Optimisation convexe non-lisse

Nelly Pustelnik

ENS Lyon – Laboratoire de Physique – CNRS UMR 5672
nelly.pustelnik@ens-lyon.fr

Optimisation : pourquoi ?

We think that convex optimization is an important enough topic that everyone who uses computational mathematics should know at least a little bit about it. In our opinion, convex optimization is a natural next topic after advanced linear algebra and linear programming.
(Stephen Boyd and Lieven Vandenberghe)

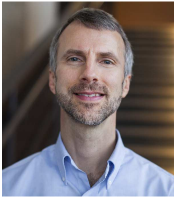

Optimisation : quoi ?

Optimization problems arise naturally in many application fields. Whatever people do, at some point they get a craving to organize things in a best possible way. This intention, converted in a mathematical form, turns out to be an optimization problem of certain type.

(Yurii Nesterov)

Optimisation : convergence ?

We believe that convergence properties are relevant to clinical medical imaging, since algorithm divergence could have unfortunate consequences.
(Jeffrey A. Fessler et Alfred Hero)



Optimization : problème de minimisation

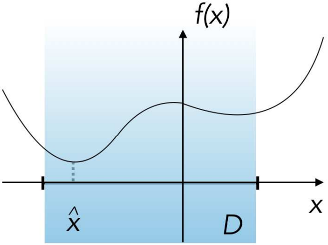
f : fonction de coût

Objectif :

$$\begin{aligned} &\text{Trouver } \widehat{x} \in D \quad \text{tel que} \quad (\forall x \in D)\ f(\widehat{x}) \leq f(x) \\ \Leftrightarrow\ &\text{Trouver } \widehat{x} \in D \quad \text{tel que} \quad f(\widehat{x}) = \inf_{x \in D} f(x) \end{aligned}$$

c'est à dire

$$\boxed{\text{Trouver } \widehat{x} \in \operatorname*{Argmin}_{x \in D} f(x).}$$



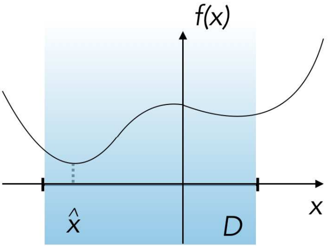
Optimization : problème de maximisation

f : fonction de satisfaction

Objectif :

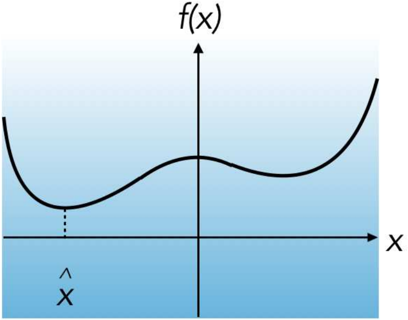
$$
\begin{aligned}
& \text{Trouver } \widehat{x} \in D \quad \text{tel que} \quad (\forall x \in D)\ f(\widehat{x}) \geq f(x) \\
\Leftrightarrow\ & \text{Trouver } \widehat{x} \in D \quad \text{tel que} \quad (\forall x \in D)\ -f(\widehat{x}) \leq -f(x) \\
\Leftrightarrow\ & \text{Trouver } \widehat{x} \in \operatorname*{Argmin}_{x \in D} \big(-f(x)\big).
\end{aligned}
$$

Sans perte de généralité, nous pouvons nous concentrer sur les problèmes de minimisation avec $f\colon D \to]-\infty, +\infty]$.



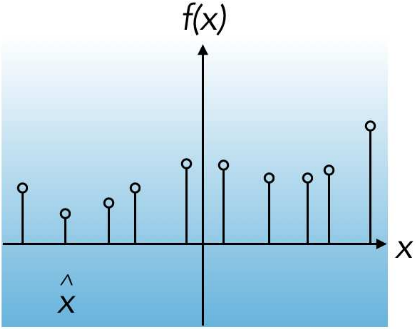
Différents types de problèmes d'optimisation

- $D = \mathbb{R}^N$: problème non contraint



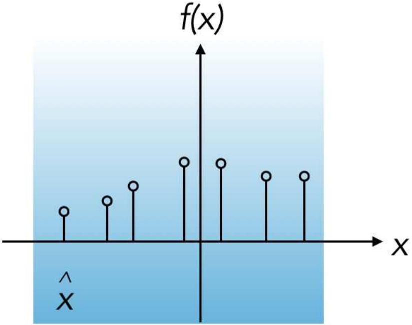
Différents types de problèmes d'optimisation

- $D = \mathbb{R}^N$: problème non contraint
- D dénombrable : problème d'optimisation discret



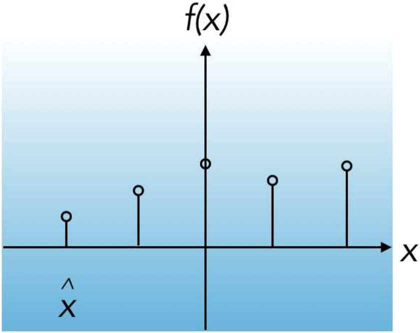
Différents types de problèmes d'optimisation

- $D = \mathbb{R}^N$: problème non contraint
- D dénombrable : problème d'optimisation discret
  - D fini : problème d'optimisation combinatoire



Différents types de problèmes d'optimisation

- $D = \mathbb{R}^N$: problème non contraint
- D dénombrable : problème d'optimisation discret
  - D fini : problème d'optimisation combinatoire
  - $D \subset \mathbb{Z}^N$: problème d'optimisation en nombre entiers

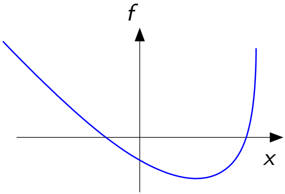


Problème non contraint

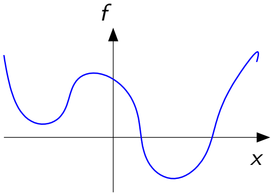
Soit $f : \mathbb{R}^N \to]-\infty, +\infty]$. Un problème d'optimisation non contraint consiste à résoudre :

$$\boxed{\widehat{x} = \arg\min_{x \in \mathbb{R}^N} f(x)}$$

- Optimisation convexe et optimisation non-convexe



Fonction convexe



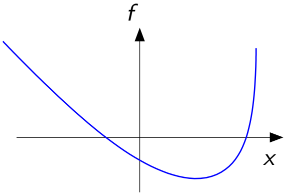
Fonction non−convexe

Problème non contraint

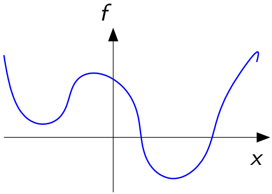
Soit $f : \mathbb{R}^N \to]-\infty, +\infty]$. Un problème d'optimisation non contraint consiste à résoudre :

$$\boxed{\widehat{x} = \arg\min_{x \in \mathbb{R}^N} f(x)}$$

- Optimisation convexe et optimisation non-convexe



Fonction convexe



Fonction non−convexe

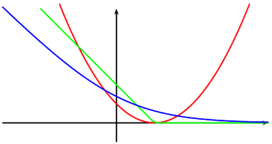
Exemples de fonctions convexes

Cf. cours de Laurent Jacob

Usual loss functions

Classification : $y \in \{0,1\}$

- 0/1 : $L(y_i, f(x_i)) = \mathbf{1}_{y_i f(x_i) \geq 0}$,
- logistic : $L(y_i, f(x_i)) = \log\left(1 + e^{-y_i f(x_i)}\right)$,
- hinge : $L(y_i, f(x_i)) = \max(0, 1 - y_i f(x_i))$.



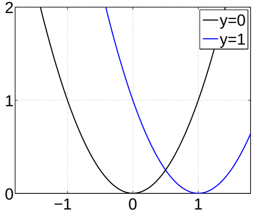
(from J. Mairal's slides)

Other problems: ranking, multi-class, survival...

Exemples de fonctions convexes

- Fonction de perte quadratique avec $y \in \mathbb{R}$

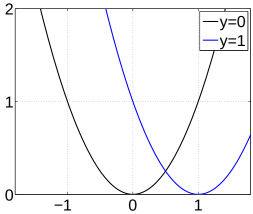
$$(\forall x \in \mathbb{R}) \quad \boxed{f(x) = (y - x)^2}$$



Exemples de fonctions convexes

- Fonction de perte quadratique avec $y \in \mathbb{R}$

$$(\forall x \in \mathbb{R}) \quad \boxed{f(x) = (y - x)^2}$$



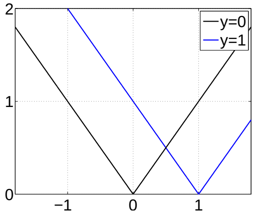
- Généralisation en plus grande dimension avec $y \in \mathbb{R}^M$ et $A \in \mathbb{R}^{M \times N}$.

$$(\forall x \in \mathbb{R}^N) \quad f(x) = \|y - Ax\|^2$$
$$= \sum_{i=1}^{M} \left(y^{(i)} - (Ax)^{(i)}\right)^2$$

Exemples de fonctions convexes

- Fonction de perte norme ℓ_1 avec $y \in \mathbb{R}$

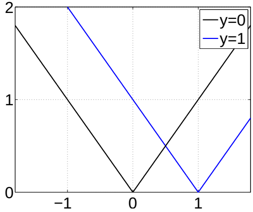
$$(\forall x \in \mathbb{R}) \quad \boxed{f(x) = |y - x|}$$



Exemples de fonctions convexes

- Fonction de perte norme ℓ_1 avec $y \in \mathbb{R}$

$$(\forall x \in \mathbb{R}) \quad \boxed{f(x) = |y - x|}$$



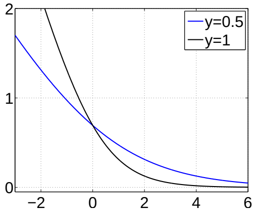
- Généralisation en plus grande dimension avec $y \in \mathbb{R}^M$ et $A \in \mathbb{R}^{M \times N}$.

$$(\forall x \in \mathbb{R}^N) \quad f(x) = \|y - Ax\|_1$$
$$= \sum_{i=1}^{M} |y^{(i)} - (Ax)^{(i)}|$$

Exemples de fonctions convexes

- Fonction logistique avec $y \in \mathbb{R}$

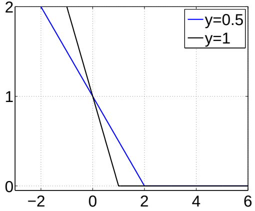
$$(\forall x \in \mathbb{R}) \quad \boxed{f(x) = \log(1 + e^{-yx})}$$



Exemples de fonctions convexes

- Fonction hinge avec $y \in \mathbb{R}$

$$(\forall x \in \mathbb{R}) \quad \boxed{f(x) = \max(0, 1 - yx)}$$



Optimisation convexe ?

Soit $f : \mathbb{R}^N \to]-\infty, +\infty]$ une fonction convexe.
Un problème d'optimisation convexe consiste à résoudre :

$$\boxed{\widehat{x} = \arg\min_{x \in \mathbb{R}^N} f(x)}$$

- $\widehat{x}$ est solution si pour tout $x \in \mathbb{R}^N$, $f(\widehat{x}) \leq f(x)$
- $\widehat{x}$: minimum global non-contraint

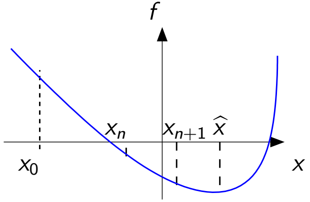
Optimisation convexe ?

Soit $f : \mathbb{R}^N \to]-\infty, +\infty]$ une fonction convexe.
Un problème d'optimisation convexe consiste à résoudre :

$$\boxed{\widehat{x} = \arg\min_{x \in \mathbb{R}^N} f(x)}$$

- $\widehat{x}$ est solution si pour tout $x \in \mathbb{R}^N$, $f(\widehat{x}) \leq f(x)$
- $\widehat{x}$: minimum global non-contraint

Objectif de ce cours : Construire une suite $(x_n)_{n \in \mathbb{Z}}$ qui converge vers $\widehat{x}$.



Une réponse simpliste

Théorème du point fixe (E. Picard, 1856-1941)

Si

- $(\forall n \in \mathbb{N})\ x_{n+1} = Tx_n$ où T opérateur de $\mathbb{R}^N$ vers $\mathbb{R}^N$,
- $\widehat{x}$ est un point fixe de T, i.e. $\widehat{x} = T\widehat{x}$,
- T est une contraction stricte, i.e. il existe $\rho \in [0,1[$ tel que

$$\left(\forall (x,x') \in \mathbb{R}^N \times \mathbb{R}^N\right) \qquad \|Tx - Tx'\| \leq \rho\|x - x'\|$$

alors $(x_n)_{n\in\mathbb{N}}$ converge vers $\widehat{x}$.

Une réponse simpliste

Théorème du point fixe (E. Picard, 1856-1941)

Si

- $(\forall n \in \mathbb{N})\ x_{n+1} = Tx_n$ où T opérateur de $\mathbb{R}^N$ vers $\mathbb{R}^N$,
- $\widehat{x}$ est un point fixe de T, i.e. $\widehat{x} = T\widehat{x}$,
- T est une contraction stricte, i.e. il existe $\rho \in [0,1[$ tel que

$$(\forall (x,x') \in \mathbb{R}^N \times \mathbb{R}^N) \qquad \|Tx - Tx'\| \leq \rho \|x - x'\|$$

alors $(x_n)_{n\in\mathbb{N}}$ converge vers $\widehat{x}$.

Preuve : Pour tout $n \in \mathbb{N}$,

$$\|x_{n+1} - \widehat{x}\| = \|Tx_n - T\widehat{x}\| \leq \rho \|x_n - \widehat{x}\|.$$

D'où $\|x_n - \widehat{x}\| \leq \rho^n \|x_0 - \widehat{x}\|$. Ceci montre que $(x_n)_{n\in\mathbb{N}}$ converge linéairement vers $\widehat{x}$.

Une réponse simpliste

Théorème du point fixe (E. Picard, 1856-1941)

Si

- $(\forall n \in \mathbb{N})\ x_{n+1} = Tx_n$ où T opérateur de $\mathbb{R}^N$ vers $\mathbb{R}^N$,
- $\widehat{x}$ est un point fixe de T, i.e. $\widehat{x} = T\widehat{x}$,
- T est une contraction stricte, i.e. il existe $\rho \in [0,1[$ tel que

$$(\forall (x,x') \in \mathbb{R}^N \times \mathbb{R}^N) \qquad \|Tx - Tx'\| \leq \rho \|x - x'\|$$

alors $(x_n)_{n\in\mathbb{N}}$ converge vers $\widehat{x}$.

- Objectif de ce cours : Comment construire T à partir de la fonction f ?

Pourquoi aller au delà ?

Difficultés :

- Il est difficile (voire parfois impossible) d'avoir un opérateur T *strictement* contractant.
- On peut préférer une récurrence du type $(\forall n \in \mathbb{N})\ x_{n+1} = T_n x_n$ où T_n est un opérateur de $\mathbb{R}^N$ vers $\mathbb{R}^N$.
- Souvent, construire T_n est complexe et il faut donc écrire T_n comme une composition d'opérateurs plus simples (*éclatement du problème*).

Philosophie du cours

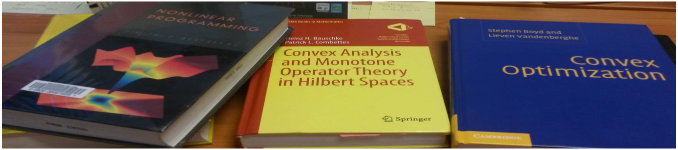
- Fournir une vision moderne de l'optimisation convexe permettant d'appréhender des problèmes non lisses (parcimonie)
  $\rightarrow$ fonctions non finies, opérateurs monotones,...

- Servir d'introduction à la littérature souvent technique sur le sujet
  $\rightarrow$ se placer dans des espaces de Hilbert de dimension infinie... même si la plupart des applications sont en dimension finie.

Philosophie du cours

- Essayer de donner une idée des raisonnements mais...
  sauter les démonstrations trop faciles ... ou celles trop longues.
- Mettre l'accent sur la convergence des itérées $(x_n)_{n\in\mathbb{N}}$ plutôt que sur celle du critère $\left(f(x_n)+g(x_n)\right)_{n\in\mathbb{N}}$.
- Mise en oeuvre de certains algorithmes en TP

Les refs



- **D. Bertsekas**, Nonlinear programming, Athena Scientic, Belmont, Massachussets, 1995.
- **Y. Nesterov**, Introductory Lectures on Convex Optimization : A Basic Course, Springer, 2004.
- **S. Boyd and L. Vandenberghe**, Convex optimization, Cambridge University Press, 2004.
- **H. H. Bauschke and P. L. Combettes**, Convex Analysis and Monotone Operator Theory in Hilbert Spaces, Springer, New York, 2011.

Plan du cours

1. (Rappels) sur les fonctions convexes différentiables

   → définitions, existence et unicité, conditions d'optimalité

2. Fonctions convexes non-différentiables

   → sous-différentielle, conjuguée, opérateur proximal

3. Algorithme de points fixes

   → convergence, opérateur contractant, opérateur α-moyenné, *Forward-Backward*

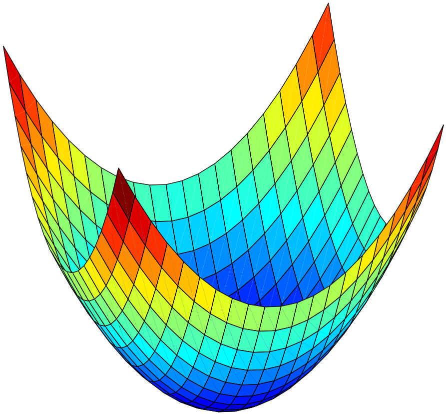
4. Conclusion

Une petite citation

Les mathématiques sont un jeu que l'on exerce selon des règles simples en manipulant des symboles ou des concepts qui n'ont en soi, aucune importance particulière.

(D. Hilbert, 1862-1943)

(Rappels) sur les fonctions convexes différentiables



Notions de base sur les espaces de Hilbert

Un espace de Hilbert $\mathcal{H}$ (réel) est un espace vectoriel réel muni d'un produit scalaire $\langle \cdot \mid \cdot \rangle$ et de la norme associée

$$(\forall x \in \mathcal{H}) \qquad \|x\| = \sqrt{\langle x \mid x \rangle}$$

qui est complet.

- Cas particulier $\mathcal{H} = \mathbb{R}^N$ (espace euclidien de dimension N).

Notions de base sur les espaces de Hilbert

Un espace de Hilbert $\mathcal{H}$ (réel) est un espace vectoriel réel muni d'un produit scalaire $\langle \cdot \mid \cdot \rangle$ et de la norme associée

$$(\forall x \in \mathcal{H}) \qquad \|x\| = \sqrt{\langle x \mid x \rangle}$$

qui est complet.

- Cas particulier $\mathcal{H} = \mathbb{R}^N$ (espace euclidien de dimension N).

$2^{\mathcal{H}}$ est l'ensemble des parties de $\mathcal{H}$, i.e. l'ensemble de tous les sous-ensembles de $\mathcal{H}$.

Définitions

Soient $\mathcal{H}$ et $\mathcal{G}$ des espaces de Hilbert. Un opérateur linéaire $L\colon \mathcal{H} \to \mathcal{G}$ est borné (ou continu) si

$$\|L\| = \sup_{x \neq 0} \frac{\|Lx\|_{\mathcal{G}}}{\|x\|_{\mathcal{H}}} < +\infty$$

Définitions

Soient $\mathcal{H}$ et $\mathcal{G}$ des espaces de Hilbert. Un opérateur linéaire $L\colon \mathcal{H} \to \mathcal{G}$ est **borné** (ou continu) si

$$\|L\| = \sup_{x \neq 0} \frac{\|Lx\|}{\|x\|} < +\infty$$

Définitions

Soient $\mathcal{H}$ et $\mathcal{G}$ des espaces de Hilbert. Un opérateur linéaire $L\colon \mathcal{H} \to \mathcal{G}$ est borné (ou continu) si

$$\|L\| = \sup_{x \neq 0} \frac{\|Lx\|}{\|x\|} < +\infty$$

- En dimension finie, tout opérateur linéaire est borné.

Définitions

Soient $\mathcal{H}$ et $\mathcal{G}$ des espaces de Hilbert. Un opérateur linéaire $L\colon \mathcal{H} \to \mathcal{G}$ est **borné** (ou continu) si

$$\|L\| = \sup_{x \neq 0} \frac{\|Lx\|}{\|x\|} < +\infty$$

- En dimension finie, tout opérateur linéaire est borné.

$\mathcal{B}(\mathcal{H}, \mathcal{G})$: espace normé des opérateurs linéaires bornés de $\mathcal{H}$ vers $\mathcal{G}$.

