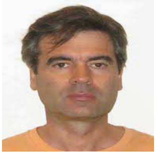
Proximal Splitting Methods in Image Processing

Jean-Christophe Pesquet

Centre de Vision Numérique – Inria
CentraleSupélec, Univ. Paris Saclay, France

Computational Methods for Inverse Problems in Imaging

In collaboration with

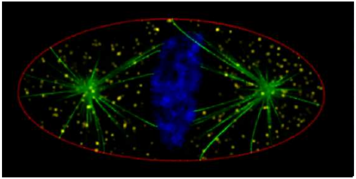


P. Combettes

N. Pustelnik

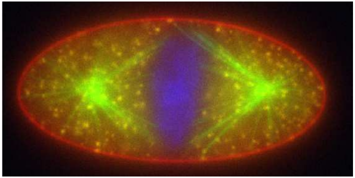
Motivation

[Microscopy, ISBI Challenge 2013, F. Soulez]

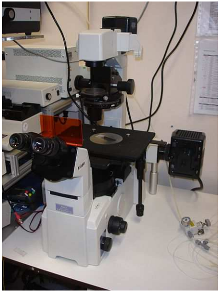


Original image

?

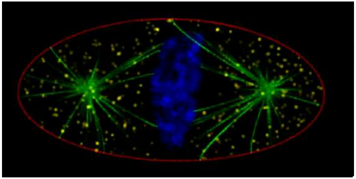


Degraded image



Motivation

[Microscopy, ISBI Challenge 2013, F. Soulez]

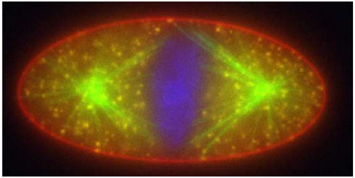


Original image

$\overline{x} \in \mathbb{R}^N$

?

$\longleftarrow$



Degraded image

$z = \mathcal{P}_\alpha(H\overline{x}) \in \mathbb{R}^M$

- $H \in \mathbb{R}^{M \times N}$: matrix associated with the degradation operator.
- $\mathcal{P}_\alpha : \mathbb{R}^M \to \mathbb{R}^M$: noise degradation with parameter α (e.g. Poisson noise).

$\rightarrow$ Find a good estimate of $\overline{x}$ from the observations z, using some a priori knowledge on H and on the noise statistics.

Motivation

Inverse problem:

Find an estimate $\widehat{x}$ close to $\overline{x}$ from the observations $z = \mathcal{P}_\alpha(H\overline{x})$.

- Inverse filtering (if $M = N$ and H is invertible)

$$\begin{aligned}\widehat{x} &= H^{-1}z \\ &= H^{-1}(H\overline{x} + b) \quad \leftarrow \quad \text{if } b \in \mathbb{R}^M \text{ is an additive noise} \\ &= \overline{x} + H^{-1}b\end{aligned}$$

$\rightarrow$ Closed form expression, but amplification of the noise if H is ill-conditioned (*ill-posed problem*).

Motivation

Inverse problem:

Find an estimate $\widehat{x}$ close to $\overline{x}$ from the observations $z = \mathcal{P}_\alpha(H\overline{x})$.

- Inverse filtering (if $M \geqslant N$ and the rank of H is N)

$$\begin{aligned}\widehat{x} &= (H^\top H)^{-1} H^\top z \\ &= (H^\top H)^{-1} H^\top (H\overline{x} + b) \quad \leftarrow \quad \text{if } b \in \mathbb{R}^M \text{ is an additive noise} \\ &= \overline{x} + (H^\top H)^{-1} H^\top b\end{aligned}$$

$\rightarrow$ Closed form expression, but amplification of the noise if H is ill-conditioned (*ill-posed problem*).

Motivation

Inverse problem:
Find an estimate $\widehat{x}$ close to $\overline{x}$ from the observations $z = \mathcal{P}_\alpha(H\overline{x})$.

- Variational approach

$$\widehat{x} \in \underset{x \in \mathbb{R}^N}{\text{Argmin}} \quad \underbrace{f_1(x)}_{\text{Data fidelity term}} \quad + \quad \underbrace{f_2(x)}_{\text{Regularization term}}$$

Data fidelity term e.g. $\|Hx - z\|_2^2$

Regularization term e.g. $\lambda \|x\|_p^p$ with $\begin{cases} p \geqslant 1 \\ \lambda \in]0, +\infty[\end{cases}$

$\rightarrow$ Often no closed form expression (e.g. if $p \neq 2$ and $H \neq \mathrm{Id}$) or solution expensive to compute (e.g. if $p = 2$, $H \neq \mathrm{Id}$ and $N \gg 1$)

$\rightarrow$ Iterative strategy.

Motivation

Inverse problem:
Find an estimate $\widehat{x}$ close to $\overline{x}$ from the observations $z = \mathcal{P}_\alpha(H\overline{x})$.

- Variational approach (more general context)

$$\widehat{x} \in \underset{x \in \mathbb{R}^N}{\operatorname{Argmin}} \sum_{i=1}^{m} f_i(x)$$

where f_i may denote a data fidelity term / a (hybrid) regularization term / constraint.

$\rightarrow$ Iterative strategy.

Motivation

Iterative strategy = Optimization algorithm:

Construct a sequence $(x_n)_{n\in\mathbb{N}}$ that converges to $\widehat{x} \in \underset{x\in\mathbb{R}^N}{\text{Argmin}} \sum_{i=1}^{m} f_i(x)$.

- Sequence such that $(\forall n \in \mathbb{N})\ x_{n+1} = Tx_n$ where T denotes an operator from $\mathbb{R}^N$ to $\mathbb{R}^N$.

  $\rightarrow$ How can we build T from the functionals $(f_i)_{1\leqslant i\leqslant m}$ involved in the minimization problem ?

  $\rightarrow$ Which properties are required by T in order to ensure the convergence of $(x_n)_{n\in\mathbb{N}}$ 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

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

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

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

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

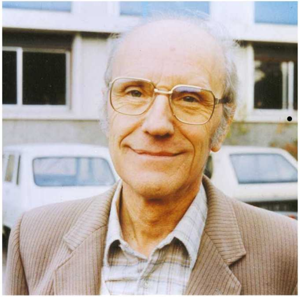
Consequently, $\|x_n - \widehat{x}\| \leqslant \rho^n \|x_0 - \widehat{x}\|$. Hence, we have proved that $(x_n)_{n\in\mathbb{N}}$ converges linearly to $\widehat{x}$.

Why do we need to go further ?

Limitations:

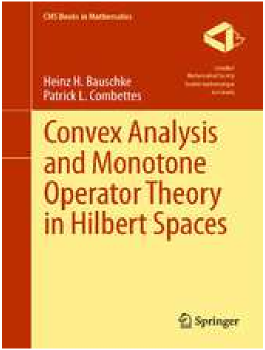
- It is difficult (even sometimes impossible) to build a *strictly* contractive operator T.
- One may prefer iterations built as $(\forall n \in \mathbb{N})\ x_{n+1} = T_n x_n$ where T_n denotes an operator from $\mathbb{R}^N$ to $\mathbb{R}^N$.
- It is often intricate to build T_n, while it may be easier to write T_n as a composition of simpler operators (*splitting techniques*).
- T_n can be multivalued, i.e. $(\forall n \in \mathbb{N})\ x_{n+1} \in T_n x_n$.

A pioneer



Jean-Jacques Moreau
(1923–2014)

Reference book



– H.H. Bauschke and P.L. Combettes –

Part 1: Background

1. Monotone operators
   - Definition
   - Properties
   - Basic operations
   - Inversion
   - Maximality
   - Usefulness for convex optimization (subdifferential)

2. Maximally monotone operators
   - Properties
   - Basic operations
   - Inversion
   - Usefulness of inversion for convex optimization (conjugate)

Hilbert spaces

A (real) Hilbert space H is a complete real vector space endowed with an inner product $\langle \cdot \mid \cdot \rangle$ whose associated norm is

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

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

Hilbert spaces

A (real) Hilbert space H is a complete real vector space endowed with an inner product $\langle \cdot \mid \cdot \rangle$ whose associated norm is

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

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

2^{H} is the power set of H, i.e. the family of all subsets of H.

Hilbert spaces

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

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

Hilbert spaces

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

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

Hilbert spaces

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

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

- In finite dimension, every linear operator is bounded.

Hilbert spaces

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

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

- In finite dimension, every linear operator is bounded.

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

Hilbert spaces

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

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

Hilbert spaces

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

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

Hilbert spaces

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

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

Hilbert spaces

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

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

Example:

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

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

Proof:

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

Hilbert spaces

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

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

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

Hilbert spaces

Let H be a Hilbert space and $L \in \mathcal{B}(\mathsf{H}, \mathsf{H})$.

- L is self-adjoint if $L^* = L$.
- L is positive if $(\forall x \in \mathsf{H})\ \langle x \mid Lx \rangle \geqslant 0$.
- L is strictly positive if L is positive and if $(\forall x \in \mathsf{H})\ \langle x \mid Lx \rangle = 0 \Leftrightarrow x = 0$.
- L is ρ-strongly positive with $\rho \in]0, +\infty[$ if $(\forall x \in \mathsf{H})\ \langle x \mid Lx \rangle \geqslant \rho \|x\|^2$.

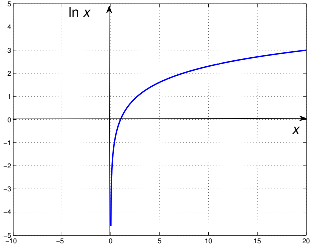
Mappings versus multivalued operators

Let H be a real Hilbert space.
A is an H-valued mapping defined on $D \subset$ H if

$$A\colon D \to \mathsf{H}$$
$$x \mapsto A(x)$$

- Example:

$$A\colon \;]0,+\infty[\; \to \mathbb{R}$$
$$x \mapsto \ln x$$



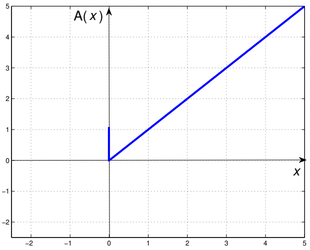
Mappings versus multivalued operators

Let H be a real Hilbert space.
A is a (multivalued) operator if

$$A \colon \mathsf{H} \to 2^{\mathsf{H}}$$
$$x \mapsto \big\{ A_i(x) \;\big|\; i \in I_x \subset \mathbb{R} \big\}$$

- Example:

$$A \colon \mathbb{R} \to 2^{\mathbb{R}}$$
$$x \mapsto \begin{cases} \{x\} & \text{if } x > 0 \\ [0,1] & \text{if } x = 0 \\ \varnothing & \text{if } x < 0 \end{cases}$$

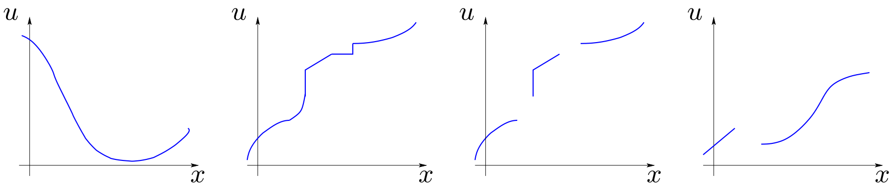


Graph

Let H be a real Hilbert space.
Let $A : \mathsf{H} \to 2^{\mathsf{H}}$.
The graph of A is

$$\operatorname{gra} A = \big\{ (x, u) \in \mathsf{H}^2 \mid u \in Ax \big\}.$$

- Graph examples:



Graph

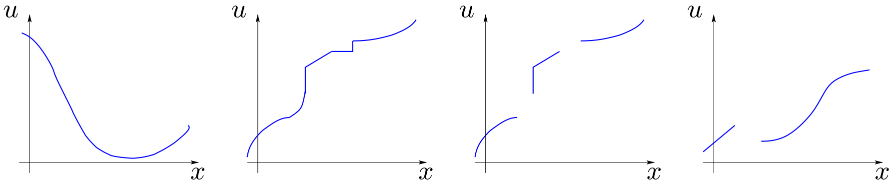
Let H be a real Hilbert space.

Let $A : \mathsf{H} \to 2^{\mathsf{H}}$.

The graph of A is

$$\operatorname{gra} A = \big\{(x,u) \in \mathsf{H}^2 \mid u \in Ax\big\}.$$

- Graph examples:



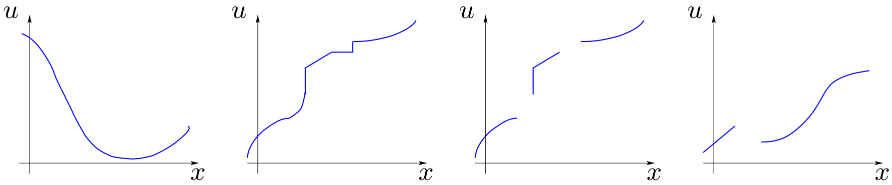
Single-valued Multivalued Multivalued Single-valued

Monotone operator: definition

Let H be a real Hilbert space.
Let $A : \mathsf{H} \to 2^{\mathsf{H}}$.
A is monotone if

$$\big(\forall(x_1, u_1) \in \operatorname{gra} A\big)\big(\forall(x_2, u_2) \in \operatorname{gra} A\big) \qquad \langle u_1 - u_2 \mid x_1 - x_2\rangle \geqslant 0.$$

- Monotone operators ?

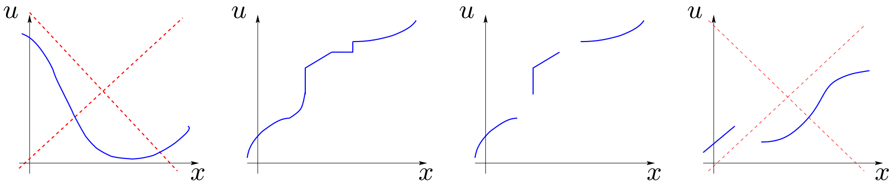


Monotone operator: definition

Let H be a real Hilbert space.
Let $A : \mathsf{H} \to 2^{\mathsf{H}}$.
A is monotone if

$$\big(\forall(x_1,u_1) \in \operatorname{gra} A\big)\big(\forall(x_2,u_2) \in \operatorname{gra} A\big) \qquad \langle u_1 - u_2 \mid x_1 - x_2\rangle \geqslant 0.$$

- Monotone operators ?

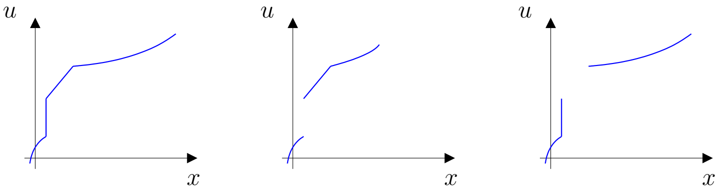


Domain

Let H be a real Hilbert space.
Let $A : \mathsf{H} \to 2^{\mathsf{H}}$.
The domain of A is

$$\operatorname{dom} A = \left\{ x \in \mathsf{H} \;\middle|\; Ax \neq \varnothing \right\}.$$

- Which domain ?

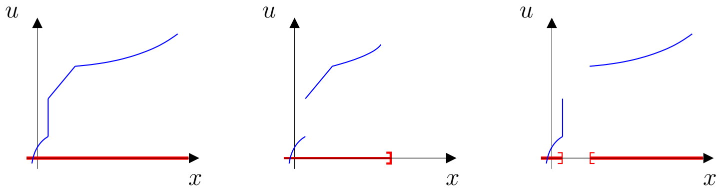


Domain

Let H be a real Hilbert space.
Let $A : \mathsf{H} \to 2^{\mathsf{H}}$.
The domain of A is

$$\operatorname{dom} A = \{x \in \mathsf{H} \mid Ax \neq \varnothing\}.$$

- Which domain ?

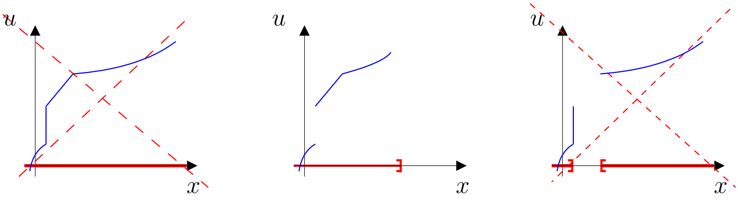


Domain

Let H be a real Hilbert space.
Let $A : \mathsf{H} \to 2^{\mathsf{H}}$.
The domain of A is

$$\operatorname{dom} A = \left\{ x \in \mathsf{H} \;\middle|\; Ax \neq \varnothing \right\}.$$

- Let $C \subset \mathsf{H}$. If $\operatorname{dom} A = C$ and for every $x \in C$, Ax is a singleton, we view A as a mapping from C to H.



Monotone operator: example

Let H be a real Hilbert space.

Let $A \in \mathcal{B}(\mathrm{H}, \mathrm{H})$.

- A is monotone $\Leftrightarrow$ A is positive
- A monotone $\Leftrightarrow$ $A + A^*$ monotone $\Leftrightarrow$ A^* monotone.

Monotone operator: example

Let H be a real Hilbert space.
Let $A \in \mathcal{B}(\mathsf{H}, \mathsf{H})$.

- A is monotone $\Leftrightarrow$ A is positive
- A monotone $\Leftrightarrow A + A^*$ monotone $\Leftrightarrow A^*$ monotone.

Proof:

$$
\begin{aligned}
A \text{ monotone} \quad & \Leftrightarrow \quad (\forall (x_1, x_2) \in \mathsf{H}^2) \qquad \langle x_1 - x_2 \mid Ax_1 - Ax_2 \rangle \geqslant 0 \\
& \Leftrightarrow \quad (\forall x \in \mathsf{H}) \qquad 2 \langle x \mid Ax \rangle \geqslant 0 \\
& \Leftrightarrow \quad (\forall x \in \mathsf{H}) \qquad \langle x \mid Ax \rangle + \langle A^* x \mid x \rangle \geqslant 0 \\
& \Leftrightarrow \quad (\forall x \in \mathsf{H}) \qquad \langle x \mid (A + A^*) x \rangle \geqslant 0 \\
& \Leftrightarrow \quad A + A^* \text{ monotone}
\end{aligned}
$$

Monotone operator: example

Let H be a real Hilbert space.
Let $A \in \mathcal{B}(\mathsf{H}, \mathsf{H})$.

- A is monotone $\Leftrightarrow$ A is positive
- A monotone $\Leftrightarrow$ $A + A^*$ monotone $\Leftrightarrow$ A^* monotone.

- For $A \in \mathcal{B}(\mathsf{H}, \mathsf{H})$ to be monotone, A is not required to be self-adjoint.

  Example : $A \in \mathcal{B}(\mathsf{H}, \mathsf{H})$ skewed (i.e. $A^* = -A$) is monotone.

Monotone operator: properties

Let H and G be two Hilbert spaces.
Let $A\colon \mathsf{H} \to 2^{\mathsf{H}}$ and $B\colon \mathsf{G} \to 2^{\mathsf{G}}$ be two monotone operators.
The following operators are monotone:

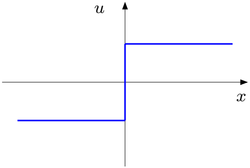
- $x \mapsto y + \gamma\rho A(\rho x + z) = \big\{ y + \gamma\rho u \;\big|\; u \in A(\rho x + z) \big\}$
  where $(y, z) \in \mathsf{H}^2$, $\gamma \in [0, +\infty[$ and $\rho \in \mathbb{R}$.
- $A \times B : \mathsf{H} \times \mathsf{G} \to 2^{\mathsf{H} \times \mathsf{G}}$
  $(x, y) \mapsto Ax \times Ay = \big\{ (u, v) \;\big|\; u \in Ax, v \in Bx \big\}$.
- $A + B : x \mapsto \big\{ u + v \;\big|\; u \in Ax, v \in Bx \big\}$ if $\mathsf{G} = \mathsf{H}$.
- $L^* B L : x \mapsto \big\{ L^* v \;\big|\; v \in B(Lx) \big\}$ if $L \in \mathcal{B}(\mathsf{H}, \mathsf{G})$.

Monotone operator: inversion

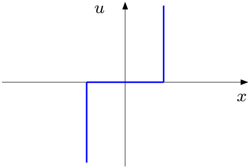
Let H be a Hilbert space.
Let $A\colon \mathsf{H} \to 2^{\mathsf{H}}$.
A^{-1} is the operator from H to 2^{H} the graph of which is

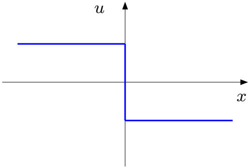
$$\operatorname{gra}(A^{-1}) = \big\{(u,x) \;\big|\; (x,u) \in \operatorname{gra} A\big\}.$$

Graph of A



Graph of A^{-1} ?



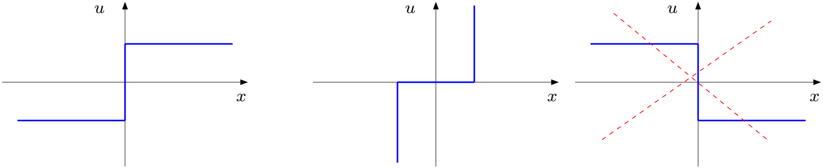


Monotone operator: inversion

Let H be a Hilbert space.
Let $A\colon \mathsf{H} \to 2^{\mathsf{H}}$.
A^{-1} is the operator from H to 2^{H} the graph of which is

$$\operatorname{gra}(A^{-1}) = \big\{(u,x) \;\big|\; (x,u) \in \operatorname{gra} A\big\}.$$

Graph of A | Graph of A^{-1} ?



Monotone operator: inversion

Let H be a Hilbert space.
Let $A\colon \mathsf{H} \to 2^{\mathsf{H}}$.
A^{-1} is the operator from H to 2^{H} the graph of which is

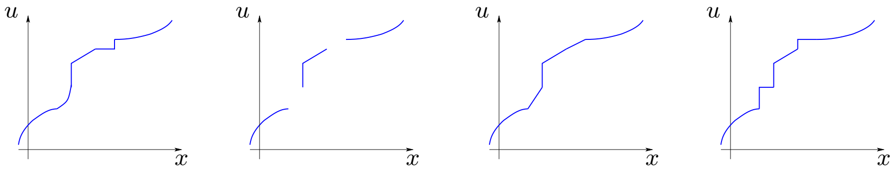
$$\operatorname{gra}(A^{-1}) = \big\{(u,x) \;\big|\; (x,u) \in \operatorname{gra} A\big\}.$$

Let H be a Hilbert space.
Let $A\colon \mathsf{H} \to 2^{\mathsf{H}}$ be a monotone operator.
A^{-1} is monotone.

Maximally monotone operator: definition

Let H be a Hilbert space.
Let $A : \mathsf{H} \to 2^{\mathsf{H}}$.
A is maximally monotone if A is monotone and if there exists no monotone operator $B\colon \mathsf{H} \to 2^{\mathsf{H}}$ (different from A) such that $\operatorname{gra} B$ properly contains $\operatorname{gra} A$.

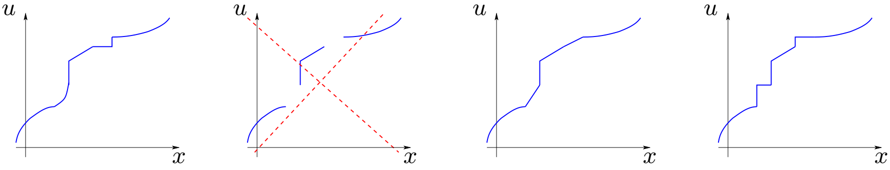
Maximally monotone operator ?



Maximally monotone operator: definition

Let H be a Hilbert space.
Let $A : \mathsf{H} \to 2^{\mathsf{H}}$.
A is **maximally monotone** if A is monotone and if there exists no monotone operator $B : \mathsf{H} \to 2^{\mathsf{H}}$ (different from A) such that $\operatorname{gra} B$ properly contains $\operatorname{gra} A$.

Maximally monotone operator ?



Maximally monotone operator: second definition

Let H be a Hilbert space.
$A : \mathsf{H} \to 2^{\mathsf{H}}$ is maximally monotone if one if the following equivalent conditions is satisfied:

(*i*) A is monotone and there exists no monotone operator $B\colon \mathsf{H} \to 2^{\mathsf{H}}$ such that $\operatorname{gra} B$ properly contains $\operatorname{gra} A$.

