Convex optimization

Nelly Pustelnik[1]

[1] ENS Lyon – Laboratoire de Physique – CNRS UMR 5672
pierre.borgnat@ens-lyon.fr, nelly.pustelnik@ens-lyon.fr

Optimization: why?

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)

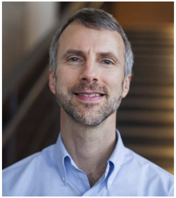
Optimization: what ?

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)

Optimization: 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)

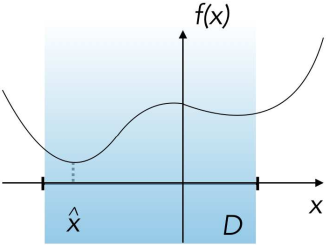
Optimization: minimization problem

- Minimization problems
  f: cost function
  We want to

$$\begin{aligned} & \text{Find } \widehat{x} \in D \quad \text{such that} \quad (\forall x \in D)\ f(\widehat{x}) \leq f(x) \\ \Leftrightarrow \quad & \text{Find } \widehat{x} \in D \quad \text{such that} \quad f(\widehat{x}) = \inf_{x \in D} f(x) \end{aligned}$$

that is

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

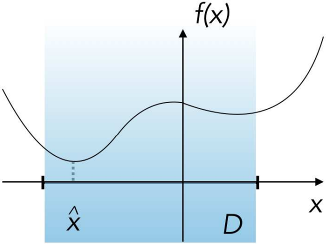


Optimization: minimization problem

- Maximization problems
  f: reward function
  We want to

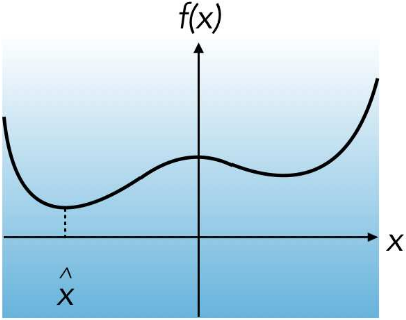
$$
\begin{aligned}
& \text{Find } \widehat{x} \in D \quad \text{such that} \quad (\forall x \in D)\ f(\widehat{x}) \geq f(x) \\
\Leftrightarrow\quad & \text{Find } \widehat{x} \in D \quad \text{such that} \quad (\forall x \in D)\ -f(\widehat{x}) \leq -f(x) \\
\Leftrightarrow\quad & \text{Find } \widehat{x} \in \underset{x \in D}{\operatorname{Argmin}} \left(-f(x)\right).
\end{aligned}
$$

Without loss of generality, we can focus on minimization problems with $f\colon D \to \left]-\infty, +\infty\right]$.



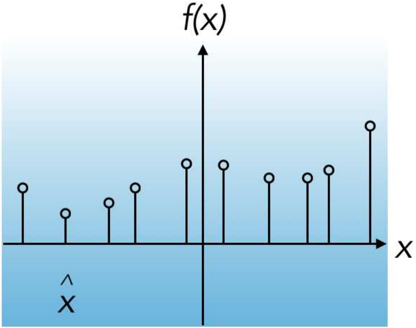
Various types of minimization problems

- $D = \mathbb{R}^N$: unconstrained problem



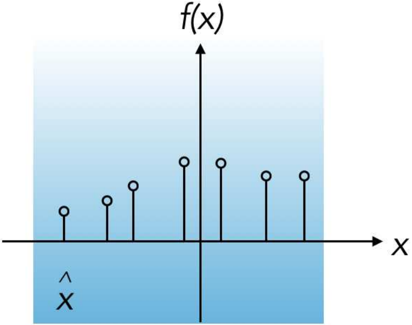
Various types of minimization problems

- $D = \mathbb{R}^N$: unconstrained problem
- D countable: discrete optimization problem



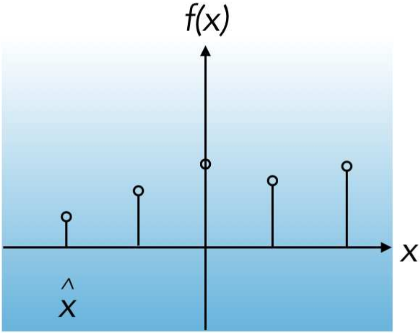
Various types of minimization problems

- $D = \mathbb{R}^N$: unconstrained problem
- D countable: discrete optimization problem



Various types of minimization problems

- $D = \mathbb{R}^N$: unconstrained problem
- D countable: discrete optimization problem



Various types of minimization problems

- D uncountable: continuous optimization problem
- Optimization problem with P equality constraints and Q inequality constraints:

$$D = \{x \in \mathbb{R}^N \mid (\forall i \in \{1, \dots, P\})\ \varphi_i(x) = \delta_i \\ \text{and } (\forall j \in \{1, \dots, Q\})\ \psi_j(x) \leq \eta_j\}$$

where $(\forall i \in \{1, \dots, P\})\ \delta_i \in \mathbb{R}$ and $\varphi_i \colon \mathbb{R}^N \to\ ]-\infty, +\infty]$,
$(\forall j \in \{1, \dots, Q\})\ \eta_j \in \mathbb{R}$ and $\psi_j \colon \mathbb{R}^N \to\ ]-\infty, +\infty]$.
If $\varphi_i \colon x \mapsto \langle x \mid u_i \rangle$ with $i \in \{1, \dots, P\}$ and $u_i \in \mathbb{R}^N$, then *linear* (or affine) equality constraint.
If $\psi_j \colon x \mapsto \langle x \mid u_j \rangle$ with $j \in \{1, \dots, Q\}$ and $u_j \in \mathbb{R}^N$, then *linear* (or affine) inequality constraint.

Various types of minimization problems

Remark:

$$\begin{aligned} & \text{Find } \widehat{x} \in \underset{x \in D}{\operatorname{Argmin}} \; f(x) \\ \Leftrightarrow \quad & \text{Find } \widehat{x} \in \underset{x \in \mathbb{R}^N}{\operatorname{Argmin}} \; \widetilde{f}(x) \end{aligned}$$

where

$$(\forall x \in \mathbb{R}^N) \quad \widetilde{f}(x) = \begin{cases} f(x) & \text{if } x \in D \\ +\infty & \text{otherwise.} \end{cases}$$

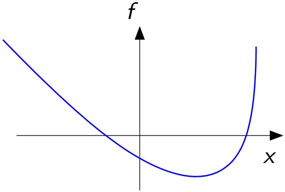
Allowing non finite valued functions leads to a unifying view of constrained and unconstrained minimization problems.

Convex/non-convex

Let $f : \mathbb{R}^N \to]-\infty, +\infty]$. An unconstrained minimization problem aims to solve:

$$\boxed{\widehat{x} \in \operatorname{Argmin}_{x \in \mathbb{R}^N} f(x)}$$

- Convex optimization and non-convex optimization



Fonction convexe

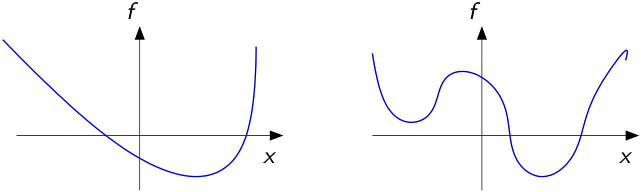
Fonction non−convexe

Convex/non-convex

Let $f : \mathbb{R}^N \to]-\infty, +\infty]$. An unconstrained minimization problem aims to solve:

$$\boxed{\widehat{x} \in \mathrm{Argmin}_{x \in \mathbb{R}^N} f(x)}$$

- Convex optimization and non-convex optimizatione



Fonction convexe Fonction non−convexe

Main questions to be addressed

1. Existence/uniqueness of a solution $\widehat{x}$?

Main questions to be addressed

1. Existence/uniqueness of a solution $\widehat{x}$?
2. Characterization of solutions: necessary/sufficient conditions for $\widehat{x}$ to be a solution.

Main questions to be addressed

1. Existence/uniqueness of a solution $\widehat{x}$?
2. Characterization of solutions: necessary/sufficient conditions for $\widehat{x}$ to be a solution.
3. Designing an algorithm to approximate a solution in the frequent case when no closed form solution is available, i.e. building a sequence $(x_n)_{n\in\mathbb{N}}$ of $\mathbb{R}^N$ such that

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

Main questions to be addressed

1. Existence/uniqueness of a solution $\widehat{x}$?
2. Characterization of solutions: necessary/sufficient conditions for $\widehat{x}$ to be a solution.
3. Designing an algorithm to approximate a solution in the frequent case when no closed form solution is available, i.e. building a sequence $(x_n)_{n\in\mathbb{N}}$ of $\mathbb{R}^N$ such that

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

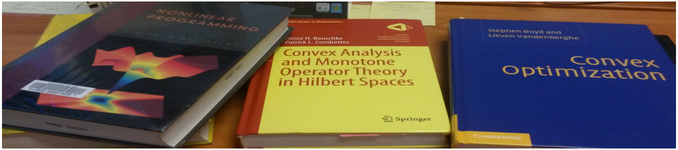
4. Evaluation of the performance of the optimization algorithm:
   - Convergence speed
     Example: If there exists $\rho \in]0,1[$ and $n^* \in \mathbb{N}$ such that $(\forall n \geq n^*)$ $\|x_{n+1} - \widehat{x}\| \leq \rho \|x_n - \widehat{x}\|$, then *Q-linear* convergence rate.
     If $\lim_{n\to+\infty} \frac{\|x_{n+1}-\widehat{x}\|}{\|x_n-\widehat{x}\|} = 0$, then *Q-superlinear* convergence rate.
   - Robustness to numerical errors
   - Amenability to parallel/distributed implementations.

Convex functions: examples

Cf. cours de Laurent Jacob

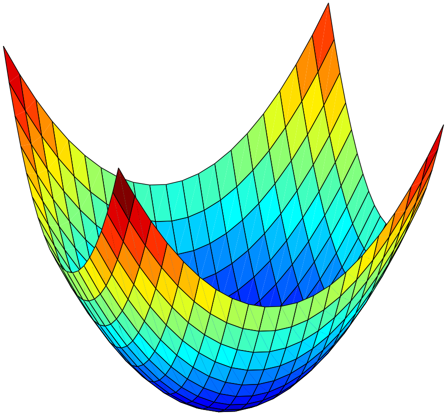
- Quadratic loss: $f(x) = \|y - Ax\|^2$ with $y \in \mathbb{R}^M$ and $A \in \mathbb{R}^{M \times N}$.
- Logistic loss: $f(x) = \log(1 + e^{-yx})$ with $y \in \{0, 1\}$
- Hinge loss: $f(x) = \max(0, 1 - yx)$ with $y \in \{0, 1\}$
- LASSO: $f(x) = \|y - Ax\|^2 + \lambda\|x\|_1$
- ...

Reference books



- **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.

Differentiable convex functions



Hilbert spaces

A (real) **Hilbert space $\mathcal{H}$** is a complete real vector space endowed with an inner product $\langle \cdot \mid \cdot \rangle$. The associated norm is

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

- Particular case: $\mathcal{H} = \mathbb{R}^N$ (Euclidean space with dimension N).

Hilbert spaces

Let $\mathcal{H}$ and $\mathcal{G}$ be two Hilbert spaces.
A linear operator $L \colon \mathcal{H} \to \mathcal{G}$ is bounded (or continuous) if

$$\|L\| = \sup_{\|x\|_{\mathcal{H}} \leq 1} \|Lx\|_{\mathcal{G}} < +\infty$$

Hilbert spaces

Let $\mathcal{H}$ and $\mathcal{G}$ be two Hilbert spaces.
A linear operator $L\colon \mathcal{H} \to \mathcal{G}$ is bounded (or continuous) if

$$\|L\| = \sup_{\|x\| \leq 1} \|Lx\| < +\infty$$

Hilbert spaces

Let $\mathcal{H}$ and $\mathcal{G}$ be two Hilbert spaces.
A linear operator $L\colon \mathcal{H} \to \mathcal{G}$ is **bounded** (or continuous) if

$$\|L\| = \sup_{\|x\| \leq 1} \|Lx\| < +\infty$$

- In finite dimension, every linear operator is bounded.

Hilbert spaces

Let $\mathcal{H}$ and $\mathcal{G}$ be two Hilbert spaces.
A linear operator $L\colon \mathcal{H} \to \mathcal{G}$ is bounded (or continuous) if

$$\|L\| = \sup_{\|x\|\leq 1} \|Lx\| < +\infty$$

- In finite dimension, every linear operator is bounded.

$\mathcal{B}(\mathcal{H}, \mathcal{G})$: Banach space of bounded linear operators from $\mathcal{H}$ to $\mathcal{G}$.

Hilbert spaces

Let $\mathcal{H}$ and $\mathcal{G}$ be two Hilbert spaces.
Let $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$. Its adjoint L^* is the operator in $\mathcal{B}(\mathcal{G}, \mathcal{H})$ defined as

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

Hilbert spaces

Let $\mathcal{H}$ and $\mathcal{G}$ be two Hilbert spaces.
Let $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$. Its adjoint L^* is the operator in $\mathcal{B}(\mathcal{G}, \mathcal{H})$ defined as

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

Hilbert spaces

Let $\mathcal{H}$ and $\mathcal{G}$ be two Hilbert spaces.
Let $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$. Its adjoint L^* is the operator in $\mathcal{B}(\mathcal{G}, \mathcal{H})$ defined as

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

Hilbert spaces

Let $\mathcal{H}$ and $\mathcal{G}$ be two Hilbert spaces.
Let $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$. Its **adjoint L^*** is the operator in $\mathcal{B}(\mathcal{G}, \mathcal{H})$ defined as

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

Example:

$$\text{If} \qquad L\colon \mathcal{H} \to \mathcal{H}^n \colon x \mapsto (x, \ldots, x)$$

$$\text{then} \qquad L^*\colon \mathcal{H}^n \to \mathcal{H} \colon y = (y_1, \ldots, y_n) \mapsto \sum_{i=1}^{n} y_i$$

Proof:

$$\langle Lx \mid y \rangle = \langle (x, \ldots, x) \mid (y_1, \ldots, y_n) \rangle = \sum_{i=1}^{n} \langle x \mid y_i \rangle = \left\langle x \mid \sum_{i=1}^{n} y_i \right\rangle$$

Hilbert spaces

Let $\mathcal{H}$ and $\mathcal{G}$ be two Hilbert spaces.
Let $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$. Its adjoint L^* is the operator in $\mathcal{B}(\mathcal{G}, \mathcal{H})$ defined as

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

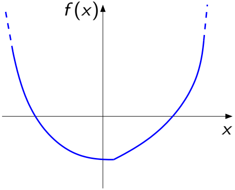
- We have $\|L^*\| = \|L\|$.
- If L is bijective (i.e. an isomorphism) then $L^{-1} \in \mathcal{B}(\mathcal{G}, \mathcal{H})$ and $(L^{-1})^* = (L^*)^{-1}$.
- If $\mathcal{H} = \mathbb{R}^N$ and $\mathcal{G} = \mathbb{R}^M$ then $L^* = L^\top$.

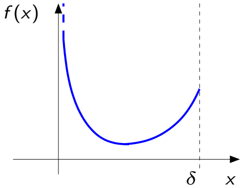
Functional analysis: definitions

Let $f : \mathcal{H} \to]-\infty, +\infty]$ where $\mathcal{H}$ is a Hilbert space.

- The domain of f is $\operatorname{dom} f = \{x \in \mathcal{H} \,|\, f(x) < +\infty\}$.
- The function f is proper if $\operatorname{dom} f \neq \varnothing$.

Domains of the functions ?



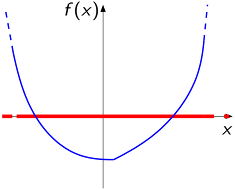


Functional analysis: definitions

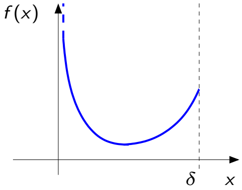
Let $f : \mathcal{H} \to \,]-\infty, +\infty]$ where $\mathcal{H}$ is a Hilbert space.

- The **domain** of f is $\operatorname{dom} f = \{x \in \mathcal{H} \mid f(x) < +\infty\}$.
- The function f is **proper** if $\operatorname{dom} f \neq \varnothing$.

<u>Domains of the functions ?</u>



$\operatorname{dom} f = \mathbb{R}$
(proper)

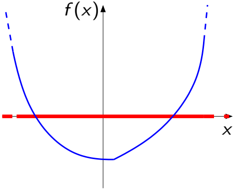


Functional analysis: definitions

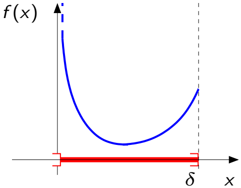
Let $f : \mathcal{H} \to]-\infty, +\infty]$ where $\mathcal{H}$ is a Hilbert space.

- The **domain** of f is $\operatorname{dom} f = \{x \in \mathcal{H} \mid f(x) < +\infty\}$.
- The function f is **proper** if $\operatorname{dom} f \neq \varnothing$.

Domains of the functions ?



$\operatorname{dom} f = \mathbb{R}$
(proper)



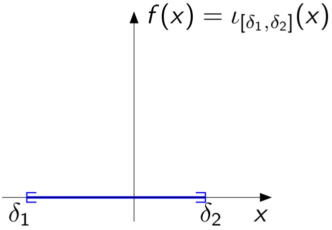
$\operatorname{dom} f =]0, \delta]$
(proper)

Functional analysis: definitions

Let $C \subset \mathcal{H}$.
The **indicator function of C** is

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

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



Epigraph

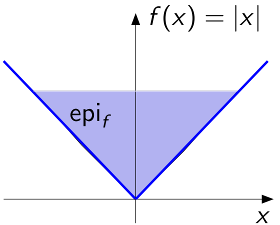
Let $f : \mathcal{H} \to \,]-\infty, +\infty]$. The **epigraph** of f is

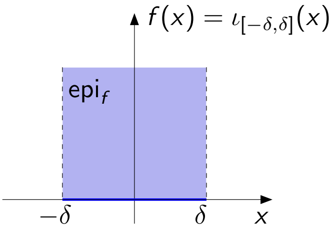
$$\operatorname{epi} f = \left\{(x, \zeta) \in \operatorname{dom} f \times \mathbb{R} \;\middle|\; f(x) \leq \zeta\right\}$$

Epigraph

Let $f : \mathcal{H} \to]-\infty, +\infty]$. The epigraph of f is

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

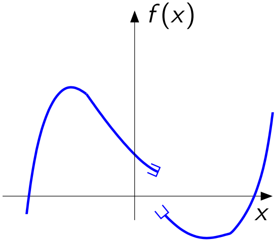


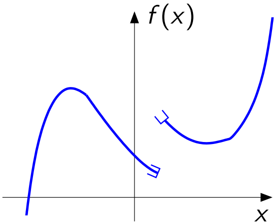


Lower semi-continuity

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

- l.s.c. functions ?

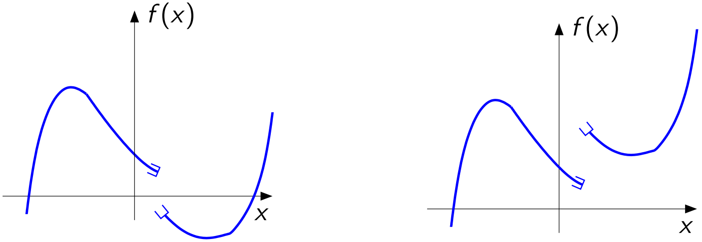




Lower semi-continuity

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

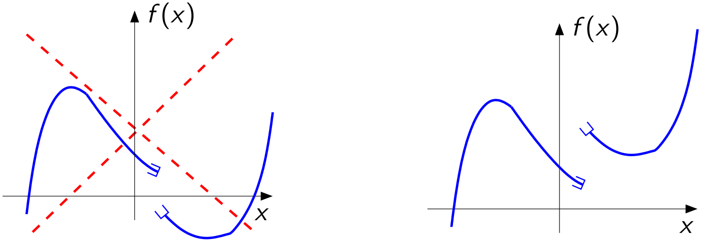
- l.s.c. functions ?



Lower semi-continuity

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

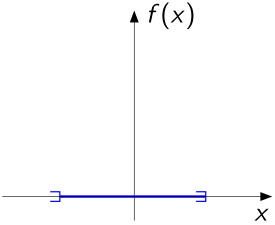
- l.s.c. functions ?

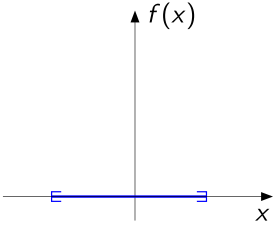


Lower semi-continuity

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

- l.s.c. functions ?

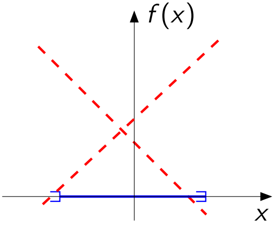


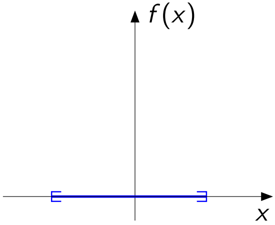


Lower semi-continuity

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

- l.s.c. functions ?

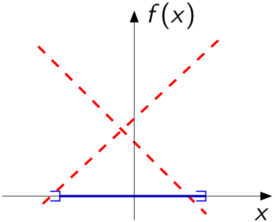


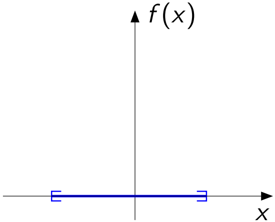


Lower semi-continuity

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

- l.s.c. functions ?





Lower semi-continuity

- Every continuous function on $\mathcal{H}$ is l.s.c.
- Every finite sum of l.s.c. functions is l.s.c.
- Let $(f_i)_{i\in I}$ be a family of l.s.c functions.
  Then, $\sup_{i\in I} f_i$ is l.s.c.