Définitions

Soit $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$. Son **adjoint L^*** est l'opérateur de $\mathcal{B}(\mathcal{G}, \mathcal{H})$ défini par

$$(\forall (x, y) \in \mathcal{G} \times \mathcal{H}) \qquad \langle x \mid Ly \rangle_{\mathcal{G}} = \langle L^* x \mid y \rangle_{\mathcal{H}} \,.$$

Définitions

Soit $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$. Son adjoint L^* est l'opérateur de $\mathcal{B}(\mathcal{G}, \mathcal{H})$ défini par

$$(\forall (x, y) \in \mathcal{G} \times \mathcal{H}) \qquad \langle x \mid Ly \rangle = \langle L^* x \mid y \rangle .$$

Définitions

Soit $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$. Son adjoint L^* est l'opérateur de $\mathcal{B}(\mathcal{G}, \mathcal{H})$ défini par

$$(\forall (x, y) \in \mathcal{G} \times \mathcal{H}) \qquad \langle Ly \mid x \rangle = \langle y \mid L^* x \rangle \,.$$

Définitions

Soit $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$. Son adjoint L^* est l'opérateur de $\mathcal{B}(\mathcal{G}, \mathcal{H})$ défini par

$$(\forall (x, y) \in \mathcal{G} \times \mathcal{H}) \qquad \langle Ly \mid x \rangle = \langle y \mid L^* x \rangle .$$

Exemple :

Si $\qquad L\colon \mathcal{H} \to \mathcal{H}^n\colon y \mapsto (y, \ldots, y)$

alors $\qquad L^*\colon \mathcal{H}^n \to \mathcal{H}\colon x = (x_1, \ldots, x_n) \mapsto \sum_{i=1}^n x_i$

Preuve :

$$\langle Ly \mid x \rangle = \langle (y, \ldots, y) \mid (x_1, \ldots, x_n) \rangle = \sum_{i=1}^n \langle y \mid x_i \rangle = \left\langle y \;\middle|\; \sum_{i=1}^n x_i \right\rangle$$

Définitions

Soit $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$. Son adjoint L^* est l'opérateur de $\mathcal{B}(\mathcal{G}, \mathcal{H})$ défini par

$$(\forall (x, y) \in \mathcal{G} \times \mathcal{H}) \qquad \langle Ly \mid x \rangle = \langle y \mid L^* x \rangle \,.$$

- On a $\|L^*\| = \|L\|$.
- Si L est bijective (i.e. un isomorphisme) alors $L^{-1} \in \mathcal{B}(\mathcal{G}, \mathcal{H})$ et $(L^{-1})^* = (L^*)^{-1}$.
- Si $\mathcal{H} = \mathbb{R}^N$ et $\mathcal{G} = \mathbb{R}^M$ alors $L^* = L^\top$.

Définitions

Soit $f : \mathcal{H} \to \,]-\infty, +\infty]$ où $\mathcal{H}$ est un espace de Hilbert.

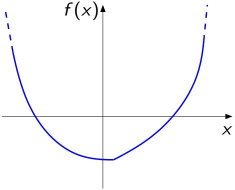
- Le domaine de f est $\operatorname{dom} f = \{x \in \mathcal{H} \mid f(x) < +\infty\}$.
- La fonction f est propre si $\operatorname{dom} f \neq \varnothing$.

Définitions

Soit $f : \mathcal{H} \to \,]-\infty, +\infty]$ où $\mathcal{H}$ est un espace de Hilbert.

- Le domaine de f est $\operatorname{dom} f = \{x \in \mathcal{H} \,|\, f(x) < +\infty\}$.
- La fonction f est propre si $\operatorname{dom} f \neq \varnothing$.

Quel est le domaine de cette fonction ?

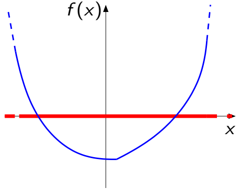


Définitions

Soit $f : \mathcal{H} \to \left]-\infty, +\infty\right]$ où $\mathcal{H}$ est un espace de Hilbert.

- Le domaine de f est $\operatorname{dom} f = \{x \in \mathcal{H} \mid f(x) < +\infty\}$.
- La fonction f est propre si $\operatorname{dom} f \neq \varnothing$.

Quel est le domaine de cette fonction ?



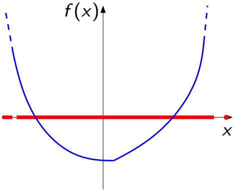
$\operatorname{dom} f = \mathbb{R}$

Définitions

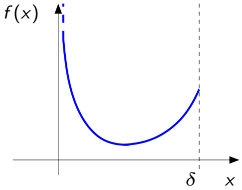
Soit $f : \mathcal{H} \to]-\infty, +\infty]$ où $\mathcal{H}$ est un espace de Hilbert.

- Le domaine de f est $\operatorname{dom} f = \{x \in \mathcal{H} \mid f(x) < +\infty\}$.
- La fonction f est propre si $\operatorname{dom} f \neq \varnothing$.

Quel est le domaine de cette fonction ?



$\operatorname{dom} f = \mathbb{R}$

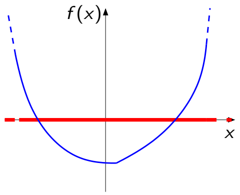


Définitions

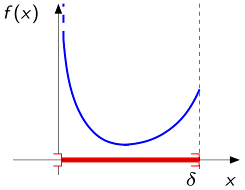
Soit $f : \mathcal{H} \to]-\infty, +\infty]$ où $\mathcal{H}$ est un espace de Hilbert.

- Le domaine de f est $\mathrm{dom}\, f = \{x \in \mathcal{H} \mid f(x) < +\infty\}$.
- La fonction f est propre si $\mathrm{dom}\, f \neq \varnothing$.

Quel est le domaine de cette fonction ?



$\mathrm{dom}\, f = \mathbb{R}$



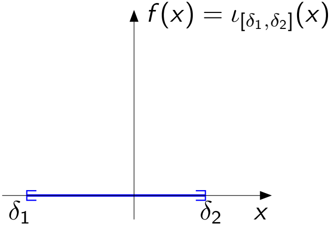
$\mathrm{dom}\, f =]0, \delta]$

Définitions

Soit $C \subset \mathcal{H}$.
La fonction indicatrice de C est

$$(\forall x \in \mathcal{H}) \qquad \iota_C(x) = \begin{cases} 0 & \text{si } x \in C \\ +\infty & \text{sinon.} \end{cases}$$

Exemple : $C = [\delta_1, \delta_2]$



Convergence dans des espaces de Hilbert

Soit $\mathcal{H}$ un espace de Hilbert.
Soit $(x_n)_{n\in\mathbb{N}}$ une suite dans $\mathcal{H}$ et $\widehat{x} \in \mathcal{H}$.

- $(x_n)_{n\in\mathbb{N}}$ **converge fortement** vers $\widehat{x}$ si

$$\lim_{n\to+\infty} \|x_n - \widehat{x}\| = 0.$$

  Elle est notée $x_n \to \widehat{x}$.
- $(x_n)_{n\in\mathbb{N}}$ **converge faiblement** vers $\widehat{x}$ if

$$(\forall y \in \mathcal{H}) \qquad \lim_{n\to+\infty} \langle y \mid x_n - \widehat{x}\rangle = 0.$$

  Elle est notée $x_n \rightharpoonup \widehat{x}$.

Remarque : $x_n \to x \Rightarrow x_n \rightharpoonup x$.
En dimension finie, les convergences forte et faible sont équivalentes.

Définitions

$C \subset \mathcal{H}$ est un ensemble convexe si

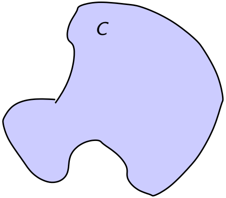
$$(\forall (x,y) \in C^2)(\forall \alpha \in]0,1[) \qquad \alpha x + (1-\alpha) y \in C$$

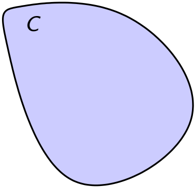
Définitions

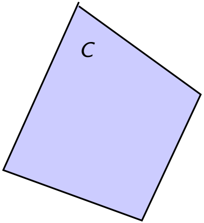
$C \subset \mathcal{H}$ est un **ensemble convexe** si

$$(\forall (x,y) \in C^2)(\forall \alpha \in]0,1[) \qquad \alpha x + (1-\alpha) y \in C$$

Quels sont les ensembles convexes ?





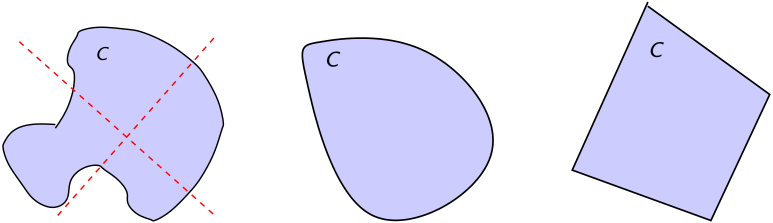


Définitions

$C \subset \mathcal{H}$ est un ensemble convexe si

$$(\forall (x,y) \in C^2)(\forall \alpha \in]0,1[) \qquad \alpha x + (1-\alpha) y \in C$$

Quels sont les ensembles convexes ?



Définitions

$f : \mathcal{H} \to \,]-\infty, +\infty]$ est une **fonction convexe** si

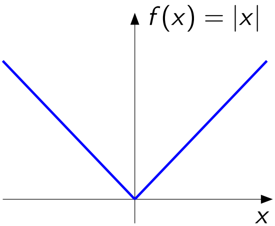
$$\big(\forall (x,y) \in \mathcal{H}^2\big)(\forall \alpha \in [0,1]) \\ f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y)$$

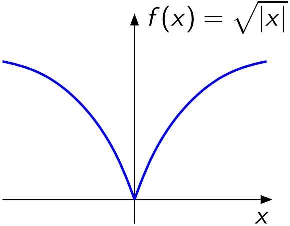
Définitions

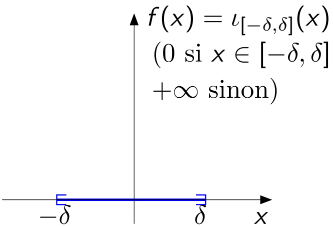
$f : \mathcal{H} \to \,]-\infty, +\infty]$ est une fonction convexe si

$$(\forall (x,y) \in \mathcal{H}^2)(\forall \alpha \in [0,1]) \quad f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y)$$

Quelles sont les fonctions convexes ?





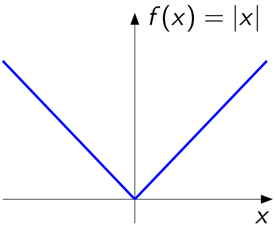


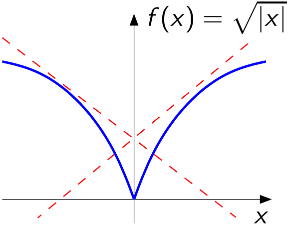
Définitions

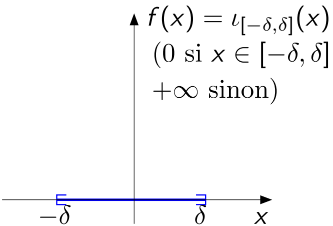
$f : \mathcal{H} \to \;]-\infty, +\infty]$ est une **fonction convexe** si

$$\big(\forall (x,y) \in \mathcal{H}^2\big)(\forall \alpha \in [0,1])$$
$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y)$$

Quelles sont les fonctions convexes ?







Définitions

$f : \mathcal{H} \to \left]-\infty, +\infty\right]$ est convexe ssi son épigraphe

$$\operatorname{epi} f = \left\{ (x, \zeta) \in \operatorname{dom} f \times \mathbb{R} \;\middle|\; f(x) \leq \zeta \right\}$$

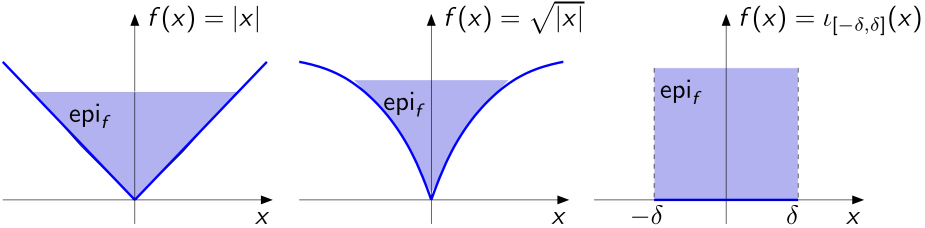
est convexe.

Définitions

$f : \mathcal{H} \to \,]-\infty, +\infty]$ est convexe ssi son épigraphe

$$\operatorname{epi} f = \left\{(x, \zeta) \in \operatorname{dom} f \times \mathbb{R} \;\middle|\; f(x) \leq \zeta\right\}$$

est convexe.

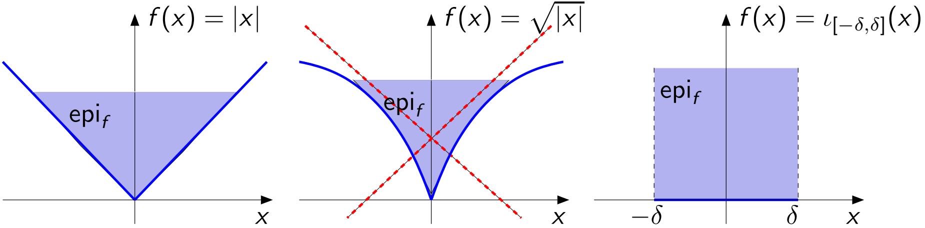


Définitions

$f : \mathcal{H} \to]-\infty, +\infty]$ est convexe ssi son épigraphe

$$\operatorname{epi} f = \left\{(x, \zeta) \in \operatorname{dom} f \times \mathbb{R} \;\middle|\; f(x) \leq \zeta\right\}$$

est convexe.

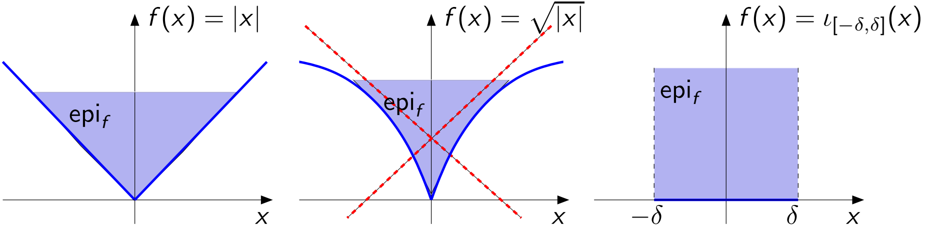


Définitions

$f : \mathcal{H} \to \left]-\infty, +\infty\right]$ est convexe ssi son épigraphe

$$\operatorname{epi} f = \left\{ (x, \zeta) \in \operatorname{dom} f \times \mathbb{R} \;\middle|\; f(x) \leq \zeta \right\}$$

est convexe.



- $f : \mathcal{H} \to \left[-\infty, +\infty\right[$ est concave si $-f$ est convexe.

Définitions

Let $f : \mathcal{H} \to \,]-\infty, +\infty]$.
f is a lower semi-continuous (l.s.c.) function at $x \in \mathcal{H}$ if, for every sequence $(x_n)_{n\in\mathbb{N}}$ of $\mathcal{H}$,

$$x_n \to x \quad \Rightarrow \quad \liminf f(x_n) \geq f(x).$$

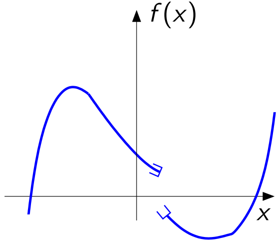
Définitions

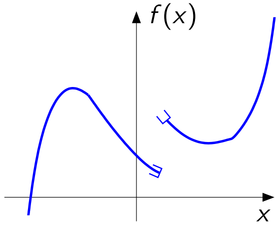
Let $f : \mathcal{H} \to \,]-\infty, +\infty]$.
f is a lower semi-continuous function on $\mathcal{H}$ if and only if $\operatorname{epi} f$ is closed

Définitions

Let $f : \mathcal{H} \to \,]-\infty, +\infty]$.
f is a lower semi-continuous function on $\mathcal{H}$ if and only if epi f is closed

Quelle est la fonction s.c.i ?

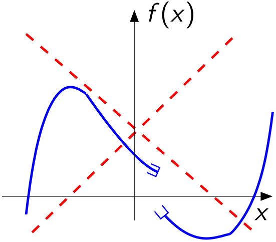


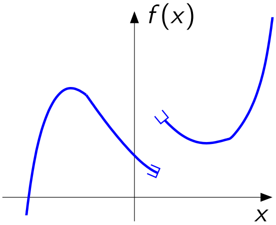


Définitions

Let $f : \mathcal{H} \to \,]-\infty, +\infty]$.
f is a lower semi-continuous function on $\mathcal{H}$ if and only if epi f is closed

Quelle est la fonction s.c.i ?

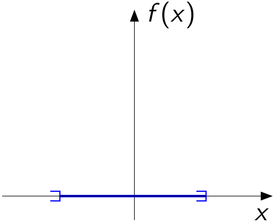


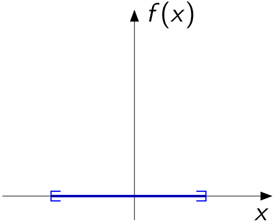


Définitions

Let $f : \mathcal{H} \to \,]-\infty, +\infty]$.
f is a lower semi-continuous function on $\mathcal{H}$ if and only if epi f is closed

Quelle est la fonction s.c.i ?

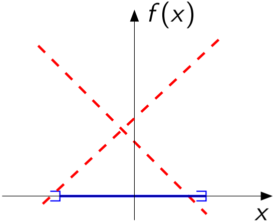


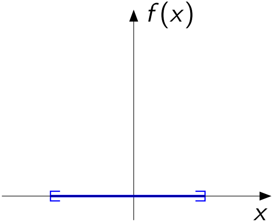


Définitions

Let $f : \mathcal{H} \to]-\infty, +\infty]$.
f is a lower semi-continuous function on $\mathcal{H}$ if and only if epi f is closed

Quelle est la fonction s.c.i ?





Propriétés

- Toute fonction continue sur $\mathcal{H}$ est s.c.i.
- $f : \mathcal{H} \to]-\infty, +\infty]$ est s.c.i. ssi son épigraphe est fermé.
- Toute somme finie de fonctions s.c.i. (convexe) est s.c.i. (convexe).
- L'ensemble des fonctions convexes, s.c.i. et propres est noté $\Gamma_0(\mathcal{H})$.
- $\iota_C \in \Gamma_0(\mathcal{H})$ ssi C est un convexe fermé non vide.
  Preuve : $\mathrm{epi}_{\iota_C} = C \times [0, +\infty[$.

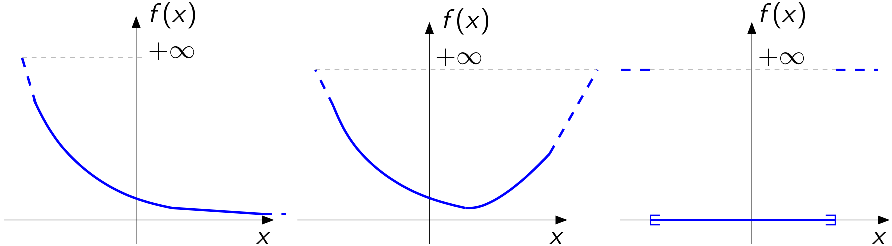
Existence et unicité du minimiseur

Soit $\mathcal{H}$ un espace de Hilbert. Soit $f : \mathcal{H} \to \ ]-\infty, +\infty]$.
f est coercive si $\lim_{\|x\| \to +\infty} f(x) = +\infty$.

Existence et unicité du minimiseur

Soit $\mathcal{H}$ un espace de Hilbert. Soit $f : \mathcal{H} \to\ ]-\infty, +\infty]$.
f est coercive si $\lim_{\|x\| \to +\infty} f(x) = +\infty$.

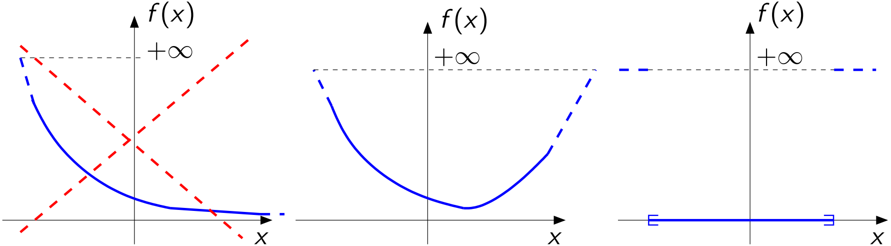
Quelles sont les fonctions coercives ?



Existence et unicité du minimiseur

Soit $\mathcal{H}$ un espace de Hilbert. Soit $f : \mathcal{H} \to]-\infty, +\infty]$.
f est coercive si $\lim_{\|x\|\to+\infty} f(x) = +\infty$.

Quelles sont les fonctions coercives ?



Existence et unicité du minimiseur

Soient $\mathcal{H}$ un espace de Hilbert. Soit $f\colon \mathcal{H} \to \,]-\infty, +\infty]$.
f est **strictement convexe** si

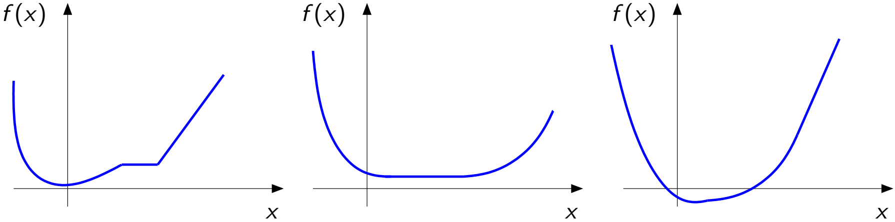
$$(\forall x \in \operatorname{dom} f)(\forall y \in \operatorname{dom} f)(\forall \alpha \in]0,1[)$$
$$x \neq y \quad \Rightarrow \quad f(\alpha x + (1-\alpha)y) < \alpha f(x) + (1-\alpha) f(y).$$

Existence et unicité du minimiseur

Soient $\mathcal{H}$ un espace de Hilbert. Soit $f\colon \mathcal{H} \to\ ]-\infty,+\infty]$.
f est strictement convexe si

$$(\forall x \in \operatorname{dom} f)(\forall y \in \operatorname{dom} f)(\forall \alpha \in]0,1[)$$
$$x \neq y \quad \Rightarrow \quad f(\alpha x + (1-\alpha)y) < \alpha f(x) + (1-\alpha) f(y).$$

Quelles sont les fonctions strictement convexes ?