(*ii*) For every $(x_1, u_1) \in \mathsf{H}^2$,

$$(x_1, u_1) \in \operatorname{gra} A \quad \Leftrightarrow \quad \big(\forall (x_2, u_2) \in \operatorname{gra} A\big) \quad \langle x_1 - x_2 \mid u_1 - u_2 \rangle \geqslant 0.$$

Equivalence of the 2 definitions:

(*ii*) $\Rightarrow$ (*i*): Condition (*ii*) ensures the monotonicity of A.
In addition, if B is monotone and $\operatorname{gra} A \subset \operatorname{gra} B$ then
$\big(\forall (x_1, u_1) \in \operatorname{gra} B\big)$

$$\big(\forall (x_2, u_2) \in \operatorname{gra} A\big) \quad \langle x_1 - x_2 \mid u_1 - u_2 \rangle \geqslant 0$$

Then, Condition (*ii*) implies that $(x_1, u_1) \in \operatorname{gra} A$. Hence, $B = A$.

Maximally monotone operator: second definition

Let H be a Hilbert space.
$A : \mathsf{H} \to 2^{\mathsf{H}}$ is maximally monotone if one if the following equivalent conditions is satisfied:

(*i*) A is monotone and there exists no monotone operator $B\colon \mathsf{H} \to 2^{\mathsf{H}}$ such that $\operatorname{gra} B$ properly contains $\operatorname{gra} A$.

(*ii*) For every $(x_1, u_1) \in \mathsf{H}^2$,

$$(x_1, u_1) \in \operatorname{gra} A \quad \Leftrightarrow \quad \big(\forall (x_2, u_2) \in \operatorname{gra} A\big) \quad \langle x_1 - x_2 \mid u_1 - u_2 \rangle \geqslant 0.$$

Equivalence of the 2 definitions:

(*i*) $\Rightarrow$ (*ii*): Let $(x_1, u_1) \in \mathsf{H}^2$ be such that the above inequality is satisfied. Let B be such that $\operatorname{gra} B = \operatorname{gra} A \cup \{(x_1, u_1)\}$. If A is monotone, B is monotone such that $\operatorname{gra} A \subset \operatorname{gra} B$. According to Condition (*i*), we have $B = A \Rightarrow (x_1, u_1) \in \operatorname{gra} A$. The reciprocal is obvious.

Continuous functions

Let H be a Hilbert space.
Let $A\colon \mathsf{H} \to \mathsf{H}$ be monotone and continuous. Then A is maximally monotone.

Continuous functions

Let H be a Hilbert space.
Let $A\colon \mathsf{H} \to \mathsf{H}$ be monotone and continuous. Then A is maximally monotone.

Proof :
Let $(x_1, u_1) \in \mathsf{H}^2$.
Assume that, for every $x_2 \in \mathsf{H}$, $\langle x_1 - x_2 \mid u_1 - Ax_2 \rangle \geqslant 0$.
Let $x_2^\alpha = x_1 + \alpha(u_1 - Ax_1)$ where $\alpha > 0$.
We have $\langle u_1 - Ax_1 \mid u_1 - Ax_2^\alpha \rangle = -\alpha^{-1} \langle x_1 - x_2^\alpha \mid u_1 - Ax_2^\alpha \rangle \leqslant 0$.
As $\alpha \to 0$, $x_2^\alpha \to x_1$ and $\|u_1 - Ax_1\|^2 \leqslant 0$. Then, $u_1 = Ax_1$.

Continuous functions

Let H be a Hilbert space.
Let $A\colon \mathsf{H} \to \mathsf{H}$ be monotone and continuous. Then A is maximally monotone.

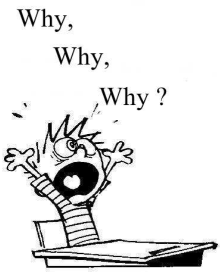
Example :
If $L \in \mathcal{B}(\mathsf{H}, \mathsf{H})$ is positive, then L is maximally monotone.

Maximally monotone operator

Maximally

monotone

operator



Maximally monotone operator

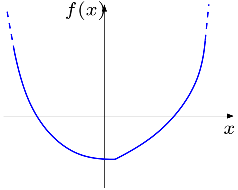
Usefulness in convex optimization

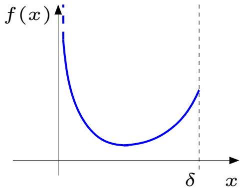
Convex analysis: definitions

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

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

Domains of the functions ?



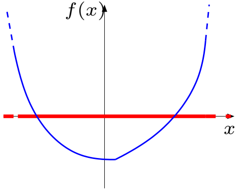


Convex analysis: definitions

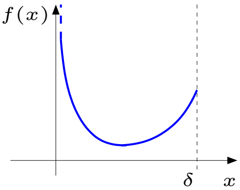
Let $f : \mathsf{H} \to \left]-\infty, +\infty\right]$ where H is a Hilbert space.

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

Domains of the functions ?



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

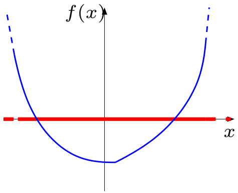


Convex analysis: definitions

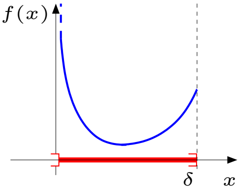
Let $f : \mathsf{H} \to]-\infty, +\infty]$ where H is a Hilbert space.

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

Domains of the functions ?



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



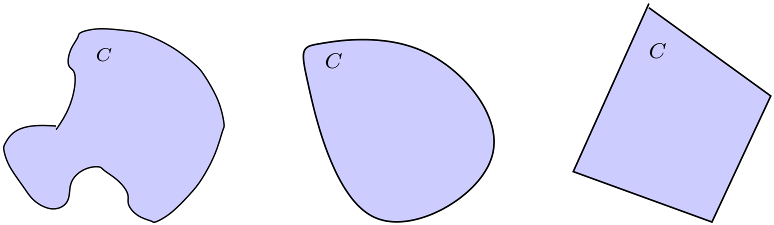
$\operatorname{dom} f =]0, \delta]$
(proper)

Convex analysis: definitions

$C \subset$ 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 ?

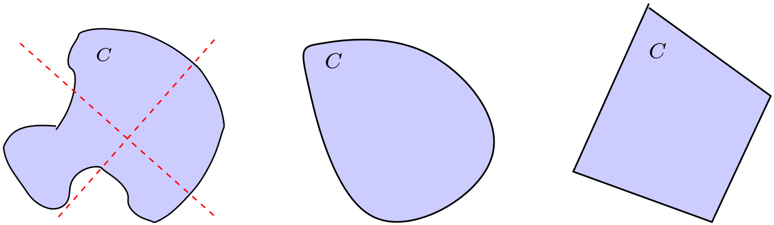


Convex analysis: definitions

$C \subset$ 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 analysis: definitions

$f : \mathsf{H} \to]-\infty, +\infty]$ is a convex fonction if

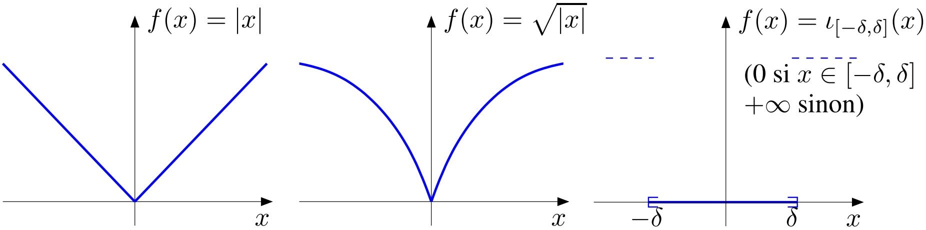
$$\big(\forall (x,y) \in \mathsf{H}^2\big)(\forall \alpha \in]0,1[) \quad f(\alpha x + (1-\alpha)y) \leqslant \alpha f(x) + (1-\alpha) f(y)$$

Convex analysis: definitions

$f : \mathsf{H} \to \left]-\infty, +\infty\right]$ is a **convex fonction** if

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

Convex functions ?

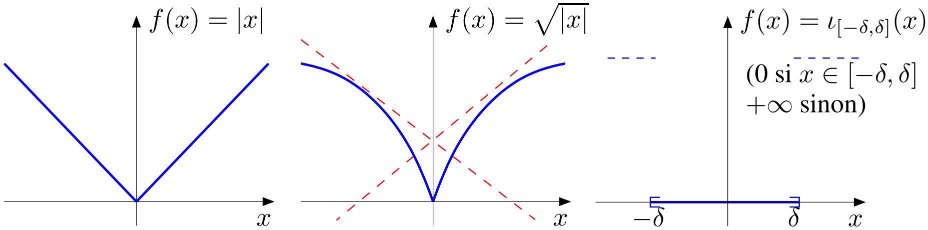


Convex analysis: definitions

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

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

Convex functions ?



Convex analysis: definitions

$f : \mathsf{H} \to]-\infty, +\infty]$ is convex $\Leftrightarrow$ the epigraph of f, i.e.

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

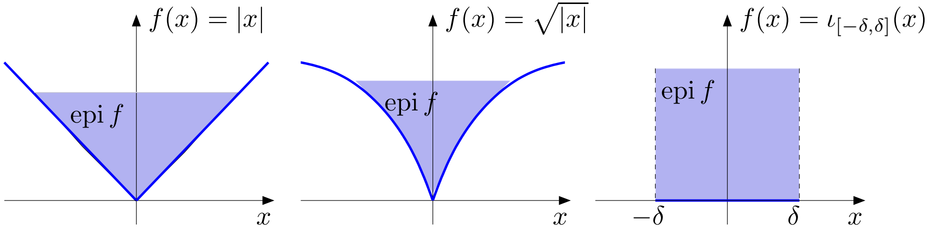
is convex.

Convex analysis: definitions

$f : \mathsf{H} \to \,]-\infty, +\infty]$ is convex $\Leftrightarrow$ the epigraph of f, i.e.

$$\operatorname{epi} f = \big\{(x,\zeta) \in \operatorname{dom} f \times \mathbb{R} \;\big|\; f(x) \leqslant \zeta\big\}$$

is convex.

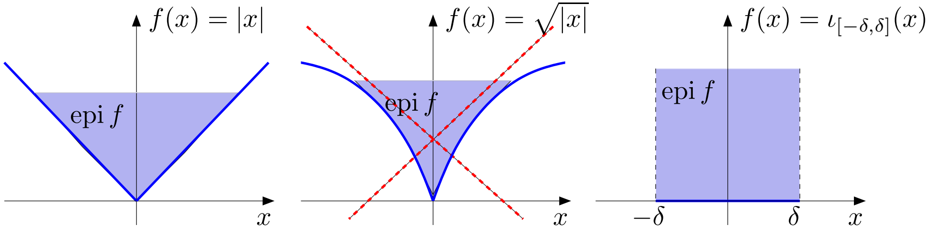


Convex analysis: definitions

$f : \mathsf{H} \to]-\infty, +\infty]$ is convex $\Leftrightarrow$ the epigraph of f, i.e.

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

is convex.

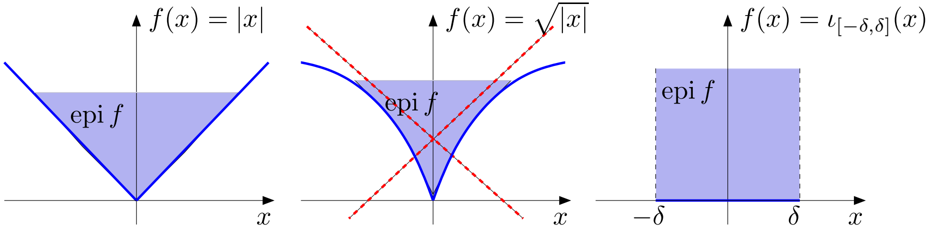


Convex analysis: definitions

$f : \mathsf{H} \to \left]-\infty, +\infty\right]$ is convex $\Leftrightarrow$ the epigraph of f, i.e.

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

is convex.

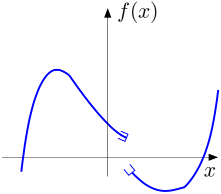


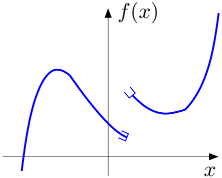
- $f : \mathsf{H} \to \left[-\infty, +\infty\right[$ is concave if $-f$ is convex.

Convex analysis: definitions

Let $f : \mathcal{H} \to]-\infty, +\infty]$.
f is a lower semi-continuous (l.s.c.) function on $\mathcal{H}$ if $\operatorname{epi} f$ is closed

- l.s.c. functions ?

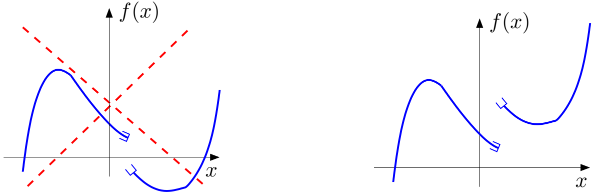




Convex analysis: definitions

Let $f : \mathcal{H} \to]-\infty, +\infty]$.
f is a lower semi-continuous (l.s.c.) function on $\mathcal{H}$ if $\operatorname{epi} f$ is closed

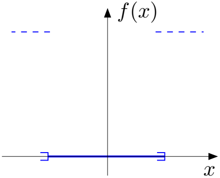
- l.s.c. functions ?

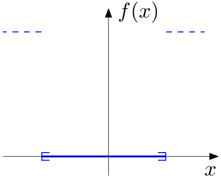


Convex analysis: definitions

Let $f : \mathcal{H} \to]-\infty, +\infty]$.
f is a lower semi-continuous (l.s.c.) function on $\mathcal{H}$ if $\operatorname{epi} f$ is closed

- l.s.c. functions ?

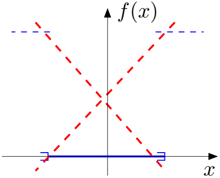


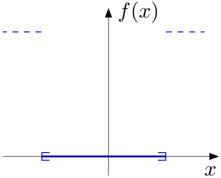


Convex analysis: definitions

Let $f : \mathcal{H} \to]-\infty, +\infty]$.
f is a lower semi-continuous (l.s.c.) function on H if $\operatorname{epi} f$ is closed

- l.s.c. functions ?





Convex analysis: definitions/properties

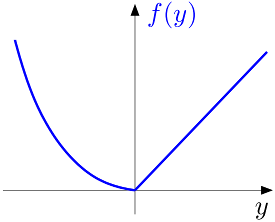
- $\Gamma_0(\mathsf{H})$: class of convex, l.s.c., and proper functions from H to $]-\infty, +\infty]$.
- Every continuous function on H is l.s.c.
- Every finite sum of l.s.c. (convex) functions is l.s.c. (convex).
- Let $(f_i)_{i\in I}$ be a family of l.s.c. (convex) functions. $\sup_{i\in I} f_i$ is l.s.c. (convex).

Subdifferential of a convex function: definition

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

Subdifferential of a convex function: definition

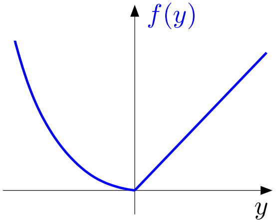
Let $f : \mathcal{H} \to]-\infty, +\infty]$ be a proper function.
The (Moreau) subdifferential of f, denoted by ∂f,

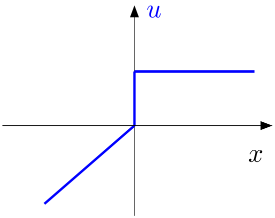


Subdifferential of a convex function: definition

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

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

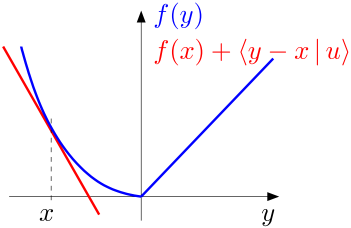


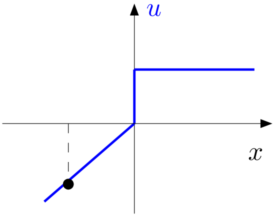


Subdifferential of a convex function: definition

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

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

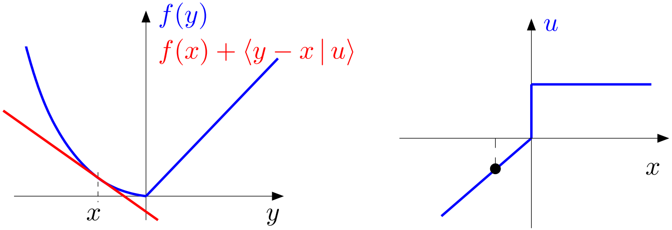




Subdifferential of a convex function: definition

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

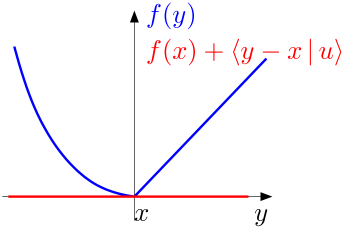
$$\partial f : \mathsf{H} \to 2^{\mathsf{H}}$$
$$x \to \{u \in \mathsf{H} \mid (\forall y \in \mathsf{H})\ \langle y - x | u \rangle + f(x) \leqslant f(y)\}$$

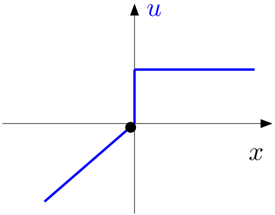


Subdifferential of a convex function: definition

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

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

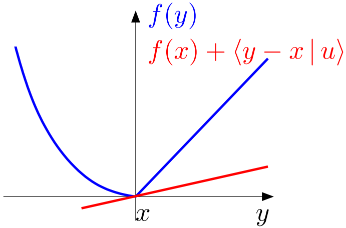


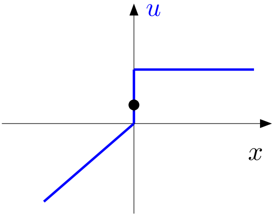


Subdifferential of a convex function: definition

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

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

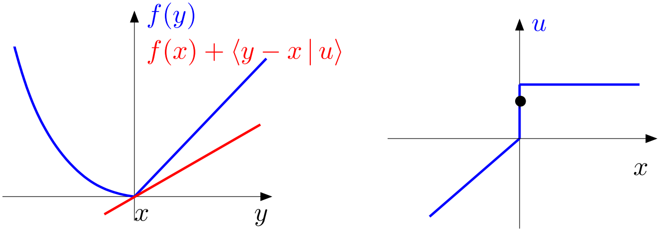




Subdifferential of a convex function: definition

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

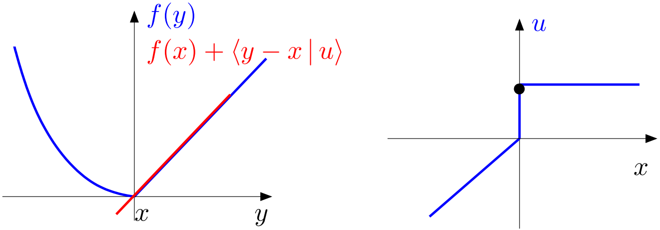
$$\begin{aligned} \partial f : \mathsf{H} &\to 2^{\mathsf{H}} \\ x &\to \{u \in \mathsf{H} \mid (\forall y \in \mathsf{H})\ \langle y - x | u \rangle + f(x) \leqslant f(y)\} \end{aligned}$$



Subdifferential of a convex function: definition

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

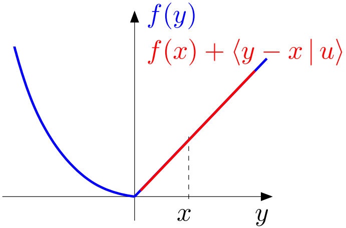
$$\partial f : \mathsf{H} \to 2^{\mathsf{H}}$$
$$x \to \{u \in \mathsf{H} \,|\, (\forall y \in \mathsf{H})\ \langle y - x | u \rangle + f(x) \leqslant f(y)\}$$

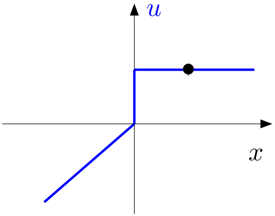


Subdifferential of a convex function: definition

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

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

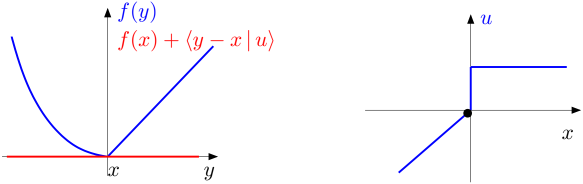




Subdifferential of a convex function: properties

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

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



- **Fermat's rule:** $$0 \in \partial f(x) \Leftrightarrow (\forall y \in \mathsf{H})\ \langle y - x \mid 0 \rangle + f(x) \leqslant f(y)$$ $$\Leftrightarrow x \in \operatorname{Argmin} f$$

Subdifferential of a convex function: properties

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

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

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

Subdifferential of a convex function: properties

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

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

- $u \in \partial f(x)$ is a subgradient of f at x.
- If $x \notin \operatorname{dom} f$, then $\partial f(x) = \varnothing$.

Subdifferential of a convex function: properties

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

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

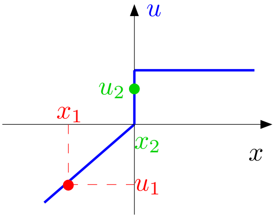
- $u \in \partial f(x)$ is a subgradient of f at x.
- If $x \notin \operatorname{dom} f$, then $\partial f(x) = \varnothing$.
- For every $x \in \operatorname{dom} f$, $\partial f(x)$ is a closed and convex set.

Subdifferential of a convex function: properties

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

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

- ∂f is a monotone operator :

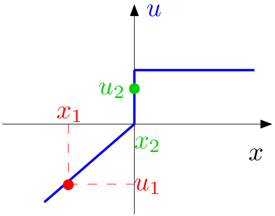


Subdifferential of a convex function: properties

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

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

- ∂f is a monotone operator :
  Let $u_1 \in \partial f(x_1)$ and $u_2 \in \partial f(x_2)$.



Subdifferential of a convex function: properties

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

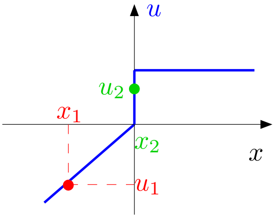
$$\begin{aligned} \partial f : \mathsf{H} &\to 2^{\mathsf{H}} \\ x &\to \{u \in \mathsf{H} \,|\, (\forall y \in \mathsf{H})\ \langle y - x | u \rangle + f(x) \leqslant f(y)\} \end{aligned}$$

- ∂f is a monotone operator :

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

  By using the subdifferential definition:

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



Subdifferential of a convex function: properties

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

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

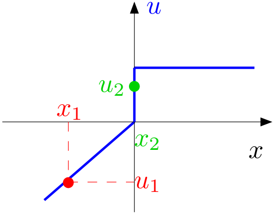
- ∂f is a monotone operator :

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

  By using the subdifferential definition:

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

  and thus $\langle x_1 - x_2 \,|\, u_1 - u_2\rangle \geqslant 0$.



Subdifferential of a convex function: properties

- The subdifferential of a convex and proper function is:
  - Monotone
  - If f is Gâteaux differentiable at x, then $\partial f(x) = \{\nabla f(x)\}$

Subdifferential of a convex function: properties

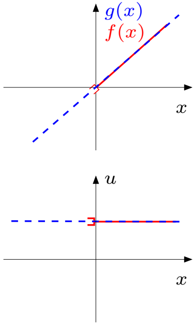
- The subdifferential of a convex and proper function is:
  - Monotone
  - If f is Gâteaux differentiable at x, then $\partial f(x) = \{\nabla f(x)\}$
  - Non necessarily maximally monotone

Counterexample: For every $x \in \mathsf{H}$,

$$f(x) = \begin{cases} x & \text{if } x > 0 \\ +\infty & \text{otherwise} \end{cases}, \qquad g(x) = x$$

$$\Rightarrow \quad \partial f(x) = \begin{cases} \{1\} & \text{if } x > 0 \\ \varnothing & \text{otherwise} \end{cases}, \qquad \partial g(x) = \{1\}.$$

Consequently, $\operatorname{gra} \partial f =]0, +\infty[\times \{1\} \subset \mathbb{R} \times \{1\}$
$\subset \operatorname{gra} \partial g$



.