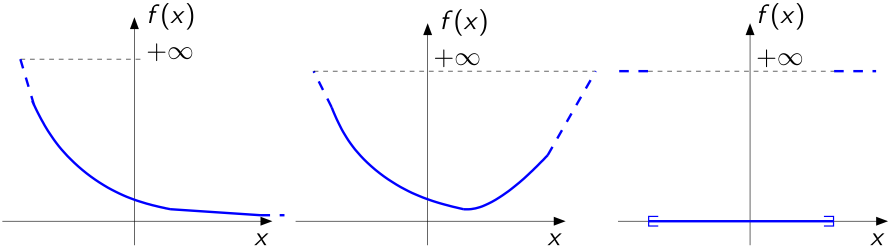
Existence of a minimizer

Let $\mathcal{H}$ be a Hilbert space. Let $f\colon \mathcal{H} \to \,]-\infty, +\infty]$.
f is **coercive** if $\lim_{\|x\|\to+\infty} f(x) = +\infty$.

Existence of a minimizer

Let $\mathcal{H}$ be a Hilbert space. Let $f : \mathcal{H} \to]-\infty, +\infty]$.
f is **coercive** if $\lim_{\|x\| \to +\infty} f(x) = +\infty$.

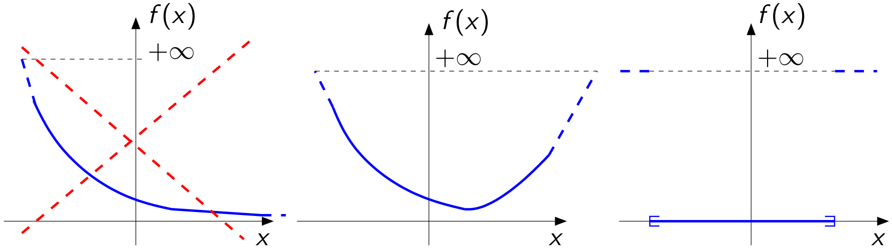
Coercive functions ?



Existence of a minimizer

Let $\mathcal{H}$ be a Hilbert space. Let $f : \mathcal{H} \to]-\infty, +\infty]$.
f is **coercive** if $\lim_{\|x\| \to +\infty} f(x) = +\infty$.

Coercive functions ?

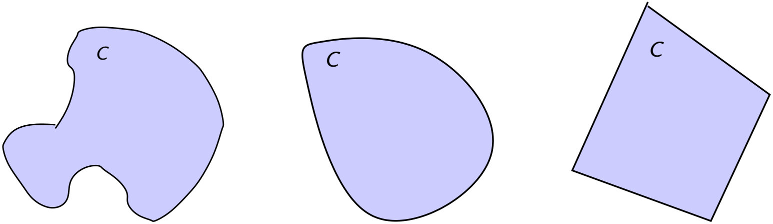


Convex set

$C \subset \mathcal{H}$ is a **convex set** if

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

Convex sets ?

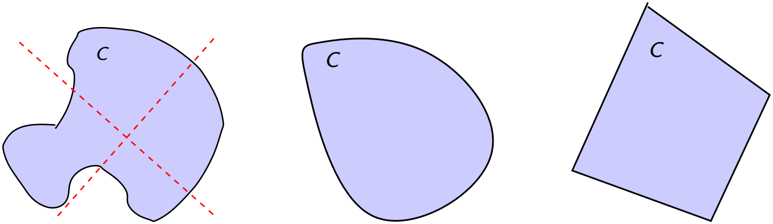


Convex set

$C \subset \mathcal{H}$ is a **convex set** if

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

Convex sets ?



Convex function: definitions

$f : \mathcal{H} \to \,]-\infty, +\infty]$ is a **convex function** if

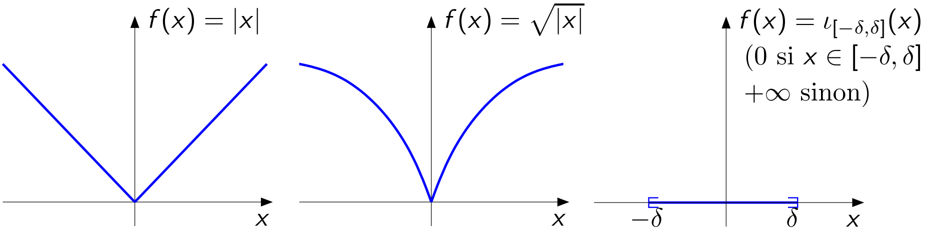
$$(\forall (x, y) \in \mathcal{H}^2)(\forall \alpha \in]0, 1[)$$
$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y)$$

Convex function: definitions

$f : \mathcal{H} \to]-\infty, +\infty]$ is a **convex function** if

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

Convex functions ?

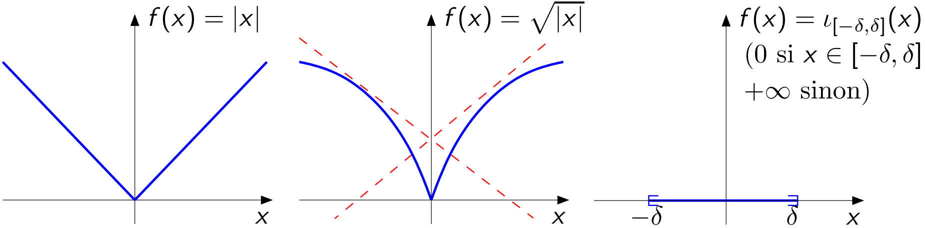


Convex function: definitions

$f : \mathcal{H} \to]-\infty, +\infty]$ is a **convex function** if

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

Convex functions ?

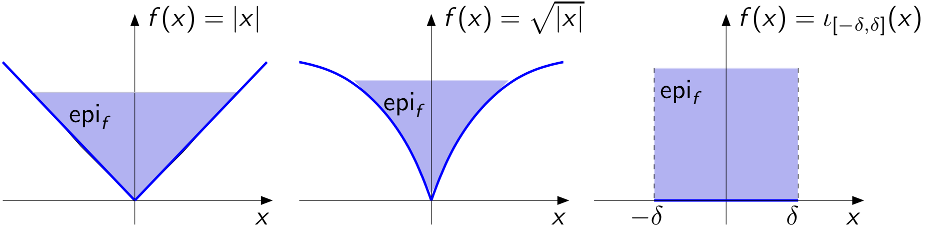


Convex functions: definition

$f : \mathcal{H} \to \,]-\infty, +\infty]$ is convex $\Leftrightarrow$ its epigraph is convex.

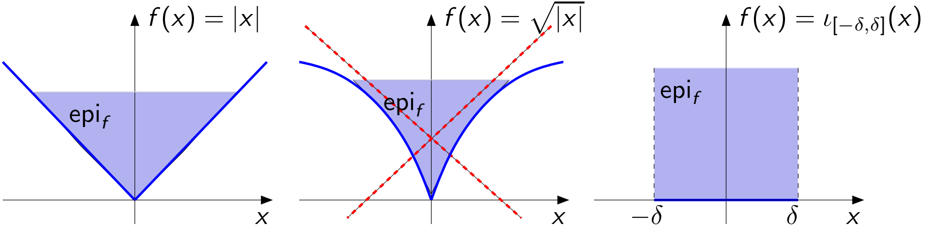
Convex functions: definition

$f : \mathcal{H} \to \,]-\infty, +\infty]$ is convex $\Leftrightarrow$ its epigraph is convex.



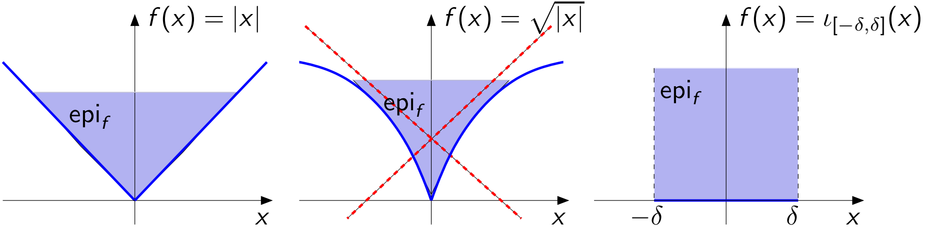
Convex functions: definition

$f : \mathcal{H} \to]-\infty, +\infty]$ is convex $\Leftrightarrow$ its epigraph is convex.



Convex functions: definition

$f : \mathcal{H} \to]-\infty, +\infty]$ is convex $\Leftrightarrow$ its epigraph is convex.



- If $f : \mathcal{H} \to]-\infty, +\infty]$ is convex, then $\operatorname{dom} f$ is convex.
- $f : \mathcal{H} \to [-\infty, +\infty[$ is concave if $-f$ is convex.

Convex functions: properties

- Every finite sum of convex functions is convex.
- Let $(f_i)_{i \in I}$ be a family of convex functions. Then, $\sup_{i \in I} f_i$ is convex.
- $\Gamma_0(\mathcal{H})$: class of convex, l.s.c., and proper functions from $\mathcal{H}$ to $]-\infty, +\infty]$.
- $\iota_C \in \Gamma_0(\mathcal{H}) \Leftrightarrow C$ is a nonempty closed convex set.
  Proof: $\text{epi}_{\iota_C} = C \times [0, +\infty[$.

Strictly convex functions

Let $\mathcal{H}$ be a Hilbert space. Let $f\colon \mathcal{H} \to \,]-\infty, +\infty]$.
f is strictly convex if

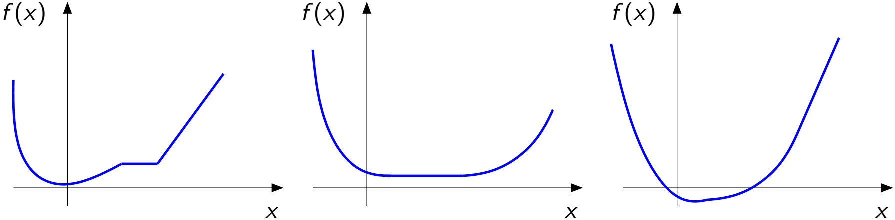
$$(\forall x \in \operatorname{dom} f)(\forall y \in \operatorname{dom} f)(\forall \alpha \in]0,1[)$$
$$x \neq y \quad \Rightarrow \quad f(\alpha x + (1-\alpha)y) < \alpha f(x) + (1-\alpha) f(y).$$

Strictly convex functions

Let $\mathcal{H}$ be a Hilbert space. Let $f : \mathcal{H} \to \,]-\infty, +\infty]$.
f is strictly convex if

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

Strictly convex functions ?

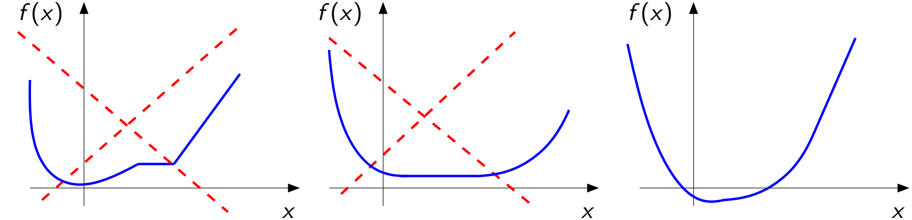


Strictly convex functions

Let $\mathcal{H}$ be a Hilbert space. Let $f\colon \mathcal{H} \to \,]-\infty, +\infty]$.
f is **strictly convex** if

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

Strictly convex functions ?



Minimizers of a convex function

Theorem

Let $\mathcal{H}$ be a Hilbert space. Let $f\colon \mathcal{H} \to \left]-\infty, +\infty\right]$ be a proper convex function such that $\mu = \inf f > -\infty$.

- $\{x \in \mathcal{H} \mid f(x) = \mu\}$ is convex.
- Every local minimizer of f is a global minimizer.
- If f is strictly convex, then there exists at most one minimizer.

Existence and uniqueness of a minimizer

Theorem

Let $\mathcal{H}$ be a Hilbert space and C a closed convex subset of $\mathcal{H}$. Let $f \in \Gamma_0(\mathcal{H})$ such that $\operatorname{dom} f \cap C \neq \varnothing$.
If f is coercive or C is bounded, then there exists $\widehat{x} \in C$ such that

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

If, moreover, f is strictly convex, this minimizer $\widehat{x}$ is unique.

Optimality condition

1st order necessary and sufficient condition (P. Fermat, 160X-1665)

Let $f \in \Gamma_0(\mathbb{R}^N)$ be continuously differentiable function on $\mathbb{R}^N$. $\widehat{x}$ is a global minimizer of f, i.e., $\widehat{x} \in \mathrm{Argmin}_{x \in \mathbb{R}^N} f(x)$, iif

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

Optimality condition

1st order necessary and sufficient condition (P. Fermat, 160X-1665)

Let $f \in \Gamma_0(\mathbb{R}^N)$ be continuously differentiable function on $\mathbb{R}^N$. $\widehat{x}$ is a global minimizer of f, i.e., $\widehat{x} \in \operatorname{Argmin}_{x\in\mathbb{R}^N} f(x)$, iif

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

Proof $(\Rightarrow)$: Let $\epsilon \in \mathbb{R}^N$. We set, for every $\alpha \in \mathbb{R}$, $g(\alpha) = f(\widehat{x} + \alpha\epsilon)$

Optimality condition

1st order necessary and sufficient condition (P. Fermat, 160X-1665)

Let $f \in \Gamma_0(\mathbb{R}^N)$ be continuously differentiable function on $\mathbb{R}^N$. $\widehat{x}$ is a global minimizer of f, i.e., $\widehat{x} \in \mathrm{Argmin}_{x \in \mathbb{R}^N} f(x)$, iif

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

Proof $(\Rightarrow)$: Let $\epsilon \in \mathbb{R}^N$. We set, for every $\alpha \in \mathbb{R}$, $g(\alpha) = f(\widehat{x} + \alpha\epsilon)$

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

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

Optimality condition

1st order necessary and sufficient condition (P. Fermat, 160X-1665)

Let $f \in \Gamma_0(\mathbb{R}^N)$ be continuously differentiable function on $\mathbb{R}^N$. $\widehat{x}$ is a global minimizer of f, i.e., $\widehat{x} \in \operatorname{Argmin}_{x\in\mathbb{R}^N} f(x)$, iif

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

Proof ($\Rightarrow$) : Let $\epsilon \in \mathbb{R}^N$. We set, for every $\alpha \in \mathbb{R}$, $g(\alpha) = f(\widehat{x} + \alpha\epsilon)$

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

Optimality condition

1st order necessary and sufficient condition (P. Fermat, 160X-1665)

Let $f \in \Gamma_0(\mathbb{R}^N)$ be continuously differentiable function on $\mathbb{R}^N$. $\widehat{x}$ is a global minimizer of f, i.e., $\widehat{x} \in \operatorname{Argmin}_{x \in \mathbb{R}^N} f(x)$, iif

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

Proof $(\Rightarrow)$: Let $\epsilon \in \mathbb{R}^N$. We set, for every $\alpha \in \mathbb{R}$, $g(\alpha) = f(\widehat{x} + \alpha\epsilon)$

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

$\rightarrow$ Because $\widehat{x}$ is a minimizer of f

Optimality condition

1st order necessary and sufficient condition (P. Fermat, 160X-1665)

Let $f \in \Gamma_0(\mathbb{R}^N)$ be continuously differentiable function on $\mathbb{R}^N$. $\widehat{x}$ is a global minimizer of f, i.e., $\widehat{x} \in \operatorname{Argmin}_{x \in \mathbb{R}^N} f(x)$, iif

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

Proof ($\Rightarrow$) : Let $\epsilon \in \mathbb{R}^N$. We set, for every $\alpha \in \mathbb{R}$, $g(\alpha) = f(\widehat{x} + \alpha\epsilon)$

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

Thus

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

Optimality condition

1st order necessary and sufficient condition (P. Fermat, 160X-1665)

Let $f \in \Gamma_0(\mathbb{R}^N)$ be continuously differentiable function on $\mathbb{R}^N$. $\widehat{x}$ is a global minimizer of f, i.e., $\widehat{x} \in \operatorname{Argmin}_{x \in \mathbb{R}^N} f(x)$, iif

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

Proof $(\Rightarrow)$: Let $\epsilon \in \mathbb{R}^N$. We set, for every $\alpha \in \mathbb{R}$, $g(\alpha) = f(\widehat{x} - \alpha\epsilon)$

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

Thus

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

Optimality condition

1st order necessary and sufficient condition (P. Fermat, 160X-1665)

Let $f \in \Gamma_0(\mathbb{R}^N)$ be continuously differentiable function on $\mathbb{R}^N$. $\widehat{x}$ is a global minimizer of f, i.e., $\widehat{x} \in \mathrm{Argmin}_{x\in\mathbb{R}^N} f(x)$, iif

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

Proof ($\Rightarrow$) : Let $\epsilon \in \mathbb{R}^N$. We set, for every $\alpha \in \mathbb{R}$, $g(\alpha) = f(\widehat{x} - \alpha\epsilon)$

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

Thus

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

Optimality condition

1st order necessary and sufficient condition (P. Fermat, 160X-1665)

Let $f \in \Gamma_0(\mathbb{R}^N)$ be continuously differentiable function on $\mathbb{R}^N$. $\widehat{x}$ is a global minimizer of f, i.e., $\widehat{x} \in \mathrm{Argmin}_{x\in\mathbb{R}^N} f(x)$, iif

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

Proof ($\Leftarrow$) : f being a convex function, this yields to

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

Optimality condition

1st order necessary and sufficient condition (P. Fermat, 160X-1665)

Let $f \in \Gamma_0(\mathbb{R}^N)$ be continuously differentiable function on $\mathbb{R}^N$. $\widehat{x}$ is a global minimizer of f, i.e., $\widehat{x} \in \operatorname{Argmin}_{x\in\mathbb{R}^N} f(x)$, iif

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

Proof ($\Leftarrow$) : f being a convex function, this yields to

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

Optimality condition

1st order necessary and sufficient condition (P. Fermat, 160X-1665)

Let $f \in \Gamma_0(\mathbb{R}^N)$ be continuously differentiable function on $\mathbb{R}^N$. $\widehat{x}$ is a global minimizer of f, i.e., $\widehat{x} \in \mathrm{Argmin}_{x\in\mathbb{R}^N} f(x)$, iif

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

Proof $(\Leftarrow)$: f being a convex function, this yields to

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

Optimality condition

1st order necessary and sufficient condition (P. Fermat, 160X-1665)

Let $f \in \Gamma_0(\mathbb{R}^N)$ be continuously differentiable function on $\mathbb{R}^N$. $\widehat{x}$ is a global minimizer of f, i.e., $\widehat{x} \in \mathrm{Argmin}_{x \in \mathbb{R}^N} f(x)$, iif

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

Proof ($\Leftarrow$) : f being a convex function, this yields to

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

Thus

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

If $\nabla f(\widehat{x}) = 0$, then

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

Optimality condition

1st order necessary and sufficient condition (P. Fermat, 160X-1665)

Let $f \in \Gamma_0(\mathbb{R}^N)$ be continuously differentiable function on $\mathbb{R}^N$. $\widehat{x}$ is a global minimizer of f, i.e., $\widehat{x} \in \mathrm{Argmin}_{x \in \mathbb{R}^N} f(x)$, iif

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

- Lead to a N equations - N unknown problem.
- Closed form expression for only few cases.
- If no closed form expression exists, an iterative procedure is required.
- Solve the optimization problem $\widehat{x} \in \mathrm{Argmin}_x f(x)$ equivalent to find a solution to $\nabla f(\widehat{x}) = 0$.

Optimality condition

- **Solving mean squares**

$$\text{Find} \quad \widehat{x} \in \operatorname{Argmin}_{x \in \mathbb{R}^N} \|Ax - y\|_2^2 \qquad \text{with} \qquad \begin{cases} A \in \mathbb{R}^{M \times N} \\ y \in \mathbb{R}^M \end{cases}$$

$\rightarrow$ Optimality condition:

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

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

$\rightarrow$ Sometimes difficult to invert $A^\top A$. Closed form expression known if A models a circulant matrix.

Optimality condition

- Logistic based criterion:

$$\text{Find} \quad \widehat{x} \in \operatorname{Argmin}_{x\in\mathbb{R}} \log\left(1+\exp(-yx)\right) \qquad \text{with} \qquad y \in \mathbb{R}$$

$\rightarrow$ Optimality condition:

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

$\rightarrow$ No closed form expression. An iterative procedure is required.

Gradient descent

Gradient descent

Let $f \in \Gamma_0(\mathbb{R}^N)$ be continuously differentiable on $\mathbb{R}^N$ and with a β-Lipschitz gradient. Let $x_0 \in \mathbb{R}^N$ and $\gamma_n \in]0, 2/\beta[$

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

thus, the sequence $(x_n)_{n\in\mathbb{N}}$ converges to a minimizer of f.

Gradient β-Lipschitz

Let $f \in \Gamma_0(\mathbb{R}^N)$ be continuously differentiable on $\mathbb{R}^N$. f is gradient β-Lipschitz with $\beta > 0$ if

$$(\forall (u, v) \in \mathbb{R}^N \times \mathbb{R}^N) \quad \|\nabla f(u) - \nabla f(v)\| \leq \beta \|u - v\|$$

Gradient descent

Gradient descent

Let $f \in \Gamma_0(\mathbb{R}^N)$ be continuously differentiable on $\mathbb{R}^N$ and with a β-Lipschitz gradient. Let $x_0 \in \mathbb{R}^N$ and $\gamma_n \in]0, 2/\beta[$

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

thus, the sequence $(x_n)_{n\in\mathbb{N}}$ converges to a minimizer of f.

- An iterative method consists to build a sequence $(x_n)_{n\in\mathbb{N}}$ such that, at each iteration k
$$f(x_{n+1}) < f(x_n)$$
- Choose γ_n for fast convergence :
$\rightarrow$ Steepest descent, mthode de Newton, ...
- Convergence proof: detailed later.

Optimality condition

1st order necessary and sufficient condition

Let C be a non-empty closed convex subset of $\mathbb{R}^N$. Let $f \in \Gamma_0(\mathbb{R}^N)$ be a continuously differentiable function on C.
$\widehat{x}$ is a minimizer of f on C, i.e, $\widehat{x} \in \mathrm{Argmin}_{x \in C} f(x)$ iif

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

- We are here interested in:

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

- Feasible x when x satisfies the constraints.

Projected gradient descent

Projected gradient descent

Let C be a non-empty closed convex subset of $\mathbb{R}^N$. Let $f \in \Gamma_0(\mathbb{R}^N)$ be a continuously differentiable function on C with β-Lipschitz gradient.
Let $x_0 \in C$ and $\gamma_n \in]0, 2/\beta[$

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

then, the sequence $(x_n)_{n\in\mathbb{N}}$ converges to a minimizer of f on C.