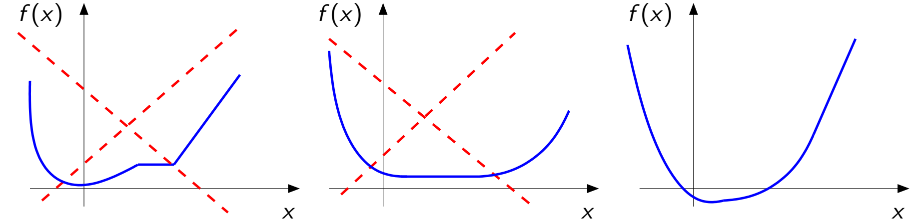


Existence et unicité du minimiseur

Soient $\mathcal{H}$ un espace de Hilbert. Soit $f\colon \mathcal{H} \to \,]-\infty,+\infty]$.
f est **strictement convexe** si

$$(\forall x \in \operatorname{dom} f)(\forall y \in \operatorname{dom} f)(\forall \alpha \in]0,1[)$$
$$x \neq y \quad \Rightarrow \quad f(\alpha x + (1-\alpha)y) < \alpha f(x) + (1-\alpha) f(y).$$

Quelles sont les fonctions strictement convexes ?



Existence et unicité du minimiseur

Soient $\mathcal{H}$ un espace de Hilbert et C un convexe fermé de $\mathcal{H}$. Soit $f \in \Gamma_0(\mathcal{H})$ tel que $\operatorname{dom} f \cap C \neq \varnothing$.
Si f est coercive ou C est borné alors il existe $p \in C$ tel que

$$f(p) = \inf_{x \in C} f(x).$$

Si, de plus, f est strictement convexe, ce minimiseur p est unique.

Conditions d'optimalité

Condition nécessaire et suffisante du 1er ordre (P. Fermat, 160X-1665)

Soit $f \in \Gamma_0(\mathbb{R}^N)$ une fonction continuement différentiable sur $\mathbb{R}^N$. $\widehat{x}$ est un minimiseur global de f, i.e., $\widehat{x} \in \mathrm{Argmin}_{x \in \mathbb{R}^N} f(x)$, ssi

$$\nabla f(\widehat{x}) = 0.$$

Conditions d'optimalité

Condition nécessaire et suffisante du 1er ordre (P. Fermat, 160X-1665)

Soit $f \in \Gamma_0(\mathbb{R}^N)$ une fonction continuement différentiable sur $\mathbb{R}^N$. $\widehat{x}$ est un minimiseur global de f, i.e., $\widehat{x} \in \mathrm{Argmin}_{x \in \mathbb{R}^N} f(x)$, ssi

$$\nabla f(\widehat{x}) = 0.$$

Preuve ($\Rightarrow$) : Soit $\epsilon \in \mathbb{R}^N$. On pose, pour tout $\alpha \in \mathbb{R}$, $g(\alpha) = f(\widehat{x} + \alpha\epsilon)$

Conditions d'optimalité

Condition nécessaire et suffisante du 1er ordre (P. Fermat, 160X-1665)

Soit $f \in \Gamma_0(\mathbb{R}^N)$ une fonction continuement différentiable sur $\mathbb{R}^N$. $\widehat{x}$ est un minimiseur global de f, i.e., $\widehat{x} \in \mathrm{Argmin}_{x\in\mathbb{R}^N} f(x)$, ssi

$$\nabla f(\widehat{x}) = 0.$$

Preuve $(\Rightarrow)$: Soit $\epsilon \in \mathbb{R}^N$. On pose, pour tout $\alpha \in \mathbb{R}$, $g(\alpha) = f(\widehat{x} + \alpha\epsilon)$

$$\frac{dg(\alpha)}{d\alpha} = \epsilon' \nabla f(\widehat{x} + \alpha\epsilon)$$

$$\frac{dg(0)}{d\alpha} = \epsilon^\top \nabla f(\widehat{x})$$

Conditions d'optimalité

Condition nécessaire et suffisante du 1er ordre (P. Fermat, 160X-1665)

Soit $f \in \Gamma_0(\mathbb{R}^N)$ une fonction continuement différentiable sur $\mathbb{R}^N$. $\widehat{x}$ est un minimiseur global de f, i.e., $\widehat{x} \in \mathrm{Argmin}_{x\in\mathbb{R}^N} f(x)$, ssi

$$\nabla f(\widehat{x}) = 0.$$

<u>Preuve</u> $(\Rightarrow)$: Soit $\epsilon \in \mathbb{R}^N$. On pose, pour tout $\alpha \in \mathbb{R}$, $g(\alpha) = f(\widehat{x} + \alpha\epsilon)$

$$\begin{aligned}\frac{dg(0)}{d\alpha} &= \epsilon^\top \nabla f(\widehat{x}) \\ &= \lim_{\alpha\to 0} \frac{f(\widehat{x}+\alpha\epsilon) - f(\widehat{x})}{\alpha}\end{aligned}$$

Conditions d'optimalité

Condition nécessaire et suffisante du 1er ordre (P. Fermat, 160X-1665)

Soit $f \in \Gamma_0(\mathbb{R}^N)$ une fonction continuement différentiable sur $\mathbb{R}^N$. $\widehat{x}$ est un minimiseur global de f, i.e., $\widehat{x} \in \mathrm{Argmin}_{x \in \mathbb{R}^N} f(x)$, ssi

$$\nabla f(\widehat{x}) = 0.$$

<u>Preuve</u> $(\Rightarrow)$: Soit $\epsilon \in \mathbb{R}^N$. On pose, pour tout $\alpha \in \mathbb{R}$, $g(\alpha) = f(\widehat{x} + \alpha\epsilon)$

$$\begin{aligned} \frac{dg(0)}{d\alpha} &= \epsilon^\top \nabla f(\widehat{x}) \\ &= \lim_{\alpha \to 0} \frac{f(\widehat{x} + \alpha\epsilon) - f(\widehat{x})}{\alpha} \geq 0 \end{aligned}$$

$\rightarrow$ Car $\widehat{x}$ est un minimiseur de f

Conditions d'optimalité

Condition nécessaire et suffisante du 1er ordre (P. Fermat, 160X-1665)

Soit $f \in \Gamma_0(\mathbb{R}^N)$ une fonction continuement différentiable sur $\mathbb{R}^N$. $\widehat{x}$ est un minimiseur global de f, i.e., $\widehat{x} \in \operatorname{Argmin}_{x \in \mathbb{R}^N} f(x)$, ssi

$$\nabla f(\widehat{x}) = 0.$$

<u>Preuve</u> $(\Rightarrow)$: Soit $\epsilon \in \mathbb{R}^N$. On pose, pour tout $\alpha \in \mathbb{R}$, $g(\alpha) = f(\widehat{x} + \alpha\epsilon)$

$$\begin{aligned}\frac{dg(0)}{d\alpha} &= \epsilon^\top \nabla f(\widehat{x}) \\ &= \lim_{\alpha \to 0} \frac{f(\widehat{x} + \alpha\epsilon) - f(\widehat{x})}{\alpha} \geq 0\end{aligned}$$

Donc

$$\boxed{\epsilon^\top \nabla f(\widehat{x}) \ \geq \ 0}$$

Conditions d'optimalité

Condition nécessaire et suffisante du 1er ordre (P. Fermat, 160X-1665)

Soit $f \in \Gamma_0(\mathbb{R}^N)$ une fonction continuement différentiable sur $\mathbb{R}^N$. $\widehat{x}$ est un minimiseur global de f, i.e., $\widehat{x} \in \mathrm{Argmin}_{x \in \mathbb{R}^N} f(x)$, ssi

$$\nabla f(\widehat{x}) = 0.$$

<u>Preuve</u> $(\Rightarrow)$: Soit $\epsilon \in \mathbb{R}^N$. On pose, pour tout $\alpha \in \mathbb{R}$, $g(\alpha) = f(\widehat{x} - \alpha\epsilon)$

$$\begin{aligned}\frac{dg(0)}{d\alpha} &= -\epsilon^\top \nabla f(\widehat{x}) \\ &= \lim_{\alpha \to 0} \frac{f(\widehat{x} - \alpha\epsilon) - f(\widehat{x})}{\alpha} \geq 0\end{aligned}$$

Donc

$$\boxed{\epsilon^\top \nabla f(\widehat{x}) \leq 0}$$

Conditions d'optimalité

Condition nécessaire et suffisante du 1er ordre (P. Fermat, 160X-1665)

Soit $f \in \Gamma_0(\mathbb{R}^N)$ une fonction continuement différentiable sur $\mathbb{R}^N$. $\widehat{x}$ est un minimiseur global de f, i.e., $\widehat{x} \in \mathrm{Argmin}_{x \in \mathbb{R}^N} f(x)$, ssi

$$\nabla f(\widehat{x}) = 0.$$

Preuve ($\Rightarrow$) : Soit $\epsilon \in \mathbb{R}^N$. On pose, pour tout $\alpha \in \mathbb{R}$, $g(\alpha) = f(\widehat{x} - \alpha\epsilon)$

$$\begin{aligned} \frac{dg(0)}{d\alpha} &= \epsilon^\top \nabla f(\widehat{x}) \\ &= \lim_{\alpha \to 0} \frac{f(\widehat{x} - \alpha\epsilon) - f(\widehat{x})}{\alpha} \geq 0 \end{aligned}$$

Donc

$$\boxed{\nabla f(\widehat{x}) = 0}$$

Conditions d'optimalité

Condition nécessaire et suffisante du 1er ordre (P. Fermat, 160X-1665)

Soit $f \in \Gamma_0(\mathbb{R}^N)$ une fonction continuement différentiable sur $\mathbb{R}^N$. $\widehat{x}$ est un minimiseur global de f, i.e., $\widehat{x} \in \mathrm{Argmin}_{x\in\mathbb{R}^N} f(x)$, ssi

$$\nabla f(\widehat{x}) = 0.$$

Preuve $(\Leftarrow)$: f est une fonction convexe donc

$$(\forall (x,z) \in \mathbb{R}^N \times \mathbb{R}^N)(\forall \alpha \in [0,1]) \quad f(\alpha z + (1-\alpha)x) \leq \alpha f(z) + (1-\alpha) f(x)$$

Conditions d'optimalité

Condition nécessaire et suffisante du 1er ordre (P. Fermat, 160X-1665)

Soit $f \in \Gamma_0(\mathbb{R}^N)$ une fonction continuement différentiable sur $\mathbb{R}^N$. $\widehat{x}$ est un minimiseur global de f, i.e., $\widehat{x} \in \mathrm{Argmin}_{x \in \mathbb{R}^N} f(x)$, ssi

$$\nabla f(\widehat{x}) = 0.$$

Preuve ($\Leftarrow$) : f est une fonction convexe donc

$$(\forall (x,z) \in \mathbb{R}^N \times \mathbb{R}^N)(\forall \alpha \in [0,1]) \quad f(x + \alpha(z-x)) \leq \alpha f(z) + (1-\alpha) f(x)$$

Conditions d'optimalité

Condition nécessaire et suffisante du 1er ordre (P. Fermat, 160X-1665)

Soit $f \in \Gamma_0(\mathbb{R}^N)$ une fonction continuement différentiable sur $\mathbb{R}^N$. $\widehat{x}$ est un minimiseur global de f, i.e., $\widehat{x} \in \mathrm{Argmin}_{x\in\mathbb{R}^N} f(x)$, ssi

$$\nabla f(\widehat{x}) = 0.$$

Preuve ($\Leftarrow$) : f est une fonction convexe donc

$$(\forall (x,z) \in \mathbb{R}^N \times \mathbb{R}^N)(\forall \alpha \in [0,1]) \quad \frac{f(x+\alpha(z-x)) - f(x)}{\alpha} \leq f(z) - f(x)$$

Conditions d'optimalité

Condition nécessaire et suffisante du 1er ordre (P. Fermat, 160X-1665)

Soit $f \in \Gamma_0(\mathbb{R}^N)$ une fonction continuement différentiable sur $\mathbb{R}^N$. $\widehat{x}$ est un minimiseur global de f, i.e., $\widehat{x} \in \mathrm{Argmin}_{x \in \mathbb{R}^N} f(x)$, ssi

$$\nabla f(\widehat{x}) = 0.$$

Preuve $(\Leftarrow)$: f est une fonction convexe donc

$$(\forall (x,z) \in \mathbb{R}^N \times \mathbb{R}^N)(\forall \alpha \in [0,1]) \quad \frac{f(x+\alpha(z-x)) - f(x)}{\alpha} \leq f(z) - f(x)$$

Par passage à la limite

$$\lim_{\alpha \to 0} \frac{f(x+\alpha(z-x)) - f(x)}{\alpha} = (z-x)^\top \nabla f(x) \leq f(z) - f(x)$$

Si $\nabla f(\widehat{x}) = 0$, alors

$$(\forall z \in \mathbb{R}^N) \quad f(z) \geq f(\widehat{x})$$

Conditions d'optimalité

Condition nécessaire et suffisante du 1er ordre (P. Fermat, 160X-1665)

Soit $f \in \Gamma_0(\mathbb{R}^N)$ une fonction continuement différentiable sur $\mathbb{R}^N$. $\widehat{x}$ est un minimiseur global de f, i.e., $\widehat{x} \in \mathrm{Argmin}_{x \in \mathbb{R}^N} f(x)$, ssi

$$\nabla f(\widehat{x}) = 0.$$

- Consiste à résoudre un problème de N équations à N inconnues.
- Forme analytique de la solution dans peu de cas.
- S'il n'existe pas de forme analytique alors méthode itérative.
- Résoudre le problème d'optimisation $\widehat{x} \in \mathrm{Argmin}_x f(x)$ est équivalent à trouver une solution de $\nabla f(\widehat{x}) = 0$.

Conditions d'optimalité

- Résolution du problème des *moindres carrés*

Trouver $\widehat{x} \in \mathrm{Argmin}_{x \in \mathbb{R}^N} \|Ax - y\|_2^2$ avec $\begin{cases} A \in \mathbb{R}^{M \times N} \\ y \in \mathbb{R}^M \end{cases}$

→ La condition d'optimalité s'écrit :

$$\nabla f(\widehat{x}) = 0 \quad \Leftrightarrow \quad A^\top (A\widehat{x} - y) = 0$$

$$\boxed{\widehat{x} = (A^\top A)^{-1}(A^\top y)}$$

→ La difficulté réside dans l'inversion de $A^\top A$. Forme explicite connue si A modélise une matrice circulante.

Conditions d'optimalité

- Résolution d'un critère basé sur la fonction logistique

Trouver $\quad \widehat{x} \in \mathrm{Argmin}_{x\in\mathbb{R}} \log\left(1+\exp(-yx)\right) \qquad$ avec $\qquad y \in \mathbb{R}$

$\rightarrow$ La condition d'optimalité s'écrit :

$$\nabla f(\widehat{x}) = 0 \quad \Leftrightarrow \quad \boxed{\frac{-y\exp(-y\widehat{x})}{1+\exp(-y\widehat{x})} = 0}$$

$\rightarrow$ Pas de solution analytique donc nécessité de mettre en œuvre un algorithme itératif.

Algorithme

Descente de gradient

Soit $f \in \Gamma_0(\mathbb{R}^N)$ continuement différentiable sur $\mathbb{R}^N$ et de gradient β-Lipschitz. Soit $x_0 \in \mathbb{R}^N$ et si $\gamma_n \in]0, 2/\beta[$

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n - \gamma_n \nabla f(x_n)$$

alors, la suite $(x_n)_{n\in\mathbb{N}}$ converge vers un minimiseur de f.

- Une méthode itérative consiste à construite une suite $(x_n)_{n\in\mathbb{N}}$ telle que, à chaque itération k
$$f(x_{n+1}) < f(x_n)$$
- Comment choisir γ_n pour converger le plus rapidement possible ?
$\rightarrow$ Steepest descent, méthode de Newton, ...
- Preuve de convergence détaillée plus loin dans le cours.

Conditions d'optimalité

Condition nécessaire et suffisante du 1er ordre

Soit C un sous ensemble convexe fermé non vide de $\mathbb{R}^N$. Soit $f \in \Gamma_0(\mathbb{R}^N)$ une fonction continuement différentiable sur C.
$\widehat{x}$ est un minimiseur de f sur C, i.e, $\widehat{x} \in \mathrm{Argmin}_{x \in C} f(x)$ ssi

$$(\forall x \in C) \quad \nabla f(\widehat{x})^\top (x - \widehat{x}) \geq 0.$$

- Le problème qui nous intéresse ici est :

$$\widehat{x} \in \mathrm{Argmin}_{x \in C} f(x) \quad \Leftrightarrow \quad (\forall x \in C) \quad f(\widehat{x}) \leq f(x)$$

- Lorsqu'un vecteur x satisfait la(es) contrainte(s), on parle de vecteur admissible (*feasible*).

Algorithme

Gradient projeté

Soit C un sous ensemble convexe, fermé, non vide de $\mathbb{R}^N$. Soit $f \in \Gamma_0(\mathbb{R}^N)$ continuement différentiable sur C et de gradient β-Lipschitz.
Soit $x_0 \in C$ et si $\gamma_n \in]0, 2/\beta[$

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = P_C(x_n - \gamma_n \nabla f(x_n))$$

alors, la suite $(x_n)_{n\in\mathbb{N}}$ converge vers un minimiseur de f sur C.

- P_C : opérateur de projection
$$(\forall x \in \mathbb{R}^N) \quad P_C(x) = \arg\min_{z\in C} \|z - x\|_2^2$$
- Soit $C = \{x = (x^{(i)})_{1\leq i\leq N} \in \mathbb{R}^N \,|\, (\forall i \in \{1,\dots,N\}) \;\; x^{(i)} \geq 0\}$, alors
$$P_C(x) = \left(\max(0, x^{(i)})\right)_{1\leq i\leq N}$$

Conditions d'optimalité

On s'intéresse au problème d'optimisation sous contrainte suivant :

$$\widehat{x} \in \operatorname{Argmin}_{x \in \mathbb{R}^N} f_0(x) \quad \text{s.t.} \quad \begin{cases} (\forall i \in \{1, \ldots, m\}) \; f_i(x) \leq 0 \\ (\forall j \in \{1, \ldots, p\}) \; g_j(x) = 0 \end{cases}$$

où,

- pour tout $i \in \{0, \ldots, m\}$, $f_i \in \Gamma_0(\mathbb{R}^N)$,
- pour tout $j \in \{0, \ldots, p\}$, $g_j \in \Gamma_0(\mathbb{R}^N)$,
- $\mathcal{D} = \bigcap_{i=0}^m \operatorname{dom} f_i \cap \bigcap_{j=1}^p \operatorname{dom} g_j \neq \emptyset$.

Lagrangien

$$(\forall (x, \lambda, \nu) \in \mathbb{R}^N \times \mathbb{R}^m \times \mathbb{R}^p) \quad L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{j=1}^p \nu_j g_j(x)$$

Conditions d'optimalité

Lagrangien

$$(\forall (x,\lambda,\nu) \in \mathbb{R}^N \times \mathbb{R}^m \times \mathbb{R}^p) \quad L(x,\lambda,\nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{p} \nu_j g_j(x)$$

- $\operatorname{dom} L = \mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^p$.
- $\lambda = (\lambda_i)_{1\le i\le m}$: multiplicateur de Lagrange associé à $f_i(x) \le 0$.
- $\nu = (\nu_j)_{1\le j\le p}$: multiplicateur de Lagrange associé à $g_j(x) = 0$.
- λ et ν sont appelés les *vecteurs multiplicateur de Lagrange* ou *variables duales*.
- *Fonction duale de Lagrange* :

$$(\forall (\lambda,\nu) \in \mathbb{R}^m \times \mathbb{R}^p) \quad d(\lambda,\nu) = \inf_{x\in\mathcal{D}} L(x,\lambda,\nu)$$

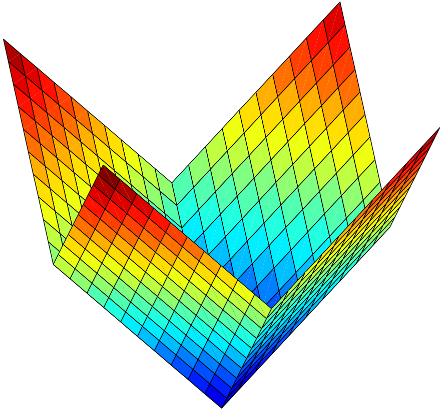
Conditions d'optimalité

Conditions nécessaires et suffisantes : Karush Kuhn Tucker (KKT)

Si $(f_i)_{0\leq i\leq m}$ and $(g_j)_{1\leq j\leq p}$ sont continuement différentiables.
$\widehat{x}$ et $(\widehat{\lambda},\widehat{\nu})$ sont les solutions primales et duales optimales ssi

$$
\begin{aligned}
(\forall i \in \{1,\ldots,m\}) \qquad & f_i(\widehat{x}) \leq 0 \\
(\forall j \in \{1,\ldots,p\}) \qquad & g_j(\widehat{x}) = 0 \\
(\forall i \in \{1,\ldots,m\}) \qquad & \widehat{\lambda}_i \geq 0 \\
(\forall i \in \{1,\ldots,m\}) \qquad & \widehat{\lambda}_i f_i(\widehat{x}) = 0 \\
\nabla f_0(\widehat{x}) + \sum_{i=1}^{m} \widehat{\lambda}_i \nabla f_i(\widehat{x}) + \sum_{j=1}^{p} \widehat{\nu}_j \nabla g_j(\widehat{x}) &= 0
\end{aligned}
$$

Fonctions convexes non-différentiables



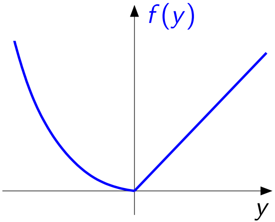
Sous-différentielle d'une fonction convexe : définitions

La sous-différentielle (de Moreau) de f, notée ∂f,

Sous-différentielle d'une fonction convexe : définitions

Soit $f : \mathcal{H} \to]-\infty, +\infty]$ une fonction propre.

