha \geq 0$$

- Tower formula. For any function $h : \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{+\infty\}$ we have

$$\inf_{(x,y) \in \mathcal{X} \times \mathcal{Y}} h(x,y) = \inf_{x \in \mathcal{X}} \big(\inf_{y \in \mathcal{Y}} h(x,y) \big) = \inf_{y \in \mathcal{Y}} \big(\inf_{x \in \mathcal{X}} h(x,y) \big)$$

- Interchange formula. For any function $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ and any function $g : \mathcal{Y} \to \mathbb{R} \cup \{+\infty\}$ we have

$$\inf_{(x,y) \in \mathcal{X} \times \mathcal{Y}} \big(f(x) + g(y) \big) = \inf_{x \in \mathcal{X}} f(x) + \inf_{y \in \mathcal{Y}} g(y)$$

*Exercise**

Give proofs for the tower and interchange formulas

Interchange in a stochastic framework

Consider a finite set $\mathbb{S}$ of scenarios equipped with a probability[1] $\{\pi_s\}_{s\in\mathbb{S}}$.
For each scenario $s \in \mathbb{S}$, we have

- a cost function $f_s : \mathcal{X}_s \to \mathbb{R} \cup \{+\infty\}$
- depending on a control $x_s \in \mathcal{X}_s$

We have

$$\inf_{\{x_s\}_{s\in\mathbb{S}} \in \prod_{s\in\mathbb{S}} \mathcal{X}_s} \left(\sum_{s\in\mathbb{S}} \pi_s f_s(x_s) \right) = \sum_{s\in\mathbb{S}} \pi_s \left(\inf_{x_s\in\mathcal{X}_s} f_s(x_s) \right)$$

that is, the operator inf and the operator $\mathbb{E}$ can be interchanged.

[1] that is, $\pi_s \geq 0$ and $\sum_{s\in\mathbb{S}} \pi_s = 1$

Outline of the presentation

Convex sets

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

- The set $C \subset \mathbb{R}^N$ is convex if we have

$$\forall (x_1, x_2) \in C \times C \ , \ \ \forall t \in [0,1] \ , \ \ t x_1 + (1-t) x_2 \in C$$

- A segment is convex
- A hyperplane is convex[2]
- An affine subspace[3] is convex
- An intersection of convex sets is convex

*Exercise**
Give the proof of the last statement

[2] Hyperplane $H = \big\{ x \in \mathbb{R}^N \,\big|\, \langle x,\, y \rangle + b = 0 \big\}$ with $y \in \mathbb{R}^N \backslash \{0\}$ and $b \in \mathbb{R}$
[3] intersection of hyperplanes

Definitions of convex functions

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

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

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

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

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

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

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

Exercises

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

- *Show that a function* $f : C \to \mathbb{R}$ *is convex if and only if its epigraph*[4] *is a convex set of* $\mathbb{R}^N \times \mathbb{R}$
- *Show that a function* $f : C \to \mathbb{R}$ *is a-strongly convex 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 if there exists a nonempty convex set* $C' \subset C$ *over which* f *is affine*

[4] $\mathrm{epi} f = \big\{(x, y) \in \mathbb{R}^N \times \mathbb{R} \,\big|\, f(x) \leq y\big\} \subset \mathbb{R}^N \times \mathbb{R}$

Convexity for differentiable multivariate functions

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

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

Proposition

Let $C \subset \mathbb{R}^N$ be an nonempty convex set 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 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 a-strongly convex on C if and only if the eigenvalues of the 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 $f(x) = \frac{1}{2}x^\top Qx$, where $x^\top$ 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 a-strongly convex

Operations preserving convexity

Proposition

Let $(f_i)_{i \in I}$ be a family of convex functions indexed by $i \in I$
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 $\sum_{i=1}^{n} \alpha_i f_i$ 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

*Exercise**

Give a proof of the first proposition

Outline of the presentation

Coercivity

Definition

A function $f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ is coercive on a set $C \subset \mathbb{R}^N$ if

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

Proposition

A a-strongly convex differentiable function is coercive

*Exercise**

Give a proof of the proposition

Minimum

Definition

We say that $u^* \in \mathcal{U}$ is a global minimum of the optimization problem

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

if we have

$$u^* \in U^{ad} \qquad \text{and} \qquad J(u^*) \leq J(u) \ , \ \ \forall u \in U^{ad}$$

In this case, we write

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

Existence and uniqueness of a minimum

We consider the finite dimensional optimization problem

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

Proposition

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

Proposition

If the constraint set U^{ad} is closed and if the criterion J is continuous and coercive on U^{ad}, then there exists a global minimum