- P_C : projection onto C

$$(\forall x \in \mathbb{R}^N) \quad P_C(x) = \arg\min_{z\in C} \|z - x\|_2^2$$

- Let $C = \{x = (x^{(i)})_{1\le i\le N} \in \mathbb{R}^N \,|\, (\forall i \in \{1,\dots,N\})\ x^{(i)} \ge 0\}$, then

$$P_C(x) = \big(\max(0, x^{(i)})\big)_{1\le i\le N}$$

Optimality conditions

Optimization problem under equality and inequality constraints:

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

where,

- for every $i \in \{0,\ldots,m\}$, $\psi_i \in \Gamma_0(\mathbb{R}^N)$,
- for every $j \in \{0,\ldots,p\}$, $\varphi_j \in \Gamma_0(\mathbb{R}^N)$,
- $\mathcal{D} = \bigcap_{i=0}^m \operatorname{dom}\psi_i \cap \bigcap_{j=1}^p \operatorname{dom}\varphi_j \neq \emptyset$.

Lagrangian

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

Optimality condition

Lagrangian

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

- $\operatorname{dom} L = \mathcal{D}\times\mathbb{R}^m\times\mathbb{R}^p$.
- $\lambda=(\lambda_i)_{1\le i\le m}$: Lagrange multiplier associated with $\psi_i(x)\le 0$.
- $\nu=(\nu_j)_{1\le j\le p}$: Lagrange multiplier associated with $\varphi_j(x)=0$.
- λ and ν are *Lagrange multiplier vectors* or *dual variables*.
- *Lagrange dual function* :

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

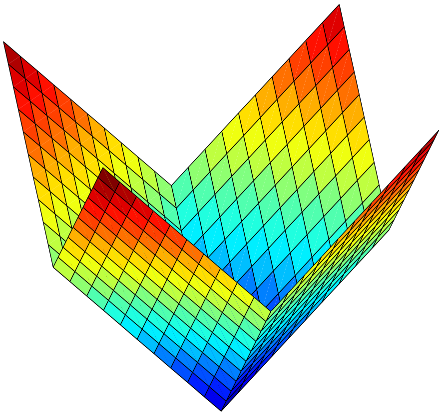
Optimality condition

Necessary and sufficient condition: Karush Kuhn Tucker (KKT)

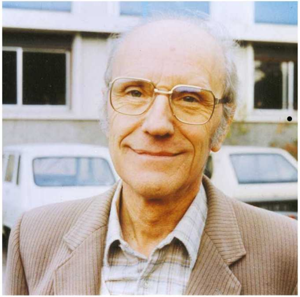
Let $(\psi_i)_{0\le i\le m}$ and $(\varphi_j)_{1\le j\le p}$ be continuously differentiable.
$\widehat{x}$ and $(\widehat{\lambda},\widehat{\nu})$ are primal and dual solutions iif

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

Non-smooth convex optimization



A pioneer



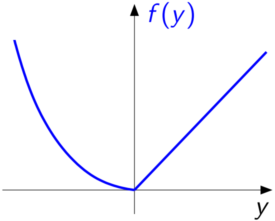
Jean-Jacques Moreau
(1923–2014)

Subdifferential of function: definition

The (Moreau) subdifferential of f, denoted by ∂f,

Subdifferential of function: definition

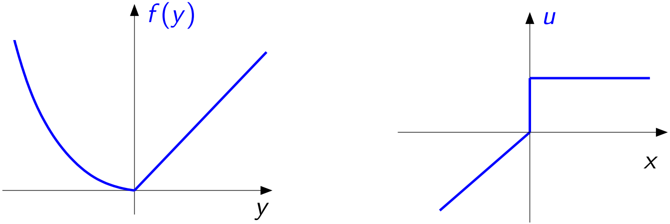
Let $f : \mathcal{H} \to]-\infty, +\infty]$ be a proper function.
The (Moreau) subdifferential of f, denoted by ∂f,



Subdifferential of function: definition

Let $f : \mathcal{H} \to]-\infty, +\infty]$ be a proper function.
The (Moreau) subdifferential of f, denoted by ∂f, is such that

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

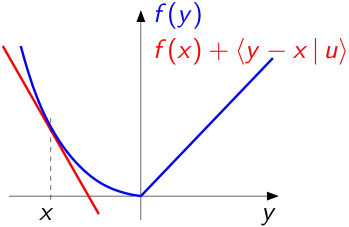


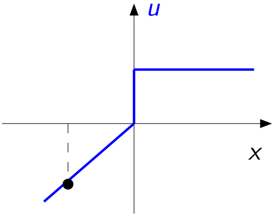
Subdifferential of function: definition

Let $f : \mathcal{H} \to \,]-\infty, +\infty]$ be a proper function.
The (Moreau) subdifferential of f, denoted by ∂f, is such that

$$\partial f : \mathcal{H} \to 2^{\mathcal{H}}$$

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

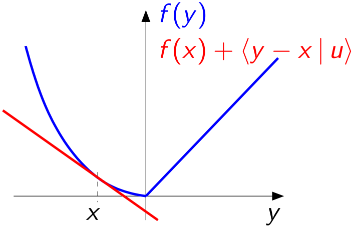


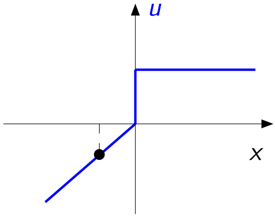


Subdifferential of function: definition

Let $f : \mathcal{H} \to \,]-\infty, +\infty]$ be a proper function.
The (Moreau) subdifferential of f, denoted by ∂f, is such that

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

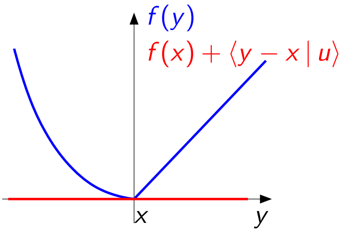


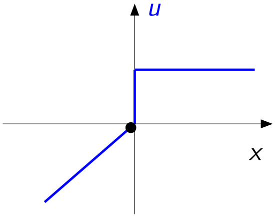


Subdifferential of function: definition

Let $f : \mathcal{H} \to]-\infty, +\infty]$ be a proper function.
The (Moreau) subdifferential of f, denoted by ∂f, is such that

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

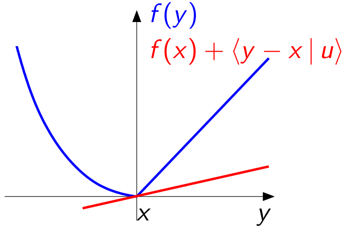


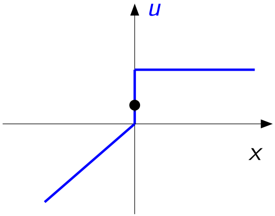


Subdifferential of function: definition

Let $f : \mathcal{H} \to \,]-\infty, +\infty]$ be a proper function.
The (Moreau) **subdifferential of f**, denoted by ∂f, is such that

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

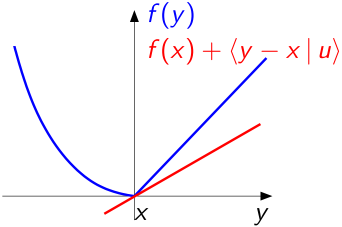


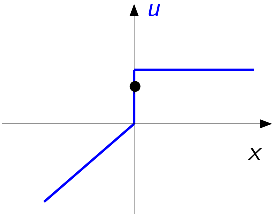


Subdifferential of function: definition

Let $f : \mathcal{H} \to]-\infty, +\infty]$ be a proper function.
The (Moreau) subdifferential of f, denoted by ∂f, is such that

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

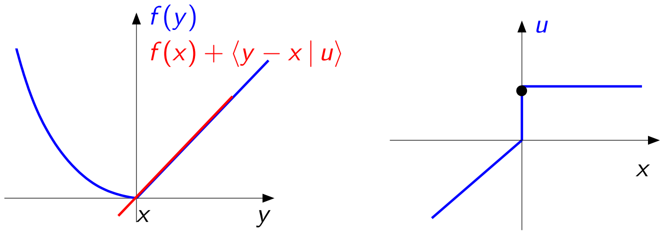




Subdifferential of function: definition

Let $f : \mathcal{H} \to]-\infty, +\infty]$ be a proper function.
The (Moreau) subdifferential of f, denoted by ∂f, is such that

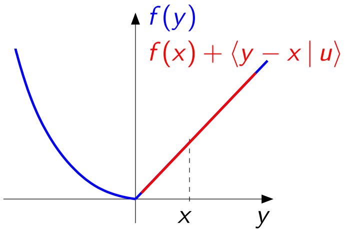
$$\partial f : \mathcal{H} \to 2^{\mathcal{H}}$$
$$x \to \{u \in \mathcal{H} \,|\, (\forall y \in \mathcal{H})\ \langle y - x | u\rangle + f(x) \leq f(y)\}$$

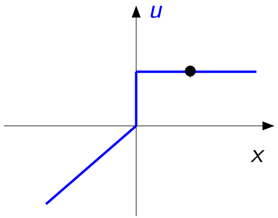


Subdifferential of function: definition

Let $f : \mathcal{H} \to]-\infty, +\infty]$ be a proper function.
The (Moreau) subdifferential of f, denoted by ∂f, is such that

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

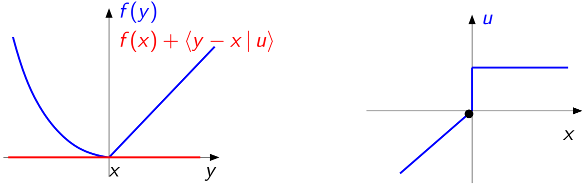




Subdifferential of a function: properties

Let $f : \mathcal{H} \to]-\infty, +\infty]$ be a proper function.
The (Moreau) subdifferential of f, denoted by ∂f, is such that

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



Fermat's rule: $0 \in \partial f(x) \Leftrightarrow x \in \operatorname{Argmin} f$

Subdifferential of a function: properties

Let $f : \mathcal{H} \to]-\infty, +\infty]$ be a proper function.
The (Moreau) subdifferential of f, denoted ∂f, is such that

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

- $u \in \partial f(x)$ is a subgradient of f at x.

Subdifferential of a convex function: properties

If $f\colon \mathcal{H} \to \,]-\infty, +\infty]$ is convex and it is Gâteaux differentiable at x, then

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

Let $\mathcal{H}$ be a Hilbert space and let $f\colon \mathcal{H} \to \,]-\infty, +\infty]$ be a proper function. f is **Gâteaux differentiable** at $x \in \operatorname{dom} f$ if there exists $\nabla f(x) \in \mathcal{H}$ such that

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

Subdifferential of a convex function: properties

If $f\colon \mathcal{H} \to \,]-\infty,+\infty]$ is convex and it is Gâteaux differentiable at x, then

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

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

Proof:
For every $\alpha \in [0,1]$ and $y \in \mathcal{H}$,

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

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

Then $\nabla f(x) \in \partial f(x)$.

Subdifferential of a convex function: properties

If $f : \mathcal{H} \to]-\infty, +\infty]$ is convex and it is Gâteaux differentiable at x, then

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

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

Proof:
Conversely, if $u \in \partial f(x)$, then, for every $\alpha \in [0, +\infty[$ and $y \in \mathcal{H}$,

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

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

By selecting $y = u - \nabla f(x)$, it results that $\|u - \nabla f(x)\|^2 \leq 0$ and then $u = \nabla f(x)$.

Subdifferential of a convex function: properties

Let $f\colon \mathcal{H} \to \,]-\infty,+\infty]$ be Gâteaux differentiable on $\operatorname{dom} f$, which is convex.
Then, f is convex if and only if

$$(\forall (x,y) \in (\operatorname{dom} f)^2) \quad f(y) \geq f(x) + \langle \nabla f(x) \mid y - x \rangle\,.$$

Subdifferential of a convex function: properties

Let $f\colon \mathcal{H} \to \,]-\infty, +\infty]$ be Gâteaux differentiable on $\operatorname{dom} f$, which is convex.
Then, f is convex if and only if

$$(\forall (x, y) \in (\operatorname{dom} f)^2) \quad f(y) \geq f(x) + \langle \nabla f(x) \mid y - x \rangle \,.$$

Proof:
We have already seen that the gradient inequality holds when f is convex and differentiable at $x \in \mathcal{H}$.

Subdifferential of a convex function: properties

Let $f\colon \mathcal{H} \to \left]-\infty, +\infty\right]$ be Gâteaux differentiable on $\operatorname{dom} f$, which is convex.
Then, f is convex if and only if

$$(\forall (x,y) \in (\operatorname{dom} f)^2) \quad f(y) \geq f(x) + \langle \nabla f(x) \mid y - x \rangle \,.$$

Proof:
Conversely, if the gradient inequality is satisfied, we have, for every $(x,y) \in (\operatorname{dom} f)^2$ and $\alpha \in [0,1]$, $\alpha x + (1-\alpha) y \in \operatorname{dom} f$, and

$$f(x) \geq f(\alpha x + (1-\alpha)y) + (1-\alpha) \langle \nabla f(\alpha x + (1-\alpha)y) \mid x - y \rangle$$
$$f(y) \geq f(\alpha x + (1-\alpha)y) + \alpha \langle \nabla f(\alpha x + (1-\alpha)y) \mid y - x \rangle \,.$$

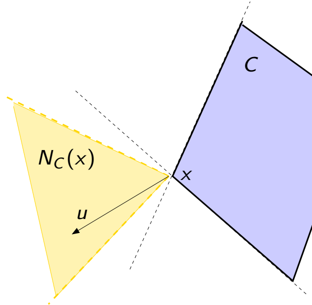
By multiplying the first inequality by α and the second one by $1-\alpha$ and summing them, we get

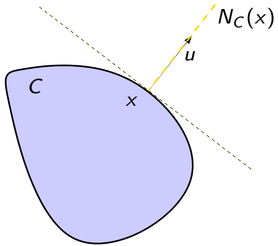
$$\alpha f(x) + (1-\alpha) f(y) \geq f(\alpha x + (1-\alpha) y).$$

Subdifferential of a convex function: example

Let C be a nonempty subset of $\mathcal{H}$.
For every $x \in \mathcal{H}$, $\partial\iota_C(x)$ is the normal cone to C at x defined by

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





Subdifferential calculus

Let $\mathcal{H}$ and $\mathcal{G}$ be two real Hilbert spaces.

- Let $f\colon \mathcal{H} \to \,]-\infty,+\infty]$ be proper, then for every $\lambda \in \,]0,+\infty[$ $\partial(\lambda f) = \lambda \partial f$.
- Let $f\colon \mathcal{H} \to \,]-\infty,+\infty]$, $g\colon \mathcal{G} \to \,]-\infty,+\infty]$, and $L \in \mathcal{B}(\mathcal{H},\mathcal{G})$. If $\mathrm{dom}\, g \cap L(\mathrm{dom}\, f) \neq \varnothing$, then

$$(\forall x \in \mathcal{H}) \qquad \partial f(x) + L^* \partial g(Lx) \subset \partial(f + g \circ L)(x).$$

Subdifferential calculus

Let $\mathcal{H}$ and $\mathcal{G}$ be two real Hilbert spaces.

- Let $f\colon \mathcal{H} \to]-\infty,+\infty]$ be proper, then for every $\lambda \in]0,+\infty[$ $\partial(\lambda f) = \lambda \partial f$.
- Let $f\colon \mathcal{H} \to]-\infty,+\infty]$, $g\colon \mathcal{G} \to]-\infty,+\infty]$, and $L \in \mathcal{B}(\mathcal{H},\mathcal{G})$. If $\operatorname{dom} g \cap L(\operatorname{dom} f) \neq \varnothing$, then

$$(\forall x \in \mathcal{H}) \qquad \partial f(x) + L^*\partial g(Lx) \subset \partial(f + g\circ L)(x).$$

Proof: Let $x \in \mathcal{H}$, $u \in \partial f(x)$ and $v \in \partial g(Lx)$. We have: $u + L^* v \in \partial f(x) + L^*\partial g(Lx)$ and

$$(\forall y \in \mathcal{H}) \qquad \begin{aligned} f(y) &\geq f(x) + \langle y - x \mid u\rangle \\ g(Ly) &\geq g(Lx) + \langle L(y-x) \mid v\rangle\,. \end{aligned}$$

Therefore, by summing,

$$f(y) + g(Ly) \geq f(x) + g(Lx) + \langle y - x \mid u + L^* v\rangle\,.$$

We deduce that $u + L^* v \in \partial(f + g\circ L)(x)$.

Subdifferential calculus

Let $\mathcal{H}$ and $\mathcal{G}$ be two real Hilbert spaces.
Let $f \in \Gamma_0(\mathcal{H})$, $g \in \Gamma_0(\mathcal{G})$, and $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$.
If $\operatorname{int}(\operatorname{dom} g) \cap L(\operatorname{dom} f) \neq \varnothing$ or $\operatorname{dom} g \cap \operatorname{int}\big(L(\operatorname{dom} f)\big) \neq \varnothing$, then

$$\partial f + L^* \partial g\, L = \partial(f + g \circ L).$$

Particular case:

- If $f \in \Gamma_0(\mathcal{H})$, $g \in \Gamma_0(\mathcal{G})$, and f is finite valued, then $\partial f + \partial g = \partial(f + g)$.
- If $g \in \Gamma_0(\mathcal{G})$, $L \in \mathcal{B}(\mathcal{G}, \mathcal{H})$, and $\operatorname{int}(\operatorname{dom} g) \cap \operatorname{ran} L \neq \varnothing$, then $L^* \partial g\, L = \partial(g \circ L)$.

Subdifferential calculus

Let $(\mathcal{H})_{i\in I}$ where $I \subset \mathbb{N}$ be Hilbert spaces and let $\mathcal{H} = \bigoplus_{i\in I} \mathcal{H}_i$.
For every $i \in I$, let $f_i \colon \mathcal{H}_i \to \left]-\infty, +\infty\right]$ be a proper function. Let

$$f \colon \mathcal{H} \to \left]-\infty, +\infty\right] : x = (x_i)_{i\in I} \mapsto \sum_{i\in I} f_i(x_i)$$

Then,

$$\big(\forall x = (x_i)_{i\in I} \in \mathcal{H}\big) \qquad \partial f(x) = {}_{i\in I} \partial f_i(x_i).$$

Subdifferential calculus

Let $(\mathcal{H})_{i\in I}$ where $I \subset \mathbb{N}$ be Hilbert spaces and let $\mathcal{H} = \bigoplus_{i\in I} \mathcal{H}_i$.
For every $i \in I$, let $f_i \colon \mathcal{H}_i \to \left]-\infty, +\infty\right]$ be a proper function. Let

$$f\colon \mathcal{H} \to \left]-\infty, +\infty\right] : x = (x_i)_{i\in I} \mapsto \sum_{i\in I} f_i(x_i)$$

Then,

$$(\forall x = (x_i)_{i\in I} \in \mathcal{H}) \qquad \partial f(x) = {}_{i\in I}\partial f_i(x_i).$$

<u>Proof</u>: Let $x = (x_i)_{i\in I} \in \mathcal{H}$. We have

$$\begin{aligned} & t = (t_i)_{i\in I} \in {}_{i\in I}\partial f_i(x_i) \\ \Leftrightarrow \quad & (\forall i \in I)(\forall y_i \in \mathcal{H}_i) \;\; f_i(y_i) \geq f_i(x_i) + \langle t_i \mid y_i - x_i \rangle \\ \Rightarrow \quad & (\forall y = (y_i)_{i\in I} \in \mathcal{H}) \;\; \sum_{i\in I} f_i(y_i) \geq \sum_{i\in I} f_i(x_i) + \sum_{i\in I} \langle t_i \mid y_i - x_i \rangle \\ \Leftrightarrow \quad & (\forall y \in \mathcal{H}) \;\; f(y) \geq f(x) + \langle t \mid y - x \rangle . \end{aligned}$$

Subdifferential calculus

Let $(\mathcal{H})_{i\in I}$ where $I \subset \mathbb{N}$ be Hilbert spaces and let $\mathcal{H} = \bigoplus_{i\in I} \mathcal{H}_i$.
For every $i \in I$, let $f_i \colon \mathcal{H}_i \to \,]-\infty, +\infty]$ be a proper function. Let

$$f \colon \mathcal{H} \to \,]-\infty, +\infty] : x = (x_i)_{i\in I} \mapsto \sum_{i\in I} f_i(x_i)$$

Then,

$$(\forall x = (x_i)_{i\in I} \in \mathcal{H}) \qquad \partial f(x) = {}_{i\in I}\partial f_i(x_i).$$

<u>Proof</u>: Conversely,

$$\begin{aligned} & t = (t_i)_{i\in I} \in \partial f(x) \\ \Leftrightarrow \quad & (\forall y = (y_i)_{i\in I} \in \mathcal{H}) \ \ \sum_{i\in I} f_i(y_i) \geq \sum_{i\in I} f_i(x_i) + \sum_{i\in I} \langle t_i \mid y_i - x_i \rangle \,. \end{aligned}$$

Let $j \in I$. By setting $(\forall i \in I \setminus \{j\})\ y_i = x_i \in \operatorname{dom} f_i$, we get

$$(\forall y_j \in \mathcal{H}_j) \ \ f_j(y_j) \geq f_j(x_j) + \langle t_j \mid y_j - x_j \rangle \,.$$

Conjugate

Adrien-Marie Legendre
(1752–1833)

Werner Fenchel
(1905–1988)

Conjugate

Adrien-Marie Legendre
(1752–1833)

Werner Fenchel
(1905–1988)

Conjugate: definition

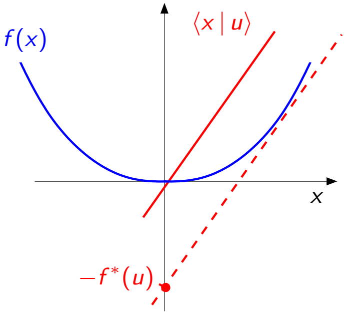
Let $\mathcal{H}$ be a Hilbert space and $f\colon \mathcal{H} \to\,]-\infty, +\infty]$.
The conjugate of f is the function $f^*\colon \mathcal{H} \to [-\infty, +\infty]$ such that

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

Conjugate: definition

Let $\mathcal{H}$ be a Hilbert space and $f\colon \mathcal{H} \to \,]-\infty, +\infty]$.
The conjugate of f is the function $f^*\colon \mathcal{H} \to [-\infty, +\infty]$ such that

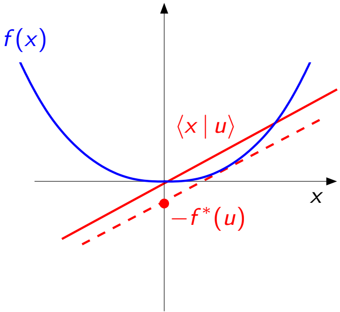
$$(\forall u \in \mathcal{H}) \qquad f^*(u) = \sup_{x \in \operatorname{dom} f} \big(\langle x \mid u \rangle - f(x) \big)$$



Conjugate: definition

Let $\mathcal{H}$ be a Hilbert space and $f\colon \mathcal{H} \to]-\infty,+\infty]$.
The conjugate of f is the function $f^*\colon \mathcal{H} \to [-\infty,+\infty]$ such that

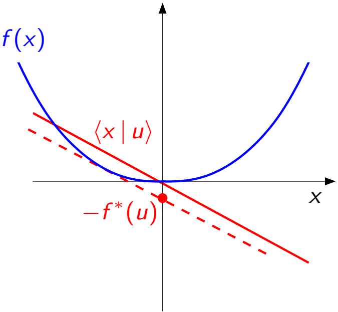
$$(\forall u \in \mathcal{H}) \qquad f^*(u) = \sup_{x \in \operatorname{dom} f} \big(\langle x \mid u \rangle - f(x) \big)$$



Conjugate: definition

Let $\mathcal{H}$ be a Hilbert space and $f\colon \mathcal{H} \to \,]-\infty,+\infty]$.
The conjugate of f is the function $f^*\colon \mathcal{H} \to [-\infty,+\infty]$ such that

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



Conjugate: definition

Let $\mathcal{H}$ be a Hilbert space and $f\colon \mathcal{H} \to \,]-\infty,+\infty]$.
The **conjugate** of f is the function $f^*\colon \mathcal{H} \to [-\infty,+\infty]$ such that

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

Examples :

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

  Proof : For every $(x,u) \in \mathcal{H}^2$, $\langle x \mid u \rangle - \frac{1}{2}\|x\|^2 = \frac{1}{2}\|u\|^2 - \frac{1}{2}\|u - x\|^2$ is maximum at $x = u$.
  Consequently, $f^*(u) = \frac{1}{2}\|u\|^2$.