Subdifferential of a convex function: properties

- The subdifferential of a convex, proper and l.s.c. function is
  - Maximally monotone

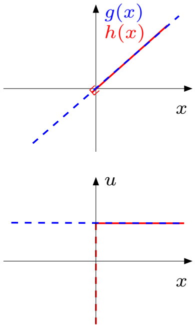
Subdifferential of a convex function: properties

- The subdifferential of a convex, proper and l.s.c. function is
  - Maximally monotone

Example: For every $x \in$ H,

$$h(x) = \begin{cases} x & \text{if } x \geqslant 0 \\ +\infty & \text{otherwise} \end{cases}, \quad g(x) = x$$

$$\Rightarrow \quad \partial h(x) = \begin{cases} \{1\} & \text{if } x > 0 \\ ]-\infty, 1] & \text{if } x = 0 \\ \varnothing & \text{otherwise} \end{cases}, \quad \partial g(x) = \{1\}.$$



Consequently, $\operatorname{gra} \partial h \not\subset \operatorname{gra} \partial g$.

Subdifferential of a convex function: properties

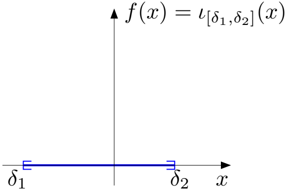
- The subdifferential of a convex, proper and l.s.c. function is
  - Maximally monotone
  - If $H = \mathbb{R}$, equivalence between both properties.

Subdifferential of a convex function: example

Let $\varnothing \neq C \subset \mathsf{H}$.
The indicator function of C is

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

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

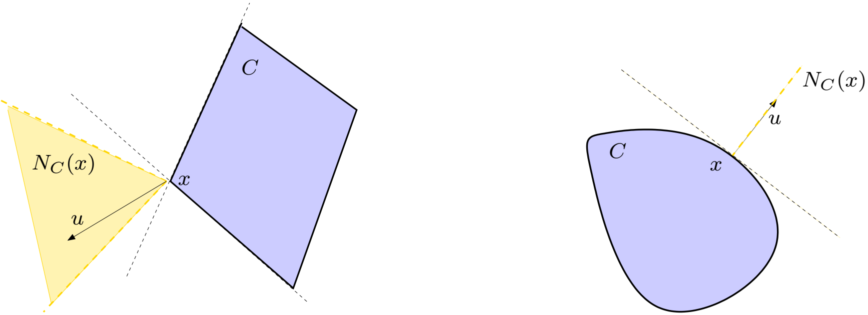


- $\iota_C \in \Gamma_0(\mathsf{H}) \Leftrightarrow C$ is a nonempty closed convex set.
  Proof: $\mathrm{epi}_{\iota_C} = C \times [0, +\infty[$.

Subdifferential of a convex function: example

For every $x \in \mathsf{H}$, $\partial\iota_C(x)$ is the normal cone to C at x defined by

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

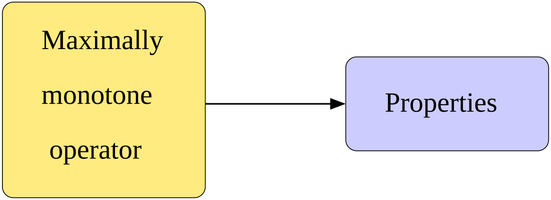


Subdifferential of a convex function: example

For every $x \in \mathsf{H}$, $\partial\iota_C(x)$ is the normal cone to C at x defined by

$$N_C(x) = \begin{cases} \big\{u \in \mathsf{H} \;\big|\; (\forall y \in C) \;\; \langle u \mid y - x\rangle \leqslant 0\big\} & \text{if } x \in C \\ \varnothing & \text{otherwise.} \end{cases}$$

- If $x \in \mathrm{int} C$ then $N_C(x) = \{0\}$.
- If C is a vector space then, for every $x \in C$, $N_C(x) = C^{\perp}$.


Maximally
monotone
operator
Properties

Maximally monotone operator: properties

Let H be a Hilbert space.
Let $A\colon \mathsf{H} \to 2^{\mathsf{H}}$ be a maximally monotone operator.
For every $x \in \mathsf{H}$, Ax is a closed convex set.

Proof:

$$Ax = \bigcap_{(x',u')\in \operatorname{gra} A} \big\{u \in \mathsf{H} \;\big|\; \big\langle x - x' \mid u - u' \big\rangle \geqslant 0\big\}.$$

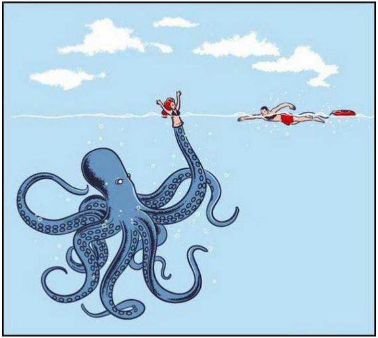
Consequently, Ax is an intersection of closed convex sets.

Maximally monotone operator: properties

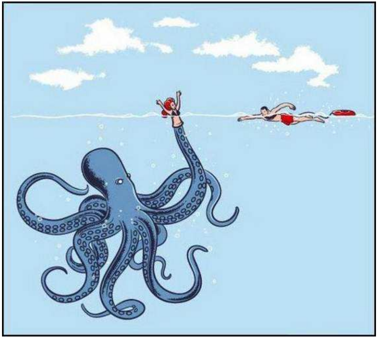
Let H and G be two Hilbert spaces.
Let $A\colon \mathsf{H} \to 2^{\mathsf{H}}$ and $B\colon \mathsf{G} \to 2^{\mathsf{G}}$ be two maximally monotone operators.
The following operators are maximally monotone:

- $y + \gamma\rho A(\rho \cdot + z)$ where $(y, z) \in \mathsf{H}^2$, $\gamma \in [0, +\infty[$ and $\rho \in \mathbb{R}$,
- $A \times B$,
- A^{-1}.

Inverse of a maximally monotone operator



Inverse of a maximally monotone operator → Usefulness ?



Conjugate: definition

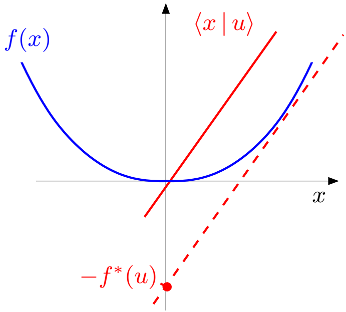
Let H be a Hilbert space and $f\colon \mathsf{H} \to \,]-\infty, +\infty]$.
The conjugate of f is the function $f^*\colon \mathsf{H} \to [-\infty, +\infty]$ such that

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

Conjugate: definition

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

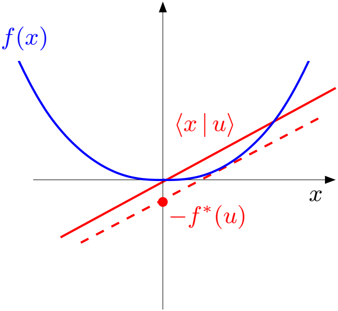
$$(\forall u \in \mathsf{H}) \qquad f^*(u) = \sup_{x\in\mathsf{H}} \big(\langle x \mid u\rangle - f(x)\big).$$



Conjugate: definition

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

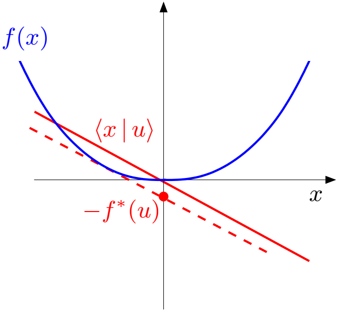
$$(\forall u \in \mathsf{H}) \qquad f^*(u) = \sup_{x \in \mathsf{H}} \big(\langle x \mid u \rangle - f(x) \big).$$



Conjugate: definition

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

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



Conjugate: definition

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

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

Example :

- $(\forall x \in \mathbb{R}^N)\ f(x) = \frac{1}{q}\|x\|_q^q$ 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 H be a Hilbert space and $f\colon \mathsf{H} \to \,]-\infty, +\infty]$.
The conjugate of f is the function $f^*\colon \mathsf{H} \to [-\infty, +\infty]$ such that

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

- f^* is l.s.c. and convex.

Conjugate: definition

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

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

Fenchel-Young inequality : If f is proper, then

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

Conjugate: properties

Moreau-Fenchel theorem

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

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

Consequence: If $f \in \Gamma_0(\mathsf{H})$, then $f^* \in \Gamma_0(\mathsf{H})$.

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

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

Consequence: $(\partial f)^{-1} = \partial f^*$.

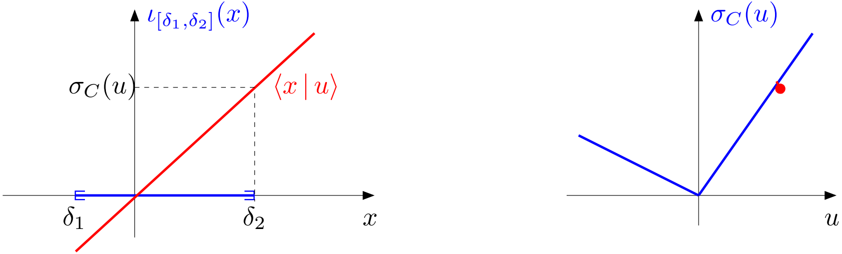
Conjugate versus Fourier transform

	conjugate		Fourier transform	
Property	$h(x)$	$h^*(u)$	$h(x)$	$\widehat{h}(\nu)$
invariant function	$\frac{1}{2}\|x\|^2$	$\frac{1}{2}\|u\|^2$	$\exp(-\pi\|x\|^2)$	$\exp(-\pi\|\nu\|^2)$
translation $c \in \mathsf{H}$	$f(x-c)$	$f^*(u) + \langle u \mid c\rangle$	$f(x-c)$	$\exp(-\jmath 2\pi\langle \nu \mid c\rangle)\widehat{f}(\nu)$
dual translation $c \in \mathsf{H}$	$f(x) + \langle x \mid c\rangle$	$f^*(u-c)$	$\exp(\jmath 2\pi\langle x \mid c\rangle) f(x-c)$	$\widehat{f}(\nu - c)$
scalar multiplication $\alpha \in]0,+\infty[$	$\alpha f(x)$	$\alpha f^*\left(\frac{u}{\alpha}\right)$	$\alpha f(x)$	$\alpha\widehat{f}(\nu)$
scaling $\alpha \in \mathbb{R}^*$	$f\left(\frac{x}{\alpha}\right)$	$f^*(\alpha u)$	$f\left(\frac{x}{\alpha}\right)$	$\|\alpha\|\widehat{f}(\alpha\nu)$
isomorphism $L \in \mathcal{B}(\mathsf{G},\mathsf{H})$	$f(Lx)$	$f^*\big((L^{-1})^\top u\big)$	$f(Lx)$	$\frac{1}{\|\det(L)\|}\widehat{f}\big((L^{-1})^\top \nu\big)$
reflection	$f(-x)$	$f^*(-u)$	$f(-x)$	$\widehat{f}(-\nu)$
separability	$\sum_{n=1}^{N} \varphi_n(x^{(n)})$ $x = (x^{(n)})_{1\leqslant n\leqslant N}$	$\sum_{n=1}^{N} \varphi_n^*(u^{(n)})$ $u = (u^{(n)})_{1\leqslant n\leqslant N}$	$\prod_{n=1}^{N} \varphi_n(x^{(n)})$ $x = (x^{(n)})_{1\leqslant n\leqslant N}$	$\prod_{n=1}^{N} \widehat{\varphi}_n(\nu^{(n)})$ $\nu = (\nu^{(n)})_{1\leqslant n\leqslant N}$
isotropy	$\psi(\|x\|)$	$\psi^*(\|u\|)$	$\psi(\|x\|)$	$\widetilde{\psi}(\|\nu\|)$
identity element of convolution	$\iota_{\{0\}}(x)$	0	$\delta(x)$	1
identity element of addition/product	0	$\iota_{\{0\}}(u)$	1	$\delta(\nu)$

Conjugate: example

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

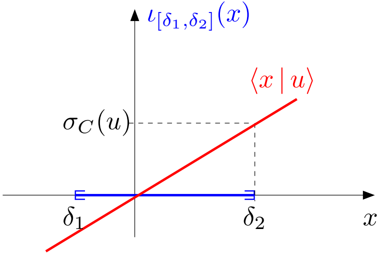
$$(\forall u \in \mathsf{H}) \qquad \sigma_C(u) = \sup_{x \in C} \langle x \mid u \rangle$$
$$= \iota_C^*(u).$$

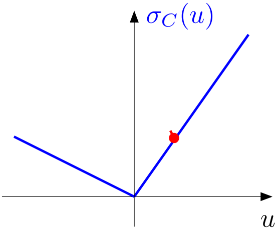


Conjugate: example

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

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

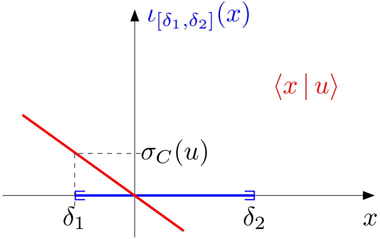


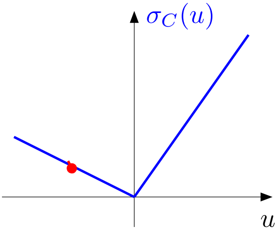


Conjugate: example

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

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





Conjugate: example

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

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

f is positively homogeneous and belongs to $\Gamma_0(\mathsf{H})$

$\Updownarrow$

$f = \sigma_C$ where C is a nonempty closed convex subset of H.

Conjugate: example

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

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

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

- Example 1: 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 $-\infty \leqslant \delta_1 < \delta_2 \leqslant +\infty$. Then, $f = \sigma_C$ where C is the closed real interval such that $\inf C = \delta_1$ and $\sup C = \delta_2$.

Conjugate: example

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

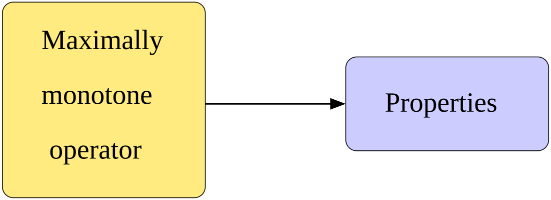
$$(\forall x \in \mathsf{H})(\forall \alpha \in \left]0,+\infty\right[) \qquad f(\alpha x) = \alpha f(x).$$

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

- Example 2: Let f be a ℓ^q norm of $\mathbb{R}^N$ with $q \in [1,+\infty]$.
  We have $f = \sigma_C$ where

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

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


Maximally
monotone
operator
Properties

Maximally monotone operator: sum

Let A and B be two maximally monotone operators.
$A + B$ is monotone but may not be maximally monotone.

Maximally monotone operator: sum

Let H be a Hilbert space.
Let A and B be two maximally monotone operators from H to 2^{H} such that one of the following assumptions is satisfied:

- $\operatorname{dom} B = \mathsf{H}$
- $\operatorname{dom} A \cap \operatorname{int}(\operatorname{dom} B) \neq \varnothing$
- $0 \in \operatorname{int}(\operatorname{dom} A - \operatorname{dom} B)$

then $A + B$ is maximally monotone.

Maximally monotone operator: sum

Let H be a Hilbert space.
Let A and B be two maximally monotone operators from H to 2^{H} such that one of the following assumptions is satisfied:

- $\operatorname{dom} B = \mathsf{H}$
- $\operatorname{dom} A \cap \operatorname{int}(\operatorname{dom} B) \neq \varnothing$
- $0 \in \operatorname{int}(\operatorname{dom} A - \operatorname{dom} B)$

then $A + B$ is maximally monotone.

Consequence: Let $\alpha \in [0, +\infty[$. If A is maximally monotone, then $A + \alpha \mathrm{Id}$ is maximally monotone.

Maximally monotone operator: linear transform

Let H and G two Hilbert spaces.
Let $B\colon \mathsf{G} \to 2^{\mathsf{G}}$ be a maximally monotone operator and let $L \in \mathcal{B}(\mathsf{H}, \mathsf{G})$ be such that one of the following assumptions is satisfied:

- L surjective
- $0 \in \operatorname{int}(\operatorname{dom} B - \operatorname{ran} L)$

then $L^* B L$ is maximally monotone.

Maximally monotone operator: linear transform

Let H and G two Hilbert spaces.
Let $B\colon \mathsf{G} \to 2^{\mathsf{G}}$ be a maximally monotone operator and let $L \in \mathcal{B}(\mathsf{H}, \mathsf{G})$ be such that one of the following assumptions is satisfied:

- L surjective
- $0 \in \operatorname{int}(\operatorname{dom} B - \operatorname{ran} L)$

then L^*BL is maximally monotone.

Consequence: Let $\mu \in \left]0, +\infty\right[$.
If B is maximally monotone and $LL^* = \mu \mathrm{Id}$, then L^*BL is maximally monotone.

Proof: $LL^* = \mu \mathrm{Id} \Rightarrow \operatorname{ran} L = \mathsf{H}$.

Part 2: Nonexpansive operators

1. Background on nonexpansive operators
   - Definition
   - Properties
   - Examples
   - Resolvent

2. Proximal operator
   - Definition
   - Properties
   - Examples

Nonexpansive operator: definition

Let H be a Hilbert space and let C be a nonempty subset of H.
Let $A\colon C \to \mathsf{H}$.
A is nonexpansive if $\big(\forall (x,y) \in C^2\big) \quad \|Ax - Ay\| \leqslant \|x - y\|$.

Nonexpansive operator: definition

Let H be a Hilbert space and let C be a nonempty subset of H.
Let $A\colon C \to$ H and $\nu \in \,]0,+\infty[$

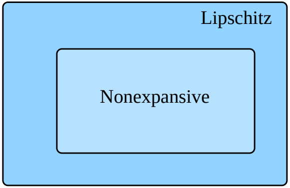
$\nu^{-1}A$ is nonexpansive if $\big(\forall (x,y) \in C^2\big) \quad \|Ax - Ay\| \leqslant \nu \|x - y\|$.

Nonexpansive operator: definition

Let H be a Hilbert space and let C be a nonempty subset of H.
Let $A\colon C \to$ H and $\nu \in \left]0,+\infty\right[$

$\nu^{-1}A$ is nonexpansive if $\big(\forall (x,y) \in C^2\big) \quad \|Ax - Ay\| \leqslant \nu \|x-y\|.$

$\nu^{-1}A$ is nonexpansive $\Leftrightarrow$ A is ν-Lipschitzian.



Nonexpansive operator: definition

Let H be a real Hilbert space.
Let $A\colon \mathsf{H} \to 2^{\mathsf{H}}$
A is firmly nonexpansive if

$$(\forall (x,u) \in \operatorname{gra} A)(\forall (y,v) \in \operatorname{gra} A) \quad \|u-v\|^2 \leqslant \langle u-v \mid x-y \rangle .$$

Nonexpansive operator: definition

Let H be a real Hilbert space and let C be a nonempty subset of H.
Let $A\colon C \to$ H.
A is firmly nonexpansive if

$$(\forall x \in C)(\forall y \in C) \quad \|Ax - Ay\|^2 \leqslant \langle Ax - Ay \mid x - y\rangle.$$

Nonexpansive operator: definition

Let H be a real Hilbert space and let C be a nonempty subset of H.
Let $A\colon C \to$ H.
A is firmly nonexpansive if

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

Nonexpansive operator: definition

Let H be a real Hilbert space and let C be a nonempty subset of H.
Let $A\colon C \to$ H.
A is **firmly nonexpansive** if

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

- A is firmly nonexpansive $\Leftrightarrow$ $\mathrm{Id} - A$ is firmly nonexpansive.
- A is firmly nonexpansive $\Leftrightarrow$ $2A - \mathrm{Id}$ is nonexpansive.

Nonexpansive operator: definition

Let H be a real Hilbert space and let C be a nonempty subset of H.
Let $A\colon C \to$ H.
A is firmly nonexpansive if

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

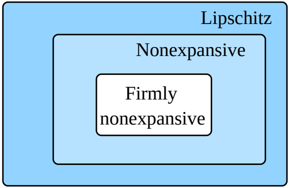
- A is firmly nonexpansive $\Leftrightarrow$ $\mathrm{Id} - A$ is firmly nonexpansive.
- A is firmly nonexpansive $\Leftrightarrow$ $\underbrace{2A - \mathrm{Id}}_{\text{Reflection of A}}$ is nonexpansive.

Nonexpansive operator: definition

Let H be a real Hilbert space and let C be a nonempty subset of H.
Let $A\colon C \to$ H.
A is **firmly nonexpansive** if

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

A is firmly nonexpansive $\Rightarrow$ A is nonexpansive.



Nonexpansive operator: definition

Let H be a real Hilbert space and let C be a nonempty subset of H.
Let $A\colon C \to$ H and $\beta \in \,]0,+\infty[$.
A is β-cocoercive if βA is firmly nonexpansive, i.e.,

$$(\forall x \in C)(\forall y \in C) \quad \beta\|Ax - Ay\|^2 \leqslant \langle x - y \mid Ax - Ay\rangle .$$

Nonexpansive operator: definition

Let H be a real Hilbert space and let C be a nonempty subset of H. Let $A\colon C \to$ H and $\beta \in]0,+\infty[$.
A is **β-cocoercive** if βA is firmly nonexpansive, i.e.,

$$(\forall x \in C)(\forall y \in C) \quad \beta\|Ax - Ay\|^2 \leqslant \langle x - y \mid Ax - Ay\rangle.$$

- Let H and G be two real Hilbert spaces, $L \in \mathcal{B}(\mathsf{G},\mathsf{H})$ nonzero, and $A\colon$ H $\to$ H. A is β-cocoercive $\Rightarrow L^*AL$ is $\|L\|^{-2}\beta$-cocoercive.
  Proof: For every $(x,y) \in \mathsf{H}^2$,
  $$\begin{aligned}\langle L^*ALx - L^*ALy \mid x - y\rangle &= \langle ALx - ALy \mid Lx - Ly\rangle \\ &\geqslant \beta\|ALx - ALy\|^2\end{aligned}$$
  Furthermore, $\|L^*ALx - L^*ALy\|^2 \leqslant \|L\|^2\|ALx - ALy\|^2$.
  Then $\langle L^*ALx - L^*ALy \mid x - y\rangle \geqslant \beta\|L^*ALx - L^*ALy\|^2/\|L\|^2$.

Nonexpansive operator: definition

Let H be a real Hilbert space and let C be a nonempty subset of H.
Let $A\colon C \to$ H and $\beta \in \left]0,+\infty\right[$.
A is β-cocoercive if βA is firmly nonexpansive, i.e.,

$$(\forall x \in C)(\forall y \in C) \quad \beta\|Ax - Ay\|^2 \leqslant \langle x - y \mid Ax - Ay\rangle.$$

- Let H and G be two real Hilbert spaces, $L \in \mathcal{B}(\mathsf{G},\mathsf{H})$ nonzero, and $A\colon$ H $\to$ H. A is β-cocoercive $\Rightarrow$ L^*AL is $\|L\|^{-2}\beta$-cocoercive.
- A is β-cocoercive $\Rightarrow$ A is β^{-1}-Lipschitzian.

Nonexpansive operator: definition

Let H be a real Hilbert space and let C be a nonempty subset of H.
Let $A\colon C \to$ H and $\beta \in \,]0,+\infty[$.
A is **β-cocoercive** if βA is firmly nonexpansive, i.e.,

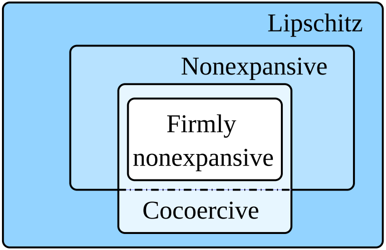
$$(\forall x \in C)(\forall y \in C) \quad \beta\|Ax - Ay\|^2 \leqslant \langle x - y \mid Ax - Ay\rangle .$$

- Let H and G be two real Hilbert spaces, $L \in \mathcal{B}(\mathsf{G},\mathsf{H})$ nonzero, and $A\colon$ H $\to$ H. A is β-cocoercive $\Rightarrow$ L^*AL is $\|L\|^{-2}\beta$-cocoercive.
- A is β-cocoercive $\Rightarrow$ A is β^{-1}-Lipschitzian.
- $A\colon$ H $\to$ H is β-cocoercive $\Rightarrow$ A is maximally monotone.