Proposition

If the constraint set U^{ad} is closed and convex and if the criterion J is strictly convex, then the global minimum (if it exists) is unique

Exercise

We consider the optimization problem

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

Give an example

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

Local minimum

Definition

We say that $u^* \in \mathcal{U}$ is a local minimum of the optimization problem

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

if there exists a neighborhood $\mathcal{V}$ of u^* in U^{ad} such that

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

Proposition

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

Outline of the presentation

Optimization problems, convex functions, local and global minima

First-order optimality conditions

Lagrangian duality and Uzawa algorithm

Outline of the presentation

Optimization over an admissible set

We consider the optimization problem

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

From now, U^{ad} is a non empty convex set of $\mathbb{R}^N$

In the case where function J is differentiable, we denote by $\nabla J(u) \in \mathbb{R}^N$ its gradient at point u:

$$\lim_{t>0, t \to 0} \frac{J(u + td) - J(u)}{t} = \nabla J(u)^\top d \ , \ \forall d \in \mathbb{R}^N$$

First-order optimality conditions

Proposition

Assume that J is differentiable and U^{ad} is a convex set
Let $u^* \in U^{ad}$. Then a necessary condition for u^* to be a local minimum of J over the set U^{ad} is that

$$\nabla J(u^*)^\top (u - u^*) \geq 0 \quad \forall u \in U^{ad}$$

*Exercise**

Show that, in the case where $U^{ad} = \mathbb{R}^N$, this optimality condition reduces to the standard stationarity condition $\nabla J(u^*) = 0$

Proposition

Assume moreover that J is a convex function
Let $u^* \in U^{ad}$. Then a sufficient condition for u^* to be a global minimum of J over the set U^{ad} is that

$$\nabla J(u^*)^\top (u - u^*) \geq 0 \quad \forall u \in U^{ad}$$

Standard projected gradient algorithm

We suppose that J is differentiable with gradient Lipschitz of modulus L,[5] a-strongly convex and that U^{ad} is a convex set

Then, for a step size ρ such that

$$0 < \rho < 2a/L^2$$

the following algorithm converges towards the unique minimum u^* of

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

Data: Initial control $u^{(0)}$, step ρ
Result: optimal control u^*
repeat
$\quad u^{(k+1)} = \mathrm{proj}_{U^{ad}}\big(u^{(k)} - \rho \nabla J(u^{(k)})\big)$ (gradient step w.r.t. u)
until *some convergence criterion is met*;

Algorithm 1: Projected gradient algorithm

[5]that is, $\big\|\nabla J(u) - \nabla J(v)\big\| \leq L \|u - v\|$

Outline of the presentation

Optimization under equality constraints

We consider the optimization problem

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

under the explicit 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 $\Theta_j : \mathbb{R}^N \to \mathbb{R}$, $j = 1, \ldots, M$

Otherwise stated, the admissible set U^{ad} is in this case

$$U^{\mathrm{ad}} = \Big\{ u \in \mathbb{R}^N \mid \Theta(u) = 0 \Big\}$$

Sufficient condition for qualification

Definition

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

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

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

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

First-order necessary optimality conditions

Proposition

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

- the criterion J and the constraints Θ are differentiable
- the equality constraints $\Theta(u) = 0$ are qualified at u^*

Then a necessary condition for u^* to be a local minimum of J over the set $U^{\mathrm{ad}} = \{u \in \mathbb{R}^N \mid \Theta(u) = 0\}$ is that there exists a vector λ^* of $\mathbb{R}^M$, called Lagrange multiplier, such that

$$\nabla J(u^*) + \left[\Theta'(u^*)\right]^\top \lambda^* = 0 \quad \text{and} \quad \Theta(u^*) = 0$$

These first-order optimality conditions are called Karush-Kuhn-Tucker (KKT) optimality conditions (specialized for equality constraints)

First-order sufficient optimality conditions

Proposition

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

- the criterion J is convex and differentiable
- the function Θ is affine

Then a sufficient condition for u^* to be a global minimum of J over the set $U^{\mathrm{ad}} = \big\{ u \in \mathbb{R}^N \mid \Theta(u) = 0 \big\}$ is that there exists a vector λ^* of $\mathbb{R}^M$, called Lagrange multiplier, such that

$$\nabla J(u^*) + \big[\Theta'(u^*)\big]^\top \lambda^* = 0 \quad \text{and} \quad \Theta(u^*) = 0$$

Remark. Using the Lagrangian function $L(u, \lambda) = J(u) + \lambda^\top \Theta(u)$,[6] the conditions above can be written as

$$\begin{aligned} \nabla_u L(u^*, \lambda^*) &= \nabla J(u^*) + \big[\Theta'(u^*)\big]^\top \lambda^* = 0 \\ \nabla_\lambda L(u^*, \lambda^*) &= \Theta(u^*) = 0 \end{aligned}$$