Conjugate: definition

Let $\mathcal{H}$ be a Hilbert space and $f\colon \mathcal{H} \to \,]-\infty,+\infty]$.
The **conjugate** of f is the function $f^*\colon \mathcal{H} \to [-\infty,+\infty]$ such that

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

Examples :

- $f = \frac{1}{2}\|\cdot\|^2 \Rightarrow f^* = \frac{1}{2}\|\cdot\|^2$.
- Let $\phi\colon \mathbb{R} \to \,]-\infty,+\infty]$ be a even function. $(\phi \circ \|\cdot\|)^* = \phi^* \circ \|\cdot\|$.
- $(\forall x \in \mathbb{R}^N)$ $f(x) = \frac{1}{q}\|x\|_q^q$ with $q \in]1,+\infty[$
  $\Rightarrow (\forall u \in \mathbb{R}^N)$ $f^*(u) = \frac{1}{q^*}\|u\|_{q^*}^{q^*}$ with $\frac{1}{q} + \frac{1}{q^*} = 1$

Conjugate: definition

Let $\mathcal{H}$ be a Hilbert space and $f\colon \mathcal{H} \to]-\infty, +\infty]$.
The conjugate of f is the function $f^*\colon \mathcal{H} \to [-\infty, +\infty]$ such that

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

- If f is even, then f^* is even.

Proof :

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

Conjugate: definition

Let $\mathcal{H}$ be a Hilbert space and $f\colon \mathcal{H} \to \,]-\infty, +\infty]$.
The **conjugate** of f is the function $f^*\colon \mathcal{H} \to [-\infty, +\infty]$ such that

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

- If f is even, then f^* is even.
- For every $\alpha \in \,]0, +\infty[$, $(\alpha f)^* = \alpha f^*(\cdot/\alpha)$.
- For every $(y, v) \in \mathcal{H}^2$ et $\alpha \in \mathbb{R}$,
  $\big(f(\cdot - y) + \langle \cdot \mid v \rangle + \alpha\big)^* = f^*(\cdot - v) + \langle y \mid \cdot - v \rangle - \alpha$.
- Let $\mathcal{G}$ be a Hilbert space and $L \in \mathcal{B}(\mathcal{G}, \mathcal{H})$ be an isomorphism.
  $(f \circ L)^* = f^* \circ (L^{-1})^*$.
- f^* is l.s.c. and convex.

Conjugate: definition

Let $\mathcal{H}$ be a Hilbert space and $f\colon \mathcal{H} \to\,]-\infty, +\infty]$.
The **conjugate** of f is the function $f^*\colon \mathcal{H} \to [-\infty, +\infty]$ such that

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

Moreau-Fenchel theorem

Let $\mathcal{H}$ be a Hilbert space and $f\colon \mathcal{H} \to\,]-\infty, +\infty]$ be a proper function.

$$f \text{ is l.s.c. and convex} \Leftrightarrow f^{**} = f.$$

Conjugate: definition

Let $\mathcal{H}$ be a Hilbert space and $f\colon \mathcal{H} \to \,]-\infty,+\infty]$.
The conjugate of f is the function $f^*\colon \mathcal{H} \to [-\infty,+\infty]$ such that

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

Moreau-Fenchel theorem

Let $\mathcal{H}$ be a Hilbert space and $f\colon \mathcal{H} \to \,]-\infty,+\infty]$ be a proper function.

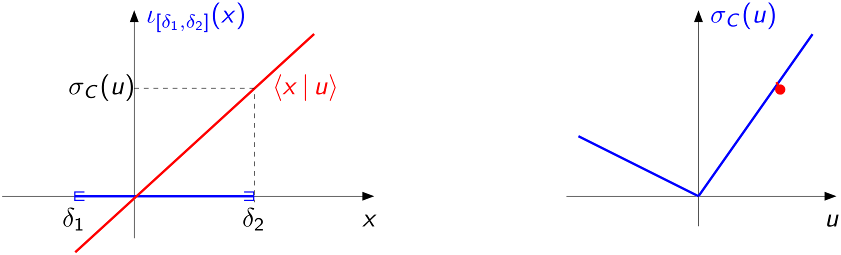
$$f \text{ is l.s.c. and convex} \Leftrightarrow f^{**} = f.$$

- Consequence: If $f \in \Gamma_0(\mathbb{R})$, then f^* is proper, hence $f^* \in \Gamma_0(\mathbb{R})$.

Conjugate: example

Let $\mathcal{H}$ be a Hilbert space and $C \subset \mathcal{H}$.
σ_C is the **support function** of C if

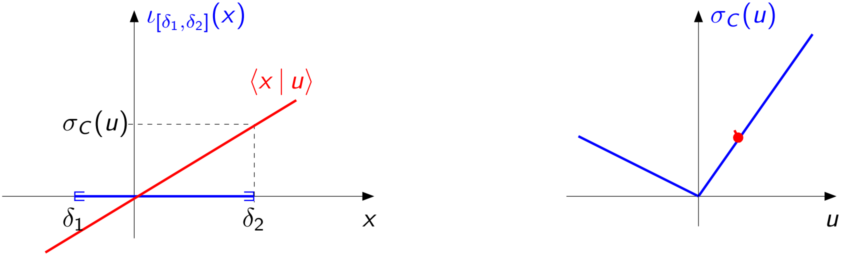
$$(\forall u \in \mathcal{H}) \qquad \sigma_C(u) = \sup_{x \in C} \langle x \mid u \rangle$$
$$= \iota_C^*(u).$$



Conjugate: example

Let $\mathcal{H}$ be a Hilbert space and $C \subset \mathcal{H}$.
σ_C is the support function of C if

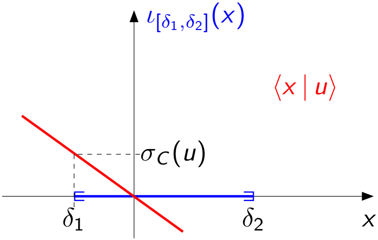
$$(\forall u \in \mathcal{H}) \qquad \sigma_C(u) = \sup_{x \in C} \langle x \mid u \rangle$$
$$= \iota_C^*(u).$$

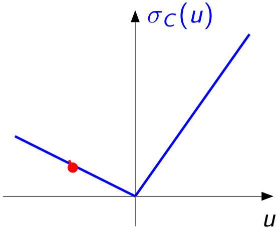


Conjugate: example

Let $\mathcal{H}$ be a Hilbert space and $C \subset \mathcal{H}$.
σ_C is the support function of C if

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





Conjugate: example

Let $\mathcal{H}$ be a Hilbert space.
$f\colon \mathcal{H} \to]-\infty, +\infty]$ is **positively homogeneous** if

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

f positively homogeneous and belongs to $\Gamma_0(\mathcal{H})$ if and only if $f = \sigma_C$ where C is a nonempty closed convex subset of $\mathcal{H}$.

Conjugate: example

Let $\mathcal{H}$ be a Hilbert space.
$f\colon \mathcal{H} \to \left]-\infty, +\infty\right]$ is **positively homogeneous** if

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

f positively homogeneous and belongs to $\Gamma_0(\mathcal{H})$ if and only if $f = \sigma_C$ where C is a nonempty closed convex subset of $\mathcal{H}$.

Proof: $(\Leftarrow)$
$f = \iota_C^*$ and $\iota_C \in \Gamma_0(\mathcal{H})$. Consequently $\sigma_C \in \Gamma_0(\mathcal{H})$.
Moreover, $(\forall x \in \mathcal{H})$ $(\forall \alpha \in \left]0, +\infty\right[)$ $\sigma_C(\alpha x) = \alpha \sigma_C(x)$.

Conjugate: example

Let $\mathcal{H}$ be a Hilbert space.
$f\colon \mathcal{H} \to \,]-\infty,+\infty]$ is positively homogeneous if

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

f positively homogeneous and belongs to $\Gamma_0(\mathcal{H})$ if and only if $f = \sigma_C$ where C is a nonempty closed convex subset of $\mathcal{H}$.

Proof : $(\Rightarrow)$
Let $y \in \operatorname{dom} f$.
$f(0) = \lim_{\alpha\to 0, \alpha\geq 0} f\big((1-\alpha)0 + \alpha y\big) = \lim_{\alpha\to 0} \alpha f(y) = 0.$
Let $C = \big\{u \in \mathcal{H} \;\big|\; (\forall x \in \mathcal{H})\, \langle x \mid u\rangle \leq f(x)\big\}$.
We have, for every $u \in C$,
$f^*(u) = \sup_{x\in\mathcal{H}} \langle x \mid u\rangle - f(x) \leq 0 = \langle 0 \mid u\rangle - f(0) \leq f^*(u).$
Consequently, $f^*(u) = 0$.

Conjugate: example

Let $\mathcal{H}$ be a Hilbert space.
$f\colon \mathcal{H} \to \,]-\infty,+\infty]$ is positively homogeneous if

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

f positively homogeneous and belongs to $\Gamma_0(\mathcal{H})$ if and only if $f = \sigma_C$ where C is a nonempty closed convex subset of $\mathcal{H}$.

Proof: $(\Rightarrow)$
Moreover, for every $u \notin C$, there exists $x \in \mathcal{H}$ such that $\langle x \mid u\rangle > f(x)$. We have then, for every $\alpha \in \,]0,+\infty[$, $f^*(u) \geq \langle \alpha x \mid u\rangle - f(\alpha x) = \alpha\big(\langle x \mid u\rangle - f(x)\big)$. By taking $\alpha \to +\infty$, we obtain $f^*(u) = +\infty$.
To conclude, $f^* = \iota_C \in \Gamma_0(\mathcal{H}) \Rightarrow f = \sigma_C$ and C is a nonempty closed convex set.

Conjugate: example

- Let $f : \mathbb{R} \to]-\infty, +\infty] : x \mapsto \begin{cases} \delta_1 x & \text{if } x < 0 \\ 0 & \text{if } x = 0 \\ \delta_2 x & \text{if } x > 0 \end{cases}$
  with $-\infty \leq \delta_1 < \delta_2 \leq +\infty$.
  Then, $f = \sigma_C$ where C is the closed real interval such that $\inf C = \delta_1$ et $\sup C = \delta_2$.

Conjugate: example

- Let $f\colon \mathbb{R} \to\,]-\infty, +\infty] : x \mapsto \begin{cases} \delta_1 x & \text{if } x < 0 \\ 0 & \text{if } x = 0 \\ \delta_2 x & \text{if } x > 0 \end{cases}$
  with $-\infty \leq \delta_1 < \delta_2 \leq +\infty$.
  Then, $f = \sigma_C$ where C is the closed real interval such that $\inf C = \delta_1$ et $\sup C = \delta_2$.

- Let f be a ℓ^q norm of $\mathbb{R}^N$ with $q \in [1, +\infty]$.
  We have $f = \sigma_C$ where

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

Conjugate: example

- Let $f\colon \mathbb{R} \to \,]-\infty,+\infty] : x \mapsto \begin{cases} \delta_1 x & \text{if } x < 0 \\ 0 & \text{if } x = 0 \\ \delta_2 x & \text{if } x > 0 \end{cases}$
  with $-\infty \leq \delta_1 < \delta_2 \leq +\infty$.
  Then, $f = \sigma_C$ where C is the closed real interval such that $\inf C = \delta_1$ et $\sup C = \delta_2$.

- Let f be a ℓ^q norm of $\mathbb{R}^N$ with $q \in [1,+\infty]$.
  We have $f = \sigma_C$ where

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

<u>Particular case</u>: ℓ^1 norm of $\mathbb{R}^N$: $C = [-1,1]^N$.

Conjugate: properties

Fenchel-Young inequality : If f is proper, then

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

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

If $f \in \Gamma_0(\mathcal{H})$, then

$$\big(\forall (x,u) \in \mathcal{H}^2\big) \qquad u \in \partial f(x) \;\Leftrightarrow\; x \in \partial f^*(u).$$

Conjugate: properties

Let $(\mathcal{H})_{i\in I}$ where $I \subset \mathbb{N}$ be Hilbert spaces and let $\mathcal{H} = \bigoplus_{i\in I} \mathcal{H}_i$.
For every $i \in I$, let $f_i \colon \mathcal{H}_i \to \,]-\infty, +\infty]$. Let

$$f \colon \mathcal{H} \to \,]-\infty, +\infty] : x = (x_i)_{i\in I} \mapsto \sum_{i\in I} f_i(x_i)$$

Then,

$$(\forall u = (u_i)_{i\in I} \in \mathcal{H}) \qquad f^*(u) = \sum_{i\in I} f_i^*(u_i).$$

Conjugate: properties

Let $(\mathcal{H})_{i\in I}$ where $I \subset \mathbb{N}$ be Hilbert spaces and let $\mathcal{H} = \bigoplus_{i\in I} \mathcal{H}_i$.
For every $i \in I$, let $f_i \colon \mathcal{H}_i \to \,]-\infty, +\infty]$. Let

$$f \colon \mathcal{H} \to \,]-\infty, +\infty] : x = (x_i)_{i\in I} \mapsto \sum_{i\in I} f_i(x_i)$$

Then,

$$(\forall u = (u_i)_{i\in I} \in \mathcal{H}) \qquad f^*(u) = \sum_{i\in I} f_i^*(u_i).$$

Proof: Let $u = (u_i)_{i\in I} \in \mathcal{H}$. We have

$$\begin{aligned} f^*(u) &= \sup_{x\in\mathcal{H}} \langle x \mid u\rangle - f(x) \\ &= \sup_{x=(x_i)_{i\in I}\in\mathcal{H}} \sum_{i\in I} \langle x_i \mid u_i\rangle - f_i(x_i) \\ &= \sum_{i\in I} \sup_{x_i\in\mathcal{H}_i} \langle x_i \mid u_i\rangle - f_i(x_i) \\ &= \sum_{i\in I} f_i^*(u_i). \end{aligned}$$

Fenchel-Rockafellar duality

Primal problem

Let $\mathcal{H}$ and $\mathcal{G}$ be two real Hilbert spaces.
Let $f\colon \mathcal{H} \to]-\infty, +\infty]$, $g\colon \mathcal{G} \to]-\infty, +\infty]$. Let $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$.
We want to

$$\underset{x\in\mathcal{H}}{\text{minimize}} \;\; f(x) + g(Lx).$$

Dual problem

Let $\mathcal{H}$ and $\mathcal{G}$ be two real Hilbert spaces.
Let $f\colon \mathcal{H} \to]-\infty, +\infty]$, $g\colon \mathcal{G} \to]-\infty, +\infty]$. Let $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$.
We want to

$$\underset{v\in\mathcal{G}}{\text{minimize}} \;\; f^*(-L^*v) + g^*(v).$$

Fenchel-Rockafellar duality

Weak duality

Let $\mathcal{H}$ and $\mathcal{G}$ be two real Hilbert spaces.
Let f be a proper fonction from $\mathcal{H}$ to $]-\infty, +\infty]$, g be a proper function from $\mathcal{G}$ to $]-\infty, +\infty]$, and $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$. Let

$$\mu = \inf_{x \in \mathcal{H}} f(x) + g(Lx) \quad \text{and} \quad \mu^* = \inf_{v \in \mathcal{G}} f^*(-L^* v) + g^*(v).$$

We have $\mu \geq -\mu^*$. If $\mu \in \mathbb{R}$, $\mu + \mu^*$ is called the duality gap.

Fenchel-Rockafellar duality

Weak duality

Let $\mathcal{H}$ and $\mathcal{G}$ be two real Hilbert spaces.
Let f be a proper fonction from $\mathcal{H}$ to $]-\infty, +\infty]$, g be a proper function from $\mathcal{G}$ to $]-\infty, +\infty]$, and $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$. Let

$$\mu = \inf_{x \in \mathcal{H}} f(x) + g(Lx) \quad \text{and} \quad \mu^* = \inf_{v \in \mathcal{G}} f^*(-L^* v) + g^*(v).$$

We have $\mu \geq -\mu^*$. If $\mu \in \mathbb{R}$, $\mu + \mu^*$ is called the **duality gap**.

Proof: According to Fenchel-Young inequality, for every $x \in \mathcal{H}$ and $v \in \mathcal{G}$,

$$f(x) + g(Lx) + f^*(-L^* v) + g^*(v) \geq \langle x \mid -L^* v \rangle + \langle Lx \mid v \rangle = 0$$

Fenchel-Rockafellar duality

Strong duality

Let $\mathcal{H}$ and $\mathcal{G}$ be two real Hilbert spaces.
Let $f \in \Gamma_0(\mathcal{H})$, $g \in \Gamma_0(\mathcal{G})$, and $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$.
If $\mathrm{int}\,(\mathrm{dom}\, g) \cap L(\mathrm{dom}\, f) \neq \varnothing$ or $\mathrm{dom}\, g \cap \mathrm{int}\,\big(L(\mathrm{dom}\, f)\big) \neq \varnothing$, then

$$\mu = \inf_{x\in\mathcal{H}} f(x) + g(Lx) = -\min_{v\in\mathcal{G}} f^*(-L^*v) + g^*(v) = -\mu^* .$$

Example 1: Linear programming

Let $L \in \mathbb{R}^{K \times N}$, $b \in \mathbb{R}^K$, and $c \in \mathbb{R}^N$.
The primal problem

$$\text{Primal-LP :} \qquad \underset{x \in [0,+\infty[^N}{\text{minimize}} \; \langle c \mid x \rangle \quad \text{s.t.} \quad Lx \geq b$$

is associated with the the dual problem

$$\text{Dual-LP :} \qquad y \in [0,+\infty[^K \langle b \mid y \rangle \quad \text{s.t.} \quad L^\top y \leq c.$$

In addition, if the primal problem has a solution, then strong duality holds.

Example 1: Linear programming

Let $L \in \mathbb{R}^{K\times N}$, $b \in \mathbb{R}^K$, and $c \in \mathbb{R}^N$.
The primal problem

$$\text{Primal-LP}: \qquad \underset{x\in[0,+\infty[^N}{\text{minimize}} \; \langle c \mid x\rangle \quad \text{s.t.} \quad Lx \geq b$$

is associated with the the dual problem

$$\text{Dual-LP}: \qquad y \in [0,+\infty[^K \langle b \mid y\rangle \quad \text{s.t.} \quad L^\top y \leq c.$$

In addition, if the primal problem has a solution, then strong duality holds.

Proof: Set
$$\begin{cases} (\forall x \in \mathcal{H} = \mathbb{R}^N) \quad f(x) = \langle c \mid x\rangle + \iota_{[0,+\infty[^N}(x), \\ (\forall z \in \mathcal{G} = \mathbb{R}^K) \quad g(z) = \iota_{[0,+\infty[^K}(z-b), \\ y = -v \end{cases}$$

Example 2: Consensus and sharing

Let $\mathcal{H}$ be a real Hilbert space.
For every $i \in \{1, \dots, m\}$, let $g_i \colon \mathcal{H} \to]-\infty, +\infty]$ and $h_i \colon \mathcal{H} \to]-\infty, +\infty]$.
The consensus problem is given by

$$\underset{\substack{(x_1,\dots,x_m)\in\mathcal{H}^m \\ x_1=\cdots=x_m}}{\text{minimize}} \quad \sum_{i=1}^{m} g_i(x_i).$$

The sharing problem is given by

$$\underset{\substack{(u_1,\dots,u_m)\in\mathcal{H}^m \\ u_1+\cdots+u_m=u}}{} \sum_{i=1}^{m} h_i(u_i), \qquad u \in \mathcal{H}.$$

If, for every $i \in \{1, \dots, m\}$, $h_i = -g_i^*(\cdot - u/m)$, then sharing is the dual problem of consensus.

Example 2: Consensus and sharing

Let $\mathcal{H}$ be a real Hilbert space.
For every $i \in \{1, \dots, m\}$, let $g_i \colon \mathcal{H} \to]-\infty, +\infty]$ and $h_i \colon \mathcal{H} \to]-\infty, +\infty]$.
The consensus problem is given by

$$\underset{\substack{(x_1,\dots,x_m)\in\mathcal{H}^m\\ x_1=\dots=x_m}}{\text{minimize}} \quad \sum_{i=1}^m g_i(x_i).$$

The sharing problem is given by

$$\underset{\substack{(u_1,\dots,u_m)\in\mathcal{H}^m\\ u_1+\dots+u_m=u}}{} \sum_{i=1}^m h_i(u_i), \qquad u \in \mathcal{H}.$$

If, for every $i \in \{1, \dots, m\}$, $h_i = -g_i^*(\cdot - u/m)$, then sharing is the dual problem of consensus.