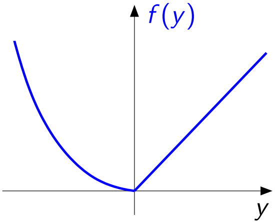
La sous-différentielle (de Moreau) de f, notée ∂f,

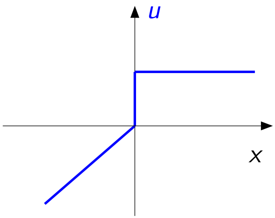


Sous-différentielle d'une fonction convexe : définitions

Soit $f : \mathcal{H} \to]-\infty, +\infty]$ une fonction propre.
La sous-différentielle (de Moreau) de f, notée ∂f, est telle que

$$\begin{aligned} \partial f : \mathcal{H} &\to 2^{\mathcal{H}} \\ x &\to \{u \in \mathcal{H} \,|\, (\forall y \in \mathcal{H})\ \langle y - x | u \rangle + f(x) \leq f(y)\} \end{aligned}$$

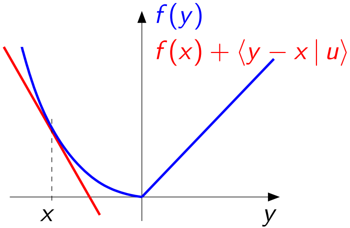


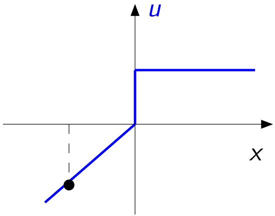


Sous-différentielle d'une fonction convexe : définitions

Soit $f : \mathcal{H} \to \,]-\infty, +\infty]$ une fonction propre.
La **sous-différentielle (de Moreau) de f**, notée ∂f, est telle que

$$\begin{aligned} \partial f : \mathcal{H} &\to 2^{\mathcal{H}} \\ x &\to \{u \in \mathcal{H} \,|\, (\forall y \in \mathcal{H})\ \langle y - x | u \rangle + f(x) \leq f(y)\} \end{aligned}$$

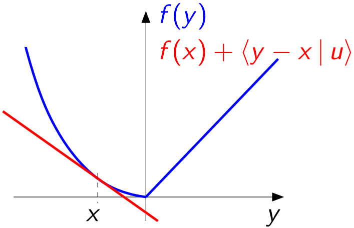


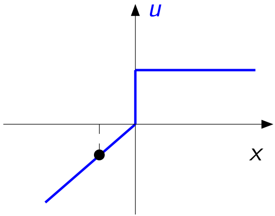


Sous-différentielle d'une fonction convexe : définitions

Soit $f : \mathcal{H} \to]-\infty, +\infty]$ une fonction propre.
La sous-différentielle (de Moreau) de f, notée ∂f, est telle que

$$\begin{aligned} \partial f : \mathcal{H} &\to 2^{\mathcal{H}} \\ x &\to \{u \in \mathcal{H} \,|\, (\forall y \in \mathcal{H})\ \langle y - x | u \rangle + f(x) \le f(y)\} \end{aligned}$$

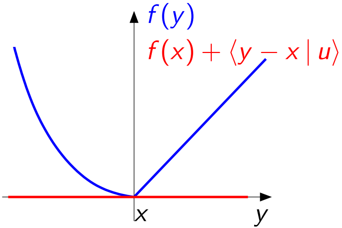


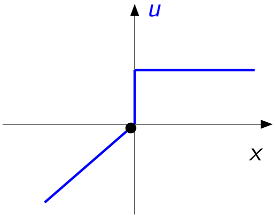


Sous-différentielle d'une fonction convexe : définitions

Soit $f : \mathcal{H} \to \,]-\infty, +\infty]$ une fonction propre.
La sous-différentielle (de Moreau) de f, notée ∂f, est telle que

$$\begin{aligned} \partial f : \mathcal{H} &\to 2^{\mathcal{H}} \\ x &\to \{u \in \mathcal{H} \mid (\forall y \in \mathcal{H})\ \langle y - x | u \rangle + f(x) \leq f(y)\} \end{aligned}$$

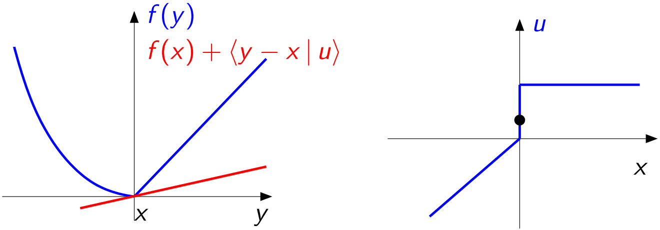




Sous-différentielle d'une fonction convexe : définitions

Soit $f : \mathcal{H} \to]-\infty, +\infty]$ une fonction propre.
La sous-différentielle (de Moreau) de f, notée ∂f, est telle que

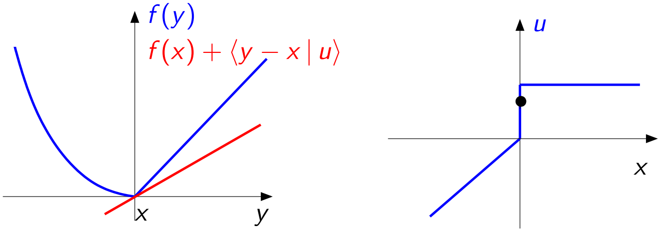
$$\begin{aligned} \partial f : \mathcal{H} &\to 2^{\mathcal{H}} \\ x &\to \{u \in \mathcal{H} \,|\, (\forall y \in \mathcal{H})\ \langle y - x | u \rangle + f(x) \leq f(y)\} \end{aligned}$$



Sous-différentielle d'une fonction convexe : définitions

Soit $f : \mathcal{H} \to]-\infty, +\infty]$ une fonction propre.
La sous-différentielle (de Moreau) de f, notée ∂f, est telle que

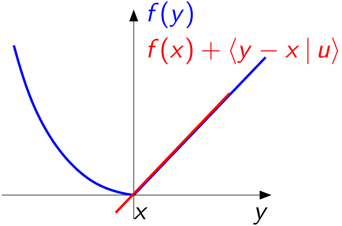
$$\partial f : \mathcal{H} \to 2^{\mathcal{H}}$$
$$x \to \{u \in \mathcal{H} \mid (\forall y \in \mathcal{H})\ \langle y - x | u \rangle + f(x) \leq f(y)\}$$

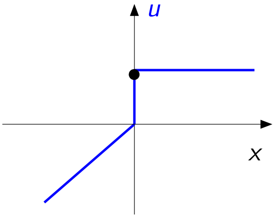


Sous-différentielle d'une fonction convexe : définitions

Soit $f : \mathcal{H} \to]-\infty, +\infty]$ une fonction propre.
La sous-différentielle (de Moreau) de f, notée ∂f, est telle que

$$\begin{aligned}\partial f : \mathcal{H} &\to 2^{\mathcal{H}}\\ x &\to \{u \in \mathcal{H} \,|\, (\forall y \in \mathcal{H})\ \langle y - x | u\rangle + f(x) \leq f(y)\}\end{aligned}$$

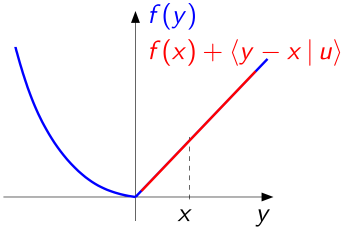


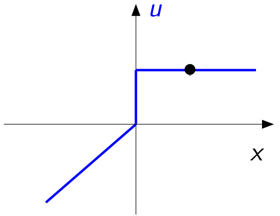


Sous-différentielle d'une fonction convexe : définitions

Soit $f : \mathcal{H} \to]-\infty, +\infty]$ une fonction propre.
La sous-différentielle (de Moreau) de f, notée ∂f, est telle que

$$\begin{aligned} \partial f : \mathcal{H} &\to 2^{\mathcal{H}} \\ x &\to \{u \in \mathcal{H} \,|\, (\forall y \in \mathcal{H})\ \langle y - x | u\rangle + f(x) \leq f(y)\} \end{aligned}$$

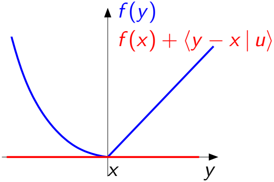


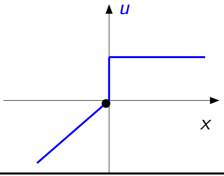


Sous-différentielle d'une fonction convexe : propriétés

Soit $f : \mathcal{H} \to]-\infty, +\infty]$ une fonction propre.
La sous-différentielle de f, notée ∂f, est telle que

$$\begin{aligned} \partial f : \mathcal{H} &\to 2^{\mathcal{H}} \\ x &\to \{u \in \mathcal{H} \,|\, (\forall y \in \mathcal{H})\ \langle y - x | u \rangle + f(x) \leq f(y)\} \end{aligned}$$





Règle de Fermat : $0 \in \partial f(x) \Leftrightarrow x \in \operatorname{Argmin} f$

Sous-différentielle d'une fonction convexe : propriétés

Soit $f : \mathcal{H} \to \,]-\infty, +\infty]$ une fonction propre.
La sous-différentielle de f, notée ∂f, est telle que

$$\begin{aligned} \partial f : \mathcal{H} &\to 2^{\mathcal{H}} \\ x &\to \{u \in \mathcal{H} \,|\, (\forall y \in \mathcal{H})\ \langle y - x | u \rangle + f(x) \leq f(y)\} \end{aligned}$$

- $u \in \partial f(x)$ est un sous-gradient de f en x.

Sous-différentielle d'une fonction convexe : propriétés

Soit $f : \mathcal{H} \to]-\infty, +\infty]$ une fonction propre.
La sous-différentielle de f, notée ∂f, est telle que

$$\begin{aligned} \partial f : \mathcal{H} &\to 2^{\mathcal{H}} \\ x &\to \{u \in \mathcal{H} \,|\, (\forall y \in \mathcal{H})\ \langle y - x | u \rangle + f(x) \leq f(y)\} \end{aligned}$$

- $u \in \partial f(x)$ est un sous-gradient de f en x.
- Si $x \notin \operatorname{dom} f$ alors $\partial f(x) = \varnothing$.

Sous-différentielle d'une fonction convexe : propriétés

Soit $f : \mathcal{H} \to \,]-\infty, +\infty]$ une fonction propre.
La sous-différentielle de f, notée ∂f, est telle que

$$\begin{aligned} \partial f : \mathcal{H} &\to 2^{\mathcal{H}} \\ x &\to \{u \in \mathcal{H} \,|\, (\forall y \in \mathcal{H})\ \langle y - x | u \rangle + f(x) \leq f(y)\} \end{aligned}$$

- $u \in \partial f(x)$ est un sous-gradient de f en x.
- Si $x \notin \mathrm{dom}\, f$ alors $\partial f(x) = \varnothing$.
- Pour tout $x \in \mathrm{dom}\, f$, $\partial f(x)$ est un convexe fermé.

Sous-différentielle d'une fonction convexe : propriétés

Si $f \in \Gamma_0(\mathcal{H})$ est différentiable au sens de Gâteaux en $x \in \mathcal{H}$ alors

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

Sous-différentielle d'une fonction convexe : propriétés

Si $f \in \Gamma_0(\mathcal{H})$ est différentiable au sens de Gâteaux en $x \in \mathcal{H}$ alors

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

$$(\forall y \in \mathcal{H}) \qquad \langle \nabla f(x) \mid y \rangle = \lim_{\substack{\alpha \to 0 \\ \alpha \neq 0}} \frac{f(x + \alpha y) - f(x)}{\alpha}.$$

Preuve :
Pour tout $\alpha \in [0, 1]$ et $y \in \mathcal{H}$,

$$f\big(x + \alpha(y - x)\big) \leq (1 - \alpha) f(x) + \alpha f(y)$$

$$\Rightarrow \quad \langle \nabla f(x) \mid y - x \rangle = \lim_{\substack{\alpha \to 0 \\ \alpha \neq 0}} \frac{f\big(x + \alpha(y - x)\big) - f(x)}{\alpha} \leq f(y) - f(x)$$

D'où $\nabla f(x) \in \partial f(x)$.

Sous-différentielle d'une fonction convexe : propriétés

Si $f \in \Gamma_0(\mathcal{H})$ est différentiable au sens de Gâteaux en $x \in \mathcal{H}$ alors

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

$$(\forall y \in \mathcal{H}) \qquad \langle \nabla f(x) \mid y \rangle = \lim_{\substack{\alpha \to 0 \\ \alpha \neq 0}} \frac{f(x+\alpha y) - f(x)}{\alpha}.$$

<u>Preuve</u> :
Inversement, si $u \in \partial f(x)$, alors, pour tout $\alpha \in [0, +\infty[$ et $y \in \mathcal{H}$,

$$f(x+\alpha y) \geq f(x) + \langle u \mid x + \alpha y - x \rangle$$

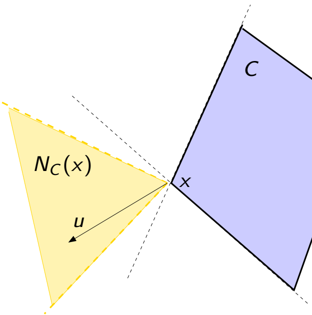
$$\Rightarrow \quad \langle \nabla f(x) \mid y \rangle = \lim_{\substack{\alpha \to 0 \\ \alpha \neq 0}} \frac{f(x+\alpha y) - f(x)}{\alpha} \geq \langle u \mid y \rangle$$

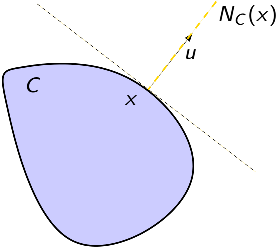
En choisissant $y = u - \nabla f(x)$, on en déduit que $\|u - \nabla f(x)\|^2 \leq 0$.
D'où $u = \nabla f(x)$.

Sous-différentielle d'une fonction convexe : exemple

Pour tout $x \in \mathcal{H}$, $\partial\iota_C(x)$ est le cône normal à C en x défini par

$$N_C(x) = \begin{cases} \{u \in \mathcal{H} \mid (\forall y \in C) \ \ \langle u \mid y - x\rangle \leq 0\} & \text{si } x \in C \\ \varnothing & \text{sinon.} \end{cases}$$





Sous-différentielle d'une fonction convexe : exemple

Pour tout $x \in \mathcal{H}$, $\partial \iota_C(x)$ est le **cône normal** à C en x défini par

$$N_C(x) = \begin{cases} \{u \in \mathcal{H} \mid (\forall y \in C) \;\; \langle u \mid y - x \rangle \leq 0\} & \text{si } x \in C \\ \varnothing & \text{sinon.} \end{cases}$$

- Soit $c \in \mathcal{H}$, $\rho \in \,]0, +\infty[$ et $C = \overline{\mathcal{B}}(c, \rho) = \{y \in \mathcal{H} \mid \|y - c\| \leq \rho\}$. Pour tout $x \in C$,

$$N_C(x) = \begin{cases} \{\alpha(x - c) \mid \alpha \in [0, +\infty[\} & \text{si } \|x - c\| = \rho \\ \{0\} & \text{si } \|x - c\| < \rho. \end{cases}$$

Conjugate

Adrien-Marie Legendre
(1752–1833)

Werner Fenchel
(1905–1988)

Conjugate

Adrien-Marie Legendre
(1752–1833)

Werner Fenchel
(1905–1988)

Conjuguée : définition

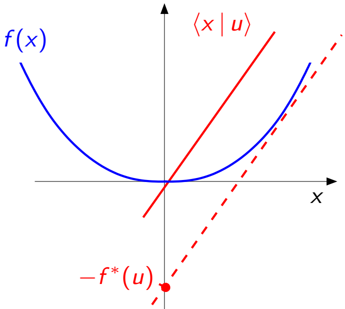
Soient $\mathcal{H}$ un espace de Hilbert et $f \colon \mathcal{H} \to \,]-\infty, +\infty]$.
La conjuguée de f est $f^* \colon \mathcal{H} \to [-\infty, +\infty]$ telle que

$$(\forall u \in \mathcal{H}) \qquad f^*(u) = \sup_{x \in \mathcal{H}} \big(\langle x \mid u \rangle - f(x) \big).$$

Conjuguée : définition

Soient $\mathcal{H}$ un espace de Hilbert et $f\colon \mathcal{H} \to \,]-\infty, +\infty]$.
La conjuguée de f est $f^*\colon \mathcal{H} \to [-\infty, +\infty]$ telle que

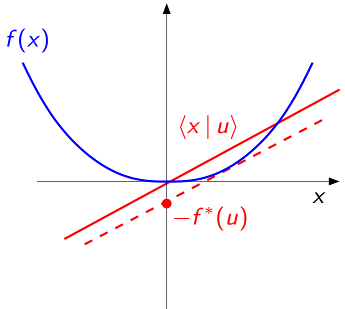
$$(\forall u \in \mathcal{H}) \qquad f^*(u) = \sup_{x \in \operatorname{dom} f} \big(\langle x \mid u\rangle - f(x)\big)$$



Conjuguée : définition

Soient $\mathcal{H}$ un espace de Hilbert et $f\colon \mathcal{H} \to \,]-\infty,+\infty]$.
La conjuguée de f est $f^*\colon \mathcal{H} \to [-\infty,+\infty]$ telle que

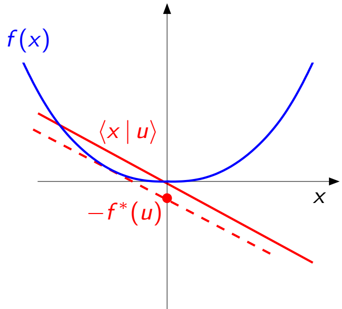
$$(\forall u \in \mathcal{H}) \qquad f^*(u) = \sup_{x \in \operatorname{dom} f} \big(\langle x \mid u \rangle - f(x) \big)$$



Conjuguée : définition

Soient $\mathcal{H}$ un espace de Hilbert et $f\colon \mathcal{H} \to \,]-\infty,+\infty]$.
La conjuguée de f est $f^*\colon \mathcal{H} \to [-\infty,+\infty]$ telle que

$$(\forall u \in \mathcal{H}) \qquad f^*(u) = \sup_{x \in \operatorname{dom} f} \big(\langle x \mid u \rangle - f(x) \big)$$



Conjuguée : définition

Soient $\mathcal{H}$ un espace de Hilbert et $f\colon \mathcal{H} \to\ ]-\infty, +\infty]$.
La **conjuguée** de f est $f^*\colon \mathcal{H} \to [-\infty, +\infty]$ telle que

$$(\forall u \in \mathcal{H}) \qquad f^*(u) = \sup_{x \in \operatorname{dom} f} \big(\langle x \mid u \rangle - f(x) \big)$$

Exemples :

- $f = \frac{1}{2}\|\cdot\|^2 \Rightarrow f^* = \frac{1}{2}\|\cdot\|^2$

  Preuve : Pour tout $(x,u) \in \mathcal{H}^2$, $\langle x \mid u \rangle - \frac{1}{2}\|x\|^2 = \frac{1}{2}\|u\|^2 - \frac{1}{2}\|u - x\|^2$ est maximum en $x = u$.
  Par conséquent, $f^*(u) = \frac{1}{2}\|u\|^2$.

Conjuguée : définition

Soient $\mathcal{H}$ un espace de Hilbert et $f\colon \mathcal{H} \to \,]-\infty, +\infty]$.
La conjuguée de f est $f^*\colon \mathcal{H} \to [-\infty, +\infty]$ telle que

$$(\forall u \in \mathcal{H}) \qquad f^*(u) = \sup_{x \in \operatorname{dom} f} \big(\langle x \mid u \rangle - f(x) \big)$$

Exemples :

- $f = \frac{1}{2}\|\cdot\|^2 \Rightarrow f^* = \frac{1}{2}\|\cdot\|^2$.
- Soit $\phi\colon \mathbb{R} \to \,]-\infty, +\infty]$ une fonction paire. $(\phi \circ \|\cdot\|)^* = \phi^* \circ \|\cdot\|$.
- $(\forall x \in \mathbb{R}^N)\ f(x) = \frac{1}{q}\|x\|_q^q$ avec $q \in]1, +\infty[$
$\Rightarrow (\forall u \in \mathbb{R}^N)\ f^*(u) = \frac{1}{q^*}\|u\|_{q^*}^{q^*}$ avec $\frac{1}{q} + \frac{1}{q^*} = 1$

Conjuguée : définition

Soient $\mathcal{H}$ un espace de Hilbert et $f\colon \mathcal{H} \to \,]-\infty, +\infty]$.
La conjuguée de f est $f^*\colon \mathcal{H} \to [-\infty, +\infty]$ telle que

$$(\forall u \in \mathcal{H}) \qquad f^*(u) = \sup_{x \in \operatorname{dom} f} \big(\langle x \mid u \rangle - f(x) \big)$$

- Si f est paire alors f^* est paire.