The first-order optimality conditions express the stationarity of the Lagrangian

[6] introduced in the next part of the course

Outline of the presentation

Optimization problems, convex functions, local and global minima

First-order optimality conditions

Lagrangian duality and Uzawa algorithm

Outline of the presentation

Duality gap

Consider a function $\phi : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ and two sets $X \subset \mathcal{X}$ and $Y \subset \mathcal{Y}$

We minimize in the variable x and maximize in the variable y

Definition

$$\inf_{x \in X} \sup_{y \in Y} \phi(x, y) - \sup_{y \in Y} \inf_{x \in X} \phi(x, y)$$

is called the duality gap

Proposition

The duality gap is always positive, that is,

$$\inf_{x \in X} \sup_{y \in Y} \phi(x, y) \geq \sup_{y \in Y} \inf_{x \in X} \phi(x, y)$$

*Exercise**

Give a proof of the proposition

Saddle-point

Consider a function $\phi : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ and two sets $X \subset \mathcal{X}$ and $Y \subset \mathcal{Y}$

Definition

We say that $(\bar{x}, \bar{y}) \in X \times Y$ is a saddle-point of ϕ on $X \times Y$ if

- $y \mapsto \phi(\bar{x}, y)$ achieves its maximum on Y at $\bar{y}$
- $x \mapsto \phi(x, \bar{y})$ achieves its minimum on X at $\bar{x}$

or, equivalently

$$\phi(x, \bar{y}) \geq \phi(\bar{x}, \bar{y}) \geq \phi(\bar{x}, y) \ , \ \ \forall (x, y) \in X \times Y$$

Proposition

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

*Exercise**

Give a proof of the proposition

Existence of a saddle point

Consider a function $\phi : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ and two sets $X \subset \mathcal{X}$ and $Y \subset \mathcal{Y}$

Proposition

Suppose that function ϕ is

- continuous
- convex in the variable x
- concave in the variable y

and that X and Y are convex closed sets such that

- there exists a $\hat{y} \in Y$ such that $\lim_{\|x\| \to +\infty} \phi(x, \hat{y}) = +\infty$, or the set X is bounded
- there exists a $\hat{x} \in X$ such that $\lim_{\|y\| \to +\infty} \phi(\hat{x}, y) = -\infty$, or the set Y is bounded

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

Outline of the presentation

Optimization under equality constraints

We consider the **original** optimization problem

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

under the equality constraint

$$\Theta(u) = 0$$

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

Lagrangian

Definition

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

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

The variables $\lambda \in \mathbb{R}^M$ are called (Lagrange) multipliers

Primal problem

Definition

The **primal** optimization problem is

$$\inf_{u \in \mathbb{R}^N} \sup_{\lambda \in \mathbb{R}^M} L(u, \lambda) = \inf_{u \in \mathbb{R}^N} \sup_{\lambda \in \mathbb{R}^M} \Big(J(u) + \sum_{j=1}^{M} \lambda_j \Theta_j(u) \Big)$$

Proposition

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

*Exercise**

Give a proof of the proposition

Dual problem

Definition

The **dual** optimization problem is

$$\sup_{\lambda \in \mathbb{R}^M} \inf_{u \in \mathbb{R}^N} L(u, \lambda) = \sup_{\lambda \in \mathbb{R}^M} \inf_{u \in \mathbb{R}^N} \left(J(u) + \sum_{j=1}^{M} \lambda_j \Theta_j(u) \right)$$

Definition

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

$$\psi(\lambda) = \inf_{u \in \mathbb{R}^N} L(u, \lambda) \tag{1}$$

hence is concave

Proposition

When there exists a saddle-point for the Lagrangian,
primal and dual problems are equivalent (no duality gap)

First-order optimality conditions and saddle point

Proposition

We suppose that

- the criterion J is differentiable and convex
- the function Θ is affine

Let $u^* \in \mathbb{R}^N$ be a minimum of J on the set $\{u \in \mathbb{R}^N \mid \Theta(u) = 0\}$
Then, there exists a vector λ^* of $\mathbb{R}^M$ (Lagrange multiplier) such that

- $u \mapsto L(u, \lambda^*)$ achieves a minimum at u^* over $\mathbb{R}^N$
- $\lambda \mapsto L(u^*, \lambda)$ achieves a maximum at λ^* over $\mathbb{R}^M$

that is, (u^*, λ^*) is a saddle point of the Lagrangian L

These two conditions are equivalent to

$$\nabla_u L(u^*, \lambda^*) = 0 = \nabla J(u^*) + \left[\Theta'(u^*)\right]^\top \lambda^*$$
$$\nabla_\lambda L(u^*, \lambda^*) = 0 = \Theta(u^*)$$

