

Convex nonsmooth optimization

Part II: Proximity operator

Barbara Pascal

LS2N, CNRS, Centrale Nantes, Nantes University, Nantes, France
barbara.pascal@cnrs.fr

<http://bpascal-fr.github.io>

Collaboration

This course is a direct adaptation of the course built by Jean-Christophe Pesquet (CentraleSupélec) and Nelly Pustelnik (LPENSL)



Motivation

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

Find

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

► Gradient descent algorithm

Set $\gamma \in]0, +\infty[$ and $x_0 \in \mathcal{H}$.

For $n = 0, 1 \dots$

$$\lfloor x_{n+1} = x_n - \gamma \nabla f(x_n).$$

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

Motivation

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

Find

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

► Alternative algorithm

Set $\gamma \in]0, +\infty[$ and $x_0 \in \mathcal{H}$.

For $n = 0, 1 \dots$

$$\lfloor x_{n+1} = x_n - \gamma \nabla f(x_{n+1}).$$

Questions:

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

Proximity operator: definition

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

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

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

Proximity operator: definition

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

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

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

- ▶ The **proximity operator** of f is

$$\text{prox}_{\gamma f}: \mathcal{H} \rightarrow \mathcal{H}: x \mapsto \operatorname{argmin}_{y \in \mathcal{H}} f(y) + \frac{1}{2\gamma} \|y - x\|^2.$$

Proximity operator: definition

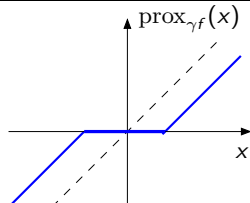
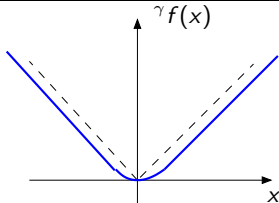
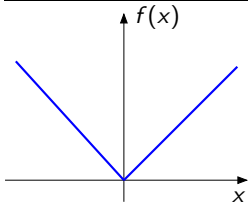
Let \mathcal{H} be a Hilbert space. Let $f \in \Gamma_0(\mathcal{H})$.

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

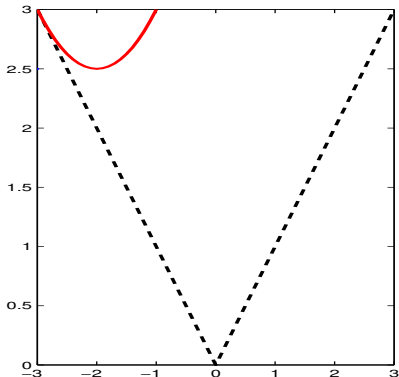
$$\gamma f: \mathcal{H} \rightarrow \mathbb{R}: x \mapsto \inf_{y \in \mathcal{H}} f(y) + \frac{1}{2\gamma} \|y - x\|^2.$$

- ▶ The **proximity operator** of f is

$$\text{prox}_{\gamma f}: \mathcal{H} \rightarrow \mathcal{H}: x \mapsto \operatorname{argmin}_{y \in \mathcal{H}} f(y) + \frac{1}{2\gamma} \|y - x\|^2.$$

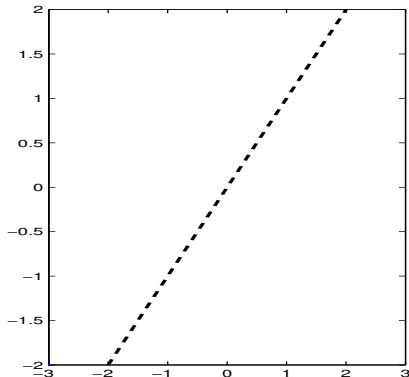


Proximity operator: definition



Moreau envelope

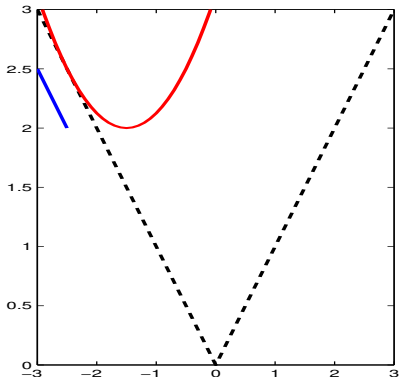
$$\gamma f(x) = \inf_{y \in \mathcal{H}} f(y) + \frac{1}{2\gamma} \|y - x\|^2$$