Preuve :

$$\begin{aligned}(\forall u \in \mathcal{H}) \quad f^*(-u) &= \sup_{x \in \mathcal{H}} \big(\langle x \mid -u \rangle - f(x) \big) \\ &= \sup_{x \in \mathcal{H}} \big(\langle x \mid u \rangle - f(-x) \big) \\ &= \sup_{x \in \mathcal{H}} \big(\langle x \mid u \rangle - f(x) \big) \\ &= f^*(u)\end{aligned}$$

Conjuguée : définition

Soient $\mathcal{H}$ un espace de Hilbert et $f\colon \mathcal{H} \to \,]-\infty,+\infty]$.
La conjuguée de f est $f^*\colon \mathcal{H} \to [-\infty,+\infty]$ telle que

$$(\forall u \in \mathcal{H}) \qquad f^*(u) = \sup_{x \in \operatorname{dom} f} \big(\langle x \mid u \rangle - f(x) \big)$$

- Si f est paire alors f^* est paire.
- Pour tout $\alpha \in \,]0,+\infty[$, $(\alpha f)^* = \alpha f^*(\cdot/\alpha)$.
- Pour tout $(y,v) \in \mathcal{H}^2$ et $\alpha \in \mathbb{R}$,
  $\big(f(\cdot - y) + \langle \cdot \mid v \rangle + \alpha\big)^* = f^*(\cdot - v) + \langle y \mid \cdot - v \rangle - \alpha$.
- Soit $\mathcal{G}$ un espace de Hilbert et $L \in \mathcal{B}(\mathcal{G},\mathcal{H})$ un isomorphisme.
  $(f \circ L)^* = f^* \circ (L^{-1})^*$.
- f^* est s.c.i. et convexe.

Conjuguée : définition

Soient $\mathcal{H}$ un espace de Hilbert et $f\colon \mathcal{H} \to \,]-\infty,+\infty]$.
La conjuguée de f est $f^*\colon \mathcal{H} \to [-\infty,+\infty]$ telle que

$$(\forall u \in \mathcal{H}) \qquad f^*(u) = \sup_{x \in \operatorname{dom} f} \big(\langle x \mid u \rangle - f(x) \big)$$

Théorème de Moreau-Fenchel
Soient $\mathcal{H}$ un espace de Hilbert et $f\colon \mathbb{R} \to \,]-\infty,+\infty]$ une fonction propre.
f est s.c.i. et convexe ssi $f^{**} = f$.

Conjuguée : définition

Soient $\mathcal{H}$ un espace de Hilbert et $f\colon \mathcal{H} \to \left]-\infty, +\infty\right]$.
La conjuguée de f est $f^*\colon \mathcal{H} \to [-\infty, +\infty]$ telle que

$$(\forall u \in \mathcal{H}) \qquad f^*(u) = \sup_{x \in \operatorname{dom} f} \big(\langle x \mid u \rangle - f(x) \big)$$

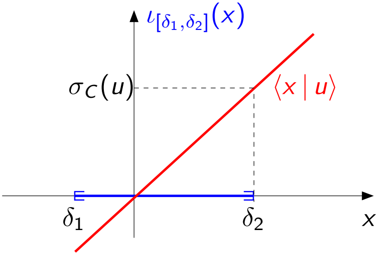
Théorème de Moreau-Fenchel

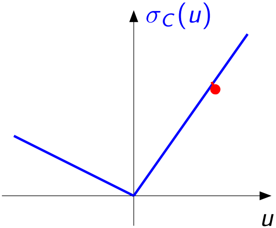
Soient $\mathcal{H}$ un espace de Hilbert et $f\colon \mathbb{R} \to \left]-\infty, +\infty\right]$ une fonction propre.
f est s.c.i. et convexe ssi $f^{**} = f$.

Conjuguée : exemple

Soit $\mathcal{H}$ un espace de Hilbert et $C \subset \mathcal{H}$.
σ_C est la fonction d'appui de C si

$$(\forall u \in \mathcal{H}) \qquad \sigma_C(u) = \sup_{x \in C} \langle x \mid u \rangle = \iota_C^*(u).$$

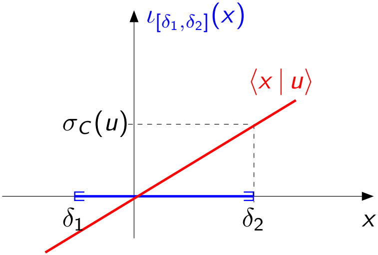


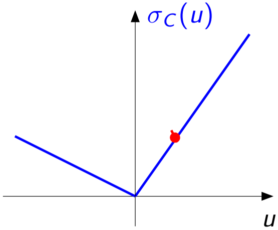


Conjuguée : exemple

Soit $\mathcal{H}$ un espace de Hilbert et $C \subset \mathcal{H}$.
σ_C est la fonction d'appui de C si

$$(\forall u \in \mathcal{H}) \qquad \sigma_C(u) = \sup_{x \in C} \langle x \mid u \rangle = \iota_C^*(u).$$

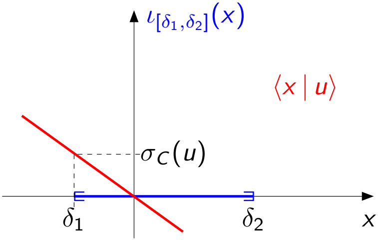


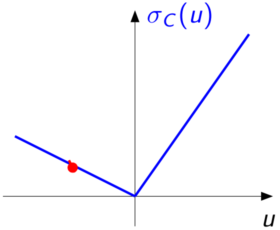


Conjuguée : exemple

Soit $\mathcal{H}$ un espace de Hilbert et $C \subset \mathcal{H}$.
σ_C est la fonction d'appui de C si

$$(\forall u \in \mathcal{H}) \qquad \sigma_C(u) = \sup_{x \in C} \langle x \mid u \rangle$$
$$= \iota_C^*(u).$$





Conjuguée : exemple

Soit $\mathcal{H}$ un espace de Hilbert.
$f : \mathcal{H} \to]-\infty, +\infty]$ est positive homogène si

$$(\forall x \in \mathcal{H})(\forall \alpha \in]0, +\infty[) \qquad f(\alpha x) = \alpha f(x)$$

f est positive homogène et appartient à $\Gamma_0(\mathcal{H})$ ssi $f = \sigma_C$ où C est un convexe fermé non vide de $\mathcal{H}$.

Conjuguée : exemple

Soit $\mathcal{H}$ un espace de Hilbert.
$f\colon \mathcal{H} \to \,]-\infty,+\infty]$ est **positive homogène** si

$$(\forall x \in \mathcal{H})(\forall \alpha \in \,]0,+\infty[) \qquad f(\alpha x) = \alpha f(x)$$

f est positive homogène et appartient à $\Gamma_0(\mathcal{H})$ ssi $f = \sigma_C$ où C est un convexe fermé non vide de $\mathcal{H}$.

Preuve : $(\Leftarrow)$
$f = \iota_C^*$ et $\iota_C \in \Gamma_0(\mathcal{H})$. D'où $\sigma_C \in \Gamma_0(\mathcal{H})$.
De plus, $(\forall x \in \mathcal{H})$ $(\forall \alpha \in \,]0,+\infty[)$ $\sigma_C(\alpha x) = \alpha \sigma_C(x)$.

Conjuguée : exemple

Soit $\mathcal{H}$ un espace de Hilbert.
$f : \mathcal{H} \to \,]-\infty, +\infty]$ est **positive homogène** si

$$(\forall x \in \mathcal{H})(\forall \alpha \in \,]0, +\infty[) \qquad f(\alpha x) = \alpha f(x)$$

f est positive homogène et appartient à $\Gamma_0(\mathcal{H})$ ssi $f = \sigma_C$ où C est un convexe fermé non vide de $\mathcal{H}$.

Preuve : $(\Rightarrow)$
Soit $y \in \operatorname{dom} f$.
$f(0) = \lim_{\alpha \to 0, \alpha \geq 0} f\big((1-\alpha)0 + \alpha y\big) = \lim_{\alpha \to 0} \alpha f(y) = 0$.
Soit $C = \big\{ u \in \mathcal{H} \; \big| \; (\forall x \in \mathcal{H}) \, \langle x \mid u \rangle \leq f(x) \big\}$.
On a, pour tout $u \in C$,
$f^*(u) = \sup_{x \in \mathcal{H}} \langle x \mid u \rangle - f(x) \leq 0 = \langle 0 \mid u \rangle - f(0) \leq f^*(u)$.
D'où $f^*(u) = 0$.

Conjuguée : exemple

Soit $\mathcal{H}$ un espace de Hilbert.
$f \colon \mathcal{H} \to \,]-\infty, +\infty]$ est **positive homogène** si

$$(\forall x \in \mathcal{H})(\forall \alpha \in \,]0, +\infty[) \qquad f(\alpha x) = \alpha f(x)$$

f est positive homogène et appartient à $\Gamma_0(\mathcal{H})$ ssi $f = \sigma_C$ où C est un convexe fermé non vide de $\mathcal{H}$.

Preuve : $(\Rightarrow)$
De plus, pour tout $u \notin C$, il existe $x \in \mathcal{H}$ tel que $\langle x \mid u \rangle > f(x)$. On a alors, pour tout $\alpha \in \,]0, +\infty[$,
$f^*(u) \geq \langle \alpha x \mid u \rangle - f(\alpha x) = \alpha\big(\langle x \mid u \rangle - f(x)\big)$. En faisant tendre α vers $+\infty$, on en deduit que $f^*(u) = +\infty$.
En conclusion, $f^* = \iota_C \in \Gamma_0(\mathcal{H}) \Rightarrow f = \sigma_C$ et C est un convexe fermé non vide.

Conjuguée : exemples de fonctions d'appui

- Soit f une norme ℓ^q de $\mathbb{R}^N$ avec $q \in [1, +\infty]$.
  On a $f = \sigma_C$ où

$$C = \left\{ y \in \mathbb{R}^N \mid \|y\|_{q^*} \leq 1 \right\} \qquad \text{avec } \frac{1}{q} + \frac{1}{q^*} = 1.$$

Conjuguée : exemples de fonctions d'appui

- Soit f une norme ℓ^q de $\mathbb{R}^N$ avec $q \in [1, +\infty]$.
  On a $f = \sigma_C$ où

$$C = \left\{ y \in \mathbb{R}^N \;\middle|\; \|y\|_{q^*} \leq 1 \right\} \qquad \text{avec } \frac{1}{q} + \frac{1}{q^*} = 1.$$

<u>Cas particulier</u> : norme ℓ^1 de $\mathbb{R}^N$: $C = [-1, 1]^N$.

Conjuguée : propriétés

Inégalité de Fenchel-Young : si f est propre alors

1. $(\forall (x,u) \in \mathcal{H}^2) \qquad f(x) + f^*(u) \geq \langle x \mid u \rangle$

2. $(\forall (x,u) \in \mathcal{H}^2) \qquad u \in \partial f(x) \quad \Leftrightarrow \quad f(x) + f^*(u) = \langle x \mid u \rangle\,.$

Conjuguée : Théorème de Fenchel-Rockafellar

Si $f \in \Gamma_0(\mathbb{R}^N)$, $g \in \Gamma_0(\mathbb{R}^N)$ et $L \in \mathbb{R}^{M\times N}$ alors

$$\min_{x\in\mathbb{R}^N} f(x) + g(Lx) = -\min_{u\in\mathbb{R}^M} f^*(L^* u) + g^*(u)$$

Si de plus il existe une solution $\widehat{u}$ du problème dual tel que f^* est différentiable en $L^*\widehat{u}$, alors

$$\widehat{x} = \nabla f^*(L^*\widehat{u})$$

Exercice : Sous-différentielle et conjuguée

On s'intéresse au problème de minimisation de la variation totale qui consiste à résoudre

$$\widehat{x} = \arg\min_{x \in \mathbb{R}^N} \frac{1}{2}\|x - y\|_2^2 + \lambda \|Lx\|_1$$

où $(Lx)^{(i)} = x^{(i+1)} - x^{(i)}$ pour tout $i \in \{1, \ldots, N-1\}$. En d'autres termes $L \in \mathbb{R}^{(N-1 \times N}$ désigne un opérateur de différences finies.

- Montrer que le problème dual associé peut s'écrire

$$\min_{u \in \mathbb{R}^{N+1}} \frac{1}{2}\|y + L^* u\|_2^2 \quad \text{s.t.} \quad \begin{cases} (\forall i \in \{1, \ldots, N-1\}) & |u^{(i)}| \leq \lambda \\ & u^{(0)} = u^{(N)} = 0 \end{cases}$$

  et que la relation entre les solutions primale et duale est

$$\widehat{x} = y + L^* \widehat{u}$$

Exercice : Sous-différentielle et conjuguée

On s'intéresse au problème de minimisation de la variation totale qui consiste à résoudre

$$\widehat{x} = \arg\min_{x \in \mathbb{R}^N} \frac{1}{2}\|x - y\|_2^2 + \lambda \|Lx\|_1$$

où $(Lx)^{(i)} = x^{(i+1)} - x^{(i)}$ pour tout $i \in \{1, \dots, N-1\}$. En d'autres termes $L \in \mathbb{R}^{(N-1 \times N}$ désigne un opérateur de différences finies.

- En combinant les conditions KKT de la formulation duale et la relation $\widehat{x} = y + L^* \widehat{u}$, montrer que les conditions d'optimalité peuvent s'écrire

$$\begin{cases} \widehat{u}^{(i)} = -\lambda & \text{si} \quad \widehat{x}^{(i+1)} > \widehat{x}^{(i)} \\ \widehat{u}^{(i)} = +\lambda & \text{si} \quad \widehat{x}^{(i+1)} < \widehat{x}^{(i)} \\ \widehat{u}^{(i)} \in [-\lambda, +\lambda] & \text{si} \quad \widehat{x}^{(i+1)} = \widehat{x}^{(i)} \end{cases}$$

Opérateur proximal : motivation

Soit $\mathcal{H}$ un espace de Hilbert réel. Soit $f \in \Gamma_0(\mathcal{H})$ une fonction de gradient Lipschitz de constante $\beta > 0$.
Find
$$\widehat{x} \in \underset{x \in \mathcal{H}}{\mathrm{Argmin}}\ f(x).$$

- **Algorithme de descente de gradient**

  Soit $\gamma \in\ ]0, +\infty[$ et $x_0 \in \mathcal{H}$.
  Pour $n = 0, 1 \ldots$
  $\lfloor\ x_{n+1} = x_n - \gamma \nabla f(x_n).$

La suite $(x_n)_{n \in \mathbb{N}}$ générée par ce schéma *explicite* converge vers un minimiseur de f si celui-ci existe et si $\gamma \in]0, 2/\beta[$.

Opérateur proximal : motivation

Soit $\mathcal{H}$ un espace de Hilbert réel. Soit $f \in \Gamma_0(\mathcal{H})$ une fonction de gradient Lipschitz de constante $\beta > 0$.
Find

$$\widehat{x} \in \underset{x \in \mathcal{H}}{\mathrm{Argmin}}\, f(x).$$

- **Algorithme alternatif**

  Soit $\gamma \in \,]0, +\infty[$ et $x_0 \in \mathcal{H}$.
  Pour $n = 0, 1 \ldots$
  $\lfloor\; x_{n+1} = x_n - \gamma \nabla f(x_{n+1})$.

Questions :

Opérateur proximal : motivation

Soit $\mathcal{H}$ un espace de Hilbert réel. Soit $f \in \Gamma_0(\mathcal{H})$ une fonction de gradient Lipschitz de constante $\beta > 0$.
Find

$$\widehat{x} \in \underset{x \in \mathcal{H}}{\mathrm{Argmin}}\ f(x).$$

- **Algorithme alternatif**

  Soit $\gamma \in \,]0, +\infty[$ et $x_0 \in \mathcal{H}$.
  Pour $n = 0, 1 \ldots$
  $\lfloor\ x_{n+1} = x_n - \gamma \nabla f(x_{n+1})$.

Questions :

- Comment déterminer x_{n+1} à chaque itération n de ce schéma *implicite* ?
- Quelles valeurs de γ garantissent la convergence de $(x_n)_{n \in \mathbb{N}}$?
- Que faire si f is non-lisse ?

Opérateur proximal : définition

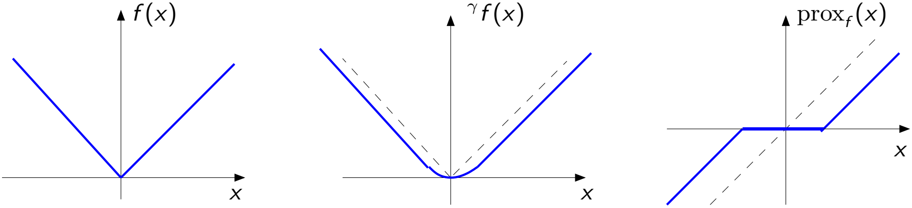
Soit $\mathcal{H}$ un espace de Hilbert. Soit $f \in \Gamma_0(\mathcal{H})$.

- L' **enveloppe de Moreau** de paramètre $\gamma \in \left]0, +\infty\right[$ de f est

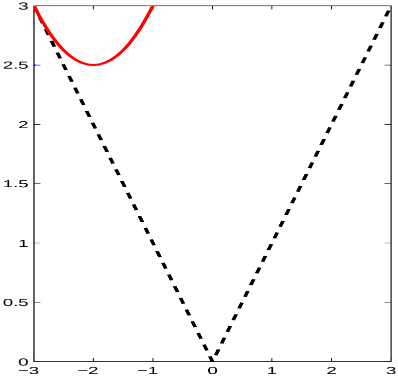
$$^{\gamma}f : \mathcal{H} \to \mathbb{R} : x \mapsto \inf_{y \in \mathcal{H}} f(y) + \frac{1}{2\gamma}\|y - x\|^2.$$

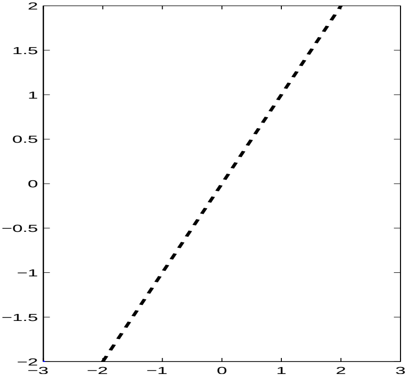
- L' **opérateur proximal** de f est

$$\mathrm{prox}_f : \mathcal{H} \to \mathcal{H} : x \mapsto \underset{y \in \mathcal{H}}{\mathrm{argmin}}\ f(y) + \frac{1}{2}\|y - x\|^2.$$