Nonexpansive operator: definition

Let H be a real Hilbert space and let C be a nonempty subset of H.
Let $A\colon C \to$ H and $\beta \in \,]0,+\infty[$.
A is β-cocoercive if βA is firmly nonexpansive, i.e.,

$$(\forall x \in C)(\forall y \in C) \quad \beta\|Ax - Ay\|^2 \leqslant \langle x - y \mid Ax - Ay\rangle .$$



Nonexpansive operator: definition

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

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

Nonexpansive operator: definition

Let H be a real Hilbert space and let C be a nonempty subset of H.
Let $A : C \to$ H and let $\alpha \in]0,1[$.
A is α-averaged if

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

Nonexpansive operator: definition

Let H be a real Hilbert space and let C be a nonempty subset of H.
Let $A : C \to$ H and let $\alpha \in]0,1[$.
A is α-averaged if

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

- A is α-averaged $\Rightarrow$ A is nonexpansive.
- A is $\frac{1}{2}$-averaged $\Leftrightarrow$ A is firmly nonexpansive.
- A is α-averaged $\Rightarrow$ A is α'-averaged for every $\alpha' \in [\alpha, 1[$.
- Let $\lambda \in]0, 1/\alpha[$. A is α-averaged $\Rightarrow$ $(1-\lambda)\mathrm{Id} + \lambda A$ is $\lambda\alpha$-averaged.

Nonexpansive operator: definition

Let H be a real Hilbert space and let C be a nonempty subset of H.
Let $A : C \to$ H and let $\alpha \in]0,1[$.
A is α-averaged if

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

- Let $(\omega_i)_{1\leqslant i\leqslant n} \in]0,1]^n$ be such that $\sum_{i=1}^n \omega_i = 1$ and let $(\alpha_i)_{1\leqslant i\leqslant n} \in]0,1[^n$. If, for every $i \in \{1,\ldots,n\}$, $A_i : C \to$ H is α_i-averaged, then $\sum_{i=1}^n \omega_i A_i$ is α-averaged with $\alpha = \max_{1\leqslant i\leqslant n} \alpha_i$.

- Let $(\alpha_i)_{1\leqslant i\leqslant n} \in]0,1[^n$. If, for every $i \in \{1,\ldots,n\}$, $A_i : C \to C$ is α_i-averaged, then $A_1 \cdots A_n$ is α-averaged with
$$\alpha = \frac{n}{n-1+\frac{1}{\max_{1\leqslant i\leqslant n} \alpha_i}}.$$

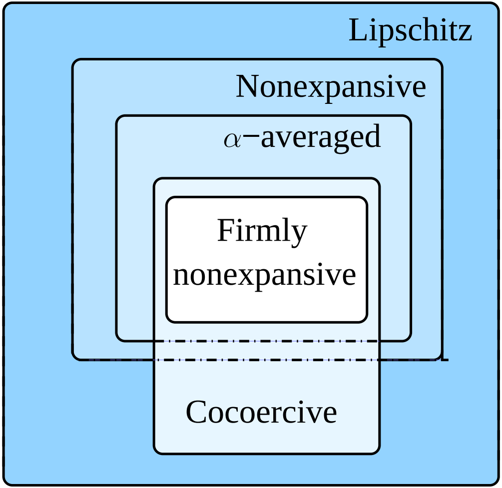
Nonexpansive operator: definition

Let H be a real Hilbert space and let C be a nonempty subset of H.
Let $A : C \to$ H and let $\alpha \in]0,1[$.
A is **α-averaged** if

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

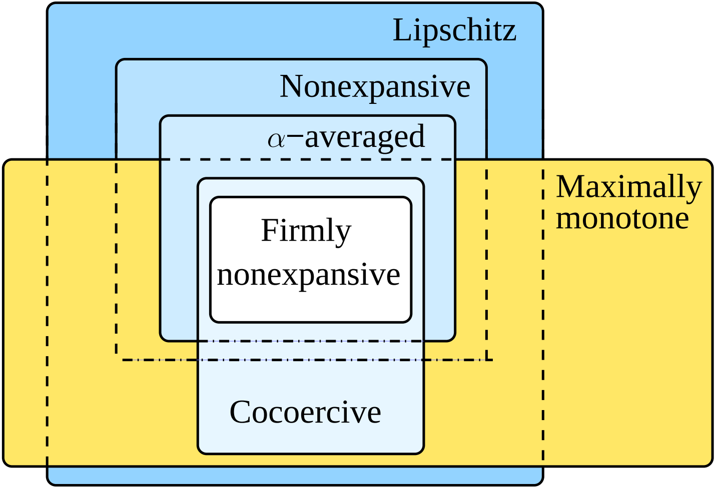
A : H $\to$ H is α-averaged with $\alpha \in]0,1/2] \Rightarrow A$ is maximally monotone.

Nonexpansive operator: recap



Nonexpansive operator: recap

(if the domain C is equal to H)



Nonexpansive operator: properties

Let H be a real Hilbert space and let C be a nonempty subset of H.
Let $A \colon C \to$ H.
Let $\beta \in \left]0,+\infty\right[$ and $\gamma \in]0,2\beta[$.
If A is β-cocoercive, then $\mathrm{Id} - \gamma A$ is $\gamma/(2\beta)$-averaged.

Proof :
A β-cocoercive $\Leftrightarrow$ βA firmly nonexpansive.
There exists a nonexpansive operator $R \colon C \to$ H such that $\beta A = (\mathrm{Id} + R)/2$.
Thus

$$\mathrm{Id} - \gamma A = \Big(1 - \frac{\gamma}{2\beta}\Big)\mathrm{Id} + \frac{\gamma}{2\beta}(-R).$$

$(-R)$ being nonexpansive, $\mathrm{Id} - \gamma A$ is $\gamma/(2\beta)$-averaged.

Nonexpansive operators

Nonexpansive operators → What is their use ?

Nonexpansive operator: example

Descent lemma

Let H be a real Hilbert space, $f\colon \mathsf{H} \to \mathbb{R}$ and $\nu \in \,]0,+\infty[$.
If f is differentiable and its gradient is ν-Lipschitzian, then

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

Baillon-Haddad theorem

Let H be a real Hilbert space, $f \in \Gamma_0(\mathsf{H})$ and $\nu \in \,]0,+\infty[$.
If f is differentiable, then ∇f ν-Lipschitzian $\Leftrightarrow$ ∇f ν^{-1}-cocoercive.

Nonexpansive operator: example

Descent lemma

Let H be a real Hilbert space, $f\colon \mathsf{H} \to \mathbb{R}$ and $\nu \in \left]0,+\infty\right[$.
If f is differentiable and its gradient is ν-Lipschitzian, then

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

Baillon-Haddad theorem

Let H be a real Hilbert space, $f \in \Gamma_0(\mathsf{H})$ and $\nu \in \left]0,+\infty\right[$.
If f is differentiable, then ∇f ν-Lipschitzian $\Leftrightarrow$ ∇f ν^{-1}-cocoercive.

Let H be a Hilbert space, $f \in \Gamma_0(\mathsf{H})$, $\nu \in \left]0,+\infty\right[$ and $\gamma \in]0,2\nu^{-1}[$.
f differentiable and ∇f ν-Lipschitzian $\Rightarrow$ $\mathrm{Id} - \gamma\nabla f$ is $\gamma\nu/2$-averaged.

Nonexpansive operator: example

Descent lemma

Let H be a real Hilbert space, $f\colon \mathsf{H} \to \mathbb{R}$ and $\nu \in \,]0,+\infty[$.
If f is differentiable and its gradient is ν-Lipschitzian, then

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

Baillon-Haddad theorem

Let H be a real Hilbert space, $f \in \Gamma_0(\mathsf{H})$ and $\nu \in \,]0,+\infty[$.
If f is differentiable, then ∇f ν-Lipschitzian $\Leftrightarrow$ ∇f ν^{-1}-cocoercive.

Let H be a Hilbert space, $f \in \Gamma_0(\mathsf{H})$, $\nu \in \,]0,+\infty[$ and $\gamma \in]0, 2\nu^{-1}[$.
f differentiable and ∇f ν-Lipschitzian $\Rightarrow$ $\underbrace{\mathrm{Id} - \gamma\nabla f}_{\text{gradient descent operator}}$ is $\gamma\nu/2$-averaged.

Nonexpansive operator: example

Descent lemma

Let H be a real Hilbert space, $f\colon \mathsf{H} \to \mathbb{R}$ and $\nu \in \,]0,+\infty[$.
If f is differentiable and its gradient is ν-Lipschitzian, then

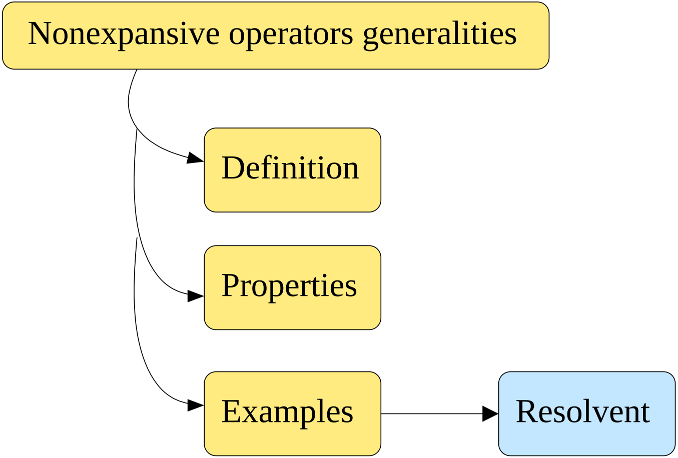
$$\big(\forall (x,y)\in \mathsf{H}^2\big) \quad f(y) \leqslant f(x) + \langle y - x \mid \nabla f(x)\rangle + \frac{\nu}{2}\|y-x\|^2.$$

Baillon-Haddad theorem

Let H be a real Hilbert space, $f \in \Gamma_0(\mathsf{H})$ and $\nu \in \,]0,+\infty[$.
If f is differentiable, then ∇f ν-Lipschitzian $\Leftrightarrow$ ∇f ν^{-1}-cocoercive.

Let H be a Hilbert space, $f \in \Gamma_0(\mathsf{H})$, and $\nu \in \,]0,+\infty[$
f differentiable and ∇f ν-Lipschitzian $\Leftrightarrow$ f^* is ν^{-1}-strongly convex.

Remark : f^* is ν^{-1}-strongly convex if $f^* - \nu^{-1}\|\cdot\|^2/2$ is convex.


Nonexpansive operators generalities
Definition
Properties
Examples
Resolvent

Resolvent: definition

Let H be a Hilbert space.
Let $A : \mathsf{H} \to 2^{\mathsf{H}}$.
The revolvent of A is

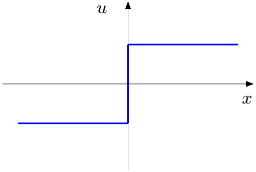
$$J_A = (\mathrm{Id} + A)^{-1}.$$

Resolvent: definition

Let H be a Hilbert space.
Let $A : \mathsf{H} \to 2^{\mathsf{H}}$.
The revolvent of A is

$$J_A = (\mathrm{Id} + A)^{-1}.$$

- Example :



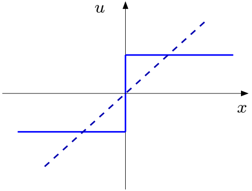
A $\qquad$ $A + \mathrm{Id}$? $\qquad$ J_A ?

Resolvent: definition

Let H be a Hilbert space.
Let $A : \mathsf{H} \to 2^{\mathsf{H}}$.
The revolvent of A is

$$J_A = (\mathrm{Id} + A)^{-1}.$$

- Example :



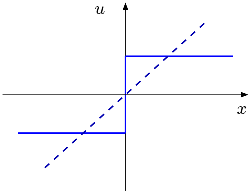
A and Id $\quad$ $A + \mathrm{Id}$? $\quad$ J_A ?

Resolvent: definition

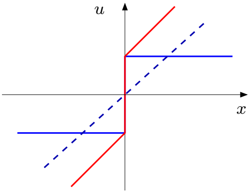
Let H be a Hilbert space.
Let $A : \mathsf{H} \to 2^{\mathsf{H}}$.
The revolvent of A is

$$J_A = (\mathrm{Id} + A)^{-1}.$$

- Example :



A and Id



$A + \mathrm{Id}$

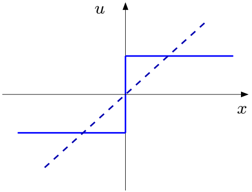
J_A ?

Resolvent: definition

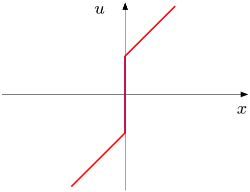
Let H be a Hilbert space.
Let $A : \mathsf{H} \to 2^{\mathsf{H}}$.
The revolvent of A is

$$J_A = (\mathrm{Id} + A)^{-1}.$$

- Example :



A and Id



$A + \mathrm{Id}$

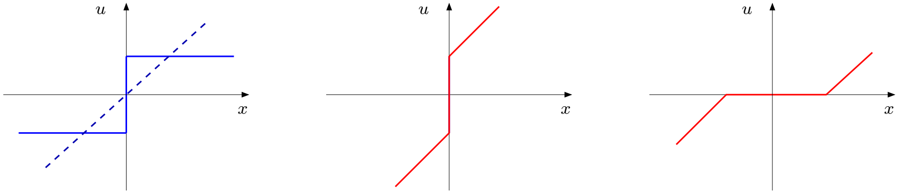
J_A ?

Resolvent: definition

Let H be a Hilbert space.
Let $A : \mathsf{H} \to 2^{\mathsf{H}}$.
The revolvent of A is

$$J_A = (\mathrm{Id} + A)^{-1}.$$

- Example :



A and Id $\qquad$ $A + \mathrm{Id}$ $\qquad$ J_A

Resolvent: definition

The range of an operator $B\colon \mathsf{H} \to 2^{\mathsf{H}}$ is

$$\operatorname{ran} B = \big\{u \in \mathsf{H} \;\big|\; \exists x \in \mathsf{H}, u \in Bx\big\}.$$

Minty theorem

Let H be a Hilbert space.
Let $A\colon \mathsf{H} \to 2^{\mathsf{H}}$ be a monotone operator.

$$\operatorname{ran}(\mathrm{Id} + A) = \mathsf{H} \quad \Rightarrow \quad A \text{ is maximally monotone.}$$

Proof: For all $(x_1, u_1) \in \mathsf{H}^2$, suppose that
$\big(\forall (x_2, u_2) \in \operatorname{gra} A\big) \quad \langle x_1 - x_2 \mid u_1 - u_2\rangle \geqslant 0.$
Since $\operatorname{ran}(\mathrm{Id} + A) = \mathsf{H}$, there exists $(x_2', u_2') \in \operatorname{gra} A$ such that
$x_1 + u_1 = x_2' + u_2'$. Consequently,
$0 \leqslant \langle x_1 - x_2' \mid u_1 - u_2'\rangle = \langle x_1 - x_2' \mid x_2' - x_1\rangle = -\|x_1 - x_2'\|^2.$
Thus, $x_1 = x_2'$ et $u_1 = u_2'$.

Resolvent: definition

The range of an operator $B\colon \mathsf{H} \to 2^{\mathsf{H}}$ is

$$\operatorname{ran} B = \left\{u \in \mathsf{H} \;\middle|\; \exists x \in \mathsf{H}, u \in Bx\right\}.$$

Minty theorem

Let H be a Hilbert space.

Let $A\colon \mathsf{H} \to 2^{\mathsf{H}}$ be a monotone operator.

$$\operatorname{ran}(\mathrm{Id} + A) = \mathsf{H} \quad \Leftrightarrow \quad A \text{ is maximally monotone.}$$

Resolvent: properties

Let H be a real Hilbert space. Let $A : \mathsf{H} \to 2^{\mathsf{H}}$.
A is monotone $\Leftrightarrow$ J_A is firmly nonexpansive.

Proof : A is monotone if and only if

$$
\begin{aligned}
&\big(\forall (x,u) \in \operatorname{gra} A\big)\big(\forall (y,v) \in \operatorname{gra} A\big) \quad \langle x-y \mid u-v\rangle \geqslant 0 \\
\Leftrightarrow &\big(\forall (x,u) \in \operatorname{gra} A\big)\big(\forall (y,v) \in \operatorname{gra} A\big) \quad \langle x-y \mid x-y+u-v\rangle \geqslant \|x-y\|^2 \\
\Leftrightarrow &\big(\forall (x,u') \in \operatorname{gra} (\mathrm{Id}\, + A)\big)\big(\forall (y,v') \in \operatorname{gra} (\mathrm{Id}\, + A)\big) \\
& \qquad\qquad\qquad\qquad \big\langle x-y \mid u'-v'\big\rangle \geqslant \|x-y\|^2 \\
\Leftrightarrow &\big(\forall (u',x) \in \operatorname{gra} J_A\big)\big(\forall (v',y) \in \operatorname{gra} J_A\big) \quad \big\langle u'-v' \mid x-y\big\rangle \geqslant \|x-y\|^2
\end{aligned}
$$

$\Leftrightarrow J_A$ is firmly nonexpansive

Resolvent: properties

Let H be a real Hilbert space. Let $A : \mathsf{H} \to 2^{\mathsf{H}}$.
A is monotone $\Leftrightarrow J_A$ is firmly nonexpansive.

Proof : A is monotone if and only if

$$
\begin{aligned}
&\big(\forall (x,u) \in \operatorname{gra} A\big)\big(\forall (y,v) \in \operatorname{gra} A\big) \quad \langle x-y \mid u-v\rangle \geqslant 0 \\
\Leftrightarrow &\big(\forall (x,u) \in \operatorname{gra} A\big)\big(\forall (y,v) \in \operatorname{gra} A\big) \quad \langle x-y \mid x-y+u-v\rangle \geqslant \|x-y\|^2 \\
\Leftrightarrow &\big(\forall (x,u') \in \operatorname{gra}(\mathrm{Id}+A)\big)\big(\forall (y,v') \in \operatorname{gra}(\mathrm{Id}+A)\big) \\
&\qquad\qquad\qquad \big\langle x-y \mid u'-v'\big\rangle \geqslant \|x-y\|^2 \\
\Leftrightarrow &\big(\forall (u',x) \in \operatorname{gra} J_A\big)\big(\forall (v',y) \in \operatorname{gra} J_A\big) \quad \big\langle u'-v' \mid x-y\big\rangle \geqslant \|x-y\|^2 \\
\Leftrightarrow &J_A \text{ is firmly nonexpansive}
\end{aligned}
$$

Remark : $J_A : \operatorname{ran}(\mathrm{Id}+A) \to \mathsf{H}$.

Resolvent: properties

Let H be a real Hilbert space. Let $A : \mathsf{H} \to 2^{\mathsf{H}}$.
A is monotone $\Leftrightarrow$ J_A is firmly nonexpansive.

Let H be a real Hilbert space. Let $A : \mathsf{H} \to 2^{\mathsf{H}}$.
A is maximally monotone $\Leftrightarrow$ $J_A : \mathsf{H} \to \mathsf{H}$ is firmly nonexpansive.

Resolvent: properties

Let H be a real Hilbert space. Let $A : \mathsf{H} \to 2^{\mathsf{H}}$.
A is monotone $\Leftrightarrow J_A$ is firmly nonexpansive.

Let H be a real Hilbert space. Let $A : \mathsf{H} \to 2^{\mathsf{H}}$.
A is maximally monotone $\Leftrightarrow J_A$: H $\to$ H is firmly nonexpansive.

Proof: A monotone $\Leftrightarrow J_A : \operatorname{ran}(\mathrm{Id} + A) \to$ H firmly nonexpansive
+ Minty theorem.

Resolvent: properties

Let H be a real Hilbert space. Let $A : \mathsf{H} \to 2^{\mathsf{H}}$.
A is monotone $\Leftrightarrow$ J_A is firmly nonexpansive.

Let H be a real Hilbert space. Let $A : \mathsf{H} \to 2^{\mathsf{H}}$.
A is maximally monotone $\Leftrightarrow$ $J_A : \mathsf{H} \to \mathsf{H}$ is firmly nonexpansive.

Let H be a Hilbert space. Let $A : \mathsf{H} \to 2^{\mathsf{H}}$ maximally monotone and $\gamma \in]0, +\infty[$. For every $x \in \mathsf{H}$, there exists a unique $p \in \mathsf{H}$ such that $x - p \in \gamma Ap$ and thus $p = J_{\gamma A}x$.

<u>Proof</u>: $x \in (\mathrm{Id} + \gamma A)(p) \Leftrightarrow p \in (\mathrm{Id} + \gamma A)^{-1}x \Leftrightarrow p = J_{\gamma A}x$

Resolvent: properties

Let H be a real Hilbert space.
Let $A : \mathsf{H} \to 2^{\mathsf{H}}$ be a maximally monotone and $\gamma \in]0, +\infty[$.

- $J_{\gamma A}$ and $\mathrm{Id} - J_{\gamma A}$ are firmly nonexpansive.
- The reflected resolvent $R_{\gamma A} = 2J_{\gamma A} - \mathrm{Id}$ is nonexpansive.
- ${}^{\gamma}A$ is γ-cocoercive.

Let H be a Hilbert space.
Let $A : \mathsf{H} \to 2^{\mathsf{H}}$ and $\gamma \in]0, +\infty[$.
The Yosida approximation of A of index γ is

$$ {}^{\gamma}A = \frac{1}{\gamma}(\mathrm{Id} - J_{\gamma A}). $$

Resolvent: properties

Let H be a Hilbert space.
Let $A\colon \mathsf{H} \to 2^{\mathsf{H}}$ be a maximally monotone operator.

- Let $z \in \mathsf{H}$ and $B = A(\cdot - z)$. Then $J_B = z + J_A(\cdot - z)$.
- Let $z \in \mathsf{H}$ and $B = z + A$. Then $J_B = J_A(\cdot - z)$.
- Let $\alpha \in [0, +\infty[$ and $B = A + \alpha \mathrm{Id}$. Then $J_B = J_{\frac{A}{1+\alpha}}\left(\frac{\cdot}{1+\alpha}\right)$

Resolvent: properties

For every $i \in \{1, \ldots, n\}$, let H_i be a Hilbert space and $A_i \colon \mathsf{H}_i \to 2^{\mathsf{H}_i}$ be a maximally monotone operator.

$$J_{A_1 \times \cdots \times A_n} = J_{A_1} \times \cdots \times J_{A_n} \colon \mathsf{H}_1 \times \cdots \times \mathsf{H}_n \to \mathsf{H}_1 \times \cdots \times \mathsf{H}_n$$
$$(x_1, \ldots, x_n) \mapsto (J_{A_1} x_1, \ldots, J_{A_n} x_n).$$

Resolvent: properties

Let H be a Hilbert space.
Let $A : \mathsf{H} \to 2^{\mathsf{H}}$ be a maximally monotone operator and $\gamma \in]0, +\infty[$.

$$J_{\gamma A^{-1}} = \mathrm{Id} - \gamma J_{\gamma^{-1} A}(\gamma^{-1} \cdot)$$

Preuve: Pour tout $x \in \mathsf{H}$,

$$\begin{aligned}
p = J_{\gamma A^{-1}} x \quad & \Leftrightarrow \quad x \in (\mathrm{Id} + \gamma A^{-1})(p) \\
& \Leftrightarrow \quad \gamma^{-1}(x - p) \in A^{-1} p \\
& \Leftrightarrow \quad p \in A\big(\gamma^{-1}(x - p)\big) \\
& \Leftrightarrow \quad \gamma^{-1} p \in \gamma^{-1} A\big(\gamma^{-1}(x - p)\big) \\
& \Leftrightarrow \quad \gamma^{-1} x \in (\mathrm{Id} + \gamma^{-1} A)\big(\gamma^{-1}(x - p)\big) \\
& \Leftrightarrow \quad \gamma^{-1}(x - p) = J_{\gamma^{-1} A}\big(\gamma^{-1} x\big) \\
& \Leftrightarrow \quad p = x - \gamma J_{\gamma^{-1} A}\big(\gamma^{-1} x\big).
\end{aligned}$$

Resolvent: properties

Let H be a Hilbert space.
Let $A : \mathsf{H} \to 2^{\mathsf{H}}$ be a maximally monotone operator and $\gamma \in]0, +\infty[$.

$$J_{\gamma A^{-1}} = \mathrm{Id} - \gamma J_{\gamma^{-1} A}(\gamma^{-1} \cdot)$$

Remarque: $J_A + J_{A^{-1}} = \mathrm{Id}$.