Existence of a minimum and of a saddle point

$$\inf_{u \in \mathbb{R}^N} J(u) \quad \text{subject to} \quad \Theta(u) = 0 \in \mathbb{R}^M$$

Proposition

We suppose that

- the criterion J is differentiable and a-strongly convex
- the function Θ is affine

Then

- there exists a unique minimum $u^* \in \mathbb{R}^N$ of J on the set $\{u \in \mathbb{R}^N \mid \Theta(u) = 0\}$
- and there exists a vector λ^* of $\mathbb{R}^M$ (Lagrange multiplier)

such that (u^*, λ^*) is a saddle point of the Lagrangian L

Outline of the presentation

Uzawa algorithm (dual gradient ascent algorithm)

We suppose that

- the criterion J is differentiable and a-strongly convex
- the function Θ is affine, with norm κ

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

$$\inf_{u \in \mathbb{R}^N} J(u) \quad \text{subject to} \quad \Theta(u) = 0$$

Data: Initial multiplier $\lambda^{(0)}$, step ρ, tolerance $\epsilon > 0$
Result: minimum and multiplier;
repeat

$u^{(k)} = \arg\min_{u \in \mathbb{R}^N} L(u, \lambda^{(k)})$ (primal minimization w.r.t. u)

$\lambda^{(k+1)} = \lambda^{(k)} + \rho\Theta(u^{(k)})$ (dual gradient step w.r.t. λ)

until $\left\| \Theta(u^{(k)}) \right\| < \epsilon$;

Algorithm 2: Uzawa algorithm

Uzawa algorithm and basic decomposition mechanism

Consider the optimization problem

$$\inf_{(u,v)\in\mathbb{R}^N\times\mathbb{R}^P} J(u) + G(v) \quad \text{subject to} \quad \Theta(u) + \Psi(v) = 0$$

whose Lagrangian is $L(u, v, \lambda) = J(u) + G(v) + \big(\Theta(u) + \Psi(v)\big)^\top \lambda$

The primal minimization w.r.t. (u, v) in Uzawa algorithm is

$$\begin{aligned} L(u^{(k)}, v^{(k)}, \lambda^{(k)}) &= \min_{(u,v)\in\mathbb{R}^N\times\mathbb{R}^P} J(u) + G(v) + \big(\Theta(u) + \Psi(v)\big)^\top \lambda^{(k)} \\ &= \underbrace{\min_{u\in\mathbb{R}^N} J(u) + \Theta(u)^\top \lambda^{(k)}}_{\text{subproblem in } u} + \underbrace{\min_{v\in\mathbb{R}^P} G(v) + \Psi(v)^\top \lambda^{(k)}}_{\text{subproblem in } v} \end{aligned}$$

by the interchange formula

The primal minimization problem splits into 2 independent subproblems!

Outline of the presentation

Augmented Lagrangian in case of equality constraints

Definition

Let $r > 0$. The augmented Lagrangian $L_r : \mathbb{R}^N \times \mathbb{R}^M \to \mathbb{R}$ is defined by

$$L_r(u, \lambda) = \max_{q \in \mathbb{R}^M} \left(L(u, q) - \frac{1}{2r} \left\| \lambda - q \right\|^2 \right) = J(u) + \Theta(u)^\top \lambda + \frac{r}{2} \left\| \Theta(u) \right\|^2$$

The associated dual function $\psi_r : \mathbb{R}^M \to \mathbb{R} \cup \{-\infty\}$

$$\psi_r(\lambda) = \inf_{u \in \mathbb{R}^N} L_r(u, \lambda)$$

is the Moreau-Yosida regularization of the dual function ψ in (1)

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

and the Lagrangian L and the augmented Lagrangian L_r have the same set of saddle points, with better mathematical properties for the augmented Lagrangian (stability, differentiability...)

Some references

R. Tyrrell Rockafellar.
Convex Analysis.
Princeton University Press, 1970.

J. Borwein, A. Lewis.
Convex Analysis.
Springer, 2006.

F. Bonnans, J.-C. Gilbert, C. Lemaréchal, C. Sagastizabal.
Numerical Optimization. Theoretical and Practical Aspects.
Springer, 2006.

J. Nocedal, S.J. Wright.
Numerical Optimization.
Springer, 2006.

A. Ruszczynski.
Nonlinear Optimization.
Princeton University Press, 2006.