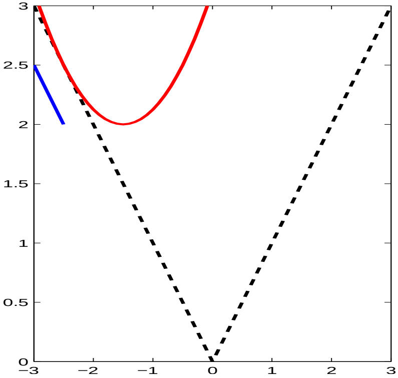


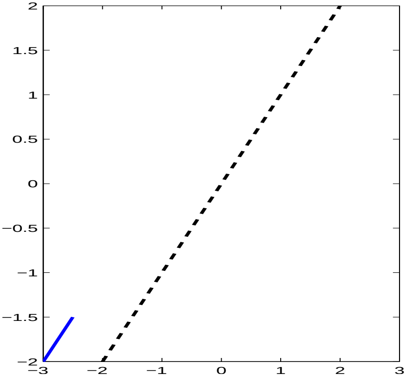
Opérateur proximal : définition



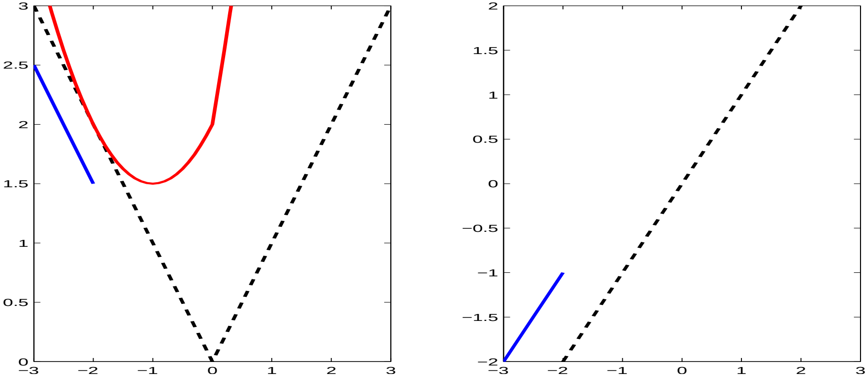


Opérateur proximal : définition

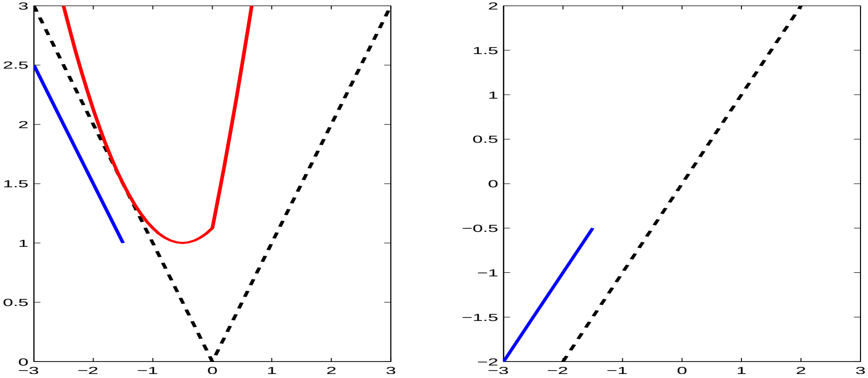




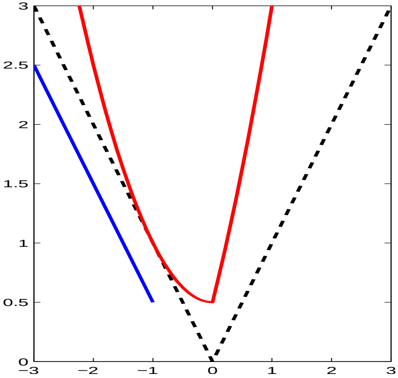
Opérateur proximal : définition

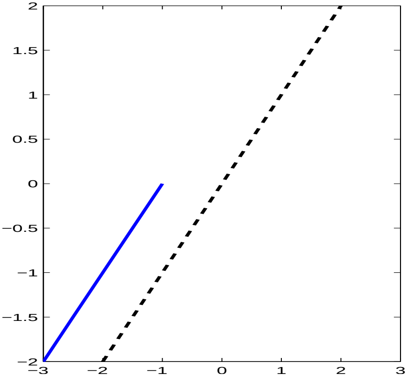


Opérateur proximal : définition

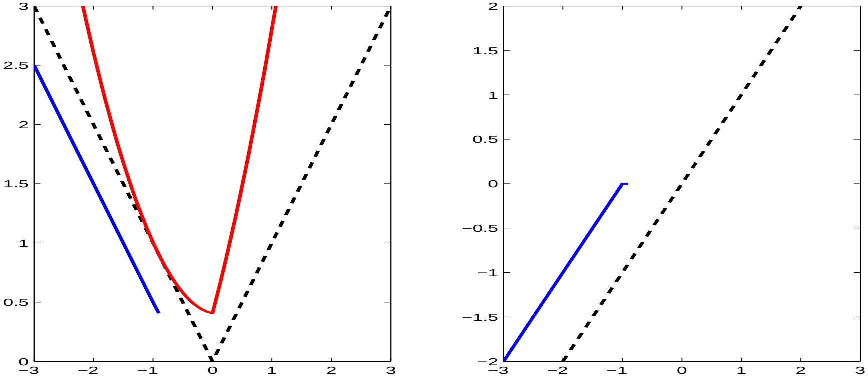


Opérateur proximal : définition

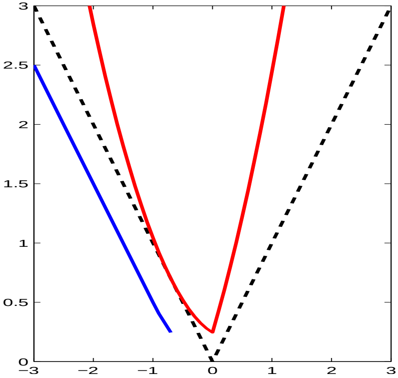


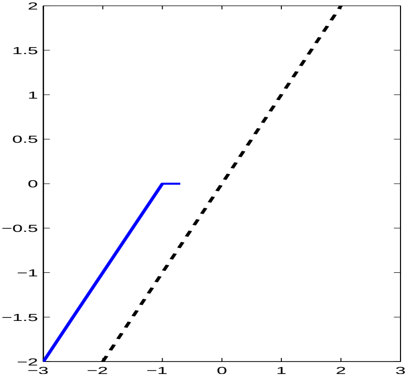


Opérateur proximal : définition

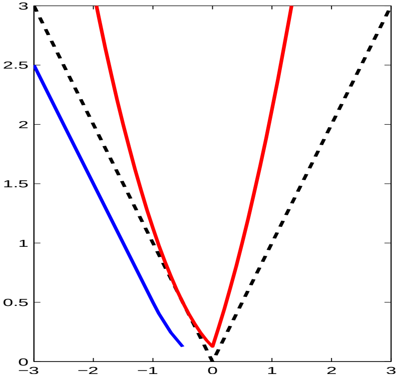


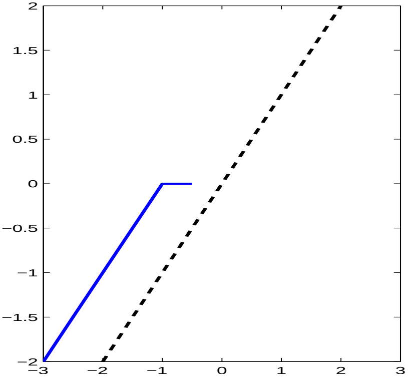
Opérateur proximal : définition



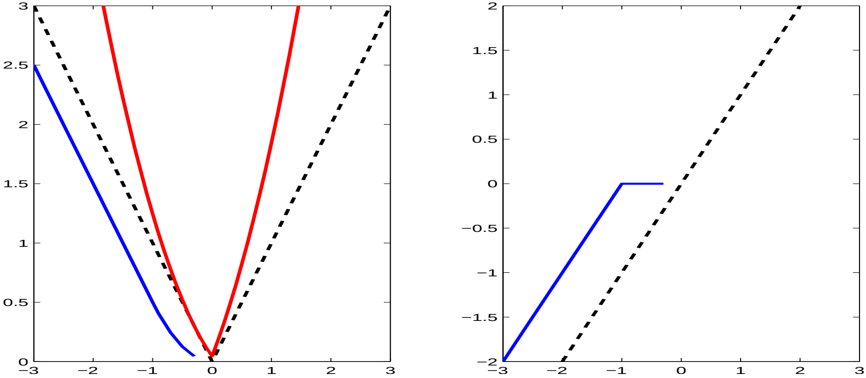


Opérateur proximal : définition

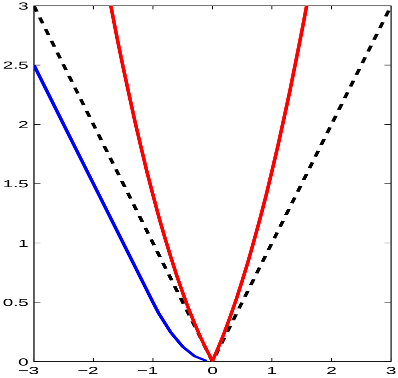


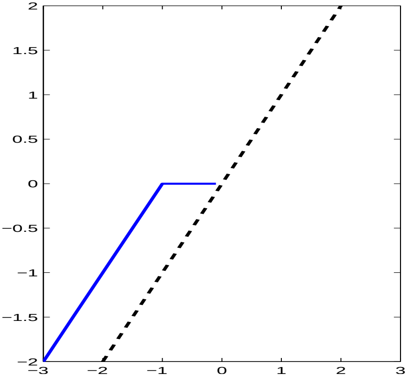


Opérateur proximal : définition

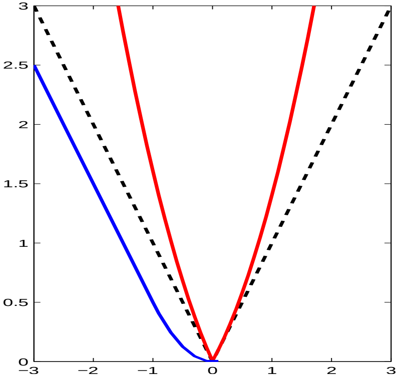


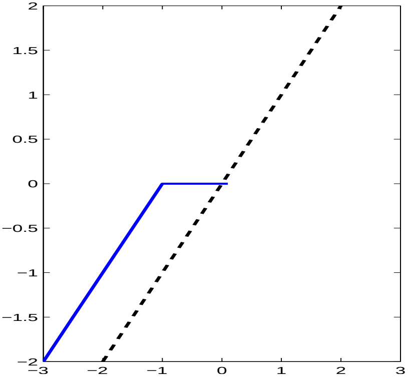
Opérateur proximal : définition



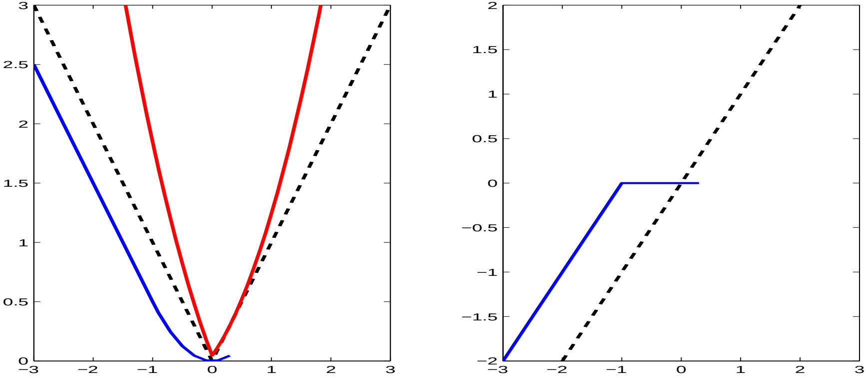


Opérateur proximal : définition

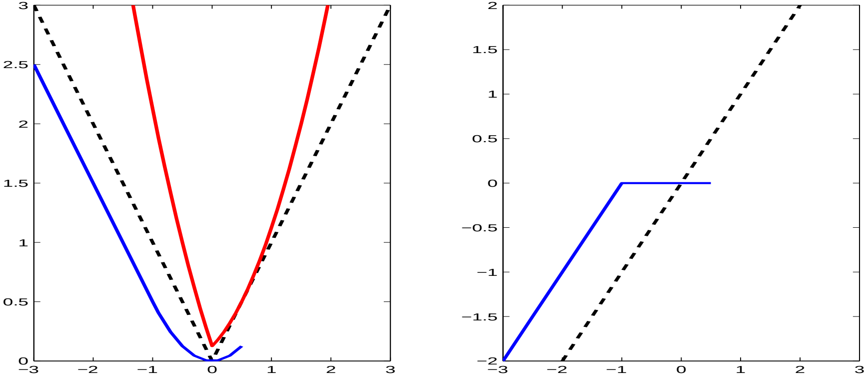




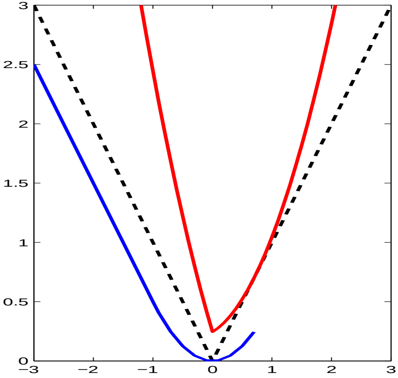
Opérateur proximal : définition

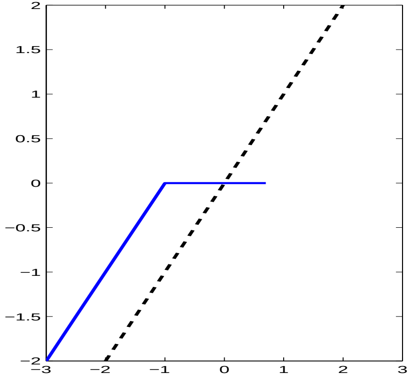


Opérateur proximal : définition

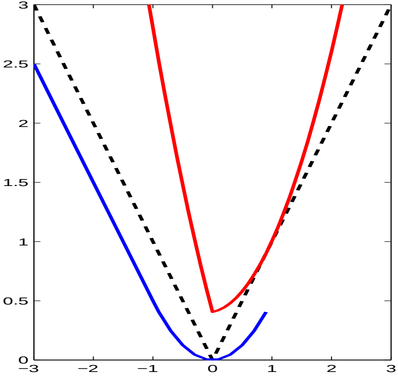


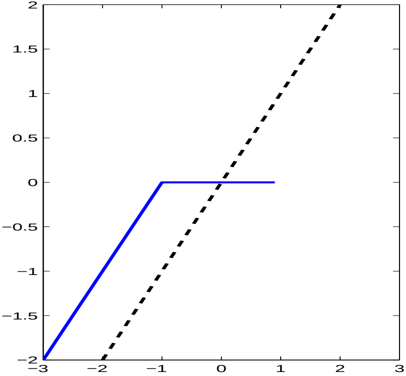
Opérateur proximal : définition



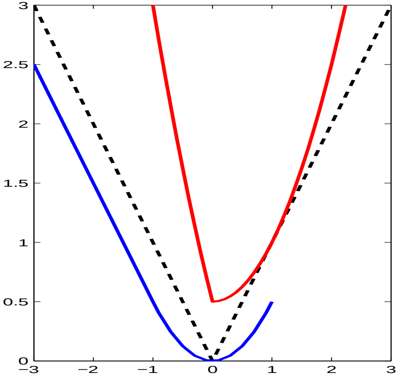


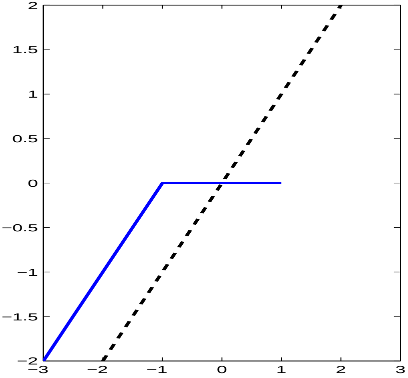
Opérateur proximal : définition



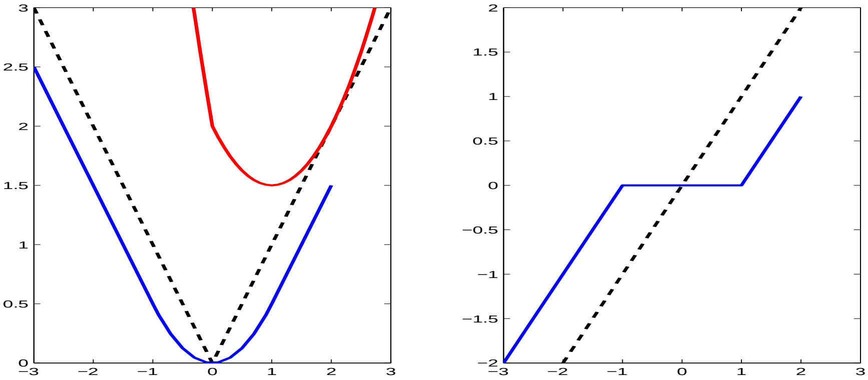


Opérateur proximal : définition

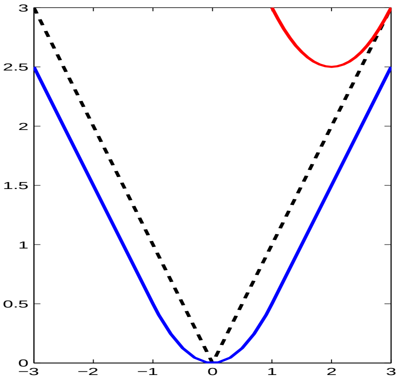


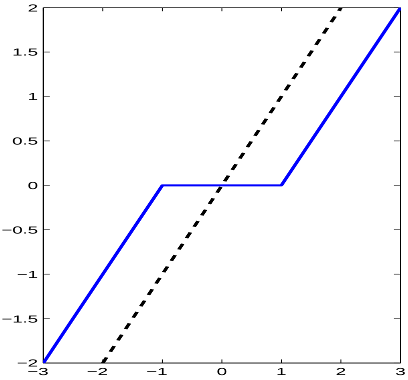


Opérateur proximal : définition



Opérateur proximal : définition





Opérateur proximal : définition

Let $\mathcal{H}$ be a Hilbert space and $f \in \Gamma_0(\mathcal{H})$.

$$(\forall x \in \mathcal{H}) \quad p = \mathrm{prox}_f(x) \quad \Leftrightarrow \quad x - p \in \partial f(p).$$

Opérateur proximal : définition

Let $\mathcal{H}$ be a Hilbert space and $f \in \Gamma_0(\mathcal{H})$.

$$(\forall x \in \mathcal{H}) \qquad \boxed{p = \mathrm{prox}_f(x) \quad \Leftrightarrow \quad x - p \in \partial f(p)}\,.$$

<u>Proof</u> : D'après la règle de Fermat, pour tout $x \in \mathcal{H}$, $p = \mathrm{prox}_f(x)$ si et seulement si

$$\begin{aligned}
& p = \underset{y \in \mathcal{H}}{\arg\min} \; f(y) + \frac{1}{2}\|y - x\|^2 \\
\Leftrightarrow \qquad & 0 \in \partial\Big(f + \frac{1}{2}\|\cdot - x\|^2\Big)(p) \\
\Leftrightarrow \qquad & 0 \in \partial f(p) + p - x \\
\Leftrightarrow \qquad & x \in (\mathrm{Id} + \partial f)(p).
\end{aligned}$$

Opérateur proximal : propriétés

Soit $\mathcal{H}$ un espace de Hilbert, $x \in \mathcal{H}$ et $f \in \Gamma_0(\mathcal{H})$.

Propriétés	$g(x)$	$\mathrm{prox}_g x$
Translation	$f(x-z), z \in \mathcal{H}$	$z + \mathrm{prox}_f(x-z)$
Perturbation quadratique	$f(x) + \alpha \parallel x \parallel^2 /2 + \langle z \mid x \rangle + \gamma$ $z \in \mathcal{H}, \alpha > 0, \gamma \in \mathbb{R}$	$\mathrm{prox}_{\frac{f}{\alpha+1}}(\frac{x-z}{\alpha+1})$
Changement d'échelle	$f(\rho x), \rho \in \mathbb{R}^*$	$\frac{1}{\rho}\mathrm{prox}_{\rho^2 f}(\rho x)$
Réflexion	$f(-x)$	$-\mathrm{prox}_f(-x)$
Enveloppe de Moreau	${}^{\gamma}f(x) = \inf_{y \in \mathcal{H}} f(y) + \frac{1}{2\gamma}\|x-y\|^2$ $\gamma > 0$	$\frac{1}{1+\gamma}\left(\gamma x + \mathrm{prox}_{(1+\gamma)f}(x)\right)$

Opérateur proximal : propriétés

Pour tout $i \in \{1, \ldots, n\}$, soient $\mathcal{H}_i$ un espace de Hilbert et $f_i \in \Gamma_0(\mathcal{H}_i)$.
Pour tout $(x_1, \ldots, x_n) \in \mathcal{H}_1 \times \cdots \times \mathcal{H}_n$,
si

$$f(x_1, \ldots, x_n) = \sum_{i=1}^{n} f_i(x_i).$$

alors

$$\mathrm{prox}_f(x_1, \ldots, x_n) = \left(\mathrm{prox}_{f_i}(x_i)\right)_{1 \leq i \leq n}.$$

Opérateur proximal : propriétés

Soient $\mathcal{H}$ un espace de Hilbert séparable et $(b_i)_{i\in I}$ une base orthonormale de $\mathcal{H}$.
Pour tout $i \in I$, soit $\varphi_i \in \Gamma_0(\mathbb{R})$ tel que $\varphi_i \geq 0$. Pour tout $x \in \mathcal{H}$, si

$$f(x) = \sum_{i\in I} \varphi_i(\langle x \mid b_i\rangle)$$

alors

$$\operatorname{prox}_f(x) = \sum_{i\in I} \operatorname{prox}_{\varphi_i}(\langle x \mid b_i\rangle) b_i.$$

Remarque : L'hypothèse $(\forall i \in I)\ \varphi_i \geq 0$ peut être relaxée si $\mathcal{H}$ est de dimension finie.

Opérateur proximal : propriétés

Soient $\mathcal{H}$ un espace de Hilbert séparable et $(b_i)_{i\in I}$ une base orthonormale de $\mathcal{H}$.
Pour tout $i \in I$, soit $\varphi_i \in \Gamma_0(\mathbb{R})$ tel que $\varphi_i \geq 0$. Pour tout $x \in \mathcal{H}$, si

$$f(x) = \sum_{i\in I} \varphi_i(\langle x \mid b_i\rangle)$$

alors

$$\operatorname{prox}_f(x) = \textstyle\sum_{i\in I} \operatorname{prox}_{\varphi_i}(\langle x \mid b_i\rangle) b_i.$$

Soit $\mathcal{H}$ un espace de Hilbert, $f \in \Gamma_0(\mathcal{H})$ et $\gamma \in \,]0, +\infty[$.

$$(\forall x \in \mathcal{H}) \qquad \operatorname{prox}_{\gamma f^*} x = x - \gamma \operatorname{prox}_{\gamma^{-1} f}(\gamma^{-1} x).$$

Opérateur proximal : propriétés

Soient $\mathcal{H}$ et $\mathcal{G}$ deux espaces de Hilbert. Soient $f \in \Gamma_0(\mathcal{H})$ et $L \in \mathcal{B}(\mathcal{G}, \mathcal{H})$ tel que $LL^* = \mu \mathrm{Id}$ où $\mu \in \,]0, +\infty[$. On a

$$\mathrm{prox}_{f \circ L} = \mathrm{Id} - \mu^{-1} L^* \circ (\mathrm{Id} - \mathrm{prox}_f) \circ L.$$

Opérateur proximal : propriétés

Soient $\mathcal{H}$ et $\mathcal{G}$ deux espaces de Hilbert. Soient $f \in \Gamma_0(\mathcal{H})$ et $L \in \mathcal{B}(\mathcal{G}, \mathcal{H})$ tel que $LL^* = \mu \mathrm{Id}$ où $\mu \in \,]0, +\infty[$. On a

$$\mathrm{prox}_{f \circ L} = \mathrm{Id} - \mu^{-1} L^* \circ (\mathrm{Id} - \mathrm{prox}_f) \circ L.$$

Remarque :

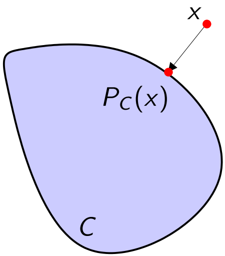
Si $L \in \mathcal{B}(\mathcal{H}, \mathcal{H})$ est une isométrie bijective, alors $\mathrm{prox}_{f \circ L} = L^* \mathrm{prox}_f L$.