Resolvent: properties

Let H and G be two Hilbert spaces.
Let $A : \mathsf{H} \to 2^{\mathsf{H}}$ be a maximally monotone operator and $L \in \mathcal{B}(\mathsf{G}, \mathsf{H})$ such that $LL^* = \mu \mathrm{Id}$ where $\mu \in]0, +\infty[$. Then

$$J_{L^*AL} = \mathrm{Id} - L^* \circ {}^{\mu}A \circ L .$$

Resolvent: properties

Let H and G be two Hilbert spaces.
Let $A : \mathsf{H} \to 2^{\mathsf{H}}$ be a maximally monotone operator and $L \in \mathcal{B}(\mathsf{G}, \mathsf{H})$ such that $LL^* = \mu \mathrm{Id}$ where $\mu \in]0, +\infty[$. Then

$$J_{L^*AL} = \mathrm{Id} - L^* \circ {}^{\mu}A \circ L.$$

Let H be a Hilbert space.
Let $A : \mathsf{H} \to 2^{\mathsf{H}}$ be a maximally monotone operator and $L \in \mathcal{B}(\mathsf{H}, \mathsf{H})$ be a unitary operator. Then

$$J_{L^*AL} = L^* J_A L.$$

Resolvent: properties

Let H be a Hilbert space.
Let $A : \mathsf{H} \to 2^{\mathsf{H}}$ be a maximally monotone operator and $B = \rho A(\rho\cdot)$ where $\rho \in \mathbb{R}^*$. Then

$$J_B = \rho^{-1} J_{\rho^2 A}(\rho\cdot).$$

Preuve: Set $L = \rho \mathrm{Id}$ and apply formula

$$J_{L^* A L} = \mathrm{Id} - L^* \circ {}^{\mu}A \circ L.$$

Resolvent: properties

Let H be a Hilbert space.
Let $A : \mathsf{H} \to 2^{\mathsf{H}}$ be a maximally monotone operator and $B = \rho A(\rho\cdot)$ where $\rho \in \mathbb{R}^*$. Then

$$J_B = \rho^{-1} J_{\rho^2 A}(\rho\cdot).$$

Let H be a Hilbert space.
Let $A : \mathsf{H} \to 2^{\mathsf{H}}$ be a maximally monotone operator and $B = -A(-\cdot)$. Then

$$J_B = -J_A(-\cdot).$$

Resolvent

Resolvent

Proximity operator

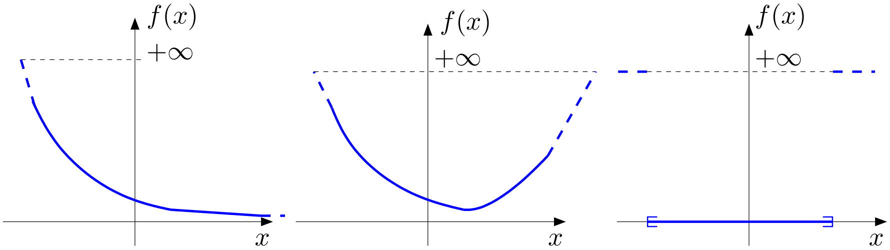
Convex analysis

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

Convex analysis

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

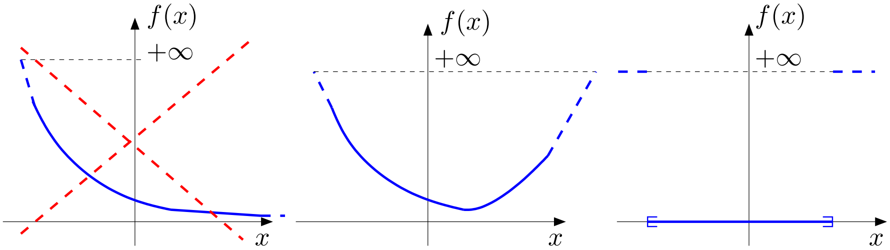
Coercive functions ?



Convex analysis

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

Coercive functions ?



Convex analysis

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

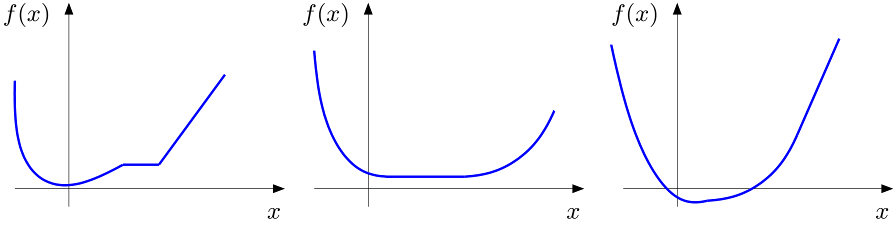
$$(\forall x \in \operatorname{dom} f)(\forall y \in \operatorname{dom} f)(\forall \alpha \in]0,1[)$$
$$x \neq y \quad \Rightarrow \quad f(\alpha x + (1-\alpha)y) < \alpha f(x) + (1-\alpha) f(y).$$

Convex analysis

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

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

Strictly convex functions ?

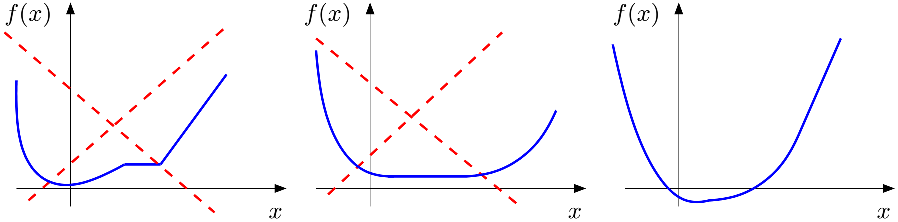


Convex analysis

Let H be a Hilbert space. Let $f\colon \mathsf{H} \to]-\infty, +\infty]$.
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 ?



Convex analysis

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

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

Moreover, if f is strictly convex, this minimizer p is unique.

Proximity operator: definition

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

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

Proximity operator: definition

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

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

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

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

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

- The proximity operator of f is

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

Proximity operator: definition

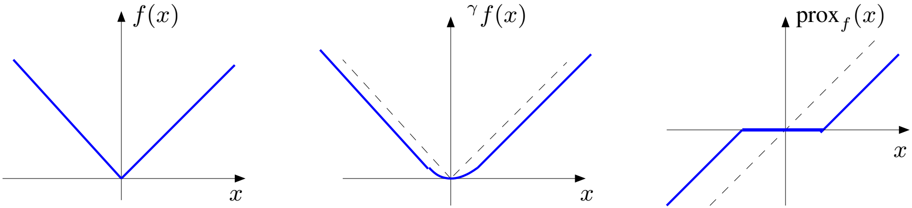
Let H be a Hilbert space. Let $f \in \Gamma_0(\mathsf{H})$.

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

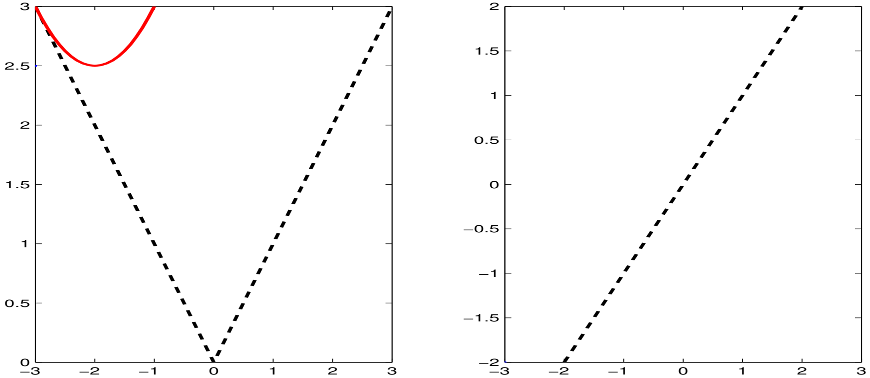
$$^{\gamma}f\colon \mathsf{H} \to \mathbb{R}\colon x \mapsto \inf_{y \in \mathsf{H}} f(y) + \frac{1}{2\gamma}\|y - x\|^2.$$

- The **proximity operator** of f is

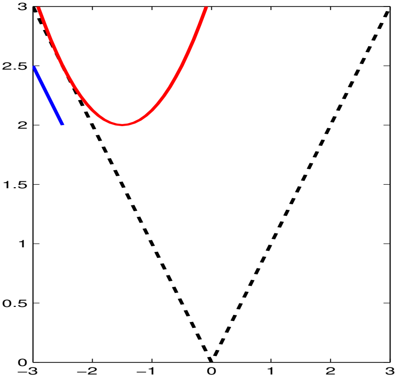
$$\operatorname{prox}_f\colon \mathsf{H} \to \mathsf{H}\colon x \mapsto \underset{y \in \mathsf{H}}{\operatorname{argmin}}\; f(y) + \frac{1}{2}\|y - x\|^2.$$

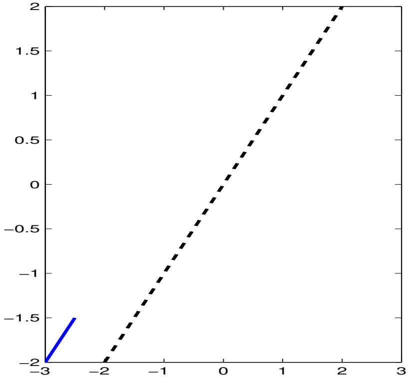


Proximity operator: definition

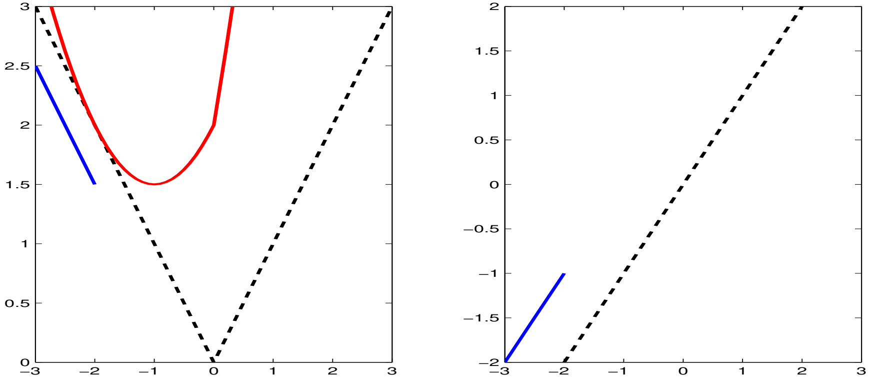


Proximity operator: definition

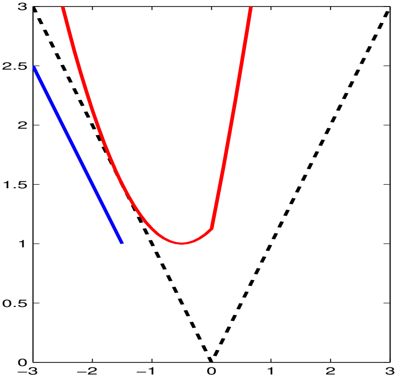


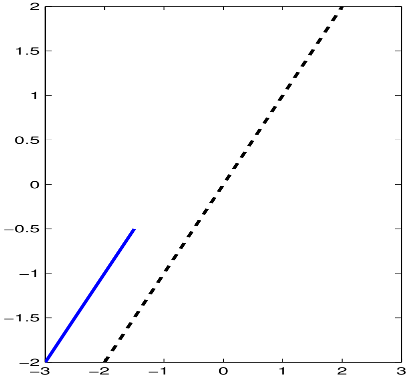


Proximity operator: definition

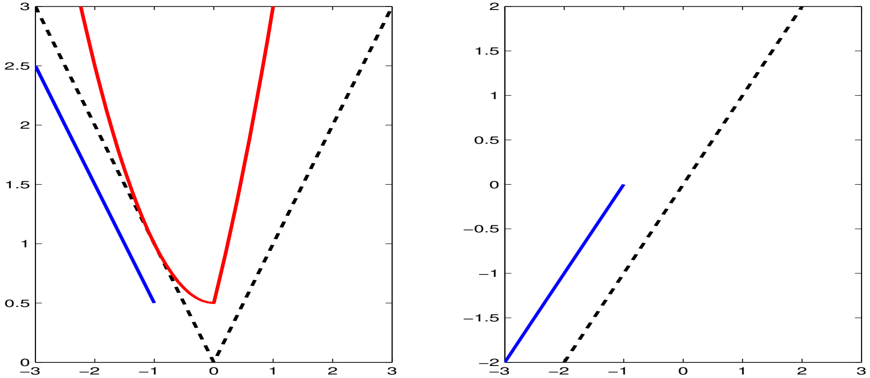


Proximity operator: definition

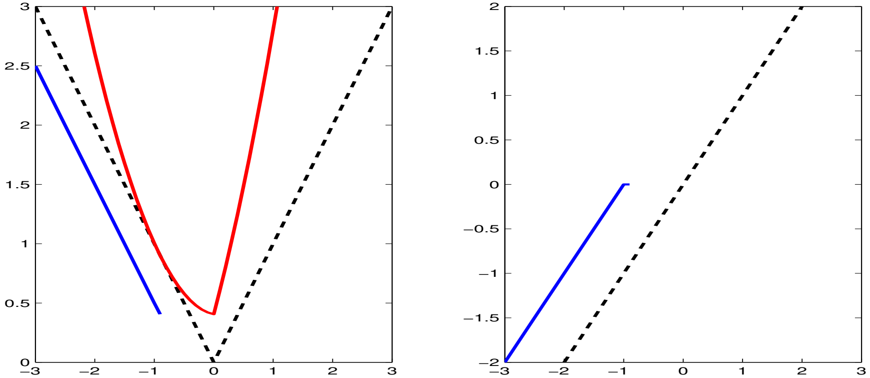




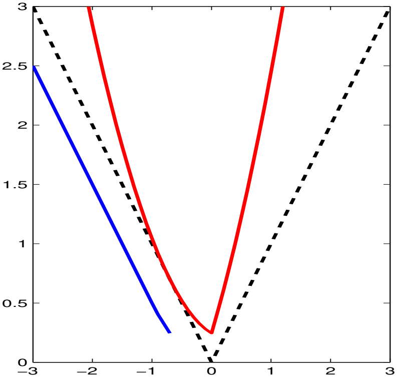
Proximity operator: definition

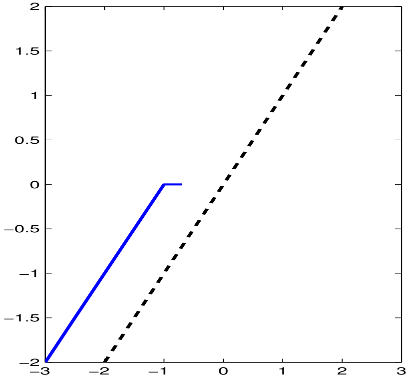


Proximity operator: definition

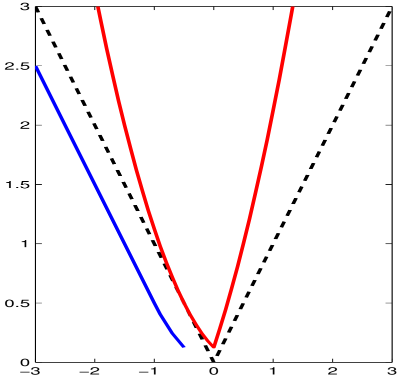


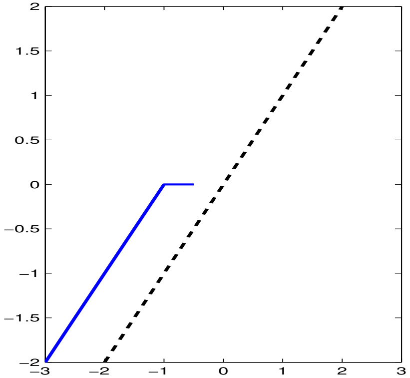
Proximity operator: definition



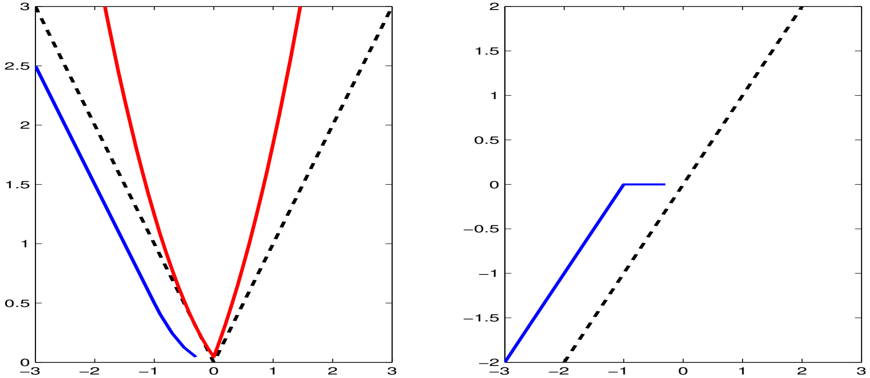


Proximity operator: definition

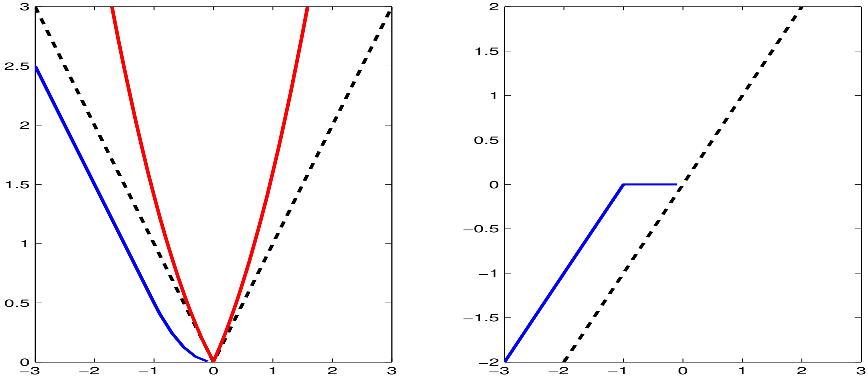




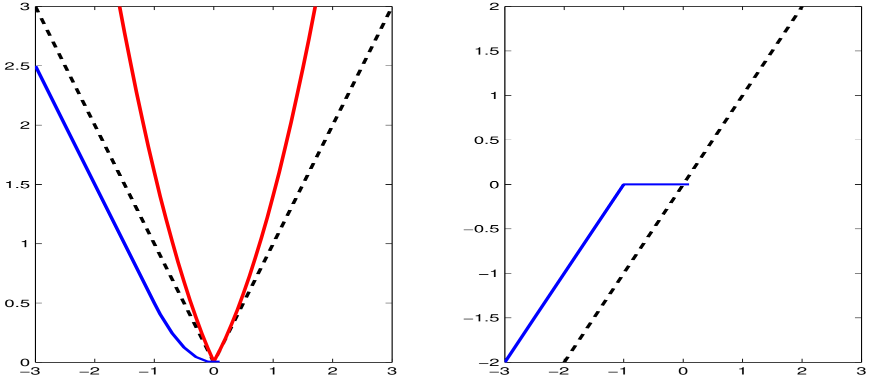
Proximity operator: definition



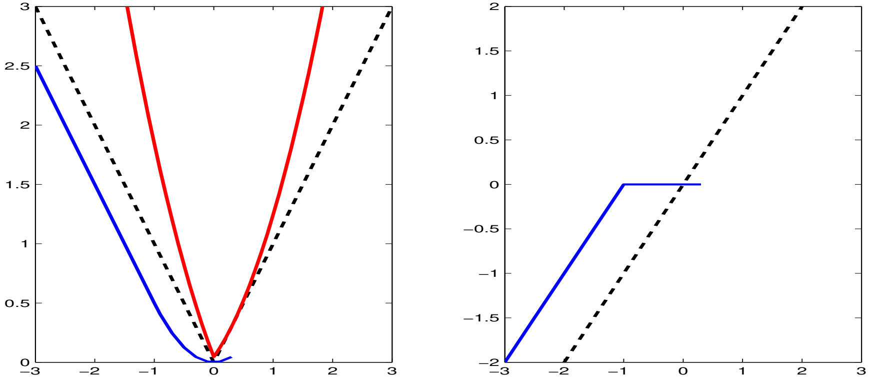
Proximity operator: definition



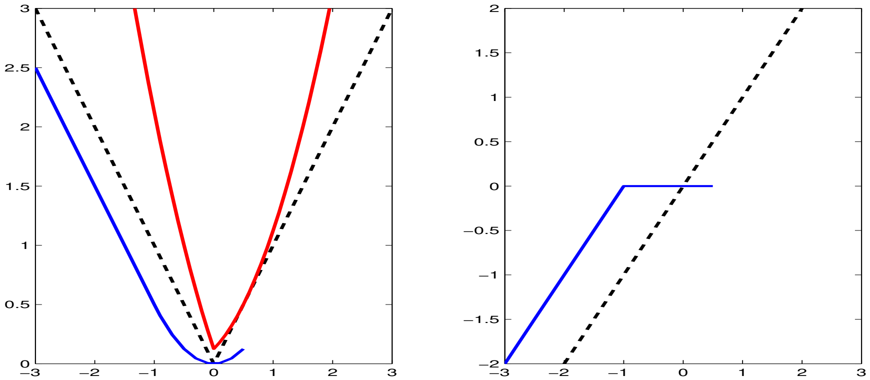
Proximity operator: definition



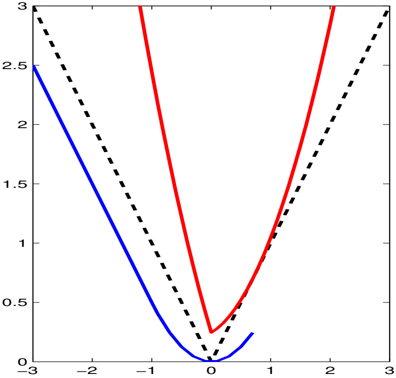
Proximity operator: definition

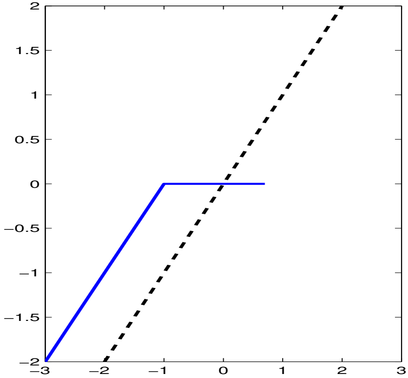


Proximity operator: definition

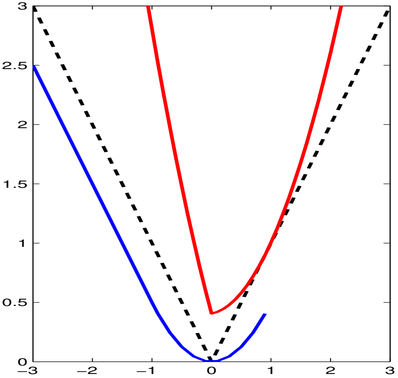


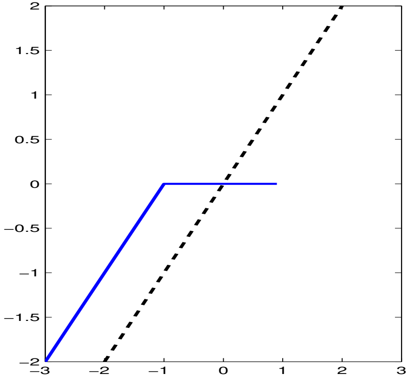
Proximity operator: definition



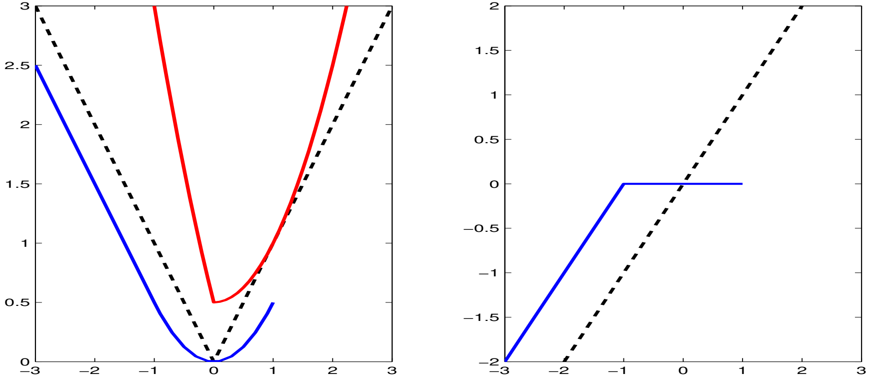


Proximity operator: definition

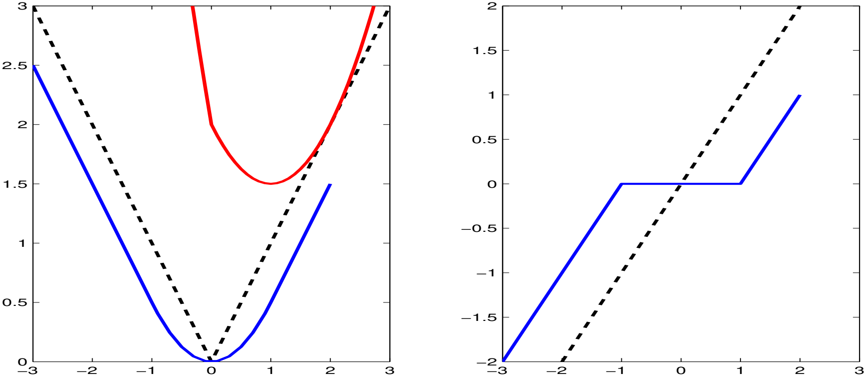




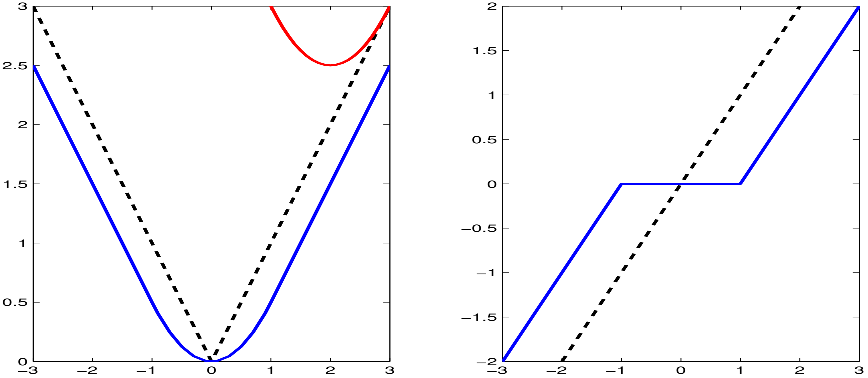
Proximity operator: definition



Proximity operator: definition



Proximity operator: definition



Proximity operator: definition

Let H be a Hilbert space. Let $f \in \Gamma_0(\mathsf{H})$ and $g \in \Gamma_0(\mathsf{H})$.
If $\operatorname{dom} f \cap \operatorname{int}(\operatorname{dom} g) \neq \varnothing$ then $\partial(f+g) = \partial f + \partial g$.