Proximity operator

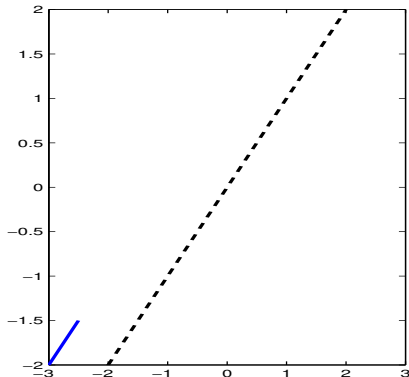
$$\text{prox}_f(x) = \underset{y \in \mathcal{H}}{\text{argmin}} f(y) + \frac{1}{2} \|y - x\|^2$$

Proximity operator: definition



Moreau envelope

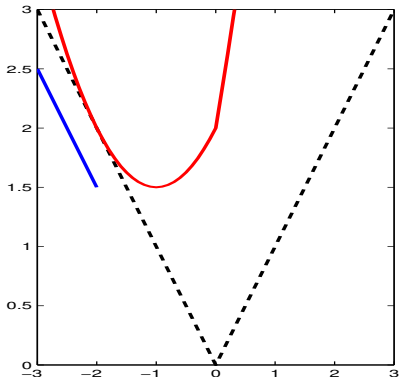
$$\gamma f(x) = \inf_{y \in \mathcal{H}} f(y) + \frac{1}{2\gamma} \|y - x\|^2$$



Proximity operator

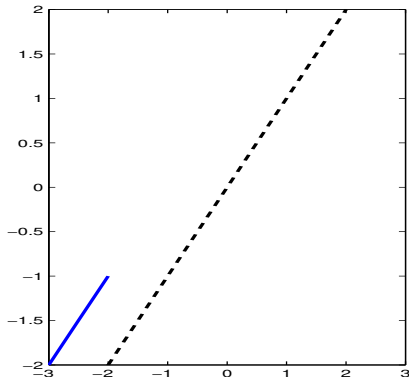
$$\text{prox}_f(x) = \underset{y \in \mathcal{H}}{\text{argmin}} f(y) + \frac{1}{2} \|y - x\|^2$$

Proximity operator: definition



Moreau envelope

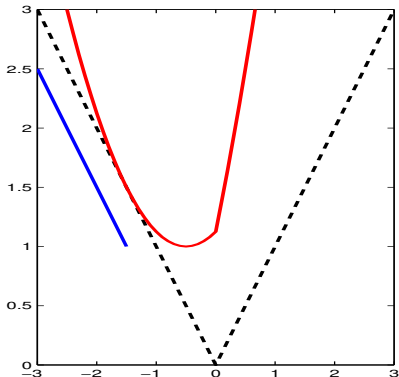
$$\gamma f(x) = \inf_{y \in \mathcal{H}} f(y) + \frac{1}{2\gamma} \|y - x\|^2$$



Proximity operator

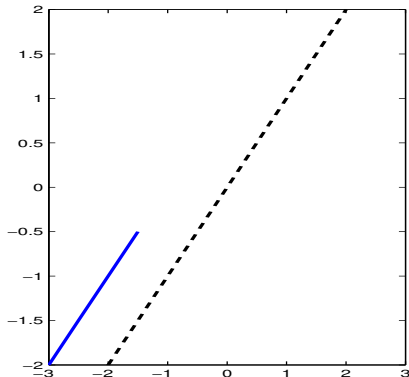
$$\text{prox}_f(x) = \underset{y \in \mathcal{H}}{\operatorname{argmin}} f(y) + \frac{1}{2} \|y - x\|^2$$