Opérateur proximal : exemples

Projection :

Soit $\mathcal{H}$ un espace de Hilbert. Soit C un convexe fermé non vide de $\mathcal{H}$.

$$(\forall x \in \mathcal{H}) \qquad \mathrm{prox}_{\iota_C}(x) = \underset{y \in C}{\mathrm{argmin}} \frac{1}{2}\|y - x\|^2 = P_C(x).$$



Opérateur proximal : exemples

Projection :
Soit $\mathcal{H}$ un espace de Hilbert. Soit C un convexe fermé non vide de $\mathcal{H}$.

$$(\forall x \in \mathcal{H}) \qquad \operatorname{prox}_{\iota_C}(x) = \underset{y \in C}{\operatorname{argmin}} \frac{1}{2}\|y - x\|^2 = P_C(x).$$

Fonction quadratique :
Soient $\mathcal{H}$ et $\mathcal{G}$ deux espaces de Hilbert.
Soient $L \in \mathcal{B}(\mathcal{G}, \mathcal{H})$, $\gamma \in \,]0, +\infty[$ et $z \in \mathcal{G}$.

$$f = \gamma \left\| L \cdot - z \right\|^2 / 2 \Rightarrow \operatorname{prox}_f = (\mathrm{Id} + \gamma L^* L)^{-1}(\cdot + \gamma L^* z).$$

Opérateur proximal : exemples

Norme ℓ_1 :

$$(\forall x = (x^{(i)})_{1 \leq i \leq N} \in \mathbb{R}^N) \qquad \mathrm{prox}_{\lambda \|\cdot\|_1} x = (\mathrm{prox}_{\lambda|\cdot|} x^{(i)})_{1 \leq i \leq N}$$

avec

$$\mathrm{prox}_{\lambda|\cdot|} x^{(i)} = \begin{cases} x^{(i)} - \lambda & \text{si } x^{(i)} < \lambda \\ 0 & \text{si } x^{(i)} \in [-\lambda, \lambda] \\ x^{(i)} + \lambda & \text{si } x^{(i)} > \lambda. \end{cases}$$

Algorithmes de points fixes

Algorithmes de points fixes : notions de convergence

Soit $\mathcal{H}$ un espace de Hilbert.
Soit $(x_n)_{n\in\mathbb{N}}$ une suite d'éléments de $\mathcal{H}$ et $\widehat{x} \in \mathcal{H}$.

- $(x_n)_{n\in\mathbb{N}}$ **converge fortement** vers $\widehat{x}$ si

$$\lim_{n\to+\infty} \|x_n - \widehat{x}\| = 0.$$

  On note $x_n \to \widehat{x}$.

- $(x_n)_{n\in\mathbb{N}}$ **converge faiblement** vers $\widehat{x}$ si

$$(\forall y \in \mathcal{H}) \qquad \lim_{n\to+\infty} \langle y \mid x_n - \widehat{x}\rangle = 0.$$

  On note $x_n \rightharpoonup \widehat{x}$.

<u>Remarque</u> : Dans un espace de Hilbert de dimension finie, les convergences forte et faible sont équivalentes.

Algorithmes de points fixes : notions de convergence

Soit $\mathcal{H}$ un espace de Hilbert.
Soit $(x_n)_{n\in\mathbb{N}}$ une suite d'éléments de $\mathcal{H}$.
$(x_n)_{n\in\mathbb{N}}$ **converge faiblement** ssi

- $(x_n)_{n\in\mathbb{N}}$ est bornée

  et

- $(x_n)_{n\in\mathbb{N}}$ possède **au plus** un **point d'accumulation** dans la topologie faible.

- $\widehat{x}$ est un point d'accumulation de $(x_n)_{n\in\mathbb{N}}$ dans la topologie faible s'il existe une sous-suite $(x_{n_k})_{k\in\mathbb{N}}$ de $(x_n)_{n\in\mathbb{N}}$ convergeant faiblement vers $\widehat{x}$.

Algorithmes de points fixes : notions de convergence

Soit $\mathcal{H}$ un espace de Hilbert.
Soit $(x_n)_{n\in\mathbb{N}}$ une suite d'éléments de $\mathcal{H}$.
$(x_n)_{n\in\mathbb{N}}$ **converge faiblement** ssi

- $(x_n)_{n\in\mathbb{N}}$ est bornée

  et

- $(x_n)_{n\in\mathbb{N}}$ possède **au plus** un **point d'accumulation** dans la topologie faible.

Illustration :

x_0	x_1	x_2	x_3	x_4	x_5	...
1	-1	1	-1	1	-1	...

$\rightarrow$ $(x_n)_{n\in\mathbb{N}}$ est bornée mais possède 2 points d'accumulations -1 et 1.
$\rightarrow$ $(x_n)_{n\in\mathbb{N}}$ ne converge pas.

Algorithmes de points fixes : notions de convergence

Lemme 1

Soit $\mathcal{H}$ un espace de Hilbert et $D \subset \mathcal{H}$ non vide.
Soit $(x_n)_{n\in\mathbb{N}}$ une suite d'éléments de D.
$(x_n)_{n\in\mathbb{N}}$ converge faiblement vers un point de D si

- pour tout $x \in D$, $(\|x_n - x\|)_{n\in\mathbb{N}}$ converge

  et

- tout point d'accumulation de $(x_n)_{n\in\mathbb{N}}$ dans la topologie faible appartient à D.

Algorithmes de points fixes : notions de convergence

Preuve :

Si $(\|x_n - x\|)_{n\in\mathbb{N}}$ converge alors $(\|x_n - x\|)_{n\in\mathbb{N}}$ et donc $(x_n)_{n\in\mathbb{N}}$ sont bornées.
Supposons que $(x_{n_k})_{k\in\mathbb{N}}$ et $(x_{n_\ell})_{\ell\in\mathbb{N}}$ soient telles que $x_{n_k} \rightharpoonup \widehat{x}$ et $x_{n_\ell} \rightharpoonup \widehat{x}'$ où $(\widehat{x}, \widehat{x}') \in D^2$. Pour tout $n \in \mathbb{N}$,

$$2\left\langle x_n \mid \widehat{x}' - \widehat{x}\right\rangle = \|x_n - \widehat{x}\|^2 - \|x_n - \widehat{x}'\|^2 - \|\widehat{x}\|^2 + \|\widehat{x}'\|^2.$$

Puisque $(\|x_n - \widehat{x}\|)_{n\in\mathbb{N}}$ et $(\|x_n - \widehat{x}'\|)_{n\in\mathbb{N}}$ convergent, il existe $\alpha \in \mathbb{R}$ tel que $\langle x_n \mid \widehat{x}' - \widehat{x}\rangle \to \alpha$ et donc $\langle x_{n_k} \mid \widehat{x}' - \widehat{x}\rangle \to \langle \widehat{x} \mid \widehat{x}' - \widehat{x}\rangle = \alpha$.
De la même façon, $\langle \widehat{x}' \mid \widehat{x}' - \widehat{x}\rangle = \alpha$. D'où $\|\widehat{x}' - \widehat{x}\|^2 = 0 \Rightarrow \widehat{x} = \widehat{x}'$.

Algorithmes de points fixes : suites Fejér-monotones

Soit $\mathcal{H}$ un espace de Hilbert et $D \subset \mathcal{H}$ non vide.
Soit $(x_n)_{n\in\mathbb{N}}$ une suite d'éléments de $\mathcal{H}$.
$(x_n)_{n\in\mathbb{N}}$ est une suite Fejér-monotone par rapport à D si

$$(\forall x \in D)(\forall n \in \mathbb{N}) \qquad \|x_{n+1} - x\| \leq \|x_n - x\|.$$

Soit $\mathcal{H}$ un espace de Hilbert et $D \subset \mathcal{H}$ non vide.
Soit $(x_n)_{n\in\mathbb{N}}$ une suite Fejér-monotone par rapport à D alors

- $(x_n)_{n\in\mathbb{N}}$ est bornée.
- pour tout $x \in D$, $(\|x_n - x\|)_{n\in\mathbb{N}}$ converge .

Algorithmes de points fixes : suites Fejér-monotones

Convergence d'une suite Fejér-monotone

Soit $\mathcal{H}$ un espace de Hilbert et $D \subset \mathcal{H}$ non vide.
Soit $(x_n)_{n\in\mathbb{N}}$ une suite d'éléments de $\mathcal{H}$.
$(x_n)_{n\in\mathbb{N}}$ **converge faiblement** vers un point de D si

- $(x_n)_{n\in\mathbb{N}}$ une suite Fejér-monotone par rapport à D

  et
- tout point d'accumulation de $(x_n)_{n\in\mathbb{N}}$ dans la topologie faible appartient à D.

Algorithmes de points fixes : suites Fejér-monotones

Lemme 2

Soit $\mathcal{H}$ un espace de Hilbert et C un convexe fermé non vide de $\mathcal{H}$.
Si $(x_n)_{n\in\mathbb{N}}$ est une suite de C convergeant faiblement vers $\widehat{x}$ alors $\widehat{x} \in C$.

Algorithmes de points fixes : suites Fejér-monotones

Lemme 2

Soit $\mathcal{H}$ un espace de Hilbert et C un convexe fermé non vide de $\mathcal{H}$.
Si $(x_n)_{n\in\mathbb{N}}$ est une suite de C convergeant faiblement vers $\widehat{x}$ alors $\widehat{x} \in C$.

<u>Preuve</u> :

On a $\widehat{x} - P_C\widehat{x} \in N_C\big(P_C\widehat{x}\big)$.

Puisque $(\forall n \in \mathbb{N})\ x_n \in C$, on a

$$\langle x_n - P_C\widehat{x} \mid \widehat{x} - P_C\widehat{x}\rangle \leq 0.$$

En utilisant le fait que $x_n \rightharpoonup \widehat{x}$, on en déduit que $\|\widehat{x} - P_C\widehat{x}\|^2 = 0$, d'où $\widehat{x} = P_C(\widehat{x}) \in C$.

Opérateur contractant : définitions

Soit $\mathcal{H}$ un espace de Hilbert. Soit $T : \mathcal{H} \to 2^{\mathcal{H}}$.
L'ensemble des points fixes de B est défini

$$\operatorname{Fix} T = \{x \in \mathcal{H} \mid x \in Tx\}$$

Soit $\mathcal{H}$ un espace de Hilbert et soit $C \subset \mathcal{H}$ non-vide.
Soit $T : C \to \mathcal{H}$.
T est une contraction si $\left(\forall (x,y) \in C^2\right) \quad \| Tx - Ty\| \leq \|x - y\|$.

Opérateur contractant : algorithmes de points fixes

Principe de demi-fermeture

Soient $\mathcal{H}$ un espace de Hilbert et C un convexe fermé non vide de $\mathcal{H}$.
Soit $T\colon C \to \mathcal{H}$ un opérateur contractant.
Si $(x_n)_{n\in\mathbb{N}}$ est une suite de C convergeant faiblement vers $\widehat{x}$ et si $Tx_n - x_n \to 0$ alors $\widehat{x} \in \operatorname{Fix} T$.

Opérateur contractant : algorithmes de points fixes

Principe de demi-fermeture

Soient $\mathcal{H}$ un espace de Hilbert et C un convexe fermé non vide de $\mathcal{H}$.
Soit $T : C \to \mathcal{H}$ un opérateur contractant.
Si $(x_n)_{n\in\mathbb{N}}$ est une suite de C convergeant faiblement vers $\widehat{x}$ et si $Tx_n - x_n \to 0$ alors $\widehat{x} \in \operatorname{Fix} T$.

Preuve :

$x_n \rightharpoonup \widehat{x} \Rightarrow \widehat{x} \in C$ et $T\widehat{x}$ défini. Pour tout $n \in \mathbb{N}$,

$$\|x_n - T\widehat{x}\|^2 = \|x_n - \widehat{x}\|^2 + \|\widehat{x} - T\widehat{x}\|^2 + 2\left\langle x_n - \widehat{x} \mid \widehat{x} - T\widehat{x}\right\rangle$$

$$\|x_n - T\widehat{x}\|^2 = \|x_n - Tx_n\|^2 + \|Tx_n - T\widehat{x}\|^2 + 2\left\langle x_n - Tx_n \mid Tx_n - T\widehat{x}\right\rangle$$

$$\Rightarrow \|\widehat{x} - T\widehat{x}\|^2 = \|x_n - Tx_n\|^2 + \|Tx_n - T\widehat{x}\|^2 - \|x_n - \widehat{x}\|^2 + 2\left\langle x_n - Tx_n \mid Tx_n - T\widehat{x}\right\rangle - 2\left\langle x_n - \widehat{x} \mid \widehat{x} - T\widehat{x}\right\rangle$$

Opérateur contractant : algorithmes de points fixes

Principe de demi-fermeture

Soient $\mathcal{H}$ un espace de Hilbert et C un convexe fermé non vide de $\mathcal{H}$.
Soit $T\colon C \to \mathcal{H}$ un opérateur contractant.
Si $(x_n)_{n\in\mathbb{N}}$ est une suite de C convergeant faiblement vers $\widehat{x}$ et si $Tx_n - x_n \to 0$ alors $\widehat{x} \in \mathrm{Fix}\, T$.

Preuve :

$$\begin{aligned}\|\widehat{x} - T\widehat{x}\|^2 =& \|x_n - Tx_n\|^2 + \|Tx_n - T\widehat{x}\|^2 - \|x_n - \widehat{x}\|^2 \\ &+ 2\left\langle x_n - Tx_n \mid Tx_n - T\widehat{x}\right\rangle - 2\left\langle x_n - \widehat{x} \mid \widehat{x} - T\widehat{x}\right\rangle .\end{aligned}$$

T étant une contraction et, en utilisant l'inégalité de Cauchy-Schwarz,

$$\begin{aligned}\|\widehat{x} - T\widehat{x}\|^2 &\leq \|x_n - Tx_n\|^2 + 2\|x_n - Tx_n\|\|Tx_n - T\widehat{x}\| - 2\left\langle x_n - \widehat{x} \mid \widehat{x} - T\widehat{x}\right\rangle \\ &\leq \|x_n - Tx_n\|^2 + 2\|x_n - Tx_n\|\|x_n - \widehat{x}\| - 2\left\langle x_n - \widehat{x} \mid \widehat{x} - T\widehat{x}\right\rangle .\end{aligned}$$

$x_n \rightharpoonup \widehat{x} \Rightarrow (x_n)_{n\in\mathbb{N}}$ borné, d'où le résultat par passage à la limite.

Opérateur contractant : algorithmes de points fixes

Soit $\mathcal{H}$ un espace de Hilbert et C un convexe fermé non vide de $\mathcal{H}$.
Soit $T\colon C \to C$ un opérateur contractant tel que $\operatorname{Fix} T \neq \varnothing$.
Soit $x_0 \in C$,

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = Tx_n.$$

Si $x_n - Tx_n \to 0$, alors $(x_n)_{n\in\mathbb{N}}$ converge faiblement vers un point de $\operatorname{Fix} T$.

Opérateur contractant : algorithmes de points fixes

Soit $\mathcal{H}$ un espace de Hilbert et C un convexe fermé non vide de $\mathcal{H}$.
Soit $T : C \to C$ un opérateur contractant tel que $\operatorname{Fix} T \neq \varnothing$.
Soit $x_0 \in C$,

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = Tx_n.$$

Si $x_n - Tx_n \to 0$, alors $(x_n)_{n\in\mathbb{N}}$ converge faiblement vers un point de $\operatorname{Fix} T$.

Preuve :
Pour tout $n \in \mathbb{N}$ et $y \in \operatorname{Fix} T$, $\|x_{n+1} - y\| \leq \|Tx_n - Ty\| \leq \|x_n - y\|$.
$(x_n)_{n\in\mathbb{N}}$ est donc Fejér-monotone par rapport à $\operatorname{Fix} T$.
Soit $(x_{n_k})_{k\in\mathbb{N}}$ une sous-suite de $(x_n)_{n\in\mathbb{N}}$ telle $x_{n_k} \rightharpoonup \widehat{x}$ où $\widehat{x} \in \mathcal{H}$.
Par hypothèse $x_{n_k} - Tx_{n_k} \to 0$ et donc, d'après le principe de demi-fermeture, $\widehat{x} \in \operatorname{Fix} T$.
Ceci assure la convergence faible de $(x_n)_{n\in\mathbb{N}}$.

Algorithmes de points fixes : suites Fejér-monotones

Algorithme de Krasnosel'skii-Mann

Soit $\mathcal{H}$ un espace de Hilbert et C un convexe fermé non vide de $\mathcal{H}$.
Soit $T\colon C \to C$ un opérateur contractant tel que $\operatorname{Fix} T \neq \varnothing$.
Soit $(\lambda_n)_{n\in\mathbb{N}}$ une suite dans $[0,1]$ telle que

$$\sum_{n\in\mathbb{N}} \lambda_n(1-\lambda_n) = +\infty.$$

Soit $x_0 \in C$ et $(\forall n \in \mathbb{N})\ x_{n+1} = x_n + \lambda_n(Tx_n - x_n)$. Les propriétés suivantes sont satisfaites :

- $(x_n)_{n\in\mathbb{N}}$ est Fejér-monotone par rapport à $\operatorname{Fix} T$.
- $(Tx_n - x_n)_{n\in\mathbb{N}}$ converge fortement vers 0.
- $(x_n)_{n\in\mathbb{N}}$ converge faiblement vers un point de $\operatorname{Fix} T$.

Algorithmes de points fixes : suites Fejér-monotones

Preuve :
Pour tout $n \in \mathbb{N}$, par combinaison convexe, $x_n \in C$.

Fejér-monotonicité par rapport à $\operatorname{Fix} T$: $(\forall x \in \operatorname{Fix} T)(\forall n \in \mathbb{N})$

$$
\begin{aligned}
& \|x_{n+1} - x\|^2 \\
=& \|x_n + \lambda_n (Tx_n - x_n) - x\|^2 \\
=& \|(1-\lambda_n)(x_n - x) + \lambda_n (Tx_n - x)\|^2 \\
=& (1-\lambda_n)^2 \|x_n - x\|^2 + \lambda_n^2 \|Tx_n - x\|^2 - 2\lambda_n (1-\lambda_n) \langle x - x_n \mid Tx_n - x \rangle \\
=& (1-\lambda_n) \|x_n - x\|^2 + \lambda_n \|Tx_n - x\|^2 - \lambda_n (1-\lambda_n) \|Tx_n - x + x - x_n\|^2 \\
=& (1-\lambda_n) \|x_n - x\|^2 + \lambda_n \|Tx_n - Tx\|^2 - \lambda_n (1-\lambda_n) \|Tx_n - x_n\|^2 \\
\leq& (1-\lambda_n) \|x_n - x\|^2 + \lambda_n \|x_n - x\|^2 - \lambda_n (1-\lambda_n) \|Tx_n - x_n\|^2 \\
\leq& \|x_n - x\|^2 .
\end{aligned}
$$

Algorithmes de points fixes : suites Fejér-monotones

Preuve :
Montrons que $Tx_n - x_n \to 0$.
On déduit de $\|x_{n+1} - x\|^2 \leq \|x_n - x\|^2 - \lambda_n(1-\lambda_n)\|Tx_n - x_n\|^2$ que

$$\sum_{n\in\mathbb{N}} \lambda_n(1-\lambda_n)\|Tx_n - x_n\|^2 \leq \|x_0 - x\|^2$$

$$\Rightarrow (\forall n \in \mathbb{N}) \qquad \inf_{k\geq n} \|Tx_k - x_k\|^2 \sum_{k=n}^{+\infty} \lambda_k(1-\lambda_k) \to 0.$$

Les hypothèses sur la suite $(\lambda_n)_{n\in\mathbb{N}}$ conduisent à $\liminf_{n\to+\infty} \|Tx_n - x_n\| = 0$.

Algorithmes de points fixes : suites Fejér-monotones

Preuve :
Les hypothèses sur la suite $(\lambda_n)_{n\in\mathbb{N}}$ conduisent à $\liminf_{n\to+\infty} \|Tx_n - x_n\| = 0$. De plus, en utilisant l'hypothèse que T est une contraction

$$\begin{aligned} \|Tx_{n+1} - x_{n+1}\| &= \|Tx_{n+1} - Tx_n + (1-\lambda_n)(Tx_n - x_n)\| \\ &\leq \|x_{n+1} - x_n\| + (1-\lambda_n)\|Tx_n - x_n\| \\ &= \|Tx_n - x_n\|. \end{aligned}$$

Par conséquent, $(\|Tx_n - x_n\|)_{n\in\mathbb{N}}$ converge et

$$Tx_n - x_n \to 0.$$

Algorithmes de points fixes : suites Fejér-monotones

Preuve :
Soit $(x_{n_k})_{k\in\mathbb{N}}$ une sous-suite de $(x_n)_{n\in\mathbb{N}}$ telle que $x_{n_k} \rightharpoonup \widehat{x}$. D'après le principe de demi-fermeture, $Tx_{n_k} - x_{n_k} \to 0$, implique que $\widehat{x} \in \mathrm{Fix}\, T$.
La convergence faible de $(x_n)_{n\in\mathbb{N}}$ vers x se déduit de la Fejér-monotonicité de $(x_n)_{n\in\mathbb{N}}$ par rapport à $\mathrm{Fix}\, T$.