Proof: Set $L = \mathrm{Id}$ and $(\forall x = (x_1, \dots, x_m) \in \mathcal{H}^N)$ $\begin{cases} f(x) = \iota_{\Lambda_m}(x), \\ g(x) = \sum_{i=1}^m g_i(x_i) \end{cases}$

where $\Lambda_m = \{(x_1, \dots, x_m) \in \mathcal{H}^m \mid x_1 = \dots = x_m\}$.

Fenchel-Rockafellar duality

Let $\mathcal{H}$ and $\mathcal{G}$ be two real Hilbert spaces.
Let $f\colon \mathcal{H} \to \,]-\infty, +\infty]$, $g\colon \mathcal{G} \to \,]-\infty, +\infty]$, and $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$.
If $\operatorname{dom} g \cap L(\operatorname{dom} f) \neq \varnothing$, then

$$(\forall x \in \mathcal{H}) \qquad \partial f(x) + L^*\partial g(Lx) \subset \partial(f + g \circ L)(x)$$

Let $\mathcal{H}$ and $\mathcal{G}$ be two real Hilbert spaces.
Let $f \in \Gamma_0(\mathcal{H})$, $g \in \Gamma_0(\mathcal{G})$, and $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$. If $\operatorname{int}(\operatorname{dom} g) \cap L(\operatorname{dom} f) \neq \varnothing$ and $\operatorname{dom} g \cap \operatorname{int}\big(L(\operatorname{dom} f)\big) \neq \varnothing$, then

$$\partial f + L^*\partial g L = \partial(f + g \circ L)$$

Fenchel-Rockafellar duality

Duality theorem (1)

Let $\mathcal{H}$ and $\mathcal{G}$ be two real Hilbert spaces.
Let $f \in \Gamma_0(\mathcal{H})$, $g \in \Gamma_0(\mathcal{G})$, and $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$.

$$\operatorname{zer}(\partial f + L \partial g L^*) \neq \varnothing \quad \Leftrightarrow \quad \operatorname{zer}\left((-L)\partial f^*(-L^*) + \partial g^*\right) \neq \varnothing$$

Fenchel-Rockafellar duality

Duality theorem (1)

Let $\mathcal{H}$ and $\mathcal{G}$ be two real Hilbert spaces.
Let $f \in \Gamma_0(\mathcal{H})$, $g \in \Gamma_0(\mathcal{G})$, and $L \in \mathcal{B}(\mathcal{H},\mathcal{G})$.

$$\operatorname{zer}\left(\partial f + L\partial g L^*\right) \neq \varnothing \quad \Leftrightarrow \quad \operatorname{zer}\left((-L)\partial f^*(-L^*) + \partial g^*\right) \neq \varnothing$$

Proof:

$$\begin{aligned}
(\exists x \in \mathcal{H})\ \ 0 \in \partial f(x) + L^*\partial g(Lx) \quad &\Leftrightarrow \quad (\exists x \in \mathcal{H})(\exists v \in \mathcal{G}) \quad \begin{cases} -L^*v \in \partial f(x) \\ v \in \partial g(Lx) \end{cases} \\
&\Leftrightarrow \quad (\exists x \in \mathcal{H})(\exists v \in \mathcal{G}) \quad \begin{cases} x \in \partial f^*(-L^*v) \\ Lx \in \partial g^*(v) \end{cases} \\
&\Leftrightarrow \quad (\exists v \in \mathcal{G}) \qquad 0 \in -L\partial f^*(-L^*v) + \partial g^*(v).
\end{aligned}$$

Fenchel-Rockafellar duality

Duality theorem (2)

Let $\mathcal{H}$ and $\mathcal{G}$ be two real Hilbert spaces.
Let $f \in \Gamma_0(\mathcal{H})$, $g \in \Gamma_0(\mathcal{G})$, and $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$.

- If there exists $\widehat{x} \in \mathcal{H}$ such that $0 \in \partial f(\widehat{x}) + L^* \partial g(L\widehat{x})$, then $\widehat{x}$ is a solution to the primal problem. Moreover, there exists a solution $\widehat{v}$ to the dual problem such that $-L^*\widehat{v} \in \partial f(\widehat{x})$ and $L\widehat{x} \in \partial g^*(\widehat{v})$.
- If there exists $(\widehat{x}, \widehat{v}) \in \mathcal{H} \times \mathcal{G}$ such that $-L^*\widehat{v} \in \partial f(\widehat{x})$ and $L\widehat{x} \in \partial g^*(\widehat{v})$ then $\widehat{x}$ (resp. $\widehat{v}$) is a solution to the primal (resp. dual) problem.

If $(\widehat{x}, \widehat{v}) \in \mathcal{H} \times \mathcal{G}$ is such that $-L^*\widehat{v} \in \partial f(\widehat{x})$ and $L\widehat{x} \in \partial g^*(\widehat{v})$, then $(\widehat{x}, \widehat{v})$ is called a **Kuhn-Tucker point**.

Fenchel-Rockafellar duality

Proof:

$$0 \in \partial f(\widehat{x}) + L^* \partial g(L\widehat{x}) \subset \partial(f + g \circ L)(\widehat{x}).$$

Then, according to Fermat rule, $\widehat{x}$ is a solution to the primal problem.
In addition, there exists $\widehat{v} \in \mathcal{G}$ such that

$$\begin{cases} 0 \in \partial f(\widehat{x}) + L^* \widehat{v} \\ \widehat{v} \in \partial g(L\widehat{x}) \end{cases} \quad \Leftrightarrow \quad \begin{cases} -L^* \widehat{v} \in \partial f(\widehat{x}) \\ L\widehat{x} \in \partial g^*(\widehat{v}). \end{cases}$$

We have also $\widehat{x} \in \partial f^*(-L^* \widehat{v})$, which implies that

$$0 \in -L \partial f^*(-L^* \widehat{v}) + \partial g^*(\widehat{v}).$$

On the other hand,

$$0 \in -L \partial f^*(-L^* \widehat{v}) + \partial g^*(\widehat{v}) \subset \partial\big(f^* \circ (-L^*) + g^*\big)(\widehat{v})$$

$\Rightarrow$ $\widehat{v}$ solution to the dual problem.
The second assertion is shown in a similar manner.

Fenchel-Rockafellar duality

Particular case:
If $f = \varphi + \frac{1}{2}\|\cdot - z\|^2$ where $\varphi \in \Gamma_0(\mathcal{H})$ and $z \in \mathcal{H}$, then

$$\begin{aligned} -L^*\widehat{v} \in \partial f(\widehat{x}) &\Leftrightarrow -L^*\widehat{v} \in \partial\varphi(\widehat{x}) + \widehat{x} - z \\ &\Leftrightarrow 0 \in \widehat{x} + L^*\widehat{v} - z + \partial\varphi(\widehat{x}). \end{aligned}$$

Hence,

$$\widehat{x} = \mathrm{prox}_{\varphi}(-L^*\widehat{v} + z).$$

Link with Lagrange duality

Minimax problem

Let $\mathcal{H}$ and $\mathcal{G}$ be two real Hilbert spaces.
Let $f\colon \mathcal{H} \to \,]-\infty,+\infty]$, $g\colon \mathcal{G} \to \,]-\infty,+\infty]$, and $L \in \mathcal{B}(\mathcal{H},\mathcal{G})$.
The primal problem is equivalent to finding

$$\mu = \inf_{(x,y)\in\mathcal{H}\times\mathcal{G}} \sup_{v\in\mathcal{G}} \mathcal{L}(x,y,v)$$

where $\mathcal{L}$ is the **Lagrange function** defined as

$$\big(\forall (x,y,v)\in\mathcal{H}\times\mathcal{G}^2\big) \quad \mathcal{L}(x,y,v) = f(x) + g(y) + \langle v \mid Lx - y\rangle$$

Link with Lagrange duality

Minimax problem

Let $\mathcal{H}$ and $\mathcal{G}$ be two real Hilbert spaces.
Let $f\colon \mathcal{H} \to \,]-\infty,+\infty]$, $g\colon \mathcal{G} \to \,]-\infty,+\infty]$, and $L \in \mathcal{B}(\mathcal{H},\mathcal{G})$.
The primal problem is equivalent to finding

$$\mu = \inf_{(x,y)\in\mathcal{H}\times\mathcal{G}} \sup_{v\in\mathcal{G}} \mathcal{L}(x,y,v)$$

where $\mathcal{L}$ is the **Lagrange function** defined as

$$(\forall(x,y,v)\in\mathcal{H}\times\mathcal{G}^2) \quad \mathcal{L}(x,y,v) = f(x)+g(y)+\langle v \mid Lx-y\rangle$$

Proof:

$$\begin{aligned}\mu = \inf_{x\in\mathcal{H}} f(x)+g(Lx) &= \inf_{(x,y)\in\mathcal{H}\times\mathcal{G}} f(x)+g(y)+\iota_{\{0\}}(Lx-y)\\ &= \inf_{(x,y)\in\mathcal{H}\times\mathcal{G}} f(x)+g(y)+\sup_{v\in\mathcal{G}}\langle v \mid Lx-y\rangle\\ &= \inf_{(x,y)\in\mathcal{H}\times\mathcal{G}} \sup_{v\in\mathcal{G}} f(x)+g(y)+\langle v \mid Lx-y\rangle\,.\end{aligned}$$

Link with Lagrange duality

Maximin problem

Let $\mathcal{H}$ and $\mathcal{G}$ be two real Hilbert spaces.
Let $f\colon \mathcal{H} \to]-\infty,+\infty]$, $g\colon \mathcal{G} \to]-\infty,+\infty]$, and $L \in \mathcal{B}(\mathcal{H},\mathcal{G})$.
The dual problem is equivalent to finding

$$-\mu^* = \sup_{v\in\mathcal{G}} \inf_{(x,y)\in\mathcal{H}\times\mathcal{G}} \mathcal{L}(x,y,v)$$

where $\mathcal{L}$ is the **Lagrange function** defined as

$$(\forall (x,y,v)\in\mathcal{H}\times\mathcal{G}^2) \quad \mathcal{L}(x,y,v) = f(x)+g(y)+\langle v \mid Lx-y\rangle$$

Proof:

$$\begin{aligned}\mu^* = \inf_{v\in\mathcal{G}} f^*(-L^*v)+g^*(v) &= \inf_{v\in\mathcal{G}} \Big(\sup_{x\in\mathcal{H}} \langle x \mid -L^*v\rangle - f(x)\Big) + \Big(\sup_{y\in\mathcal{G}} \langle y \mid v\rangle - g(y)\Big)\\ &= \inf_{v\in\mathcal{G}} -\Big(\inf_{(x,y)\in\mathcal{H}\times\mathcal{G}} f(x)+g(y)+\langle v \mid Lx-y\rangle\Big)\\ &= -\sup_{v\in\mathcal{G}} \inf_{(x,y)\in\mathcal{H}\times\mathcal{G}} f(x)+g(y)+\langle v \mid Lx-y\rangle\,.\end{aligned}$$

Link with Lagrange duality

Maximin problem

Let $\mathcal{H}$ and $\mathcal{G}$ be two real Hilbert spaces.
Let $f\colon \mathcal{H} \to]-\infty,+\infty]$, $g\colon \mathcal{G} \to]-\infty,+\infty]$, and $L \in \mathcal{B}(\mathcal{H},\mathcal{G})$.
The dual problem is equivalent to finding

$$-\mu^* = \sup_{v\in\mathcal{G}} \inf_{(x,y)\in\mathcal{H}\times\mathcal{G}} \mathcal{L}(x,y,v)$$

where $\mathcal{L}$ is the **Lagrange function** defined as

$$(\forall (x,y,v) \in \mathcal{H}\times\mathcal{G}^2) \quad \mathcal{L}(x,y,v) = f(x) + g(y) + \langle v \mid Lx - y\rangle$$

Remark: v is called the Lagrange multiplier associated with the constraint $Lx = y$.

Link with Lagrange duality

Let $(\widehat{x},\widehat{y},\widehat{v}) \in \mathcal{H}\times\mathcal{G}^2$.
$(\widehat{x},\widehat{y},\widehat{v})$ is a saddle point of the Lagrange function $\mathcal{L}$ if

$$\big(\forall (x,y,v)\in\mathcal{H}\times\mathcal{G}^2\big)\qquad \mathcal{L}(\widehat{x},\widehat{y},v)\le\mathcal{L}(\widehat{x},\widehat{y},\widehat{v})\le\mathcal{L}(x,y,\widehat{v}).$$

Let $\mathcal{H}$ and $\mathcal{G}$ be two real Hilbert spaces.
Let $f\in\Gamma_0(\mathcal{H})$, $g\in\Gamma_0(\mathcal{G})$, and $L\in\mathcal{B}(\mathcal{H},\mathcal{G})$. Let $(\widehat{x},\widehat{y},\widehat{v})\in\mathcal{H}\times\mathcal{G}^2$.
Assume that $\operatorname{int}(\operatorname{dom} g)\cap L(\operatorname{dom} f)\neq\varnothing$ or $\operatorname{dom} g\cap\operatorname{int}\big(L(\operatorname{dom} f)\big)\neq\varnothing$.

$$(\widehat{x},\widehat{y},\widehat{v})\quad \text{is a saddle point of the Lagrange function}$$
$$\Updownarrow$$
$$(\widehat{x},\widehat{v})\quad \text{is a Kuhn-Tucker point and } \widehat{y}=L\widehat{x}.$$

Link with Lagrange duality

Proof (⇒): If $(\widehat{x}, \widehat{y}, \widehat{v})$ is a saddle point of $\mathcal{L}$, then it is a critical point of $\mathcal{L}$, that is

$$\begin{cases} 0 \in \partial_x \mathcal{L}(\widehat{x}, \widehat{y}, \widehat{v}) = \partial f(\widehat{x}) + L^* \widehat{v} \\ 0 \in \partial_y \mathcal{L}(\widehat{x}, \widehat{y}, \widehat{v}) = \partial g(\widehat{y}) - \widehat{v} \\ 0 = \nabla_v \mathcal{L}(\widehat{x}, \widehat{y}, \widehat{v}) = L\widehat{x} - \widehat{y} \end{cases}$$

$$\Leftrightarrow \quad \begin{cases} -L^* \widehat{v} \in \partial f(\widehat{x}) \\ \widehat{v} \in \partial g(\widehat{y}) \\ \widehat{y} = L\widehat{x} \end{cases}$$

$$\Leftrightarrow \quad \begin{cases} -L^* \widehat{v} \in \partial f(\widehat{x}) \\ L\widehat{x} \in \partial g^*(\widehat{v}) \\ \widehat{y} = L\widehat{x}. \end{cases}$$

Link with Lagrange duality

<u>Proof</u> (⇐): Conversely, assume that $(\widehat{x}, \widehat{v})$ is a Kuhn-Tucker point and $\widehat{y} = L\widehat{x}$. Since $\widehat{x}$ (resp. $\widehat{v}$) is a solution to the primal (resp. dual) problem, then

$$\mu = \inf_{(x,y)\in\mathcal{H}\times\mathcal{G}} \sup_{v\in\mathcal{G}} \mathcal{L}(x,y,v) = \sup_{v\in\mathcal{G}} \mathcal{L}(\widehat{x},\widehat{y},v)$$

$$-\mu^* = \sup_{v\in\mathcal{G}} \inf_{(x,y)\in\mathcal{H}\times\mathcal{G}} \mathcal{L}(x,y,v) = \inf_{(x,y)\in\mathcal{H}\times\mathcal{G}} \mathcal{L}(x,y,\widehat{v}).$$

By strong duality, $\sup_{v\in\mathcal{G}} \mathcal{L}(\widehat{x},\widehat{y},v) = \inf_{(x,y)\in\mathcal{H}\times\mathcal{G}} \mathcal{L}(x,y,\widehat{v})$, which can be rewritten as

$$\big(\forall(x,y,v)\in\mathcal{H}\times\mathcal{G}^2\big) \qquad \mathcal{L}(\widehat{x},\widehat{y},v) \le \mathcal{L}(x,y,\widehat{v})$$

or equivalently

$$\big(\forall(x,y,v)\in\mathcal{H}\times\mathcal{G}^2\big) \qquad \mathcal{L}(\widehat{x},\widehat{y},v) \le \mathcal{L}(\widehat{x},\widehat{y},\widehat{v}) \le \mathcal{L}(x,y,\widehat{v}).$$

Exercise

We are interested in the estimation of a piecewise estimate x by means of total variation leading to the minimization problem:

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

where $(Lx)^{(i)} = x^{(i+1)} - x^{(i)}$ for every $i \in \{1, \ldots, N-1\}$ such that $L \in \mathbb{R}^{(N-1 \times N}$ denoted a finite difference operator.

- Prove that the dual problem can be written as

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

and that:

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

Exercise

We are interested in the estimation of a piecewise estimate x by means of total variation leading to the minimization problem:

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

where $(Lx)^{(i)} = x^{(i+1)} - x^{(i)}$ for every $i \in \{1, \ldots, N-1\}$ such that $L \in \mathbb{R}^{(N-1 \times N}$ denoted a finite difference operator.

- Prove that optimality conditions can be written

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

Proximity operator: motivation

Let $\mathcal{H}$ be a real Hilbert space. Let $f \in \Gamma_0(\mathcal{H})$ have a Lipschitz gradient with Lipschitz constant $\beta > 0$.
Find

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

- **Gradient descent algorithm**

  Set $\gamma \in \,]0, +\infty[$ and $x_0 \in \mathcal{H}$.
  For $n = 0, 1 \dots$
  $\lfloor\; x_{n+1} = x_n - \gamma \nabla f(x_n)$.

The sequence $(x_n)_{n \in \mathbb{N}}$ generated by this *explicit* scheme converges to a minimizer of f provided that such a minimizer exists and $\gamma \in]0, 2/\beta[$.

Proximity operator: motivation

Let $\mathcal{H}$ be a real Hilbert space. Let $f \in \Gamma_0(\mathcal{H})$ have a Lipschitz gradient with Lipschitz constant $\beta > 0$.
Find

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

- **Alternative algorithm**

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

Questions:

- How to determine x_{n+1} at each iteration n of this *implicit* scheme ?
- Which values of γ guarantee the convergence of $(x_n)_{n \in \mathbb{N}}$?
- What to do if f is nonsmooth ?

Proximity operator: definition

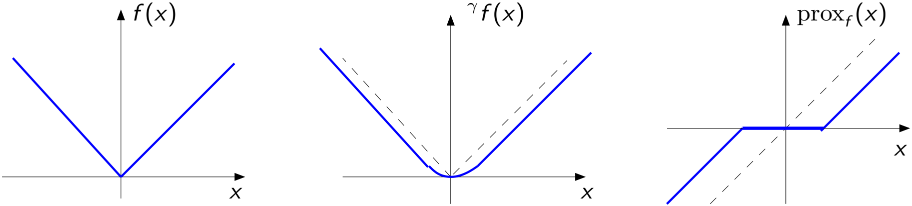
Let $\mathcal{H}$ be a Hilbert space. Let $f \in \Gamma_0(\mathcal{H})$.