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

$$\operatorname{prox}_f = J_{\partial f} \, .$$

Proximity operator: definition

Let H be a Hilbert space. Let $f \in \Gamma_0(\mathsf{H})$ and $g \in \Gamma_0(\mathsf{H})$.
If $\operatorname{dom} f \cap \operatorname{int}(\operatorname{dom} g) \neq \varnothing$ then $\partial(f+g) = \partial f + \partial g$.

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

$$\operatorname{prox}_f = J_{\partial f}\,.$$

Proof: By using Fermat's rule, for every $x \in \mathsf{H}$,

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

Proximity operator: definition

Let H be a Hilbert space. Let $f \in \Gamma_0(\mathsf{H})$ and $g \in \Gamma_0(\mathsf{H})$.
If $\mathrm{dom}\, f \cap \mathrm{int}(\mathrm{dom}\, g) \neq \varnothing$ then $\partial(f+g) = \partial f + \partial g$.

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

$$\mathrm{prox}_f = J_{\partial f}.$$

Remark: As $\mathrm{dom}\,(\mathrm{prox}_f) = \mathsf{H}$, this provides a proof that ∂f is maximally monotone !

Proximity operator: properties

Let H be a Hilbert space, $f \in \Gamma_0(\mathsf{H})$ and $(x, p) \in \mathsf{H}^2$.

$$p = \mathrm{prox}_{\gamma f} x \quad \Leftrightarrow \quad (\forall y \in \mathsf{H}) \;\; \langle y - p \mid x - p \rangle + f(p) \leqslant f(y).$$

Proximity operator: properties

Let H be a Hilbert space, $f \in \Gamma_0(\mathsf{H})$ and $(x,p) \in \mathsf{H}^2$.

$$p = \operatorname{prox}_{\gamma f} x \quad \Leftrightarrow \quad (\forall y \in \mathsf{H}) \;\; \langle y - p \mid x - p \rangle + f(p) \leqslant f(y).$$

Let H be a Hilbert space, $f \in \Gamma_0(\mathsf{H})$ and $\gamma \in \,]0,+\infty[$.
${}^{\gamma}f$ is differentiable and $\nabla\, {}^{\gamma}f$ is γ^{-1}-Lipschitzian

$$(\forall x \in \mathsf{H}) \quad \nabla \underbrace{{}^{\gamma}f}_{\text{Moreau envelope}} = \gamma^{-1}(\mathrm{Id} - \operatorname{prox}_{\gamma f}) = \underbrace{{}^{\gamma}\partial f}_{\text{Yosida approximation}} .$$

Proximity operator: properties

Let H be a Hilbert space, $f \in \Gamma_0(\mathsf{H})$ and $(x, p) \in \mathsf{H}^2$.

$$p = \mathrm{prox}_{\gamma f} x \quad \Leftrightarrow \quad (\forall y \in \mathsf{H}) \quad \langle y - p \mid x - p \rangle + f(p) \leqslant f(y).$$

Let H be a Hilbert space, $f \in \Gamma_0(\mathsf{H})$ and $\gamma \in \,]0, +\infty[$.
${}^{\gamma}f$ is differentiable and $\nabla\, {}^{\gamma}f$ is γ^{-1}-Lipschitzian

$$(\forall x \in \mathsf{H}) \quad \nabla \underbrace{{}^{\gamma}f}_{\text{Moreau envelope}} = \gamma^{-1}(\mathrm{Id} - \mathrm{prox}_{\gamma f}) = \underbrace{{}^{\gamma}\partial f}_{\text{Yosida approximation}}.$$

Proof: Previous property + ... calculations.

Proximity operator: properties

Let H be a Hilbert space, $f \in \Gamma_0(\mathsf{H})$ and $(x,p) \in \mathsf{H}^2$.

$$p = \operatorname{prox}_{\gamma f} x \quad \Leftrightarrow \quad (\forall y \in \mathsf{H}) \quad \langle y - p \mid x - p \rangle + f(p) \leqslant f(y).$$

Let H be a Hilbert space, $f \in \Gamma_0(\mathsf{H})$ and $\gamma \in]0,+\infty[$.
${}^{\gamma}f$ is differentiable and $\nabla\, {}^{\gamma}f$ is γ^{-1}-Lipschitzian

$$(\forall x \in \mathsf{H}) \quad \nabla \underbrace{{}^{\gamma}f}_{\text{Moreau envelope}} = \gamma^{-1}(\mathrm{Id} - \operatorname{prox}_{\gamma f}) = \underbrace{{}^{\gamma}\partial f}_{\text{Yosida approximation}} .$$

Interpretation: ${}^{\gamma}f$ is a smooth approximation of f.

Proximity operator: properties

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

Properties	$g(x)$	$\operatorname{prox}_g x$
Translation	$f(x-z), z \in \mathsf{H}$	$z + \operatorname{prox}_f(x-z)$
Quadratic perturbation	$f(x) + \alpha \parallel x \parallel^2 /2 + \langle z \mid x \rangle + \gamma$ $z \in \mathsf{H}, \alpha > 0, \gamma \in \mathbb{R}$	$\operatorname{prox}_{\frac{f}{\alpha+1}}(\frac{x-z}{\alpha+1})$
Scale change	$f(\rho x), \rho \in \mathbb{R}^*$	$\frac{1}{\rho}\operatorname{prox}_{\rho^2 f}(\rho x)$
Reflection	$f(-x)$	$-\operatorname{prox}_f(-x)$
Moreau envelope	${}^{\gamma}f(x) = \inf_{y \in \mathsf{H}} f(y) + \dfrac{1}{2\gamma}\|x-y\|^2$ $\gamma > 0$	$\frac{1}{1+\gamma}\left(\gamma x + \operatorname{prox}_{(1+\gamma)f}(x)\right)$

Proximity operator: properties

For every $i \in \{1, \ldots, n\}$, let H_i be a Hilbert space and $f_i \in \Gamma_0(\mathsf{H}_i)$.
For all $(x_1, \ldots, x_n) \in \mathsf{H}_1 \times \cdots \times \mathsf{H}_n$,
if

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

then

$$\operatorname{prox}_f(x_1, \ldots, x_n) = \big(\operatorname{prox}_{f_i}(x_i)\big)_{1 \leqslant i \leqslant n}.$$

Proximity operator: properties

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

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

then

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

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

Proximity operator: properties

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

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

then

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

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

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

Proximity operator: properties

Moreau decomposition formula

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

$$(\forall x \in \mathsf{H}) \qquad J_{\gamma(\partial f)^{-1}} x = x - \gamma J_{\gamma^{-1}\partial f}(\gamma^{-1} x) \,.$$

Proximity operator: properties

Moreau decomposition formula

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

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

Proximity operator: properties

Moreau decomposition formula

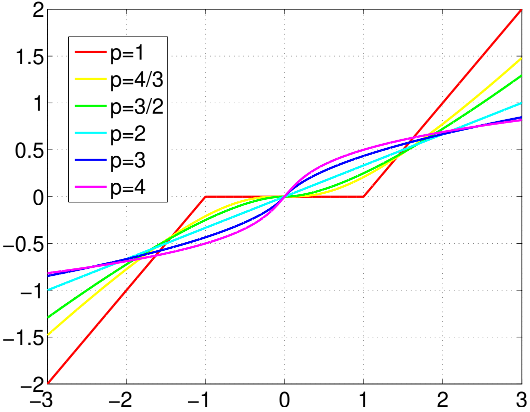
Let H be a Hilbert space, $f \in \Gamma_0(\mathsf{H})$ and $\gamma \in \,]0,+\infty[$.

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

Example: If $\mathsf{H} = \mathbb{R}^N$, $f = \frac{1}{q}\|\cdot\|_q^q$ 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: properties



Proximity operator: properties

Let H and G be two Hilbert spaces. Let $f \in \Gamma_0(\mathsf{H})$ and $L \in \mathcal{B}(\mathsf{G}, \mathsf{H})$ such that $\operatorname{ran} L = \mathsf{H}$. Then

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

Proximity operator: properties

Let H and G be two Hilbert spaces. Let $f \in \Gamma_0(\mathsf{H})$ and $L \in \mathcal{B}(\mathsf{G}, \mathsf{H})$ such that $\operatorname{ran} L = \mathsf{H}$. Then

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

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

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

Proximity operator: properties

Let H and G be two Hilbert spaces. Let $f \in \Gamma_0(\mathsf{H})$ and $L \in \mathcal{B}(\mathsf{G},\mathsf{H})$ such that $\operatorname{ran} L = \mathsf{H}$. Then

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

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

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

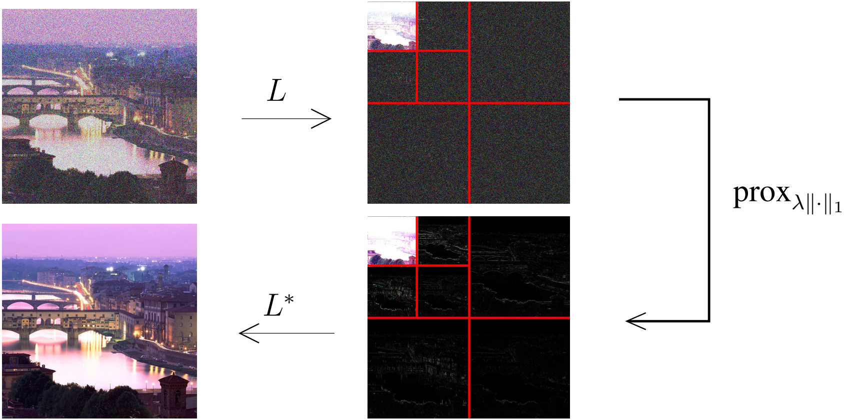
Remark :

Useful property for data fidelity terms involving a neg-log-likelihood f and a synthesis tight frame operator L.

Proximity operator: properties

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

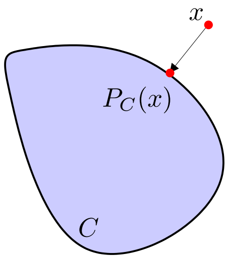
- Illustration: denoising using an ℓ_1 penalty on the coefficients resulting from an orthogonal wavelet transform L.



Proximity operator: examples

Projection :
Let H be a Hilbert space. Let C be a nonempty closed convex subset of H.

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



Proximity operator: examples

Projection :

Let H be a Hilbert space. Let C be a nonempty closed convex subset of H.

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

Remark :

- $p = P_C(x) \Leftrightarrow x - p \in \partial \iota_C(p) = N_C(p)$

  $$\Rightarrow (\forall y \in C)\ \langle y - p \mid x - p \rangle \leqslant 0.$$

  Particular case: if C is a vector space: $p = P_C(x) \Leftrightarrow \begin{cases} x - p \in C^\perp \\ p \in C. \end{cases}$

- ${}^{\gamma}\iota_C = (2\gamma)^{-1} d_C^2$ where d_C distance to the convex set C is defined by $d_C \colon x \mapsto \inf_{y \in C} \|y - x\| = \|x - P_C x\|$. We have then $\nabla d_C^2 = \nabla(\frac{1}{2}\iota_C) = 2(\mathrm{Id} - P_C)$.

Proximity operator: examples

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

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

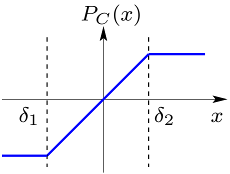
Proximity operator: examples

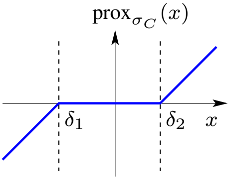
Support function :
Let H be a Hilbert space and $C \subset$ H be nonempty closed convex.

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

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

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





Proximity operator: Bayesian interpretation

- If $\mathsf{H} = \mathbb{R}^N$ and

$$x = \overline{y} + w$$

  where $\overline{y}$ is a realization of a random vector with probability density function $\exp(-f)$ and w is a realization of a $\mathcal{N}(0, \mathbb{I})$ noise, then $\operatorname{prox}_f(x)$ is a Maximum A Posteriori estimate of $\overline{y}$.
- Explicit form for objective functions associated with usual log-concave probability densities

  - Laplace
  - Generalized Gaussian
  - maximum entropy
  - gamma
  - uniform
  - Weibull
  - Generalized inverse Gaussian
  - Gaussian
  - Huber
  - Smoothed Laplace
  - chi
  - triangular
  - Pearson type I

  ...

- And many other functions ! http://proximity-operator.net

Part 3: Fixed point algorithms

1. Convergence
   - Definition
   - Fejér monotonicity
   - Demiclosedness principle
2. Algorithms
   - Krasnosel'skii Mann
   - Douglas-Rachford
   - PPXA
   - Forward-Backward

Fixed point algorithms

Fixed point algorithms: convergence

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

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

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

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

Fixed point algorithms: convergence

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

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

  and

- $(x_n)_{n\in\mathbb{N}}$ possesses 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 algorithms: convergence

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

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

  and

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

<u>Illustration</u>:

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 algorithms: convergence

Opial's lemma

Let D be a nonempty subset of H.
Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in H.
$(x_n)_{n\in\mathbb{N}}$ 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 algorithms: convergence

Opial's lemma

Let D be a nonempty subset of H.
Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in H.
$(x_n)_{n\in\mathbb{N}}$ 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.

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 algorithms: Fejér-monotone sequence

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

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

Let $D \subset$ H.
Let $(x_n)_{n\in\mathbb{N}}$ be Fejér-monotone with respect to D then

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

Fixed point algorithms: Fejér-monotone sequence

Fejér-monotone convergence

Let D be a nonempty subset of a Hilbert space H.
Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in 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 algorithms: Fejér-monotone sequence

Lemma

Let C be a nonempty closed convex subset of 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 algorithms: Fejér-monotone sequence

Lemma

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

Let C be a nonempty set of a Hilbert space H. Let $T\colon C \to$ H.
The set of fixed points of T is

$$\operatorname{Fix} T = \{x \in C \mid x = Tx\}.$$

Let $S\colon C \to 2^{\mathsf{H}}$. The set of zeros of S is

$$\operatorname{zer} S = \{x \in C \mid 0 \in Sx\}.$$

Nonexpansive operator: fixed point algorithm

Demiclosedness principle

Let C be a nonempty closed convex subset of a Hilbert space H.
Let $T: C \to \mathsf{H}$ 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 C be a nonempty closed convex subset of a Hilbert space 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 C be a nonempty closed convex subset of a Hilbert space H.
Let $T\colon C \to$ 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,

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

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

Nonexpansive operator: fixed point algorithm

Let C be a nonempty closed convex subset of a Hilbert space H.
Let $T\colon C \to C$ be a nonexpansive operator

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

Nonexpansive operator: fixed point algorithm

Let C be a nonempty closed convex subset of a Hilbert space H.
Let $T\colon C \to C$ 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 C be a nonempty closed convex subset of a Hilbert space 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 $y \in \operatorname{Fix} T$,
$\|x_{n+1} - y\| = \|Tx_n - Ty\| \leqslant \|x_n - y\|$.
$(x_n)_{n\in\mathbb{N}}$ 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 $\widehat{x} \in$ 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 algorithms: Fejér-monotone sequence

Krasnosel'skii-Mann algorithm

Let C be a nonempty closed convex subset of a Hilbert space 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 $\operatorname{Fix} T$.

Typical choice: $(\forall n \in \mathbb{N})\ \lambda_n = \lambda \in]0,1[$.

Fixed point algorithms: Fejér-monotone sequence

Proof :

For every $n \in \mathbb{N}$, by convex combination, $x_n \in C$.

Fejér-monotonicity with respect to $\operatorname{Fix} T$: $(\forall x \in \operatorname{Fix} T)(\forall n \in \mathbb{N})$

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

Fixed point algorithms: Fejér-monotone sequence

Proof :
We want to prove that $Tx_n - x_n \to 0$.
We deduce from $\|x_{n+1} - x\|^2 \leqslant \|x_n - x\|^2 - \lambda_n(1 - \lambda_n)\|Tx_n - x_n\|^2$ that

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

$$\Rightarrow \inf_{k \geqslant n} \|Tx_k - x_k\|^2 \sum_{k=n}^{+\infty} \lambda_k(1 - \lambda_k) \to 0.$$

The assumptions over the sequence $(\lambda_n)_{n \in \mathbb{N}}$ lead to $\liminf_{n \to +\infty} \|Tx_n - x_n\| = 0$.

Fixed point algorithms: Fejér-monotone sequence

Proof :

The assumptions on $(\lambda_n)_{n\in\mathbb{N}}$ lead to $\liminf_{n\to+\infty}\|Tx_n - x_n\| = 0$.
Moreover, as T is a nonexpansive operator

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

Consequently, $(\|Tx_n - x_n\|)_{n\in\mathbb{N}}$ converges and

$$Tx_n - x_n \to 0.$$

Fixed point algorithms: Fejér-monotone sequence

Proof :
Let $(x_{n_k})_{k\in\mathbb{N}}$ be a subsequence of $(x_n)_{n\in\mathbb{N}}$ such that $x_{n_k} \rightharpoonup \widehat{x}$. According to the demiclosedness principle, $Tx_{n_k} - x_{n_k} \to 0$, yields $\widehat{x} \in \operatorname{Fix} T$. The weak convergence of $(x_n)_{n\in\mathbb{N}}$ to $\widehat{x}$ results from the Fejér-monotonicity of $(x_n)_{n\in\mathbb{N}}$ with respect to $\operatorname{Fix} T$.

α-averaged operator: recall

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

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

Remark : When $\alpha = 1/2$, A is firmly nonexpansive.

Fixed point algorithms: α-averaged operator

Let $T\colon \mathsf{H} \to \mathsf{H}$ 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 $x_0 \in \mathsf{H}$ 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 $\operatorname{Fix} T$.
- $(Tx_n - x_n)_{n\in\mathbb{N}}$ converges strongly to 0.
- $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in $\operatorname{Fix} T$.

Fixed point algorithms: α-averaged operator

<u>Proof</u> :
Since T is α-averaged, 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

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

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

Optimization algorithms: *Forward-Backward*

Let $f \in \Gamma_0(\mathsf{H})$.
Let $g \in \Gamma_0(\mathsf{H})$ be differentiable with a ν-Lipschitzian gradient where $\nu \in]0, +\infty[$.
Let $\gamma \in]0, 2/\nu[$ and $\delta = 2 - \gamma\nu/2 \in]1, 2[$.
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 $\operatorname{Argmin}(f + g) \neq \varnothing$. Let $x_0 \in \mathsf{H}$ and

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

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

Optimization algorithms: *Forward-Backward*

Proof: Let $T = \mathrm{prox}_{\gamma f} \circ (\mathrm{Id} - \gamma \nabla g)$. For every $x \in$ 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, $\mathrm{Fix}\, T = \mathrm{zer}\,(\nabla g + \partial f) = \mathrm{zer}\,\big(\partial(g+f)\big) \neq \varnothing$.
Moreover, for every $n \in \mathbb{N}$,

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

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

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

Optimization algorithms: *Forward-Backward* with varying stepsize

Let $f \in \Gamma_0(\mathsf{H})$.
Let $g \in \Gamma_0(\mathsf{H})$ be differentiable with a ν-Lipschitzian gradient where $\nu \in]0, +\infty[$.
Let $(\gamma_n)_{n\in\mathbb{N}}$ in $[\underline{\gamma}, \overline{\gamma}]$ where $0 < \underline{\gamma} < \overline{\gamma} < 2/\nu$ and
let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[\underline{\lambda}, 1]$ with $0 < \underline{\lambda} \leqslant 1$.
We assume that $\operatorname{Argmin}(f+g) \neq \varnothing$. Let $x_0 \in \mathsf{H}$ and

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

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

Optimization algorithms: projected gradient

Let C be a nonempty closed convex subset of H.
Let $g \in \Gamma_0(\mathsf{H})$ be differentiable with a ν-Lipschitzian gradient where $\nu \in]0, +\infty[$.
Let $\gamma \in]0, 2/\nu[$ and $\delta = 2 - \gamma\nu/2 \in]1, 2[$.
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 $\underset{x \in C}{\operatorname{Argmin}}\; g(x) \neq \varnothing$. Let $x_0 \in \mathsf{H}$ 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 algorithms: gradient descent

Let $g \in \Gamma_0(\mathsf{H})$ be a differentiable function with a ν-Lipschitzian gradient where $\nu \in]0, +\infty[$.
Let $(\gamma_n)_{n\in\mathbb{N}}$ in $[\underline{\gamma}, \overline{\gamma}]$ where $0 < \underline{\gamma} < \overline{\gamma} < 2/\nu$.
We assume that $\operatorname{Argmin} g \neq \varnothing$. Let $x_0 \in \mathsf{H}$ and

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

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

Optimization algorithms: proximal point algorithm

Let $f \in \Gamma_0(\mathsf{H})$.
Let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $]0,+\infty[$ such that $\sum_{n\in\mathbb{N}} \gamma_n = +\infty$. We assume that $\operatorname{Argmin} f \neq \varnothing$. Let $x_0 \in \mathsf{H}$ and

$$(\forall n \in \mathbb{N}) \qquad x_{n+1} = \operatorname{prox}_{\gamma_n f} x_n.$$

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

Optimization algorithms: Douglas-Rachford

Motivation

Let $f \in \Gamma_0(\mathsf{H})$ and $g \in \Gamma_0(\mathsf{H})$. We want to

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

Possible solutions :

- gradient descent algorithm $\Rightarrow f + g$ needs to be smooth
- proximal point algorithm $\Rightarrow f + g$ needs to be “proximable”
- Forward-Backward algorithm $\Rightarrow g$ needs to be smooth

Can we find a splitting algorithm when both f and g are nonsmooth?

Optimization algorithms: Douglas-Rachford

Let $\gamma \in]0, +\infty[$ and let $f \in \Gamma_0(\mathrm{H})$.
The reflection of the proximal operator defined as

$$\mathrm{rprox}_{\gamma f} = 2\mathrm{prox}_{\gamma f} - \mathrm{Id}$$

is nonexpansive.