Opérateur α-moyenné : définition

Soit $\mathcal{H}$ un espace de Hilbert et soit $C \subset \mathcal{H}$ non-vide.
Soit $A : C \to \mathcal{H}$ et soit $\alpha \in]0,1[$.
A est α-moyenné s'il existe une contraction $R\colon C \to \mathcal{H}$ tel que

$$A = (1-\alpha)\mathrm{Id} + \alpha R.$$

Soit $\mathcal{H}$ un espace de Hilbert et soit $C \subset \mathcal{H}$ non-vide.
Soit $A : C \to \mathcal{H}$ et soit $\alpha \in]0,1[$.
A est α-moyenné si

$$(\forall (x,y) \in C^2) \quad \|Ax - Ay\|^2 + \frac{1-\alpha}{\alpha}\|(\mathrm{Id}-A)x - (\mathrm{Id}-A)y\|^2 \leq \|x-y\|^2.$$

Opérateur α-moyenné : exemples

Soient $\mathcal{H}$ un espace de Hilbert, $f \in \Gamma_0(\mathcal{H})$, $\nu \in \,]0, +\infty[$.
Si f est différentiable et de gradient ν-lipschitzien alors $\mathrm{Id} - \nabla f$ est $\nu/2$-moyenné.

Remarque : $\mathrm{Id} - \nabla f$ est l'opérateur de descente de gradient.

Opérateur α-moyenné : exemples

Preuve : 1) Lemme de descente

Pour tout $(x, y) \in \mathcal{H}^2$ et $t \in \mathbb{R}$, soit $\varphi(t) = f(x + t(y - x))$.
φ est différentiable et $\varphi'(t) = \langle y - x \mid \nabla f(x + t(y - x))\rangle$. On a alors

$$\varphi(1) - \varphi(0) = \int_0^1 \varphi'(t)dt$$

$$\Leftrightarrow \; f(y) - f(x) - \langle y - x \mid \nabla f(x)\rangle = \int_0^1 \langle y - x \mid \nabla f(x + t(y - x)) - \nabla f(x)\rangle \, dt.$$

Or, d'après l'inégalité de Cauchy-Schwarz,

$$\begin{aligned}&\langle y - x \mid \nabla f(x + t(y - x)) - \nabla f(x)\rangle \\ \leq\; &\|y - x\| \|\nabla f(x + t(y - x)) - \nabla f(x)\| \leq t\nu\|y - x\|^2.\end{aligned}$$

Par conséquent,

$$\boxed{\big(\forall (x, y) \in \mathcal{H}^2\big) \quad f(y) \leq f(x) + \langle y - x \mid \nabla f(x)\rangle + \frac{\nu}{2}\|y - x\|^2.}$$

Opérateur α-moyenné : exemples

Preuve : 2) $\mathrm{Id} - \nabla f$ est α-moyenné

D'après le lemme de descente, pour tout $(x, y, z) \in \mathcal{H}^3$,

$$\begin{aligned} f^*(\nabla f(y)) &\geq \langle z \mid \nabla f(y) \rangle - f(z) \\ &\geq \langle z \mid \nabla f(y) - \nabla f(x) \rangle + \langle x \mid \nabla f(x) \rangle - f(x) - \frac{\nu}{2}\|z - x\|^2. \end{aligned}$$

De plus, d'après l'inégalité de Fenchel-Young,

$$\langle x \mid \nabla f(x) \rangle - f(x) = f^*(\nabla f(x)).$$

D'où

$$f^*(\nabla f(y)) \geq \langle z \mid \nabla f(y) - \nabla f(x) \rangle + f^*(\nabla f(x)) - \frac{\nu}{2}\|z - x\|^2$$

Opérateur α-moyenné : exemples

Preuve : 2) $\mathrm{Id} - \nabla f$ est α-moyenné

D'où

$$f^*(\nabla f(y)) \geq \langle z \mid \nabla f(y) - \nabla f(x)\rangle + f^*(\nabla f(x)) - \frac{\nu}{2}\|z - x\|^2$$

$$= f^*(\nabla f(x)) + \langle x \mid \nabla f(y) - \nabla f(x)\rangle + \langle z - x \mid \nabla f(y) - \nabla f(x)\rangle - \frac{\nu}{2}\|z - x\|^2.$$

On en déduit que

$$\begin{aligned} f^*(\nabla f(y)) \geq & f^*(\nabla f(x)) + \langle x \mid \nabla f(y) - \nabla f(x)\rangle \\ & + \left(\nu\|\cdot\|^2/2\right)^*\left(\nabla f(y) - \nabla f(x)\right) \\ \geq & f^*(\nabla f(x)) + \langle x \mid \nabla f(y) - \nabla f(x)\rangle + \frac{1}{2\nu}\|\nabla f(y) - \nabla f(x)\|^2. \end{aligned}$$

Par conséquent,

$$\boxed{f^*(\nabla f(y)) \geq f^*(\nabla f(x)) + \langle x \mid \nabla f(y) - \nabla f(x)\rangle + \frac{1}{2\nu}\|\nabla f(y) - \nabla f(x)\|^2.}$$

Opérateur α-moyenné : exemples

Preuve : 2) $\mathrm{Id} - \nabla f$ est α-moyenné

Pour tout $(x, y) \in \mathcal{H}^2$,

$$f^*(\nabla f(y)) \geq f^*(\nabla f(x)) + \langle x \mid \nabla f(y) - \nabla f(x) \rangle + \frac{1}{2\nu}\|\nabla f(y) - \nabla f(x)\|^2$$

et symétriquement

$$f^*(\nabla f(x)) \geq f^*(\nabla f(y)) + \langle y \mid \nabla f(x) - \nabla f(y) \rangle + \frac{1}{2\nu}\|\nabla f(x) - \nabla f(y)\|^2.$$

En sommant,

$$-\langle y - x \mid \nabla f(y) - \nabla f(x) \rangle + \frac{1}{\nu}\|\nabla f(x) - \nabla f(y)\|^2 \leq 0.$$

Par conséquent

$$\boxed{\|(\mathrm{Id} - \nabla f)x - (\mathrm{Id} - \nabla f)y\|^2 + \frac{1 - \nu/2}{\nu/2}\|\nabla f(x) - \nabla f(y)\|^2 \leq \|x - y\|^2.}$$

Opérateur α-moyenné : exemples

Soient $\mathcal{H}$ un espace de Hilbert, $f \in \Gamma_0(\mathcal{H})$, $\nu \in \,]0,+\infty[$ et $\gamma \in \,]0,+\infty[$. Si f est différentiable et de gradient ν-lipschitzien alors $\mathrm{Id} - \gamma\nabla f$ est $\gamma\nu/2$-moyenné.

Remarque : $\mathrm{Id} - \gamma\nabla f$ est l'opérateur de descente de gradient.

Opérateur α-moyenné : exemples

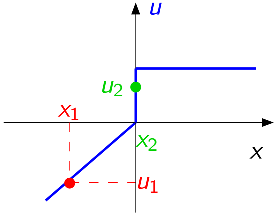
Soient $\mathcal{H}$ un espace de Hilbert, $f \in \Gamma_0(\mathcal{H})$.
$\mathrm{prox}_f = (\mathrm{Id} + \partial f)^{-1}$ est 1/2-moyenné.

Opérateur α-moyenné : exemples

Soient $\mathcal{H}$ un espace de Hilbert, $f \in \Gamma_0(\mathcal{H})$.
$\mathrm{prox}_f = (\mathrm{Id} + \partial f)^{-1}$ est 1/2-moyenné.

Preuve :

- Rappel : $\partial f(x) = \{u \in \mathcal{H} \,|\, (\forall y \in \mathcal{H})\ \langle y - x | u\rangle + f(x) \leq f(y)\}$



Opérateur α-moyenné : exemples

Soient $\mathcal{H}$ un espace de Hilbert, $f \in \Gamma_0(\mathcal{H})$.
$\mathrm{prox}_f = (\mathrm{Id} + \partial f)^{-1}$ est 1/2-moyenné.

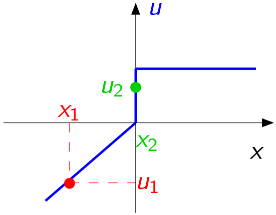
Preuve :

- Rappel : $\partial f(x) = \{u \in \mathcal{H} \,|\, (\forall y \in \mathcal{H})\ \langle y - x | u \rangle + f(x) \leq f(y)\}$
- Soient $u_1 \in \partial f(x_1)$ et $u_2 \in \partial f(x_2)$.

  D'après la définition :

$$\langle x_2 - x_1 | u_1 \rangle + f(x_1) \leq f(x_2)$$
$$\langle x_1 - x_2 | u_2 \rangle + f(x_2) \leq f(x_1)$$

  ce qui conduit à $\boxed{\langle x_1 - x_2 \,|\, u_1 - u_2 \rangle \geq 0}$.

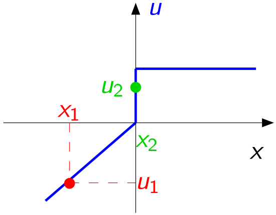


Opérateur α-moyenné : exemples

Soient $\mathcal{H}$ un espace de Hilbert, $f \in \Gamma_0(\mathcal{H})$.
$\mathrm{prox}_f = (\mathrm{Id} + \partial f)^{-1}$ est 1/2-moyenné.

Preuve :

- Rappel : $\partial f(x) = \{u \in \mathcal{H} \,|\, (\forall y \in \mathcal{H})\ \langle y - x | u\rangle + f(x) \leq f(y)\}$
- Soient $u_1 \in \partial f(x_1)$ et $u_2 \in \partial f(x_2)$.
  D'après la définition :
  $$\langle x_2 - x_1 | u_1\rangle + f(x_1) \leq f(x_2)$$
  $$\langle x_1 - x_2 | u_2\rangle + f(x_2) \leq f(x_1)$$
  ce qui conduit à $\boxed{\langle x_1 - x_2 \,|\, u_1 - u_2\rangle \geq 0}$.



- *Remarque* : ∂f est un opérateur monotone.

Opérateur α-moyenné : exemples

<u>Preuve</u> : (suite)
Soient $u_1 \in \partial f(x_1)$ et $u_2 \in \partial f(x_2)$

$$\langle x_1 - x_2 \,|\, u_1 - u_2 \rangle \geq 0 \Leftrightarrow \quad \langle x_1 - x_2 \,|\, x_1 - x_2 + u_1 - u_2 \rangle \geq \|x_1 - x_2\|^2$$

En posant $u_1' \in (\mathrm{Id} + \partial f)x_1$ et $u_2' \in (\mathrm{Id} + \partial f)x_2$, on obtient

$$\langle x_1 - x_2 \,|\, u_1' - u_2' \rangle \geq \|x_1 - x_2\|^2$$

Puis par définition de l'opérateur proximal

$$\langle \mathrm{prox}_f u_1' - \mathrm{prox}_f u_2' \,|\, u_1' - u_2' \rangle \geq \|\mathrm{prox}_f u_1' - \mathrm{prox}_f u_2'\|^2$$

On en déduit que prox_f est $1/2$-moyenné, i.e,

$$\|u_1' - u_2'\|^2 \geq \|\mathrm{prox}_f u_1' - \mathrm{prox}_f u_2'\|^2 + \|(\mathrm{Id} - \mathrm{prox}_f)u_1' - (\mathrm{Id} - \mathrm{prox}_f)u_2'\|^2$$

Opérateur α-moyenné : exemples

Soient $\mathcal{H}$ un espace de Hilbert, $f \in \Gamma_0(\mathcal{H})$ et soit $\gamma > 0$.
$\mathrm{prox}_{\gamma f} = (\mathrm{Id} + \gamma \partial f)^{-1}$ est $1/2$-moyenné.

Opérateur α-moyenné : algorithmes de points fixes

Soit $\mathcal{H}$ un espace de Hilbert et soit $\alpha \in]0,1[$.
Soit $T\colon \mathcal{H} \to \mathcal{H}$ un opérateur α-moyenné avec $\alpha \in]0,1[$ tel que $\operatorname{Fix} T \neq \varnothing$.
Soit $(\lambda_n)_{n\in\mathbb{N}}$ une suite dans $[0, 1/\alpha]$ telle que

$$\sum_{n\in\mathbb{N}} \lambda_n(1-\alpha\lambda_n) = +\infty.$$

Soit $x_0 \in \mathcal{H}$ et $(\forall n \in \mathbb{N})\ x_{n+1} = x_n + \lambda_n(Tx_n - x_n)$. Les propriétés suivantes sont satisfaites :

- $(x_n)_{n\in\mathbb{N}}$ est Fejér-monotone par rapport à $\operatorname{Fix} T$.
- $(Tx_n - x_n)_{n\in\mathbb{N}}$ converge fortement vers 0.
- $(x_n)_{n\in\mathbb{N}}$ converge faiblement vers un point de $\operatorname{Fix} T$.

Opérateur α-moyenné : algorithmes de points fixes

Preuve :
T étant α-moyenné, il existe une contraction R telle que $T = (1-\alpha)\mathrm{Id} + \alpha R$.
Soit $(\forall n \in \mathbb{N})$ $\mu_n = \alpha\lambda_n \in [0,1]$.
Les itérations peuvent se ré-écrire

$$(\forall n \in \mathbb{N}) \qquad \begin{aligned} x_{n+1} &= x_n + \lambda_n(Tx_n - x_n) \\ &= x_n + \mu_n(Rx_n - x_n). \end{aligned}$$

Par ailleurs, $\mathrm{Fix} R = \mathrm{Fix}\, T$.
$+$ Algorithme de Krasnosel'skii-Mann.

Algorithmes d'optimisation : descente de gradient

Soit $\mathcal{H}$ un espace de Hilbert.
Soit $g \in \Gamma_0(\mathcal{H})$ de gradient ν-lipschitzien où $\nu \in \,]0, +\infty[$.
Soit $\gamma \in]0, 2/\nu[$.
Supposons que $\operatorname{Argmin} g \neq \varnothing$. Soit $x_0 \in \mathcal{H}$ et

$$(\forall n \in \mathbb{N}) \qquad x_{n+1} = x_n - \gamma \nabla g(x_n)$$

$(x_n)_{n\in\mathbb{N}}$ converge faiblement vers un minimiseur de f.

Algorithmes d'optimisation : *Forward-Backward*

<u>Preuve</u> : Soit $T = \mathrm{Id} - \gamma\nabla g$. Pour tout $x \in \mathcal{H}$,

$$x \in \mathrm{Fix}\, T \Leftrightarrow 0 \in \nabla g(x).$$

D'où $\mathrm{Fix}\, T = \mathrm{zer}\,(\nabla g) \neq \varnothing$. De plus, pour tout $n \in \mathbb{N}$,

$$x_{n+1} = x_n + \lambda_n (T x_n - x_n).$$

avec $\lambda_n \equiv 1$. Par ailleurs, $\mathrm{Id} - \gamma\nabla g$ est $\gamma\nu/2$-moyenné.
Le résultat découle alors du précédent.

Algorithmes d'optimisation : descente de gradient

Soit $\mathcal{H}$ un espace de Hilbert.
Soit $g \in \Gamma_0(\mathcal{H})$ de gradient ν-lipschitzien où $\nu \in \,]0, +\infty[$.
Soient $(\gamma_n)_{n\in\mathbb{N}}$ une suite dans $[\underline{\gamma}, \overline{\gamma}]$ où $0 < \underline{\gamma} \leq \overline{\gamma} < 2/\nu$.
Supposons que $\mathrm{Argmin}\, g \neq \varnothing$. Soit $x_0 \in \mathcal{H}$ et

$$(\forall n \in \mathbb{N}) \qquad x_{n+1} = x_n - \gamma_n \nabla g(x_n)$$

$(x_n)_{n\in\mathbb{N}}$ converge faiblement vers un minimiseur de f.

Algorithmes d'optimisation : *Forward-Backward*

Soit $\mathcal{H}$ un espace de Hilbert.
Soit $f \in \Gamma_0(\mathcal{H})$.
Soit $g \in \Gamma_0(\mathcal{H})$ de gradient ν-lipschitzien où $\nu \in\ ]0, +\infty[$.
Soient $\gamma \in]0, 2/\nu[$ et $\delta = \min\{1, 1/(\nu\gamma)\} + 1/2$.
Soit $(\lambda_n)_{n\in\mathbb{N}}$ une suite dans $[0, \delta[$ telle que $\sum_{n\in\mathbb{N}} \lambda_n(\delta - \lambda_n) = +\infty$.
Supposons que $\operatorname{Argmin}(f+g) \neq \varnothing$. Soit $x_0 \in \mathcal{H}$ et

$$(\forall n \in \mathbb{N}) \qquad \begin{cases} y_n = x_n - \gamma \nabla g(x_n) \\ x_{n+1} = x_n + \lambda_n(\operatorname{prox}_{\gamma f} y_n - x_n). \end{cases}$$

$(x_n)_{n\in\mathbb{N}}$ converge faiblement vers un minimiseur de $f+g$.

Algorithmes d'optimisation : *Forward-Backward*

Preuve : Soit $T = (\mathrm{Id} + \gamma \partial f)^{-1}(\mathrm{Id} - \gamma \nabla g)$. Pour tout $x \in \mathcal{H}$,

$$x \in \mathrm{Fix}\, T \Leftrightarrow (\mathrm{Id} - \gamma \nabla g)x \in (\mathrm{Id} + \gamma \partial f)x \Leftrightarrow 0 \in \nabla g(x) + \partial f(x).$$

D'où $\mathrm{Fix}\, T = \mathrm{zer}\,(\nabla g + \partial f) \neq \varnothing$. De plus, pour tout $n \in \mathbb{N}$,

$$x_{n+1} = x_n + \lambda_n (Tx_n - x_n).$$

Par ailleurs, $\mathrm{prox}_{\gamma f}$ est $1/2$-moyenné et $\mathrm{Id} - \gamma \nabla g$ est $\gamma\nu/2$-moyenné. On en déduit que T est α-moyenné avec

$$\alpha = \frac{2}{1 + \frac{1}{\max\{\frac{1}{2}, \frac{\gamma\nu}{2}\}}} \quad \Leftrightarrow \quad \alpha^{-1} = \delta.$$

Le résultat découle alors du précédent.

Algorithmes d'optimisation : gradient projeté

Soit $\mathcal{H}$ un espace de Hilbert.
Soit C un convexe fermé non vide de $\mathcal{H}$.
Soit $g \in \Gamma_0(\mathcal{H})$ de gradient ν-Lipschitzien où $\nu \in \,]0, +\infty[$.
Soient $\gamma \in]0, 2/\nu[$ et $\delta = \min\{1, 1/(\nu\gamma)\} + 1/2$.
Soit $(\lambda_n)_{n\in\mathbb{N}}$ une suite dans $[0, \delta[$ telle que $\sum_{n\in\mathbb{N}} \lambda_n(\delta - \lambda_n) = +\infty$.
Supposons que $\mathrm{Argmin}_{x\in C} g(x) \neq \varnothing$. Soit $x_0 \in \mathcal{H}$ et

$$(\forall n \in \mathbb{N}) \qquad \begin{cases} y_n = x_n - \gamma \nabla g(x_n) \\ x_{n+1} = x_n + \lambda_n (P_C y_n - x_n). \end{cases}$$

$(x_n)_{n\in\mathbb{N}}$ converge faiblement vers un minimiseur de g sur C.

Conclusions

- Introduction de la sous-différentielle et de l'opérateur proximal indispensable pour gérer des critères convexes non-lisses.
- Démonstration des preuves de convergence des algorithmes de gradient, gradient,projeté, forward-backward (FB) basée sur la notion d'opérateur α-moyenné.
- Il existe des versions accélérées de FB : TWIST [Bioucas-Dias, Figueiredo, 2007], FISTA [Beck, Teboulle, 2009], ...
- Autre algorithme fondamental : Douglas-Rachford (DR) [Combettes, Pesquet, 2007] dont la démonstration de convergence se base également sur l'Algorithme de Krasnosel'skii-Mann. Algorithme adapté pour minimiser une somme de deux fonctions non-nécessairement différentiales.

Conclusions

- A partir de FB et DR, de nombreux algorithmes ont récemment été proposés pour résoudre

$$\min_{x \in HH} \sum_{i=1}^{m} f_i(L_i x)$$

avec $f_i \in \Gamma_0(\mathcal{G}_i)$ et $L_i \in \mathcal{B}(H, \mathcal{G}_i)$.

  - PPXA [Combettes, Pesquet, 2008]
  - ADMM, SDMM [Figueiredo, Nowak, 2009], [Goldstein, Osher, 2009], [Steidl, Teuber, 2010]
  - Algorithme Primal-Dual [Chambolle, Pock, 2010], [Briceño-Arias, Combettes, 2011], [Combettes, Pesquet, 2012], [Condat, 2013], [Vũ, 2013]