- The **Moreau envelope** of f of parameter $\gamma \in \,]0,+\infty[$ is

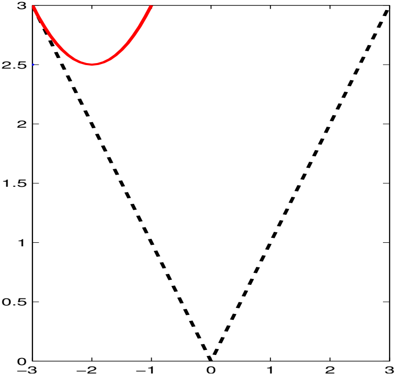
$$^{\gamma}f\colon \mathcal{H} \to \mathbb{R}\colon x \mapsto \inf_{y\in\mathcal{H}} f(y) + \frac{1}{2\gamma}\|y-x\|^2.$$

- The **proximity operator** of f is

$$\mathrm{prox}_f\colon \mathcal{H} \to \mathcal{H}\colon x \mapsto \operatorname*{argmin}_{y\in\mathcal{H}} \; f(y) + \frac{1}{2}\|y-x\|^2.$$

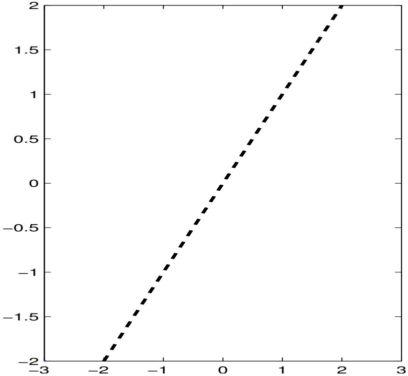


Proximity operator: definition



Moreau envelope

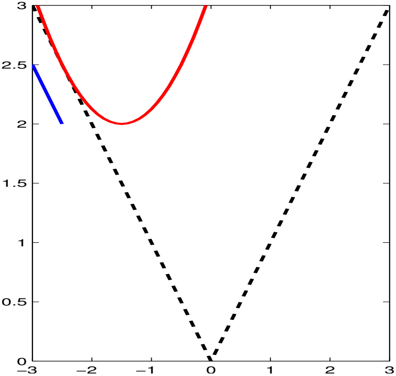
$^{\gamma}f(x) = \inf_{y \in \mathcal{H}} f(y) + \frac{1}{2\gamma}\|y - x\|^2$



Proximity operator

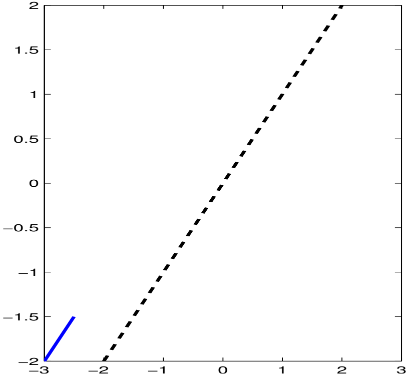
$\mathrm{prox}_f(x) = \underset{y \in \mathcal{H}}{\mathrm{argmin}} \; f(y) + \frac{1}{2}\|y - x\|^2$

Proximity operator: definition



Moreau envelope

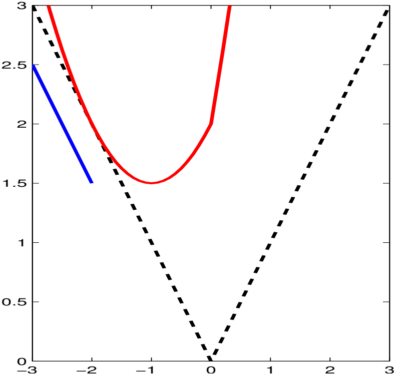
$$^{\gamma}f(x) = \inf_{y \in \mathcal{H}} f(y) + \frac{1}{2\gamma}\|y - x\|^2$$



Proximity operator

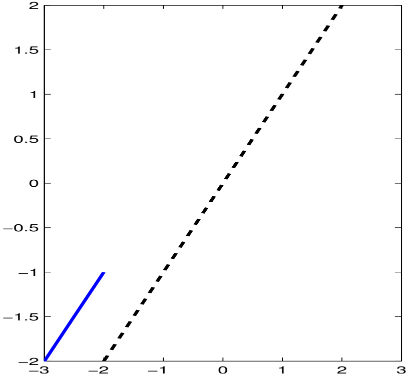
$$\mathrm{prox}_f(x) = \underset{y \in \mathcal{H}}{\mathrm{argmin}} \; f(y) + \frac{1}{2}\|y - x\|^2$$

Proximity operator: definition



Moreau envelope

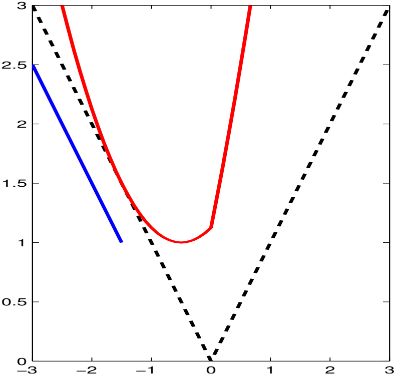
$^{\gamma}f(x) = \inf_{y \in \mathcal{H}} f(y) + \frac{1}{2\gamma}\|y - x\|^2$



Proximity operator

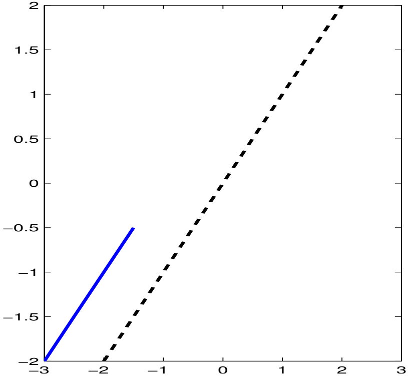
$\mathrm{prox}_f(x) = \underset{y \in \mathcal{H}}{\mathrm{argmin}} \; f(y) + \frac{1}{2}\|y - x\|^2$

Proximity operator: definition



Moreau envelope

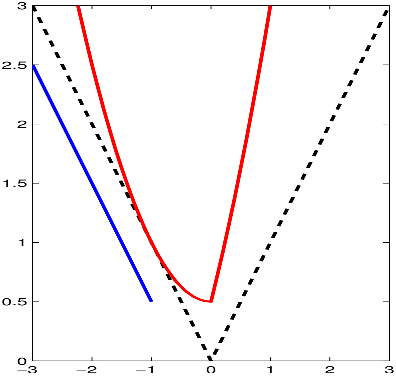
$$^{\gamma}f(x) = \inf_{y\in\mathcal{H}} f(y) + \frac{1}{2\gamma}\|y-x\|^2$$



Proximity operator

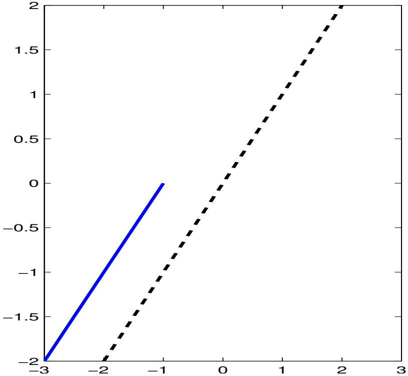
$$\mathrm{prox}_f(x) = \underset{y\in\mathcal{H}}{\mathrm{argmin}}\ f(y) + \frac{1}{2}\|y-x\|^2$$

Proximity operator: definition



Moreau envelope

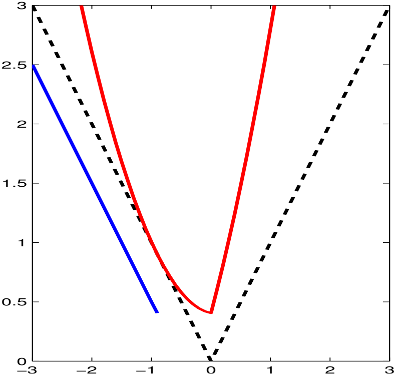
$^{\gamma}f(x) = \inf_{y\in\mathcal{H}} f(y) + \frac{1}{2\gamma}\|y-x\|^2$



Proximity operator

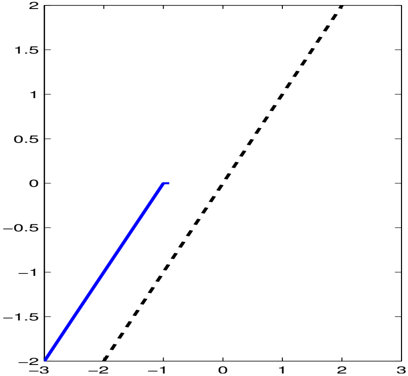
$\mathrm{prox}_f(x) = \underset{y\in\mathcal{H}}{\mathrm{argmin}}\; f(y) + \frac{1}{2}\|y-x\|^2$

Proximity operator: definition



Moreau envelope

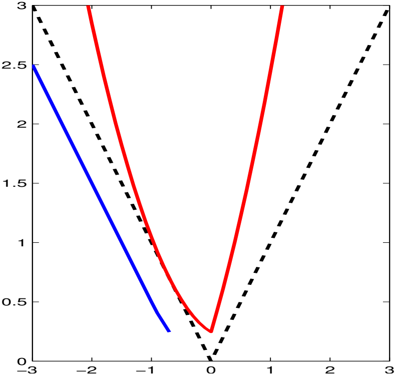
$^{\gamma}f(x) = \inf_{y\in\mathcal{H}} f(y) + \frac{1}{2\gamma}\|y-x\|^2$



Proximity operator

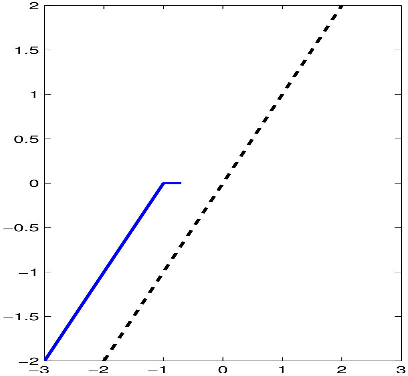
$\mathrm{prox}_f(x) = \underset{y\in\mathcal{H}}{\operatorname{argmin}} \; f(y) + \frac{1}{2}\|y-x\|^2$

Proximity operator: definition



Moreau envelope

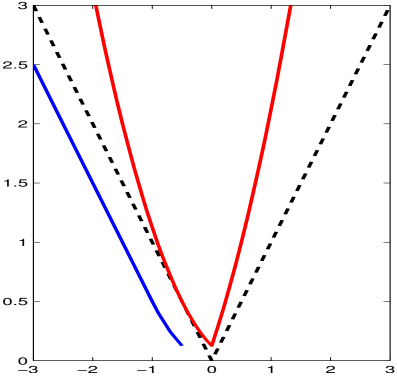
$^{\gamma}f(x) = \inf_{y \in \mathcal{H}} f(y) + \frac{1}{2\gamma}\|y - x\|^2$



Proximity operator

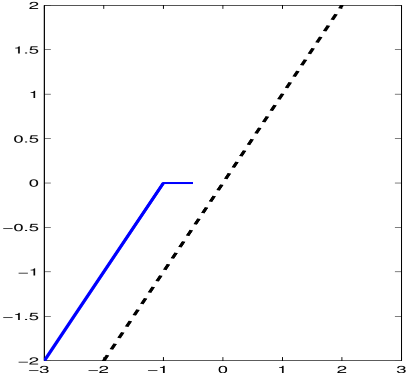
$\mathrm{prox}_f(x) = \underset{y \in \mathcal{H}}{\operatorname{argmin}} \; f(y) + \frac{1}{2}\|y - x\|^2$

Proximity operator: definition



Moreau envelope

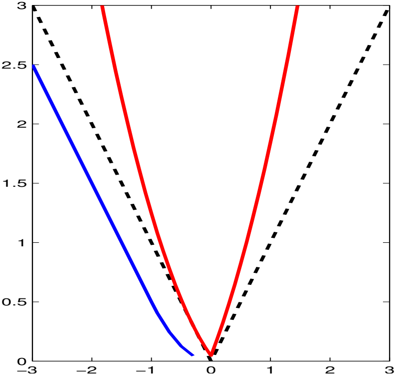
$^{\gamma}f(x) = \inf_{y\in\mathcal{H}} f(y) + \frac{1}{2\gamma}\|y-x\|^2$



Proximity operator

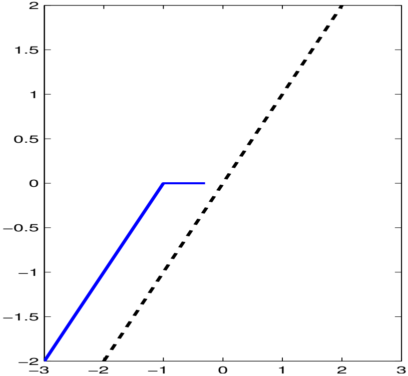
$\mathrm{prox}_f(x) = \underset{y\in\mathcal{H}}{\mathrm{argmin}}\ \ f(y) + \frac{1}{2}\|y-x\|^2$

Proximity operator: definition



Moreau envelope

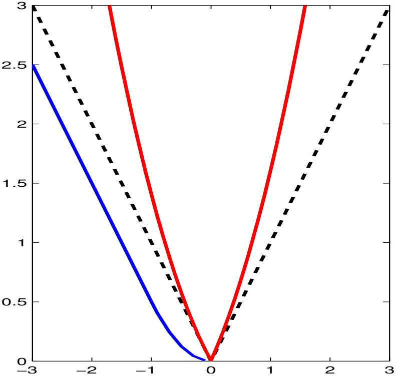
$^{\gamma}f(x) = \inf_{y\in\mathcal{H}} f(y) + \frac{1}{2\gamma}\|y-x\|^2$



Proximity operator

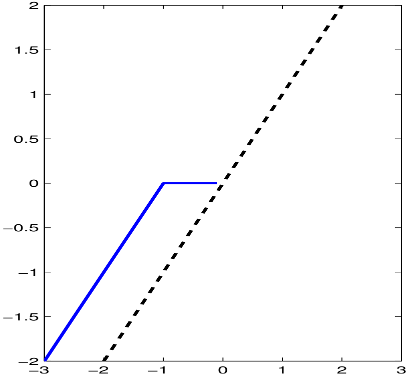
$\mathrm{prox}_f(x) = \underset{y\in\mathcal{H}}{\text{argmin}} \;\; f(y) + \frac{1}{2}\|y-x\|^2$

Proximity operator: definition



Moreau envelope

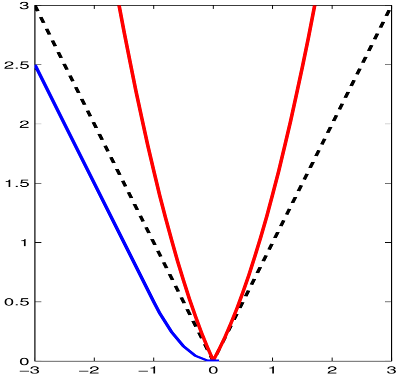
$^{\gamma}f(x) = \inf_{y \in \mathcal{H}} f(y) + \frac{1}{2\gamma}\|y - x\|^2$



Proximity operator

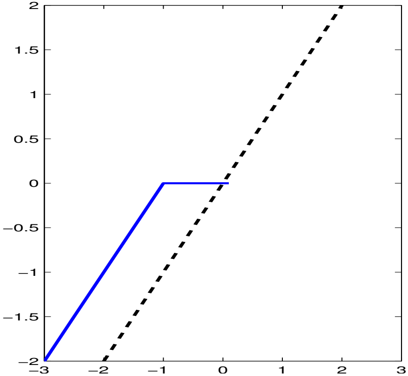
$\mathrm{prox}_f(x) = \underset{y \in \mathcal{H}}{\mathrm{argmin}} \; f(y) + \frac{1}{2}\|y - x\|^2$

Proximity operator: definition



Moreau envelope

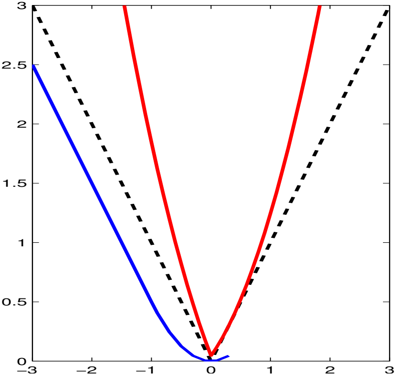
$^{\gamma}f(x) = \inf_{y \in \mathcal{H}} f(y) + \frac{1}{2\gamma}\|y - x\|^2$



Proximity operator

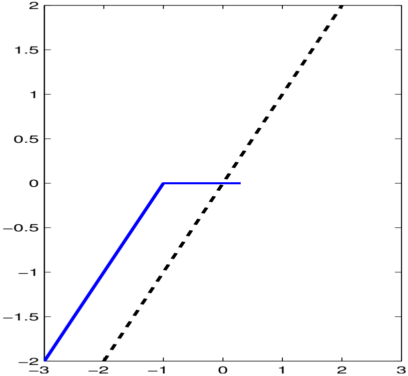
$\mathrm{prox}_f(x) = \underset{y \in \mathcal{H}}{\mathrm{argmin}}\ f(y) + \frac{1}{2}\|y - x\|^2$

Proximity operator: definition



Moreau envelope

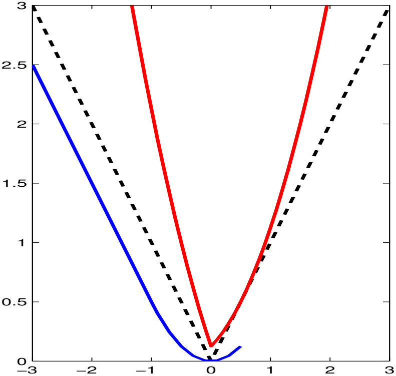
$^{\gamma}f(x) = \inf_{y\in\mathcal{H}} f(y) + \frac{1}{2\gamma}\|y-x\|^2$



Proximity operator

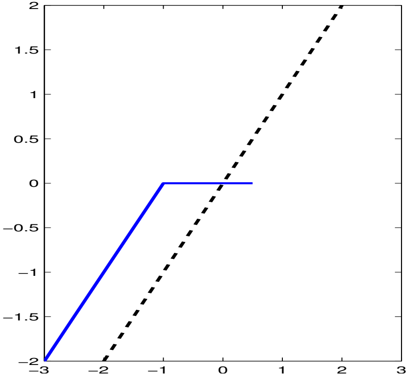
$\mathrm{prox}_f(x) = \underset{y\in\mathcal{H}}{\mathrm{argmin}} \; f(y) + \frac{1}{2}\|y-x\|^2$

Proximity operator: definition



Moreau envelope

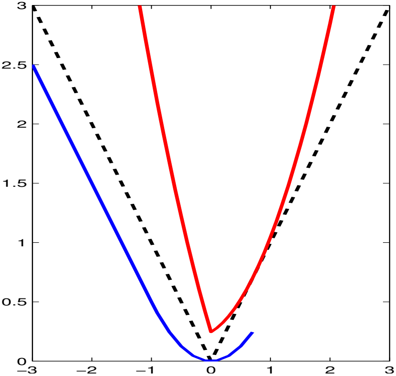
$^{\gamma}f(x) = \inf_{y\in\mathcal{H}} f(y) + \frac{1}{2\gamma}\|y-x\|^2$



Proximity operator

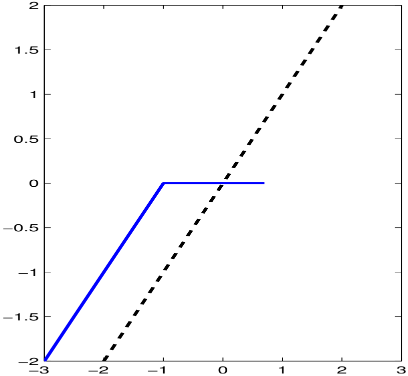
$\mathrm{prox}_f(x) = \operatorname*{argmin}_{y\in\mathcal{H}} \; f(y) + \frac{1}{2}\|y-x\|^2$

Proximity operator: definition



Moreau envelope

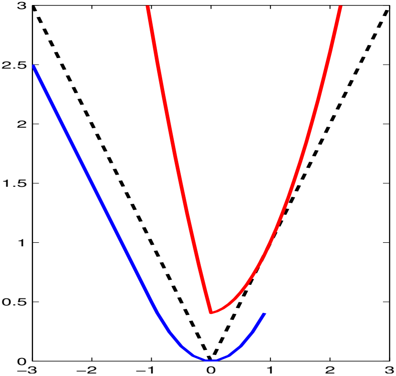
$^{\gamma}f(x) = \inf_{y\in\mathcal{H}} f(y) + \frac{1}{2\gamma}\|y - x\|^2$



Proximity operator

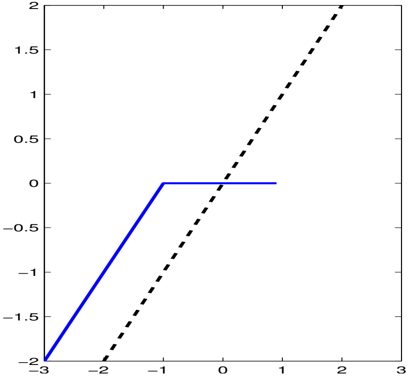
$\operatorname{prox}_f(x) = \underset{y\in\mathcal{H}}{\operatorname{argmin}}\; f(y) + \frac{1}{2}\|y - x\|^2$

Proximity operator: definition



Moreau envelope

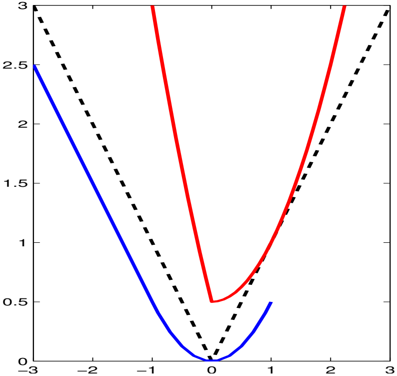
$${}^{\gamma}f(x) = \inf_{y\in\mathcal{H}} f(y) + \frac{1}{2\gamma}\|y-x\|^2$$



Proximity operator

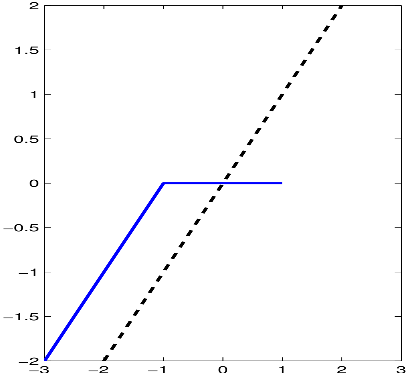
$$\mathrm{prox}_f(x) = \underset{y\in\mathcal{H}}{\operatorname{argmin}} \; f(y) + \frac{1}{2}\|y-x\|^2$$

Proximity operator: definition



Moreau envelope

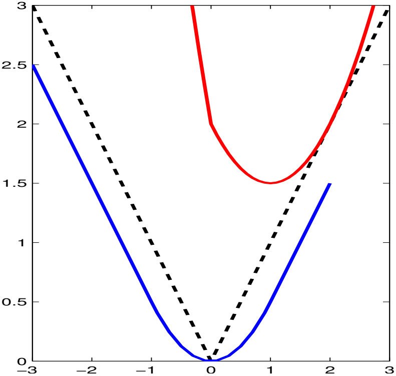
$^{\gamma}f(x) = \inf_{y\in\mathcal{H}} f(y) + \frac{1}{2\gamma}\|y-x\|^2$



Proximity operator

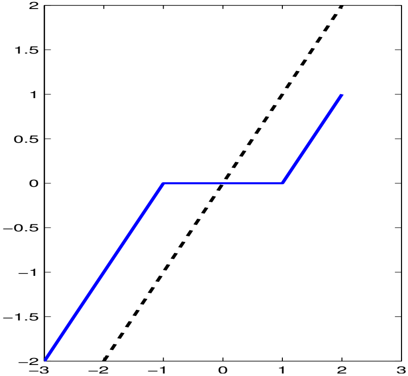
$\operatorname{prox}_f(x) = \underset{y\in\mathcal{H}}{\operatorname{argmin}} \; f(y) + \frac{1}{2}\|y-x\|^2$

Proximity operator: definition



Moreau envelope

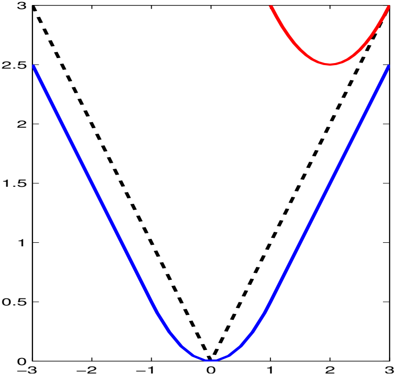
$^{\gamma}f(x) = \inf_{y\in\mathcal{H}} f(y) + \frac{1}{2\gamma}\|y-x\|^2$



Proximity operator

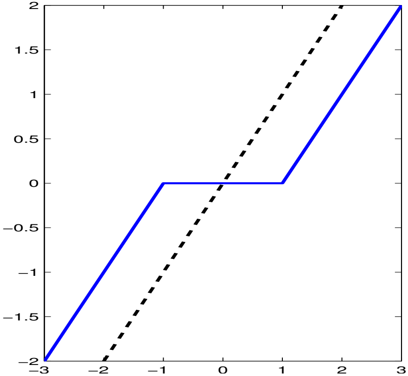
$\mathrm{prox}_f(x) = \underset{y\in\mathcal{H}}{\mathrm{argmin}}\ \ f(y) + \frac{1}{2}\|y-x\|^2$

Proximity operator: definition



Moreau envelope

$^{\gamma}f(x) = \inf_{y\in\mathcal{H}} f(y) + \frac{1}{2\gamma}\|y-x\|^2$



Proximity operator

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

Proximity operator: existence and uniqueness

Let $f \in \Gamma_0(\mathcal{H})$ and $\gamma \in \,]0,+\infty[$.
For every $x \in \mathcal{H}$, there exists a unique vector $p \in \mathcal{H}$ such that

$$f(p) + \frac{1}{2\gamma}\|p - x\|^2 = \inf_{y \in \mathcal{H}} f(y) + \frac{1}{2\gamma}\|y - x\|^2 .$$

Proof: $f \in \Gamma_0(\mathcal{H}) \Leftrightarrow f^* \in \Gamma_0(\mathcal{H})$. Thus, there exists $u \in \mathcal{H}$ such that $f^*(u) \in \mathbb{R}$. According to Fenchel-Young inequality, we have

$$(\forall y \in \mathcal{H}) \qquad f(y) \geq \langle u \mid y \rangle - f^*(u).$$

Then, $f(y) + (2\gamma)^{-1}\|y - x\|^2 \to +\infty$ when $\|y\| \to +\infty$.
Furthermore $(2\gamma)^{-1}\|\cdot - x\|^2$ being strictly convex, $f + (2\gamma)^{-1}\|\cdot - x\|^2$ is a strictly convex coercive function.