Let $\gamma \in]0, +\infty[$ and let $f \in \Gamma_0(\mathrm{H})$ and $g \in \Gamma_0(\mathrm{H})$.
We have

$$\mathrm{zer}\,(\partial f + \partial g) = \mathrm{prox}_{\gamma g}(\mathrm{Fix}\, T)$$

where $T = \mathrm{rprox}_{\gamma f} \circ \mathrm{rprox}_{\gamma g}$.

Optimization algorithms: Douglas-Rachford

Let $\gamma \in]0, +\infty[$ and let $f \in \Gamma_0(\mathsf{H})$.
The reflection of the proximal operator defined as

$$\mathrm{rprox}_{\gamma f} = 2\mathrm{prox}_{\gamma f} - \mathrm{Id}$$

is nonexpansive.

Let $\gamma \in]0, +\infty[$ and let $f \in \Gamma_0(\mathsf{H})$ and $g \in \Gamma_0(\mathsf{H})$.
We have

$$\mathrm{zer}\,(\partial f + \partial g) = \mathrm{prox}_{\gamma g}(\mathrm{Fix}\,T)$$

where $T = \mathrm{rprox}_{\gamma f} \circ \mathrm{rprox}_{\gamma g}$.

Proof: Let $x \in \mathsf{H}$.

$$\begin{aligned}
0 \in \gamma\big(\partial f(x) + \partial g(x)\big) &\Leftrightarrow (\exists y \in \mathsf{H})\;\; x - y \in \gamma\partial f(x) \text{ and } y - x \in \gamma\partial g(x)\\
&\Leftrightarrow (\exists y \in \mathsf{H})\;\; 2x - y \in (\mathrm{Id} + \gamma\partial f)x\\
&\qquad\qquad \text{and } x = \mathrm{prox}_{\gamma g}(y)\\
&\Leftrightarrow (\exists y \in \mathsf{H})\;\; x = \mathrm{prox}_{\gamma f}\big(\mathrm{rprox}_{\gamma g}(y)\big) \text{ and } x = \mathrm{prox}_{\gamma g}(y)\\
&\Leftrightarrow (\exists y \in \mathsf{H})\;\; \mathrm{rprox}_{\gamma f}\big(\mathrm{rprox}_{\gamma g}(y)\big) = 2x - \mathrm{rprox}_{\gamma g}(y) = 2x - (2\mathrm{prox}_{\gamma g}(y) - y) = y\\
&\qquad\qquad \text{and } x = \mathrm{prox}_{\gamma g}(y)\\
&\Leftrightarrow (\exists y \in \mathrm{Fix}\,T)\;\; x = \mathrm{prox}_{\gamma g}(y).
\end{aligned}$$

Optimization algorithms: Douglas-Rachford

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

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

Optimization algorithms: Douglas-Rachford

Let H be a Hilbert space.
Let $f \in \Gamma_0(\mathsf{H})$ and $g \in \Gamma_0(\mathsf{H})$.
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 $\operatorname{zer}(\partial f + \partial g) \neq \varnothing$. Let $x_0 \in \mathsf{H}$ and

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

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 $\widehat{y} = \operatorname{prox}_{\gamma g}\widehat{x} \in \operatorname{Argmin}(f+g)$.

Optimization algorithms: Douglas-Rachford

Proof: For simplicity, assume that H is finite dimensional.
Let $T = \operatorname{rprox}_{\gamma f} \circ \operatorname{rprox}_{\gamma g}$. T is nonexpansive and $\varnothing \neq \operatorname{zer}(\partial f + \partial g) = \operatorname{prox}_{\gamma g}(\operatorname{Fix} T) \Rightarrow \operatorname{Fix} T \neq \varnothing$. Moreover, for every $n \in \mathbb{N}$,

$$\begin{aligned}
x_{n+1} &= x_n + \lambda_n \big(\operatorname{prox}_{\gamma f}(2\operatorname{prox}_{\gamma g}(x_n) - x_n) - \operatorname{prox}_{\gamma g}(x_n)\big)\\
&= x_n + \frac{\lambda_n}{2}\big(2\operatorname{prox}_{\gamma f}(2\operatorname{prox}_{\gamma g}(x_n) - x_n) - 2\operatorname{prox}_{\gamma g}(x_n) + x_n - x_n\big)\\
&= x_n + \frac{\lambda_n}{2}\big(2\operatorname{prox}_{\gamma f}(\operatorname{rprox}_{\gamma g}(x_n)) - \operatorname{rprox}_{\gamma g}(x_n) - x_n\big)\\
&= x_n + \frac{\lambda_n}{2}\big(Tx_n - x_n\big).
\end{aligned}$$

$\Rightarrow$ Krasnosel'skii-Mann algorithm with relaxation factors $(\lambda_n/2)_{n\in\mathbb{N}}$.
We deduce that $Tx_n - x_n \to 0$ and $x_n \to \widehat{x} \in \operatorname{Fix} T$.

Optimization algorithms: Douglas-Rachford

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

$$z_n - y_n = \mathrm{prox}_{\gamma f}(2\mathrm{prox}_{\gamma g}(x_n) - x_n) - \mathrm{prox}_{\gamma g}(x_n) = \frac{1}{2}(Tx_n - x_n) \to 0.$$

Moreover, $\mathrm{prox}_{\gamma g}$ being continuous (since nonexpansive), we have $y_n \to \mathrm{prox}_{\gamma g}\widehat{x} \in \mathrm{zer}\,(\partial f + \partial g)$.
Because $z_n - y_n \to 0$, we have: $z_n \to \mathrm{prox}_{\gamma g}\widehat{x} \in \mathrm{zer}\,(\partial f + \partial g)$.
In addition,

$$\begin{aligned} & \varnothing \neq \partial f + \partial g \subset \partial(f+g) \\ \Rightarrow \quad & \widehat{y} = \mathrm{prox}_{\gamma g}(\widehat{x}) \in \mathrm{zer}\,\big(\partial(f+g)\big) = \mathrm{Argmin}(f+g). \end{aligned}$$

Optimization algorithms: Parallel form of Douglas-Rachford

Let H and G be two Hilbert spaces.
Let $g \in \Gamma_0(\mathsf{H})$ and $L \in \mathcal{B}(\mathsf{G}, \mathsf{H})$ be such that $\operatorname{ran} L$ is closed and L^*L is an 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 $x_0 \in \mathsf{H}$, $v_0 = (L^*L)^{-1}L^*x_0$ and

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

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

Optimization algorithms: Parallel form of Douglas-Rachford

Sketch of proof:

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

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

Optimization algorithms: Parallel form of DR

Particular case of Douglas-Rachford algorithm:

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

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

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

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

PPXA+ algorithm

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

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

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

Optimization algorithm: PPXA+

Image restoration : Variational approach

$$\widehat{x} \in \underset{x\in\mathbb{R}^N}{\mathrm{Argmin}} \|Ax - z\|_2^2 + \eta\|[H^* \quad V^*]^* x\|_{2,1} + \iota_C(x) \text{ with } \begin{cases} \eta \in]0,+\infty[\\ H, V \in \mathbb{R}^{N\times N} \\ C = [0,255]^N \end{cases}$$

- PPXA+ with $g_1 = \|A \cdot - z\|_2^2$ and $L_1 = \mathrm{Id}$
  $g_2 = \eta\|\cdot\|_{1,2}$ and $L_2 = [H^* \quad V^*]^*$
  $g_3 = \iota_C$ and $L_3 = \mathrm{Id}$

$$(\forall n \in \mathbb{N}) \quad \begin{cases} y_{n,i} = \mathrm{prox}_{\gamma g_i} x_{n,i}, i \in \{1,2,3\} \\ c_n = (\sum_{i=1}^3 L_i^* L_i)^{-1} \sum_{i=1}^3 L_i^* y_{n,i} \\ x_{n+1,i} = x_{n,i} + \lambda_n\big(L_i(2c_n - v_n) - y_{n,i}\big), \ i \in \{1,2,3\} \\ v_{n+1} = v_n + \lambda_n(c_n - v_n). \end{cases}$$

Optimization algorithm: PPXA+

Image restoration : Variational approach

$$\widehat{x} \in \underset{x \in \mathbb{R}^N}{\text{Argmin}} \|Ax - z\|_2^2 + \eta \|[H^* \quad V^*]^* x\|_{2,1} + \iota_C(x) \text{ with } \begin{cases} \eta \in]0, +\infty[\\ H, V \in \mathbb{R}^{N \times N} \\ C = [0, 255]^N \end{cases}$$

- PPXA+ with $g_1 = \|A \cdot - z\|_2^2$ and $L_1 = \mathrm{Id}$
  $g_2 = \eta \| \cdot \|_{1,2}$ and $L_2 = [H^* \quad V^*]^*$
  $g_3 = \iota_C$ and $L_3 = \mathrm{Id}$

$$(\forall n \in \mathbb{N}) \quad \begin{cases} y_{n,i} = \text{prox}_{\gamma g_i} x_{n,i}, i \in \{1,2,3\} & \rightarrow \quad \text{Closed form} \\ c_n = (\sum_{i=1}^3 L_i^* L_i)^{-1} \sum_{i=1}^3 L_i^* y_{n,i} & \rightarrow \quad \text{Closed form} \\ x_{n+1,i} = x_{n,i} + \lambda_n \big(L_i(2c_n - v_n) - y_{n,i} \big), \; i \in \{1,2,3\} \\ v_{n+1} = v_n + \lambda_n (c_n - v_n). \end{cases}$$

Optimization algorithm: PPXA+

Image restoration : Variational approach

$$\widehat{x} \in \underset{x \in \mathbb{R}^N}{\operatorname{Argmin}} \|Ax - z\|_2^2 + \eta \|[H^* \quad V^*]^* x\|_{2,1} + \iota_C(x) \text{ with } \begin{cases} \eta \in]0, +\infty[\\ H, V \in \mathbb{R}^{N \times N} \\ C = [0, 255]^N \end{cases}$$

- PPXA+

Degraded image z

Restored image $\widehat{x}$ [PPXA – TV]

Optimization algorithm: PPXA+

Image restoration : Variational approach

$$\widehat{x} \in \underset{x \in \mathbb{R}^N}{\text{Argmin}} \|Ax - z\|_2^2 + \eta \|[H^* \quad V^*]^* x\|_{2,1} + \iota_C(x) \text{ with } \begin{cases} \eta \in]0, +\infty[\\ H, V \in \mathbb{R}^{N \times N} \\ C = [0, 255]^N \end{cases}$$

- PPXA+

Degraded image z

Restored image $\widehat{x}$ [DR – DWT]

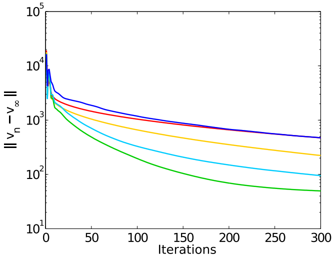
Optimization algorithm: PPXA+

Image restoration : Variational approach

$$\widehat{x} \in \underset{x\in\mathbb{R}^N}{\operatorname{Argmin}} \|Ax - z\|_2^2 + \eta\|[H^* \quad V^*]^* x\|_{2,1} + \iota_C(x) \text{ with } \begin{cases} \eta \in \,]0,+\infty[\\ H, V \in \mathbb{R}^{N\times N} \\ C = [0,255]^N \end{cases}$$

- PPXA+

$$\begin{cases} y_{n,i} = \operatorname{prox}_{\gamma g_i} x_{n,i}, \\ c_n = \left(\sum_{i=1}^3 L_i^* L_i\right)^{-1} \sum_{i=1}^3 L_i^* y_{n,i} \\ x_{n+1,i} = x_{n,i} + \lambda_n\big(L_i(2c_n - v_n) - y_{n,i}\big), \\ v_{n+1} = v_n + \lambda_n(c_n - v_n). \end{cases}$$



$$\gamma \in \{5.10^2, 10^3, 5.10^3, 10^4, 5.10^4\}$$

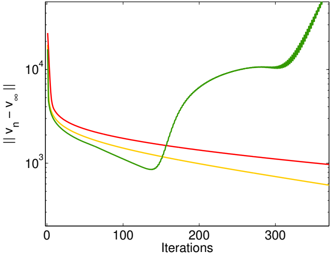
Optimization algorithm: PPXA+

Image restoration : Variational approach

$$\widehat{x} \in \underset{x\in\mathbb{R}^N}{\mathrm{Argmin}} \|Ax - z\|_2^2 + \eta\|[H^* \quad V^*]^* x\|_{2,1} + \iota_C(x) \text{ with } \begin{cases} \eta \in \,]0,+\infty[\\ H, V \in \mathbb{R}^{N\times N} \\ C = [0,255]^N \end{cases}$$

- PPXA+

$$\begin{cases} y_{n,i} = \mathrm{prox}_{\gamma g_i} x_{n,i}, \\ c_n = (\sum_{i=1}^3 L_i^* L_i)^{-1} \sum_{i=1}^3 L_i^* y_{n,i} \\ x_{n+1,i} = x_{n,i} + \lambda_n \big(L_i(2c_n - v_n) - y_{n,i}\big), \\ v_{n+1} = v_n + \lambda_n(c_n - v_n). \end{cases}$$



$\lambda_n \in \{1, 1.8, 2.1\}$

Part 4: Duality

1. General duality concepts
   - Inf-convolution
   - Primal and dual problems
   - Duality theorems
2. Augmented Lagrangian algorithms
3. Primal-dual algorithms
   - FB-based PD algorithm

Duality

Let H be a Hilbert space.
Let $A\colon \mathsf{H} \to 2^{\mathsf{H}}$ and $B\colon \mathsf{H} \to 2^{\mathsf{H}}$.
The parallel sum of A and B is

$$A \,\Box\, B = (A^{-1} + B^{-1})^{-1}.$$

- If A and B are monotone, then $A \,\Box\, B$ is monotone.
- $A \,\Box\, N_{\{0\}} = A$ where $N_{\{0\}} = \partial\iota_{\{0\}}$ and $\operatorname{gra} N_{\{0\}} = \{0\} \times \mathsf{H}$.
- $A \,\Box\, B = B \,\Box\, A$.

Duality

Primal problem

Let H and G be two Hilbert spaces. Let $A\colon \mathsf{H} \to 2^{\mathsf{H}}$, $C\colon \mathsf{H} \to 2^{\mathsf{H}}$, $B\colon \mathsf{G} \to 2^{\mathsf{G}}$ and $D\colon \mathsf{G} \to 2^{\mathsf{G}}$ be monotone operators. Let $L \in \mathcal{B}(\mathsf{H}, \mathsf{G})$. Find $\widehat{x} \in \mathsf{H}$ such that

$$0 \in A\widehat{x} + C\widehat{x} + L^*(B \,\square\, D)L\widehat{x}\,.$$

Dual problem

Let H and G be two Hilbert spaces. Let $A\colon \mathsf{H} \to 2^{\mathsf{H}}$, $C\colon \mathsf{H} \to 2^{\mathsf{H}}$, $B\colon \mathsf{G} \to 2^{\mathsf{G}}$ and $D\colon \mathsf{G} \to 2^{\mathsf{G}}$ be monotone operators. Let $L \in \mathcal{B}(\mathsf{H}, \mathsf{G})$. Find $\widehat{v} \in \mathsf{G}$ such that

$$0 \in -L(A^{-1} \,\square\, C^{-1})(-L^*\widehat{v}) + B^{-1}\widehat{v} + D^{-1}\widehat{v}\,.$$

Duality theorem

Let $\widehat{x} \in \mathsf{H}$ and $\widehat{v} \in \mathsf{G}$. $(\widehat{x}, \widehat{v})$ is a **Kuhn-Tucker point** if

$$\begin{cases} -L^*\widehat{v} \in (A+C)\widehat{x} \\ L\widehat{x} \in (B^{-1}+D^{-1})\widehat{v}. \end{cases}$$

If $\widehat{x} \in \mathsf{H}$ is a solution to the primal problem, then there exists a solution $\widehat{v}$ to the dual problem such that $(\widehat{x}, \widehat{v})$ is a Kuhn-Tucker point.

Proof:

$$0 \in A\widehat{x}+C\widehat{x}+L^*(B \,\square\, D)L\widehat{x} \quad \Leftrightarrow (\exists \widehat{v} \in \mathsf{G}) \quad \begin{cases} 0 \in A\widehat{x}+C\widehat{x}+L^*\widehat{v} \\ \widehat{v} \in (B \,\square\, D)L\widehat{x} \end{cases}$$

$$\Leftrightarrow (\exists \widehat{v} \in \mathsf{G}) \quad \begin{cases} -L^*\widehat{v} \in A\widehat{x}+C\widehat{x} \\ L\widehat{x} \in (B \,\square\, D)^{-1}\widehat{v} = (B^{-1}+D^{-1})\widehat{v} \end{cases}$$

$$\Rightarrow (\exists \widehat{v} \in \mathsf{G}) \quad \begin{cases} \widehat{x} \in (A+C)^{-1}(-L^*\widehat{v}) \\ 0 \in -L(A+C)^{-1}(-L^*\widehat{v})+B^{-1}\widehat{v}+D^{-1}\widehat{v}. \end{cases}$$

Duality theorem

Let $\widehat{x} \in$ H and $\widehat{v} \in$ G. $(\widehat{x}, \widehat{v})$ is a **Kuhn-Tucker point** if

$$\begin{cases} -L^* \widehat{v} \in (A + C)\widehat{x} \\ \quad L\widehat{x} \in (B^{-1} + D^{-1})\widehat{v}. \end{cases}$$

If $\widehat{v} \in$ G is a solution to the dual problem, then there exists a solution $\widehat{x}$ to the primal problem such that $(\widehat{x}, \widehat{v})$ is a Kuhn-Tucker point.

Duality theorem

Let $\widehat{x} \in$ H and $\widehat{v} \in$ G. $(\widehat{x}, \widehat{v})$ is a Kuhn-Tucker point if

$$\begin{cases} -L^*\widehat{v} \in (A+C)\widehat{x} \\ \quad L\widehat{x} \in (B^{-1}+D^{-1})\widehat{v}. \end{cases}$$

If $(\widehat{x}, \widehat{v})$ is a Kuhn-Tucker point, then $\widehat{x}$ is a solution to the primal problem and $\widehat{v}$ is a solution to the dual problem.

Inf-convolution

Let H be a Hilbert space.
Let $f\colon \mathsf{H} \to \,]-\infty, +\infty]$ and $g\colon \mathsf{H} \to \,]-\infty, +\infty]$.
The inf-convolution of f and g is

$$f \,\square\, g\colon \mathsf{H} \to [-\infty, +\infty]\colon x \mapsto \inf_{y \in \mathsf{H}} f(y) + g(x-y)$$

Inf-convolution

Let H be a Hilbert space.
Let $f\colon \mathsf{H} \to \left]-\infty,+\infty\right]$ and $g\colon \mathsf{H} \to \left]-\infty,+\infty\right]$.
The inf-convolution of f and g is

$$f \,\square\, g\colon \mathsf{H} \to [-\infty,+\infty]\colon x \mapsto \inf_{\substack{(u,v)\in\mathsf{H}^2 \\ u+v=x}} f(u) + g(v)$$

Inf-convolution

Let H be a Hilbert space.
Let $f\colon \mathsf{H} \to \left]-\infty,+\infty\right]$ and $g\colon \mathsf{H} \to \left]-\infty,+\infty\right]$.
The inf-convolution of f and g is

$$f \,\square\, g\colon \mathsf{H} \to [-\infty,+\infty]\colon x \mapsto \inf_{\substack{(u,v)\in\mathsf{H}^2\\ u+v=x}} f(u)+g(v)$$

- $f \,\square\, \iota_{\{0\}} = f$
- $f \,\square\, g = g \,\square\, f$
- $\operatorname{dom}(f \,\square\, g) = \operatorname{dom} f + \operatorname{dom} g$
- ${}^{\gamma}f = f \,\square\, \frac{1}{2\gamma}\|\cdot\|^2$, $\gamma \in \left]0,+\infty\right[$.

Inf-convolution

Let H be a Hilbert space.
Let $f\colon \mathsf{H} \to \left]-\infty,+\infty\right]$ and $g\colon \mathsf{H} \to \left]-\infty,+\infty\right]$.
The inf-convolution of f and g is

$$f \,\Box\, g\colon \mathsf{H} \to [-\infty,+\infty]\colon x \mapsto \inf_{y\in\mathsf{H}} f(y) + g(x-y)$$

If $f\colon \mathsf{H} \to \left]-\infty,+\infty\right]$ and $g\colon \mathsf{H} \to \left]-\infty,+\infty\right]$ are convex,
then $f \,\Box\, g$ is convex.

Inf-convolution

	conjugate		Fourier transform (H finite dimensional)	
Property	$h(x)$	$h^*(u)$	$h(x)$	$\widehat{h}(\nu)$
inf-convolution /convolution	$(f \,\Box\, g)(x)$ $= \inf_{y\in \mathsf{H}} f(y) + g(x-y)$	$f^*(u) + g^*(u)$	$(f \star g)(x)$ $= \int_{\mathsf{H}} f(y)g(x-y)dy$	$\widehat{f}(\nu)\widehat{g}(\nu)$
sum/product	$f(x) + g(x)$ $f \in \Gamma_0(\mathsf{H})$ $g \in \Gamma_0(\mathsf{H})$ $\mathrm{dom}\, f \cap \mathrm{dom}\, g \neq \varnothing$	$(f^* \,\Box\, g^*)(u)$	$f(x)g(x)$	$(\widehat{f} \star \widehat{g})(\nu)$

Inf-convolution

Let H be a Hilbert space.
Let $f \in \Gamma_0(\mathsf{H})$ and $g \in \Gamma_0(\mathsf{H})$.
If $\operatorname{dom} f^* \cap \operatorname{int}(\operatorname{dom} g^*) \neq \varnothing$, then

$$\partial(f \underbrace{\,\square\,}_{\text{inf-convolution}} g) = \partial f \underbrace{\,\square\,}_{\text{parallel sum}} \partial g.$$

Inf-convolution

Let H be a Hilbert space.
Let $f \in \Gamma_0(\mathrm{H})$ and $g \in \Gamma_0(\mathrm{H})$.
If $\mathrm{dom}\, f^* \cap \mathrm{int}(\mathrm{dom}\, g^*) \neq \varnothing$, then

$$\partial(f \underbrace{\,\square\,}_{\text{inf-convolution}} g) = \partial f \underbrace{\,\square\,}_{\text{parallel sum}} \partial g.$$

<u>Proof</u>:

$$\begin{aligned}
\partial(f \,\square\, g) &= \partial(f^* + g^*)^* \\
&= \big(\partial(f^* + g^*)\big)^{-1} \\
&= \big(\partial f^* + \partial g^*\big)^{-1} \\
&= \big((\partial f)^{-1} + (\partial g)^{-1}\big)^{-1} \\
&= \partial f \,\square\, \partial g.
\end{aligned}$$

Fenchel-Rockafellar duality

Let H and G be two Hilbert spaces and let

- $f \in \Gamma_0(\mathsf{H})$, $g \in \Gamma_0(\mathsf{G})$
- $h\colon \mathsf{H} \to \mathbb{R}$ convex, μ-Lipschitz differentiable function with $\mu \in]0,+\infty[$
- $l \in \Gamma_0(\mathsf{G})$ ν-strongly convex with $\nu \in]0,+\infty[$ $\Leftrightarrow l^* \in \Gamma_0(\mathsf{G})$ ν-Lipschitz differentiable
- $L\colon \mathsf{H} \to \mathsf{G}$ linear and bounded.

Primal problem

We want to

$$\underset{x\in\mathsf{H}}{\text{minimize}} \;\; f(x) + h(x) + (g \,\square\, l)(Lx).$$

Dual problem

We want to

$$\underset{v\in\mathsf{G}}{\text{minimize}} \;\; (f^* \,\square\, h^*)(-L^*v) + g^*(v) + l^*(v).$$

Fenchel-Rockafellar duality

Primal problem

We want to

$$\underset{x\in\mathsf{H}}{\text{minimize}} \;\; f(x) + h(x) + (g \,\square\, l)(Lx).$$

Dual problem

We want to

$$\underset{v\in\mathsf{G}}{\text{minimize}} \;\; (f^* \,\square\, h^*)(-L^*v) + g^*(v) + l^*(v).$$

Strong duality

If $\operatorname{int}(\operatorname{dom} g + \operatorname{dom} l) \cap L(\operatorname{dom} f) \neq \varnothing$ or $(\operatorname{dom} g + \operatorname{dom} l) \cap \operatorname{int}(L(\operatorname{dom} f)) \neq \varnothing$, then

$$\inf_{x\in\mathsf{H}} f(x) + h(x) + (g \,\square\, l)(Lx) = -\min_{v\in\mathsf{G}} (f^* \,\square\, h^*)(-L^*v) + g^*(v) + l^*(v)$$

Fenchel-Rockafellar duality

Primal problem

We want to

$$\underset{x \in \mathsf{H}}{\text{minimize}} \;\; f(x) + h(x) + (g \,\square\, l)(Lx).$$

Dual problem

We want to

$$\underset{v \in \mathsf{G}}{\text{minimize}} \;\; (f^* \,\square\, h^*)(-L^* v) + g^*(v) + l^*(v).$$

If $(\widehat{x}, \widehat{v}) \in \mathsf{H} \times \mathsf{G}$ is such that

$$-L^*\widehat{v} - \nabla h(\widehat{x}) \in \partial f(\widehat{x}) \;\; \text{and} \;\; L\widehat{x} - \nabla l^*(\widehat{v}) \in \partial g^*(\widehat{v}),$$

then $(\widehat{x}, \widehat{v})$ is a **Kuhn-Tucker point**.

- If $(\widehat{x}, \widehat{v})$ is a **Kuhn-Tucker point** then $\widehat{x}$ (resp. $\widehat{v}$) is a solution to the primal (resp. dual) problem.

Fenchel-Rockafellar duality

Primal problem

We want to

$$\underset{x \in \mathsf{H}}{\text{minimize}} \;\; f(x) + h(x) + (g \,\Box\, l)(Lx).$$

Dual problem

We want to

$$\underset{v \in \mathsf{G}}{\text{minimize}} \;\; (f^* \,\Box\, h^*)(-L^* v) + g^*(v) + l^*(v).$$

If $(\widehat{x}, \widehat{v}) \in \mathsf{H} \times \mathsf{G}$ is such that

$$-L^*\widehat{v} - \nabla h(\widehat{x}) \in \partial f(\widehat{x}) \;\; \text{and} \;\; L\widehat{x} - \nabla l^*(\widehat{v}) \in \partial g^*(\widehat{v}),$$

then $(\widehat{x}, \widehat{v})$ is a **Kuhn-Tucker point**.

- If $(\widehat{x}, \widehat{v})$ is a **Kuhn-Tucker point** then $\widehat{x}$ (resp. $\widehat{v}$) is a solution to the primal (resp. dual) problem.

Proof: deduced from duality theorem for monotone operators by setting $A = \partial f$, $B = \partial g$, $C = \nabla h$, and $D = \partial l$

Augmented Lagrangian method

- Augmented Lagrange function

  Let $\gamma \in \,]0,+\infty[$, define, for every $(x,y,z) \in \mathsf{H} \times \mathsf{G}^2$,

$$\mathcal{L}(x,y,z) = f(x) + h(x) + (g \,\Box\, l)(y) + \gamma \left\langle z \mid Lx - y \right\rangle$$

  The Lagrange multiplier is $v = \gamma z$.

Augmented Lagrangian method

- **Augmented Lagrange function**

  Let $\gamma \in]0, +\infty[$, define, for every $(x, y, z) \in \mathsf{H} \times \mathsf{G}^2$,

$$\widetilde{\mathcal{L}}(x, y, z) = f(x) + h(x) + (g \,\square\, l)(y) + \gamma \langle z \mid Lx - y \rangle + \frac{\gamma}{2}\|Lx - y\|^2$$

  The Lagrange multiplier is $v = \gamma z$.

Let $(\widehat{x}, \widehat{y}, \widehat{z}) \in \mathsf{H} \times \mathsf{G}^2$.
Assume that $\operatorname{int}(\operatorname{dom} g + \operatorname{dom} l) \cap L(\operatorname{dom} f) \neq \varnothing$
or $(\operatorname{dom} g + \operatorname{dom} l) \cap \operatorname{int}\big(L(\operatorname{dom} f)\big) \neq \varnothing$.

$(\widehat{x}, \widehat{y}, \widehat{z})$ is a saddle point of the augmented Lagrange function

$$\Updownarrow$$

$(\widehat{x}, \gamma\widehat{z})$ is a Kuhn-Tucker point and $\widehat{y} = L\widehat{x}$.

Augmented Lagrangian method

- **Augmented Lagrange function**

Let $\gamma \in]0,+\infty[$, define, for every $(x,y,z) \in \mathsf{H} \times \mathsf{G}^2$,

$$\widetilde{\mathcal{L}}(x,y,z) = f(x) + h(x) + (g \,\square\, l)(y) + \gamma \langle z \mid Lx - y \rangle + \frac{\gamma}{2}\|Lx - y\|^2$$

The Lagrange multiplier is $v = \gamma z$.

Algorithm for searching $(\widehat{x}, \widehat{y}, \widehat{z})$:

$$(\forall n \in \mathbb{N}) \quad \begin{cases} x_n = \underset{x \in \mathsf{H}}{\operatorname{argmin}} \; \widetilde{\mathcal{L}}(x, y_n, z_n) \\ y_{n+1} = \underset{y \in \mathsf{G}}{\operatorname{argmin}} \; \widetilde{\mathcal{L}}(x_n, y, z_n) \\ z_{n+1} \text{ tel que } \widetilde{\mathcal{L}}(x_n, y_{n+1}, z_{n+1}) \geqslant \widetilde{\mathcal{L}}(x_n, y_{n+1}, z_n). \end{cases}$$

Augmented Lagrangian method

- **Augmented Lagrange function**

  Let $\gamma \in \,]0,+\infty[$, define, for every $(x,y,z) \in \mathsf{H} \times \mathsf{G}^2$,

$$\widetilde{\mathcal{L}}(x,y,z) = f(x) + h(x) + (g \,\square\, l)(y) + \gamma \langle z \mid Lx - y \rangle + \frac{\gamma}{2}\|Lx - y\|^2$$

  The Lagrange multiplier is $v = \gamma z$.

Algorithme de recherche de $(\widehat{x}, \widehat{y}, \widehat{z})$:

$$(\forall n \in \mathbb{N}) \qquad \begin{cases} x_n = \underset{x \in \mathsf{H}}{\operatorname{argmin}} \;\; \widetilde{\mathcal{L}}(x, y_n, z_n) \\ y_{n+1} = \underset{y \in \mathsf{G}}{\operatorname{argmin}} \;\; \widetilde{\mathcal{L}}(x_n, y, z_n) \\ z_{n+1} = z_n + \frac{1}{\gamma} \nabla_z \widetilde{\mathcal{L}}(x_n, y_{n+1}, z_n). \end{cases}$$

Augmented Lagrangian method

ADMM (*Alternating-direction method of multipliers*)

Let L be such that L^*L is an isomorphism and let $\gamma \in]0, +\infty[$.
Assume that $\operatorname{int}(\operatorname{dom} g + \operatorname{dom} l) \cap L(\operatorname{dom} f) \neq \varnothing$
or $(\operatorname{dom} g + \operatorname{dom} l) \cap \operatorname{int}\big(L(\operatorname{dom} f)\big) \neq \varnothing$, and that $\operatorname{Argmin}(f + h + (g \,\square\, l) \circ L) \neq \varnothing$.
Let $(y_0, z_0) \in \mathsf{G}^2$ and

$$(\forall n \in \mathbb{N}) \quad \begin{cases} x_n = \underset{x \in \mathsf{H}}{\operatorname{argmin}} \ \frac{1}{2} \|Lx - y_n + z_n\|^2 + \frac{1}{\gamma}(f(x) + h(x)) \\ s_n = Lx_n \\ y_{n+1} = \operatorname{prox}_{\frac{1}{\gamma} g \,\square\, l}(z_n + s_n) \\ z_{n+1} = z_n + s_n - y_{n+1}. \end{cases}$$

We have:

- $x_n \rightharpoonup \widehat{x}$ where $\widehat{x}$ is a solution to the primal problem
- $\gamma z_n \rightharpoonup \widehat{v}$ where $\widehat{v}$ is a solution to the dual problem.

Limitations:

- Computation of x_n at iteration $n \in \mathbb{N}$ may be complicated.
- Convergence requires L^*L to be invertible.
- The smoothness of h and l^* is not exploited.

Augmented Lagrangian method

ADMM algorithm

$\equiv$ Douglas-Rachford for the dual problem

Let H and G be two Hilbert spaces. Let $f \in \Gamma_0(\mathsf{H})$ and $g \in \Gamma_0(\mathsf{G})$.
Let $L \in \mathcal{B}(\mathsf{H}, \mathsf{G})$ such that L^*L is an isomorphism and let $\gamma \in]0, +\infty[$.
The Douglas-Rachford iterations to minimize $f^* \circ (-L^*) + g^*$ are

$$(\forall n \in \mathbb{N}) \qquad \begin{cases} v_n = \operatorname{prox}_{\gamma g^*} u_n \\ w_n = \operatorname{prox}_{\gamma f^* \circ (-L^*)}(2v_n - u_n) \\ u_{n+1} = u_n + w_n - v_n \end{cases}$$

Augmented Lagrangian method

ADMM algorithm

$\equiv$ Douglas-Rachford for the dual problem

Let H and G be two Hilbert spaces. Let $f \in \Gamma_0(\mathsf{H})$ and $g \in \Gamma_0(\mathsf{G})$.
Let $L \in \mathcal{B}(\mathsf{H}, \mathsf{G})$ such that L^*L is an isomorphism and let $\gamma \in]0, +\infty[$.
The Douglas-Rachford iterations to minimize $f^* \circ (-L^*) + g^*$ are

$$(\forall n \in \mathbb{N}) \qquad \begin{cases} v_n = u_n - \gamma \mathrm{prox}_{\gamma^{-1}g}(\gamma^{-1}u_n) \\ 2v_n - u_n - w_n \in \gamma\partial\big(f^* \circ (-L^*)\big)w_n \\ u_{n+1} = u_n + w_n - v_n \end{cases}$$

Augmented Lagrangian method

ADMM algorithm

$\equiv$ Douglas-Rachford for the dual problem

Let H and G be two Hilbert spaces. Let $f \in \Gamma_0(\mathsf{H})$ and $g \in \Gamma_0(\mathsf{G})$.
Let $L \in \mathcal{B}(\mathsf{H}, \mathsf{G})$ such that L^*L is an isomorphism and let $\gamma \in]0, +\infty[$.
The Douglas-Rachford iterations to minimize $f^* \circ (-L^*) + g^*$ are

$$(\forall n \in \mathbb{N}) \qquad \begin{cases} v_n = u_n - \gamma \mathrm{prox}_{\gamma^{-1}g}(\gamma^{-1}u_n) \\ 2v_n - u_n - w_n \in -\gamma L \circ \partial f^* \circ (-L^*) w_n \\ u_{n+1} = u_n + w_n - v_n \end{cases}$$

Augmented Lagrangian method

ADMM algorithm

$\equiv$ Douglas-Rachford for the dual problem

Let H and G be two Hilbert spaces. Let $f \in \Gamma_0(\mathsf{H})$ and $g \in \Gamma_0(\mathsf{G})$.
Let $L \in \mathcal{B}(\mathsf{H},\mathsf{G})$ such that L^*L is an isomorphism and let $\gamma \in]0,+\infty[$.
The Douglas-Rachford iterations to minimize $f^* \circ (-L^*) + g^*$ are

$$(\forall n \in \mathbb{N}) \qquad \begin{cases} v_n = u_n - \gamma \mathrm{prox}_{\gamma^{-1}g}(\gamma^{-1}u_n) \\ 2v_n - u_n - w_n \in -\gamma L \circ \underbrace{\partial f^* \circ (-L^*) w_n}_{x_n \in} \\ u_{n+1} = u_n + w_n - v_n \end{cases}$$

Augmented Lagrangian method

ADMM algorithm

$\equiv$ Douglas-Rachford for the dual problem

Let H and G be two Hilbert spaces. Let $f \in \Gamma_0(\mathsf{H})$ and $g \in \Gamma_0(\mathsf{G})$.
Let $L \in \mathcal{B}(\mathsf{H}, \mathsf{G})$ such that L^*L is an isomorphism and let $\gamma \in]0, +\infty[$.
The Douglas-Rachford iterations to minimize $f^* \circ (-L^*) + g^*$ are

$$(\forall n \in \mathbb{N}) \qquad \begin{cases} v_n = u_n - \gamma \mathrm{prox}_{\gamma^{-1} g}(\gamma^{-1} u_n) \\ 2v_n - u_n - w_n = -\gamma L x_n \\ x_n \in \partial f^*(-L^* w_n) \\ u_{n+1} = u_n + w_n - v_n \end{cases}$$

Augmented Lagrangian method

ADMM algorithm

$\equiv$ Douglas-Rachford for the dual problem

Let H and G be two Hilbert spaces. Let $f \in \Gamma_0(\mathsf{H})$ and $g \in \Gamma_0(\mathsf{G})$.
Let $L \in \mathcal{B}(\mathsf{H}, \mathsf{G})$ such that L^*L is an isomorphism and let $\gamma \in \,]0, +\infty[$.
The Douglas-Rachford iterations to minimize $f^* \circ (-L^*) + g^*$ are

$$(\forall n \in \mathbb{N}) \qquad \begin{cases} v_n = u_n - \gamma \mathrm{prox}_{\gamma^{-1} g}(\gamma^{-1} u_n) \\ 2v_n - u_n - w_n = -\gamma L x_n \\ -L^* w_n \in \partial f(x_n) \\ u_{n+1} = u_n + w_n - v_n \end{cases}$$

Augmented Lagrangian method

ADMM algorithm

$\equiv$ Douglas-Rachford for the dual problem

Let H and G be two Hilbert spaces. Let $f \in \Gamma_0(\mathsf{H})$ and $g \in \Gamma_0(\mathsf{G})$.
Let $L \in \mathcal{B}(\mathsf{H}, \mathsf{G})$ such that L^*L is an isomorphism and let $\gamma \in]0, +\infty[$.
The Douglas-Rachford iterations to minimize $f^* \circ (-L^*) + g^*$ are

$$(\forall n \in \mathbb{N}) \qquad \begin{cases} v_n = u_n - \gamma \mathrm{prox}_{\gamma^{-1} g}(\gamma^{-1} u_n) \\ L^*(u_n - 2v_n - \gamma L x_n) \in \partial f(x_n) \\ u_{n+1} = v_n + \gamma L x_n \end{cases}$$

using $y_n = \gamma^{-1}(u_n - v_n)$ and $z_n = \gamma^{-1} v_n$

Augmented Lagrangian method

ADMM algorithm

$\equiv$ Douglas-Rachford for the dual problem

Let H and G be two Hilbert spaces. Let $f \in \Gamma_0(\mathsf{H})$ and $g \in \Gamma_0(\mathsf{G})$.
Let $L \in \mathcal{B}(\mathsf{H}, \mathsf{G})$ such that L^*L is an isomorphism and let $\gamma \in \,]0, +\infty[$.
The Douglas-Rachford iterations to minimize $f^* \circ (-L^*) + g^*$ are

$$(\forall n \in \mathbb{N}) \qquad \begin{cases} y_n = \mathrm{prox}_{\gamma^{-1}g}(\gamma^{-1}u_n) \\ z_n = \gamma^{-1}u_n - y_n \\ L^*(y_n - z_n - Lx_n) \in \frac{1}{\gamma}\partial f(x_n) \\ u_{n+1} = \gamma z_n + \gamma L x_n \end{cases}$$

Augmented Lagrangian method

ADMM algorithm

$\equiv$ Douglas-Rachford for the dual problem

Let H and G be two Hilbert spaces. Let $f \in \Gamma_0(\mathsf{H})$ and $g \in \Gamma_0(\mathsf{G})$.
Let $L \in \mathcal{B}(\mathsf{H}, \mathsf{G})$ such that L^*L is an isomorphism and let $\gamma \in]0, +\infty[$.
The Douglas-Rachford iterations to minimize $f^* \circ (-L^*) + g^*$ are

$$(\forall n \in \mathbb{N}) \qquad \begin{cases} y_n = \operatorname{prox}_{\gamma^{-1} g}(\gamma^{-1} u_n) \\ z_n = \gamma^{-1} u_n - y_n \\ x_n = \underset{x \in \mathsf{H}}{\operatorname{argmin}} \ \frac{1}{2} \|Lx - y_n + z_n\|^2 + \frac{1}{\gamma} f(x) \\ u_{n+1} = \gamma z_n + \gamma L x_n \end{cases}$$

Augmented Lagrangian method

ADMM algorithm

$\equiv$ **Douglas-Rachford for the dual problem**

Let H and G be two Hilbert spaces. Let $f \in \Gamma_0(\mathsf{H})$ and $g \in \Gamma_0(\mathsf{G})$.
Let $L \in \mathcal{B}(\mathsf{H}, \mathsf{G})$ such that L^*L is an isomorphism and let $\gamma \in \,]0, +\infty[$.
The Douglas-Rachford iterations to minimize $f^* \circ (-L^*) + g^*$ are

$$(\forall n \in \mathbb{N}) \qquad \begin{cases} x_n = \underset{x \in \mathsf{H}}{\operatorname{argmin}} \;\; \frac{1}{2} \|Lx - y_n + z_n\|^2 + \frac{1}{\gamma} f(x) \\ s_n = Lx_n \\ y_{n+1} = \operatorname{prox}_{\frac{g}{\gamma}} (z_n + s_n) \\ z_{n+1} = z_n + s_n - y_{n+1}. \end{cases}$$

Primal-dual approach

- The optimization problem is reformulated as finding

$$\inf_{x \in \mathsf{H}} f(x) + h(x) + \sup_{v \in \mathsf{G}} (\langle v \mid Lx \rangle - g^*(v) - l^*(v)).$$

Primal-dual approach

- The optimization problem is reformulated as finding

$$\inf_{x \in \mathsf{H}} \sup_{v \in \mathsf{G}} f(x) + h(x) + \langle v \mid Lx \rangle - g^*(v) - l^*(v).$$

Primal-dual approach

- The optimization problem is reformulated as finding

$$\inf_{x\in\mathsf{H}} \sup_{v\in\mathsf{G}} f(x) + h(x) + \langle v \mid Lx\rangle - g^*(v) - l^*(v).$$

- Arrow-Hurwitz method : Let $(\tau_n)_{n\in\mathbb{N}}$ and $(\sigma_n)_{n\in\mathbb{N}}$ be sequences in $]0,+\infty[$.

$$(\forall n\in\mathbb{N}) \qquad \begin{cases} t_n \in \partial f(x_n) \\ x_{n+1} = x_n - \tau_n(t_n + \nabla h(x_n) + L^* v_n) \\ s_n \in \partial g^*(v_n) \\ v_{n+1} = v_n - \sigma_n(s_n + \nabla l^*(v_n) - L x_{n+1}) \end{cases}$$

$\rightsquigarrow$ requires stringent conditions on the choice of the step-sizes (e.g. decaying to zero)
... but it can be modified.

Primal-dual approach

- First modification: Use implicit updates

$$
(\forall n \in \mathbb{N}) \qquad \begin{cases} t_n \in \partial f(x_{n+1}) \\ x_{n+1} = x_n - \tau_n(t_n + \nabla h(x_n) + L^* v_n) \\ s_n \in \partial g^*(v_{n+1}) \\ v_{n+1} = v_n - \sigma_n(s_n + \nabla l^*(v_n) - L x_{n+1}) \end{cases}
$$

$$
\Leftrightarrow \quad \begin{cases} 0 \in x_{n+1} - x_n + \tau_n(\nabla h(x_n) + L^* v_n) + \tau_n \partial f(x_{n+1}) \\ 0 \in v_{n+1} - v_n - \sigma_n(L x_{n+1} - \nabla l^*(v_n)) + \sigma_n \partial g^*(v_{n+1}) \end{cases}
$$

$$
\Leftrightarrow \quad \begin{cases} x_{n+1} = \mathrm{prox}_{\tau_n f}\big(x_n - \tau_n(L^* v_n + \nabla h(x_n))\big) \\ v_{n+1} = \mathrm{prox}_{\sigma_n g^*}\big(v_n + \sigma_n(L x_{n+1} - \nabla l^*(v_n))\big) \end{cases}
$$

$\rightsquigarrow$ still does not converge for constant values of the step-size.

Primal-dual methods

- Second modification: Use the approximation $x_{n+1} \simeq 2x_{n+1} - x_n$

$$(\forall n \in \mathbb{N}) \quad \begin{cases} x_{n+1} = \operatorname{prox}_{\tau_n f}\big(x_n - \tau_n(L^* v_n + \nabla h(x_n))\big) \\ v_{n+1} = \operatorname{prox}_{\sigma_n g^*}\big(v_n + \sigma_n\big(L(2x_{n+1} - x_n) - \nabla l^*(v_n)\big)\big). \end{cases}$$

Primal-dual optimization algorithm

Modified PD algorithm

Let $\tau \in]0,+\infty[$ and $\sigma \in]0,+\infty[$ be such that

$$1-\sqrt{\tau\sigma}\|L\| > \max\{\mu\tau,\nu\sigma\}/2.$$

Assume that $\operatorname{zer}\big(\partial f + \nabla h + L^*\partial(g\,\square\, l)L\big) \neq \varnothing$.
Let $x_0 \in \mathsf{H}$, $v_0 \in \mathsf{G}$, and

$$(\forall n \in \mathbb{N}) \qquad \begin{cases} x_{n+1} = \operatorname{prox}_{\tau f}\big(x_n - \tau\big(L^* v_n + \nabla h(x_n)\big)\big) \\ v_{n+1} = \operatorname{prox}_{\sigma g^*}\big(v_n + \sigma(L(2x_{n+1}-x_n)) - \nabla l^*(v_n)\big)\big). \end{cases}$$

We have:

- $x_n \rightharpoonup \widehat{x}$ where $\widehat{x}$ is a solution to the primal problem
- $v_n \rightharpoonup \widehat{v}$ where $\widehat{v}$ is a solution to the dual problem.

Remark: If $l = \iota_{\{0\}}$, a more general convergence condition is

$$\tau^{-1} - \sigma\|L\|^2 \geqslant \mu/2.$$

Primal-dual optimization algorithm

Modified PD algorithm (symmetric form)

Let $\tau \in]0,+\infty[$ and $\sigma \in]0,+\infty[$ be such that

$$1-\sqrt{\tau\sigma}\|L\| > \max\{\mu\tau,\nu\sigma\}/2.$$

Assume that $\operatorname{zer}\big(\partial f + \nabla h + L^*\partial(g\,\square\, l)L\big) \neq \varnothing$.
Let $x_0 \in \mathsf{H}$, $v_0 \in \mathsf{G}$, and

$$(\forall n \in \mathbb{N}) \qquad \begin{cases} v_{n+1} = \operatorname{prox}_{\sigma g^*}\big(v_n + \sigma\big(Lx_n - \nabla l^*(v_n)\big)\big) \\ x_{n+1} = \operatorname{prox}_{\tau f}\big(x_n - \tau(L^*(2v_{n+1}-v_n) + \nabla h(x_n))\big). \end{cases}$$

We have:

- $x_n \rightharpoonup \widehat{x}$ where $\widehat{x}$ is a solution to the primal problem
- $v_n \rightharpoonup \widehat{v}$ where $\widehat{v}$ is a solution to the dual problem.

Proximal primal-dual algorithm

Advantages:

- No linear operator inversion.
- Use of proximable or/and differentiable functions.
- Special cases: Forward-Backward and Douglas-Rachford algorithms.
- Possible use of preconditioning linear operators.

Proximal primal-dual algorithm

Advantages:

- No linear operator inversion.
- Use of proximable or/and differentiable functions.
- Special cases: Forward-Backward and Douglas-Rachford algorithms.
- Possible use of preconditioning linear operators.

Bibliographical remarks:

- Methods based on Forward-Backward iteration
  - type I: [Vu - 2013][Condat - 2013]
    (extensions of [Esser *et al.* - 2010][Chambolle and Pock - 2011]) [Esser et al.,2010][Chambolle,Pock,2011])
  - type II: [Combettes *et al.* - 2014]
    (extensions of [Loris and Verhoeven - 2011][Chen *et al.* - 2014])
- Methods based on Forward-Backward-Forward iteration
  [Combettes and Pesquet - 2012] [Boţ and Hendrich,2014] [Combettes,Pesquet,2012]
- Projection based methods
  [Alotaibi *et al.* - 2013]
- ...