Proximity operator: characterization

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

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

Proximity operator: characterization

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

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

<u>Proof</u>: By using Fermat's rule, for every $x \in \mathcal{H}$, $p = \mathrm{prox}_f(x)$ if and only if

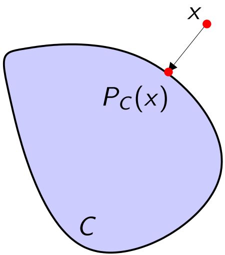
$$\begin{aligned}
& p = \underset{y \in \mathcal{H}}{\arg\min} \; f(y) + \frac{1}{2}\|y - x\|^2 \\
\Leftrightarrow \quad & 0 \in \partial\Big(f + \frac{1}{2}\|\cdot - x\|^2\Big)(p) \\
\Leftrightarrow \quad & 0 \in \partial f(p) + p - x \\
\Leftrightarrow \quad & x \in (\mathrm{Id} + \partial f)(p).
\end{aligned}$$

Proximity operator: examples

Projection :

Let $\mathcal{H}$ be a Hilbert space. Let C be a nonempty closed convex subset of $\mathcal{H}$.

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



Proximity operator: examples

Projection :
Let $\mathcal{H}$ be a Hilbert space. Let C be a nonempty closed convex subset of $\mathcal{H}$.

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

Remark :

- $p = P_C(x) \Leftrightarrow x - p \in \partial\iota_C(p) = N_C(p)$
  $\Leftrightarrow (\forall y \in C)\, \langle y - p \mid x - p \rangle \leq 0$.
  Particular case: if C is a vector space: $p = P_C(x) \Leftrightarrow x - p \in C^{\perp}$.
- ${}^{\gamma}\iota_C = (2\gamma)^{-1} d_C^2$ where d_C distance to the convex set C is defined by $d_C \colon x \mapsto \inf_{y \in C} \|y - x\| = \|x - P_C x\|$.

Proximity operator: examples

Power q function with $q \geq 1$:

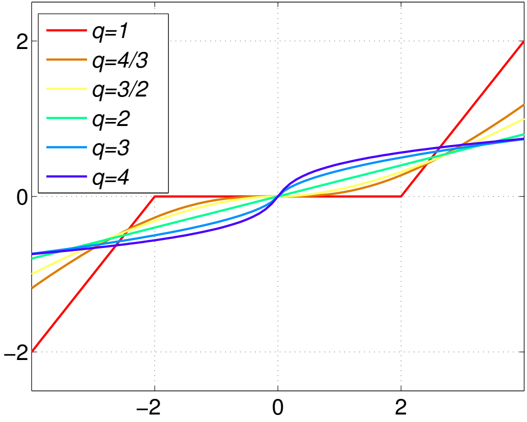
Let $\chi > 0$, $q \in [1, +\infty[$ and $\varphi\colon \mathbb{R} \to]-\infty, +\infty] : \xi \mapsto \chi|\xi|^q$.

Then, for every $\xi \in \mathbb{R}$,

$$\mathrm{prox}_{\varphi}\xi = \begin{cases} \mathrm{sign}(\xi)\max\{|\xi| - \chi, 0\} & \text{if } q = 1 \\ \xi + \frac{4\chi}{3\,.\,2^{1/3}}\left((\epsilon - \xi)^{1/3} - (\epsilon + \xi)^{1/3}\right) & \\ \qquad \text{where} \quad \epsilon = \sqrt{\xi^2 + 256\chi^3/729} & \text{if } q = \frac{4}{3} \\ \xi + \frac{9\chi^2 \mathrm{sign}(\xi)}{8}\left(1 - \sqrt{1 + \frac{16|\xi|}{9\chi^2}}\right) & \text{if } q = \frac{3}{2} \\ \frac{\xi}{1+2\chi} & \text{if } q = 2 \\ \mathrm{sign}(\xi)\frac{\sqrt{1+12\chi|\xi|}-1}{6\chi} & \text{if } q = 3 \\ \left(\frac{\epsilon+\xi}{8\chi}\right)^{1/3} - \left(\frac{\epsilon-\xi}{8\chi}\right)^{1/3} \quad \text{where} \quad \epsilon = \sqrt{\xi^2 + 1/(27\chi)} & \text{if } q = 4 \end{cases}$$

Proximity operator: examples

Power q function with $q \geq 1$ and $\chi = 2$.



Proximity operator: examples

Quadratic function :
Let $\mathcal{H}$ and $\mathcal{G}$ be two Hilbert spaces.
Let $L \in \mathcal{B}(\mathcal{G}, \mathcal{H})$, $\gamma \in \,]0, +\infty[$ and $z \in \mathcal{G}$.

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

- Exercise : Prove this property.

Proximity operator: properties

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

Properties	$g(x)$	$\mathrm{prox}_g x$
Translation	$f(x-z), z \in \mathcal{H}$	$z + \mathrm{prox}_f(x-z)$
Quadratic perturbation	$f(x) + \alpha \parallel x \parallel^2 /2 + \langle z \mid x \rangle + \gamma$ $z \in \mathcal{H}, \alpha > 0, \gamma \in \mathbb{R}$	$\mathrm{prox}_{\frac{f}{\alpha+1}}(\frac{x-z}{\alpha+1})$
Scaling	$f(\rho x), \rho \in \mathbb{R}^*$	$\frac{1}{\rho}\mathrm{prox}_{\rho^2 f}(\rho x)$
Reflexion	$f(-x)$	$-\mathrm{prox}_f(-x)$
Moreau enveloppe	${}^{\gamma}f(x) = \inf_{y\in\mathcal{H}} f(y) + \frac{1}{2\gamma}\|x-y\|^2$ $\gamma > 0$	$\frac{1}{1+\gamma}\left(\gamma x + \mathrm{prox}_{(1+\gamma)f}(x)\right)$

Exercice

Establish all the properties in the previous table.

Proximity operator: properties

For every $i \in \{1, \ldots, n\}$, let $\mathcal{H}_i$ be a Hilbert space and let $f_i \in \Gamma_0(\mathcal{H}_i)$.
If

$$\big(\forall x = (x_1, \ldots, x_n) \in \mathcal{H}_1 \times \cdots \times \mathcal{H}_n\big) \quad f(x) = \sum_{i=1}^{n} f_i(x_i),$$

then

$$\big(\forall x = (x_1, \ldots, x_n) \in \mathcal{H}_1 \times \cdots \times \mathcal{H}_n\big) \quad \mathrm{prox}_f(x) = \big(\mathrm{prox}_{f_i}(x_i)\big)_{1 \le i \le n}.$$

Proximity operator: properties

Let $\mathcal{H}$ be a separable Hilbert space.
Let $(b_i)_{i\in I}$ be an orthonormal basis of $\mathcal{H}$.
For every $i \in I$, let $\varphi_i \in \Gamma_0(\mathbb{R})$ such that $\varphi_i \geq 0$. For every $x \in \mathcal{H}$, if

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

then

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

Remark: The assumption $(\forall i \in I)\ \varphi_i \geq 0$ can be relaxed if $\mathcal{H}$ is finite dimensional.

Proximity operator: properties

Let $\mathcal{H}$ be a separable Hilbert space.
Let $(b_i)_{i\in I}$ be an orthonormal basis of $\mathcal{H}$.
For every $i \in I$, let $\varphi_i \in \Gamma_0(\mathbb{R})$ such that $\varphi_i \geq 0$. For every $x \in \mathcal{H}$, if

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

then

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

Example: $\mathcal{H} = \mathbb{R}^N$, $(b_i)_{1\leq i\leq N}$ canonical basis of $\mathbb{R}^N$, $f = \lambda\|\cdot\|_1$ with $\lambda \in [0, +\infty[$.

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

Proximity operator: properties

Moreau decomposition formula

Let $\mathcal{H}$ be a Hilbert space, $f \in \Gamma_0(\mathcal{H})$ and $\gamma \in \,]0,+\infty[$.

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

Proof:

$$\begin{aligned}
p = \mathrm{prox}_{\gamma f^*} x &\Leftrightarrow x - p \in \gamma \partial f^*(p) \\
&\Leftrightarrow p \in \partial f\left(\frac{x-p}{\gamma}\right) \\
&\Leftrightarrow \frac{x}{\gamma} - \frac{x-p}{\gamma} \in \frac{1}{\gamma}\partial f\left(\frac{x-p}{\gamma}\right) \\
&\Leftrightarrow \frac{x-p}{\gamma} = \mathrm{prox}_{\gamma^{-1} f}(\gamma^{-1} x) \\
&\Leftrightarrow p = x - \gamma \mathrm{prox}_{\gamma^{-1} f}(\gamma^{-1} x).
\end{aligned}$$

Proximity operator: properties

Moreau decomposition formula

Let $\mathcal{H}$ be a Hilbert space, $f \in \Gamma_0(\mathcal{H})$ and $\gamma \in \,]0,+\infty[$.

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

Example: If $\mathcal{H} = \mathbb{R}^N$, $f = \frac{1}{q}\|\cdot\|_q^q$ with $q \in]1,+\infty[$, then $f^* = \frac{1}{q^*}\|\cdot\|_{q^*}^{q^*}$ with $1/q + 1/q^* = 1$, and

$$(\forall x \in \mathbb{R}^N) \qquad \mathrm{prox}_{\frac{\gamma}{q^*}\|\cdot\|_{q^*}^{q^*}} x = x - \gamma \mathrm{prox}_{\frac{1}{\gamma q}\|\cdot\|_q^q}(\gamma^{-1} x).$$

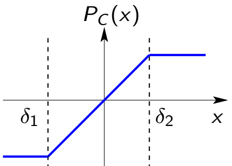
Proximity operator: examples

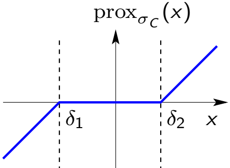
Support function :
Let $\mathcal{H}$ be a Hilbert space and $C \subset \mathcal{H}$ be nonempty closed convex.

$$(\forall x \in \mathcal{H}) \qquad \mathrm{prox}_{\sigma_C} = \mathrm{Id} - P_C.$$

Soft-thresholding : $\mathcal{H} = \mathbb{R}$, $\delta_1 = \inf C$ and $\delta_2 = \sup C$. For every $x \in \mathbb{R}$,

$$\sigma_C(x) = \begin{cases} \delta_1 x & \text{if } x < 0 \\ 0 & \text{if } x = 0 \\ \delta_2 x & \text{if } x > 0 \end{cases} \Rightarrow \ \mathrm{prox}_{\sigma_C}(x) = \mathrm{soft}_C(x) = \begin{cases} x - \delta_1 & \text{if } x < \delta_1 \\ 0 & \text{if } x \in C \\ x - \delta_2 & \text{if } x > \delta_2. \end{cases}$$





Proximity operator: properties

Let $\mathcal{H}$ and $\mathcal{G}$ be two Hilbert spaces. Let $f \in \Gamma_0(\mathcal{H})$ and $L \in \mathcal{B}(\mathcal{G}, \mathcal{H})$ such that $LL^* = \mu \mathrm{Id}$ where $\mu \in \,]0, +\infty[$. Then

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

Proximity operator: properties

Let $\mathcal{H}$ and $\mathcal{G}$ be two Hilbert spaces. Let $f \in \Gamma_0(\mathcal{H})$ and $L \in \mathcal{B}(\mathcal{G}, \mathcal{H})$ such that $LL^* = \mu \mathrm{Id}$ where $\mu \in \,]0, +\infty[$. Then

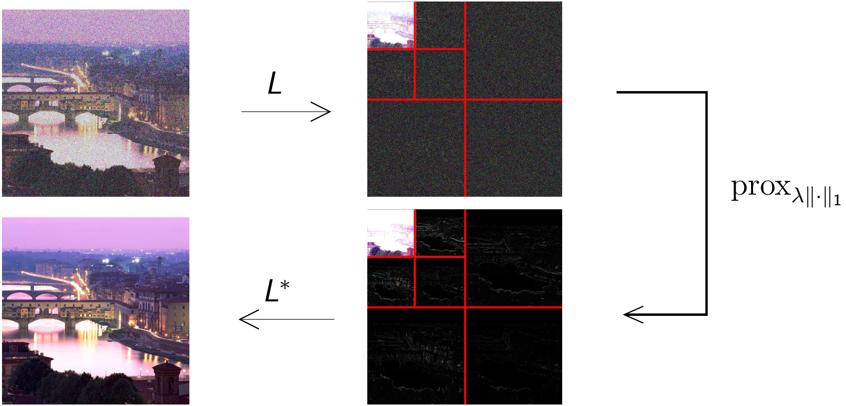
$$\mathrm{prox}_{f \circ L} = \mathrm{Id} - \mu^{-1} L^* \circ (\mathrm{Id} - \mathrm{prox}_{\mu f}) \circ L.$$

Proof: $LL^* = \mu \mathrm{Id} \Rightarrow \mathrm{ran}\, L = \mathcal{H}$ is closed, hence $V = \mathrm{ran}\,(L^*) = (\ker L)^\perp$ is closed. The orthogonal projection onto V is $P_V = L^*(LL^*)^{-1}L = \mu^{-1}L^*L$.
For every $x \in \mathcal{H}$, $p = \mathrm{prox}_{f\circ L} x \Leftrightarrow x - p \in= \partial(f \circ L)(p) = L^* \partial f(Lp)$ (since $\mathrm{ran}\, L = \mathcal{H}$). Thus, $x - p \in V$.
It can be deduced that $P_{V^\perp} p = P_{V^\perp} x = x - P_V x = x - \mu^{-1} L^* L x$.
Furthermore,
$x - p \in L^* \partial(Lp) \Rightarrow Lx - Lp \in \mu \partial f(Lp) \Leftrightarrow Lp = \mathrm{prox}_{\mu f}(Lx)$.
We have thus $P_V p = \mu^{-1} L^* L p = \mu^{-1} L^* \mathrm{prox}_{\mu f}(Lx)$ and
$p = P_V p + P_{V^\perp} p = x - \mu^{-1} L^* (\mathrm{Id} - \mathrm{prox}_{\mu f})(Lx)$.

Proximity operator: properties

Particular case : $L \in \mathcal{B}(\mathcal{H}, \mathcal{H})$ unitary, $\mathrm{prox}_{f \circ L} = L^* \mathrm{prox}_f L$.

- Illustration: denoising using an ℓ_1 penalty on the coefficients resulting from an orthogonal wavelet transform L.



Fixed point algorithm

Naive answer

Fixed point theorem (E. Picard, 1856-1941)

If

- $\widehat{x}$ is a fixed point of T, i.e. $\widehat{x} = T\widehat{x}$
- T is a strict contraction, i.e. there exists $\rho \in [0, 1[$ such that

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

then $(x_n)_{n \in \mathbb{N}}$ converges to $\widehat{x}$.

Naive answer

Fixed point theorem (E. Picard, 1856-1941)

If

- $\widehat{x}$ is a fixed point of T, i.e. $\widehat{x} = T\widehat{x}$
- T is a strict contraction, i.e. there exists $\rho \in [0,1[$ such that

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

then $(x_n)_{n\in\mathbb{N}}$ converges to $\widehat{x}$.

Proof: For all $n \in \mathbb{N}$,

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

Consequently, $\|x_n - \widehat{x}\| \leq \rho^n \|x_0 - \widehat{x}\|$. Hence, we have proved that $(x_n)_{n\in\mathbb{N}}$ converges linearly to $\widehat{x}$.

Objective of this part

- Extend this theorem to more general operators
  - not necessarily *strictly* contractive
  - possibly dependent on the iteration number n
  - built from composition of simpler operators (*splitting techniques*).

- Apply this to solve minimization problems.
  $\rightsquigarrow$ How to relate T to the objective function f ?

Fixed point algorithm: convergence

Let $\mathcal{H}$ be a Hilbert space.
Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$ and $\widehat{x} \in \mathcal{H}$.

- $(x_n)_{n\in\mathbb{N}}$ **converges strongly** to $\widehat{x}$ if

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

  It is denoted by $x_n \to \widehat{x}$.

- $(x_n)_{n\in\mathbb{N}}$ **converges weakly** to $\widehat{x}$ if

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

  It is denoted by $x_n \rightharpoonup \widehat{x}$.

Remark: In a finite dimensional Hilbert space, strong and weak convergences are equivalent.

Fixed point algorithm: convergence

Let $\mathcal{H}$ be a Hilbert space.
Let $(x_n)_{n\in\mathbb{N}}$ be a sequence of $\mathcal{H}$.
$(x_n)_{n\in\mathbb{N}}$ converges weakly if and only if

- $(x_n)_{n\in\mathbb{N}}$ is bounded

  and

- $(x_n)_{n\in\mathbb{N}}$ posseses at most one sequential cluster point in the weak topology.

- $\widehat{x}$ is a sequential cluster point of $(x_n)_{n\in\mathbb{N}}$ in the weak topology if there exists a subsequence $(x_{n_k})_{k\in\mathbb{N}}$ of $(x_n)_{n\in\mathbb{N}}$ that converges weakly to $\widehat{x}$.

Fixed point algorithm: convergence

Let $\mathcal{H}$ be a Hilbert space.
Let $(x_n)_{n\in\mathbb{N}}$ be a sequence of $\mathcal{H}$.
$(x_n)_{n\in\mathbb{N}}$ converges weakly if and only if

- $(x_n)_{n\in\mathbb{N}}$ is bounded

  and

- $(x_n)_{n\in\mathbb{N}}$ posseses at most one sequential cluster point in the weak topology.

Illustration:

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

$\rightarrow$ $(x_n)_{n\in\mathbb{N}}$ is bounded but it has 2 sequential cluster points: -1 and 1.

$\rightarrow$ $(x_n)_{n\in\mathbb{N}}$ does not converge.

Fixed point algorithm: convergence

Lemma 1

Let $\mathcal{H}$ be a Hilbert space and $D \subset \mathcal{H}$ nonempty.
Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in D.
$(x_n)_{n\in\mathbb{N}}$ weakly converges to a point in D if

- for every $x \in D$, $(\|x_n - x\|)_{n\in\mathbb{N}}$ converges

  and
- every weak sequential cluster point of $(x_n)_{n\in\mathbb{N}}$ lies in D.

Fixed point algorithm: convergence

Proof:

If $(\|x_n - x\|)_{n\in\mathbb{N}}$ converges, then $(\|x_n - x\|)_{n\in\mathbb{N}}$ and thus $(x_n)_{n\in\mathbb{N}}$ are bounded.
We assume that $(x_{n_k})_{k\in\mathbb{N}}$ and $(x_{n_\ell})_{\ell\in\mathbb{N}}$ are such that $x_{n_k} \rightharpoonup \widehat{x}$ and $x_{n_\ell} \rightharpoonup \widehat{x}'$ where $(\widehat{x}, \widehat{x}') \in D^2$. For every $n \in \mathbb{N}$,

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

Because $(\|x_n - \widehat{x}\|)_{n\in\mathbb{N}}$ and $(\|x_n - \widehat{x}'\|)_{n\in\mathbb{N}}$ converge, there exists $\alpha \in \mathbb{R}$ such that $\langle x_n \mid \widehat{x}' - \widehat{x}\rangle \to \alpha$ and thus
$\langle x_{n_k} \mid \widehat{x}' - \widehat{x}\rangle \to \langle \widehat{x} \mid \widehat{x}' - \widehat{x}\rangle = \alpha$. Similarly, $\langle \widehat{x}' \mid \widehat{x}' - \widehat{x}\rangle = \alpha$.
Consequently, $\|\widehat{x}' - \widehat{x}\|^2 = 0 \Rightarrow \widehat{x} = \widehat{x}'$.

Fixed point algorithm: Fejér-monotone sequence

Let $\mathcal{H}$ be a Hilbert space and D be a nonempty subset of $\mathcal{H}$.
Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$.
$(x_n)_{n\in\mathbb{N}}$ is Fejér-monotone with respect to D if

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

Let $\mathcal{H}$ be a Hilbert space and D be a nonempty subset of $\mathcal{H}$.
Let $(x_n)_{n\in\mathbb{N}}$ be Fejér-monotone with respect to D then

- $(x_n)_{n\in\mathbb{N}}$ is bounded.
- for every $x \in D$, $(\|x_n - x\|)_{n\in\mathbb{N}}$ converges.

Fixed point algorithm: Fejér-monotone sequence

Fejér-monotone convergence

Let $\mathcal{H}$ be a Hilbert space and let D be a nonempty subset of $\mathcal{H}$.
Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$.
$(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in D if

- $(x_n)_{n\in\mathbb{N}}$ is Fejér-monotone with respect to D

  and
- every weak sequential cluster point of $(x_n)_{n\in\mathbb{N}}$ lies in D.

Fixed point algorithm: Fejér-monotone sequence

Lemma 2

Let $\mathcal{H}$ be a Hilbert space and C be a nonempty closed convex subset of $\mathcal{H}$. If $(x_n)_{n\in\mathbb{N}}$ denotes a sequence in C that weakly converges to $\widehat{x}$ then $\widehat{x} \in C$.

Fixed point algorithm: Fejér-monotone sequence

Lemma 2

Let $\mathcal{H}$ be a Hilbert space and C be a nonempty closed convex subset of $\mathcal{H}$. If $(x_n)_{n\in\mathbb{N}}$ denotes a sequence in C that weakly converges to $\widehat{x}$ then $\widehat{x} \in C$.

Proof:

We have $\widehat{x} - P_C\widehat{x} \in N_C\big(P_C\widehat{x}\big)$.

Because $(\forall n \in \mathbb{N})$ $x_n \in C$, we have

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

By using $x_n \rightharpoonup \widehat{x}$, it results that $\|\widehat{x} - P_C\widehat{x}\|^2 = 0$, and thus $\widehat{x} = P_C(\widehat{x}) \in C$.

Nonexpansive operator: definition

Let $\mathcal{H}$ be a Hilbert space. Let $T : \mathcal{H} \to 2^{\mathcal{H}}$.
The set of fixed points of B is

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

Let $\mathcal{H}$ be a Hilbert space and let $C \subset \mathcal{H}$ be a nonempty closed convex set.
Let $T \colon C \to \mathcal{H}$.
T is a nonexpansive operator if $\left(\forall (x, y) \in C^2\right) \quad \|Tx - Ty\| \leq \|x - y\|$.

Nonexpansive operator: fixed point algorithm

Demiclosedness principle

Let $\mathcal{H}$ be a Hilbert space and C be a nonempty closed convex subset of $\mathcal{H}$. Let $T\colon C \to \mathcal{H}$ be a nonexpansive operator.
If $(x_n)_{n\in\mathbb{N}}$ is a sequence in C that converges weakly to $\widehat{x}$ and if $Tx_n - x_n \to 0$ then $\widehat{x} \in \operatorname{Fix} T$.

Nonexpansive operator: fixed point algorithm

Demiclosedness principle

Let $\mathcal{H}$ be a Hilbert space and C be a nonempty closed convex subset of $\mathcal{H}$. Let $T\colon C \to \mathcal{H}$ be a nonexpansive operator. If $(x_n)_{n\in\mathbb{N}}$ is a sequence in C that converges weakly to $\widehat{x}$ and if $Tx_n - x_n \to 0$ then $\widehat{x} \in \operatorname{Fix} T$.

Proof:

$x_n \rightharpoonup \widehat{x} \Rightarrow \widehat{x} \in C$ and $T\widehat{x}$ defined. For every $n \in \mathbb{N}$,

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

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

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

Nonexpansive operator: fixed point algorithm

Demiclosedness principle

Let $\mathcal{H}$ be a Hilbert space and C be a nonempty closed convex subset of $\mathcal{H}$. Let $T\colon C \to \mathcal{H}$ be a nonexpansive operator.
If $(x_n)_{n\in\mathbb{N}}$ is a sequence in C that converges weakly to $\widehat{x}$ and if $Tx_n - x_n \to 0$ then $\widehat{x} \in \operatorname{Fix} T$.

Proof:

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

Since T is nonexpansive, by using the Cauchy-Schwarz inequality,

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

$x_n \rightharpoonup \widehat{x} \Rightarrow (x_n)_{n\in\mathbb{N}}$ bounded. The result follows by taking the limit.

Nonexpansive operator: fixed point algorithm

Let $\mathcal{H}$ be a Hilbert space and C be a nonempty subset of $\mathcal{H}$.
Let $T: C \to C$ be a nonexpansive operator

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

Nonexpansive operator: fixed point algorithm

Let $\mathcal{H}$ be a Hilbert space and C be a nonempty subset of $\mathcal{H}$.
Let $T: C \to C$ be a nonexpansive operator

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

If $x_n - Tx_n \to 0$,

Nonexpansive operator: fixed point algorithm

Let $\mathcal{H}$ be a Hilbert space and C be a nonempty subset of $\mathcal{H}$.
Let $T\colon C \to C$ be a nonexpansive operator such that $\operatorname{Fix} T \neq \varnothing$.
Let $x_0 \in C$,

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

If $x_n - Tx_n \to 0$, then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in $\operatorname{Fix} T$.

Proof :

For every $n \in \mathbb{N}$ and $y \in \operatorname{Fix} T$, $\|x_{n+1} - y\| \leq \|Tx_n - Ty\| \leq \|x_n - y\|$.
$(x_n)_{n\in\mathbb{N}}$ is Fejér-monotone with respect to $\operatorname{Fix} T$.
Let $(x_{n_k})_{k\in\mathbb{N}}$ be a subsequence of $(x_n)_{n\in\mathbb{N}}$ such that $x_{n_k} \rightharpoonup \widehat{x}$ where $\widehat{x} \in \mathcal{H}$.
By assumption $x_{n_k} - Tx_{n_k} \to 0$ and thus, according to the demiclosedness principle, $\widehat{x} \in \operatorname{Fix} T$.
This shows the weak convergence of $(x_n)_{n\in\mathbb{N}}$.

Fixed point algorithm: Fejér-monotone sequence

Krasnosel'skii-Mann algorithm

Let $\mathcal{H}$ be a Hilbert space and C be a nonempty closed convex subset of $\mathcal{H}$.
Let $T\colon C \to C$ be a nonexpansive operator such that $\operatorname{Fix} T \neq \varnothing$.
Let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[0,1]$ such that

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

Let $x_0 \in C$ and $(\forall n \in \mathbb{N})\ x_{n+1} = x_n + \lambda_n(Tx_n - x_n)$. Then,

- $(x_n)_{n\in\mathbb{N}}$ is Fejér-monotone with respect to $\operatorname{Fix} T$.
- $(Tx_n - x_n)_{n\in\mathbb{N}}$ converges strongly to 0.
- $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in $\operatorname{Fix} T$.

α-averaged operator: definition

Let $\mathcal{H}$ be Hilbert space and let $C \subset \mathcal{H}$ nonempty closed convex set.
Let $A : C \to \mathcal{H}$ and let $\alpha \in]0,1[$.
A is a **α-averaged** operator if there exists a nonexpansive operator $R \colon C \to \mathcal{H}$ such that

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

Let $\mathcal{H}$ be Hilbert space and let $C \subset \mathcal{H}$ nonempty closed convex set.
Let $A : C \to \mathcal{H}$and let $\alpha \in]0,1[$.
A is a **α-averaged** operator if

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

α-averaged operator: example

Let $\mathcal{H}$ be a Hilbert space, $f \in \Gamma_0(\mathcal{H})$, $\nu \in]0, 2[$.
If f is differentiable with a ν-lipschitzian gradient then, $\mathrm{Id} - \nabla f$ is a $\nu/2$-averaged operator.

Remark : $\mathrm{Id} - \nabla f$ denotes the gradient descent operator.

α-averaged operator: example

Proof : 1) Descent lemma

For every $(x, y) \in \mathcal{H}^2$ and $t \in \mathbb{R}$, let $\varphi(t) = f(x + t(y - x))$.
φ is differentiable and $\varphi'(t) = \langle y - x \mid \nabla f(x + t(y - x))\rangle$. We have then

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

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

But, according to the Cauchy-Schwarz inequality,

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

This leads to

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

α-averaged operator: example

Proof : 2) $\mathrm{Id} - \nabla f$ is α-averaged

From the descent lemma, for every $(x, y, z) \in \mathcal{H}^3$,

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

Moreover, according to the Fenchel-Young inequality,

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

Thus,

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

α-averaged operator: example

Proof : 2) $\mathrm{Id} - \nabla f$ is α-averaged

Thus,

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

This yields

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

Consequently,

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

α-averaged operator: example

Proof : 2) $\mathrm{Id} - \nabla f$ is α-averaged

For every $(x, y) \in \mathcal{H}^2$,

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

and symmetrically

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

By summing

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

It results that

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

α-averaged operator: example

Let $\mathcal{H}$ be a Hilbert space, $f \in \Gamma_0(\mathcal{H})$, $\nu \in \,]0, +\infty[$ and $\gamma \in]0, 2/\nu[$.
If f is differentiable with a ν-lipschitzian gradient then $\mathrm{Id} - \gamma \nabla f$ is a $\gamma\nu/2$-averaged operator.

Remark : $\mathrm{Id} - \gamma \nabla f$ denotes the gradient descent operator.

α-averaged operator: example

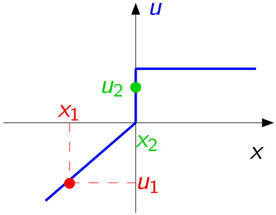
Let $\mathcal{H}$ be a Hilbert space, $f \in \Gamma_0(\mathcal{H})$.
prox_f is a 1/2-averaged operator.

α-averaged operator: example

Let $\mathcal{H}$ be a Hilbert space, $f \in \Gamma_0(\mathcal{H})$.
prox_f is a 1/2-averaged operator.

Proof:

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



α-averaged operator: example

Let $\mathcal{H}$ be a Hilbert space, $f \in \Gamma_0(\mathcal{H})$.
prox_f is a $1/2$-averaged operator.

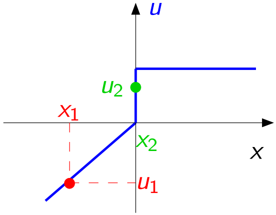
Proof:

- We recall that : $\partial f(x) = \{u \in \mathcal{H} \,|\, (\forall y \in \mathcal{H})\ \langle y - x | u \rangle + f(x) \le f(y)\}$
- Let $u_1 \in \partial f(x_1)$ and $u_2 \in \partial f(x_2)$.

  By definition:

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

  it results that $\boxed{\langle x_1 - x_2 \,|\, u_1 - u_2 \rangle \ge 0}$.

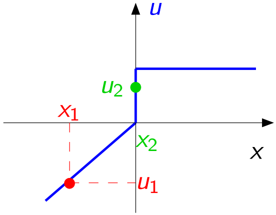


α-averaged operator: example

Let $\mathcal{H}$ be a Hilbert space, $f \in \Gamma_0(\mathcal{H})$.
prox_f is a $1/2$-averaged operator.

Proof:

- We recall that : $\partial f(x) = \{u \in \mathcal{H} \,|\, (\forall y \in \mathcal{H})\ \langle y - x | u \rangle + f(x) \leq f(y)\}$
- Let $u_1 \in \partial f(x_1)$ and $u_2 \in \partial f(x_2)$.
  By definition:
  $$\langle x_2 - x_1 | u_1 \rangle + f(x_1) \leq f(x_2)$$
  $$\langle x_1 - x_2 | u_2 \rangle + f(x_2) \leq f(x_1)$$
  it results that $\boxed{\langle x_1 - x_2 \,|\, u_1 - u_2 \rangle \geq 0}$.



- *Remark* : ∂f is a monotone operator.

α-averaged operator: example

Proof :
Let $u_1 \in \partial f(x_1)$ and $u_2 \in \partial f(x_2)$

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

We consider $u_1' \in (\mathrm{Id} + \partial f)x_1$ et $u_2' \in (\mathrm{Id} + \partial f)x_2$, it results that

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

Then, from the definition of the proximity operator,

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

We can deduce that prox_f is a 1/2-averaged operator, i.e,

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

Fixed point algorithm: α-averaged operator

Let $\mathcal{H}$ be a Hilbert space and let $\alpha \in]0,1[$.
Let $T\colon \mathcal{H} \to \mathcal{H}$ be an α-averaged operator with $\alpha \in]0,1[$ such that $\operatorname{Fix} T \neq \varnothing$.
Let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[0, 1/\alpha]$ such that

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

Let $x_0 \in \mathcal{H}$ and $(\forall n \in \mathbb{N})\ x_{n+1} = x_n + \lambda_n(Tx_n - x_n)$. The following properties are satisfied

- $(x_n)_{n\in\mathbb{N}}$ is Fejér-monotone with respect to $\operatorname{Fix} T$.
- $(Tx_n - x_n)_{n\in\mathbb{N}}$ converge strongly to 0.
- $(x_n)_{n\in\mathbb{N}}$ converge weakly to a point in $\operatorname{Fix} T$.

Fixed point algorithm: α-averaged operator

Proof :
Since T is α-average, there exists a non expansive operator R such that $T = (1-\alpha)\mathrm{Id} + \alpha R$.
Let $(\forall n \in \mathbb{N})$ $\mu_n = \alpha\lambda_n \in [0,1]$.
The iterations can be written as

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

Moreover, $\mathrm{Fix} R = \mathrm{Fix}\, T$.
$+$ Krasnosel'skii-Mann algorithm.

Optimization algorithm: *Forward-Backward*

Let $\mathcal{H}$ be a Hilbert space.
Let $f \in \Gamma_0(\mathcal{H})$.
Let $g \in \Gamma_0(\mathcal{H})$ be differentiable with a ν-Lipschitzian gradient where $\nu \in]0, +\infty[$.
Let $\gamma \in]0, 2/\nu[$ and $\delta = \min\{1, 1/(\nu\gamma)\} + 1/2$.
Let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[0, \delta[$ such that $\sum_{n\in\mathbb{N}} \lambda_n(\delta - \lambda_n) = +\infty$.
We assume that $\mathrm{Argmin}(f + g) \neq \varnothing$. Let $x_0 \in \mathcal{H}$ and

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

Then, $(x_n)_{n\in\mathbb{N}}$ converges weakly to a minimizer of $f + g$.

Optimization algorithm: *Forward-Backward*

Proof: Let $T = \mathrm{prox}_{\gamma f} \circ (\mathrm{Id} - \gamma \nabla g)$. For every $x \in \mathcal{H}$,

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

Consequently, $\mathrm{Fix}\, T = \mathrm{zer}\,(\nabla g + \partial f) \neq \varnothing$. Moreover, for every $n \in \mathbb{N}$,

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

$\mathrm{prox}_{\gamma f}$ is 1/2-average and $\mathrm{Id} - \gamma \nabla g$ is $\gamma\nu/2$-averaged.
It follows that T is α-averaged with

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

Optimization algorithm: projected gradient

Let $\mathcal{H}$ be a Hilbert space.
Let C a nonempty closed convex subset of $\mathcal{H}$.
Let $g \in \Gamma_0(\mathcal{H})$ be differentiable with a ν-Lipschitzian gradient where $\nu \in]0,+\infty[$.
Let $\gamma \in]0, 2/\nu[$ and $\delta = \min\{1, 1/(\nu\gamma)\} + 1/2$.
Let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[0, \delta[$ such that $\sum_{n\in\mathbb{N}} \lambda_n(\delta - \lambda_n) = +\infty$.
We assume that $\mathrm{Argmin}_{x\in C} g(x) \neq \varnothing$. Let $x_0 \in \mathcal{H}$ and

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

Then, $(x_n)_{n\in\mathbb{N}}$ converges weakly to a minimizer of g over C.

Optimization algorithm: gradient descent

Let $\mathcal{H}$ be a Hilbert space.
Let $g \in \Gamma_0(\mathcal{H})$ be a differentiable function with a ν-lipschitzian gradient where $\nu \in\ ]0, +\infty[$.
Le $\gamma \in]0, 2/\nu[$.
We assume that $\mathrm{Argmin}\, g \neq \varnothing$. Let $x_0 \in \mathcal{H}$ and

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

Then, $(x_n)_{n\in\mathbb{N}}$ converges weakly to a minimizer of f.

Optimization algorithm: Douglas-Rachford

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

$$(\forall n \in \mathbb{N}) \qquad \begin{cases} y_n = \operatorname{prox}_{\gamma g} x_n \\ z_n = \operatorname{prox}_{\gamma f}(2y_n - x_n) \\ x_{n+1} = x_n + \lambda_n (z_n - y_n). \end{cases}$$

Optimization algorithm: Douglas-Rachford

Let $\mathcal{H}$ be a Hilbert space.
Let $f \in \Gamma_0(\mathcal{H})$ and $g \in \Gamma_0(\mathcal{H})$.
Let $\gamma \in]0,+\infty[$ and let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[0,2]$ such that $\sum_{n\in\mathbb{N}} \lambda_n(2-\lambda_n) = +\infty$.
We assume that $\mathrm{zer}\,(\partial f + \partial g) \neq \varnothing$. Let $x_0 \in \mathcal{H}$ and

$$(\forall n \in \mathbb{N}) \qquad \begin{cases} y_n = \mathrm{prox}_{\gamma g} x_n \\ z_n = \mathrm{prox}_{\gamma f}(2y_n - x_n) \\ x_{n+1} = x_n + \lambda_n(z_n - y_n). \end{cases}$$

The following properties are satisfied:

- $x_n \rightharpoonup \widehat{x}$
- $z_n - y_n \to 0$, $y_n \rightharpoonup \widehat{y}$, $z_n \rightharpoonup \widehat{y}$ where $\widehat{y} = \mathrm{prox}_{\gamma g}\widehat{x} \in \mathrm{Argmin}(f+g)$.

Optimization algorithm: Douglas-Rachford

Proof: follows from Krasnosel'skii-Mann algorithm (skipped)

Optimization algorithm: Douglas-Rachford

Let $\mathcal{H}$ and $\mathcal{G}$ be two Hilbert spaces.
Let $g \in \Gamma_0(\mathcal{H})$ and $L \in \mathcal{B}(\mathcal{G},\mathcal{H})$ such that $\operatorname{ran} L$ is closed and L^*L is a isomorphism.
Let $\gamma \in \,]0,+\infty[$ and let $(\lambda_n)_{n\in\mathbb{N}}$ a sequence in $[0,2]$ such that $\sum_{n\in\mathbb{N}} \lambda_n(2-\lambda_n) = +\infty$.
We assume that $\operatorname{zer}(L^* \circ \partial g \circ L) \neq \varnothing$. Let $x_0 \in \mathcal{H}$, $v_0 = (L^*L)^{-1}L^*x_0$ et

$$(\forall n \in \mathbb{N}) \qquad \begin{cases} y_n = \operatorname{prox}_{\gamma g} x_n \\ c_n = (L^*L)^{-1}L^*y_n \\ x_{n+1} = x_n + \lambda_n\big(L(2c_n - v_n) - y_n\big) \\ v_{n+1} = v_n + \lambda_n(c_n - v_n). \end{cases}$$

We have then:
$v_n \rightharpoonup \widehat{v}$ where $\widehat{v} \in \operatorname{Argmin}(g \circ L)$.

Optimization algorithm: Douglas-Rachford

Sketch of proof:

$$\underset{v\in\mathcal{G}}{\text{minimize}} \; g(Lv) \quad \Leftrightarrow \quad \underset{x\in\mathcal{H}}{\text{minimize}} \; \iota_E(x) + g(x)$$

where $E = \operatorname{ran} L$.
We apply Douglas-Rachford algorithm with
$f = \iota_E \Rightarrow \operatorname{prox}_{\gamma f} = P_E$ by setting

$$(\forall n \in \mathbb{N}) \qquad P_E y_n = Lc_n \text{ and } P_E x_n = Lv_n$$

where $c_n = \underset{c\in\mathcal{H}}{\operatorname{argmin}} \; \|y_n - Lc\|^2 = (L^*L)^{-1}L^*y_n$.

Optimization algorithm: Douglas-Rachford

Particular case of Douglas-Rachford algorithm:

$\mathcal{H} = \mathcal{H}_1 \times \cdots \mathcal{H}_m$ where $\mathcal{H}_1, \ldots, \mathcal{H}_m$ Hilbert spaces

$\big(\forall x = (x_1, \ldots, x_m) \in \mathcal{H}\big)\ g(x) = \sum_{i=1}^m g_i(x_i)$

where $(\forall i \in \{1, \ldots, m\}\ g_i \in \Gamma_0(\mathcal{H}_i)$

$L\colon v \mapsto (L_1 v, \ldots, L_m v)$ where $(\forall i \in \{1, \ldots, m\})\ L_i \in \mathcal{B}(\mathcal{G}, \mathcal{H}_i)$.

PPXA+ algorithm

Let $(x_{0,i})_{1 \le i \le m} \in \mathcal{H}$, $v_0 = \left(\sum_{i=1}^m L_i^* L_i\right)^{-1} \sum_{i=1}^m L_i^* x_{0,i}$ and

$$(\forall n \in \mathbb{N}) \quad \begin{cases} y_{n,i} = \mathrm{prox}_{\gamma g_i} x_{n,i}, \quad i \in \{1, \ldots, m\} \\ c_n = \left(\sum_{i=1}^m L_i^* L_i\right)^{-1} \sum_{i=1}^m L_i^* y_{n,i} \\ x_{n+1,i} = x_{n,i} + \lambda_n\big(L_i(2c_n - v_n) - y_{n,i}\big), \quad i \in \{1, \ldots, m\} \\ v_{n+1} = v_n + \lambda_n(c_n - v_n). \end{cases}$$

We have then $v_n \rightharpoonup \widehat{v} \in \mathrm{Argmin} \sum_{i=1}^m g_i \circ L_i$.

Optimization algorithm: Douglas-Rachford

Particular case of Douglas-Rachford algorithm:

$\mathcal{H} = \mathcal{H}_1 \times \cdots \mathcal{H}_m$ where $\mathcal{H}_1 = \ldots = \mathcal{H}_m$ Hilbert spaces

$\big(\forall x = (x_1, \ldots, x_m) \in \mathcal{H}\big)\ g(x) = \sum_{i=1}^m g_i(x_i)$

where $(\forall i \in \{1, \ldots, m\}\ g_i \in \Gamma_0(\mathcal{H}_i)$

$L\colon v \mapsto (L_1 v, \ldots, L_m v)$ where $L_1 = \ldots = L_m = \mathrm{Id}.$

PPXA algorithm

Let $(x_{0,i})_{1\leq i\leq m} \in \mathcal{H}$, $v_0 = \frac{1}{m}\sum_{i=1}^m x_{0,i}$ and

$$(\forall n \in \mathbb{N}) \qquad \begin{cases} y_{n,i} = \mathrm{prox}_{\gamma g_i} x_{n,i}, \quad i \in \{1, \ldots, m\} \\ c_n = \frac{1}{m}\sum_{i=1}^m y_{n,i} \\ x_{n+1,i} = x_{n,i} + \lambda_n\big(2c_n - v_n - y_{n,i}\big), \quad i \in \{1, \ldots, m\} \\ v_{n+1} = v_n + \lambda_n(c_n - v_n). \end{cases}$$

We have then $v_n \rightharpoonup \widehat{v} \in \mathrm{Argmin} \sum_{i=1}^m g_i$.