Proximity operator: definition



Moreau envelope

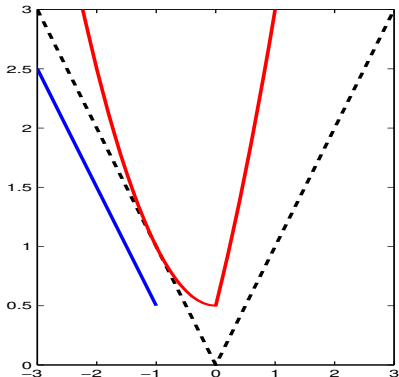
$$\gamma f(x) = \inf_{y \in \mathcal{H}} f(y) + \frac{1}{2\gamma} \|y - x\|^2$$



Proximity operator

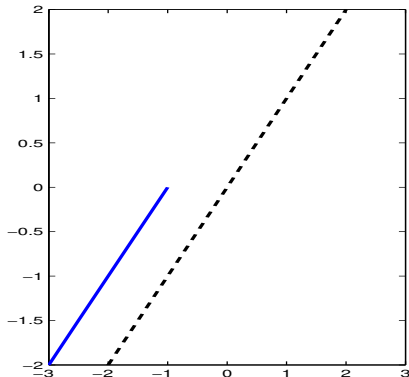
$$\text{prox}_f(x) = \underset{y \in \mathcal{H}}{\operatorname{argmin}} f(y) + \frac{1}{2} \|y - x\|^2$$

Proximity operator: definition



Moreau envelope

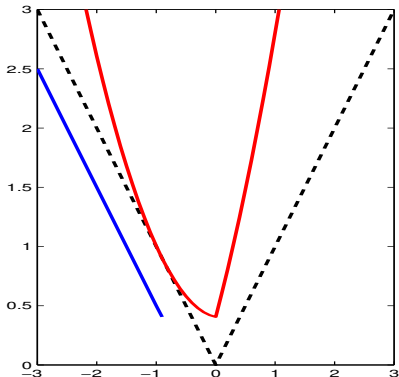
$$\gamma f(x) = \inf_{y \in \mathcal{H}} f(y) + \frac{1}{2\gamma} \|y - x\|^2$$



Proximity operator

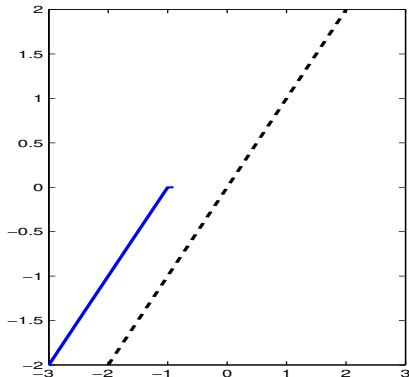
$$\text{prox}_f(x) = \underset{y \in \mathcal{H}}{\operatorname{argmin}} f(y) + \frac{1}{2} \|y - x\|^2$$

Proximity operator: definition



Moreau envelope

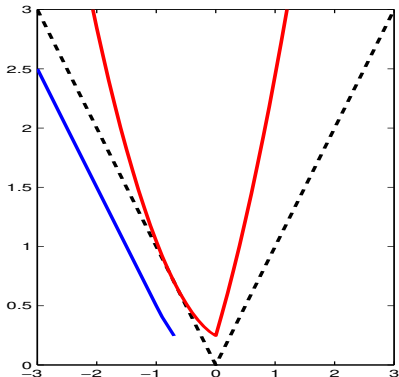
$$\gamma f(x) = \inf_{y \in \mathcal{H}} f(y) + \frac{1}{2\gamma} \|y - x\|^2$$



Proximity operator

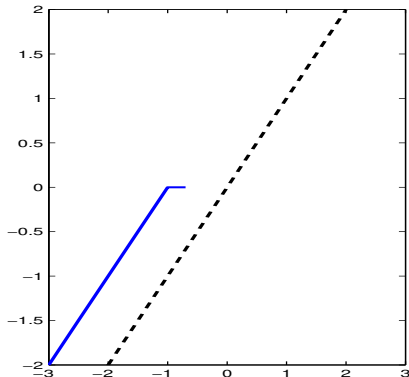
$$\text{prox}_f(x) = \underset{y \in \mathcal{H}}{\operatorname{argmin}} f(y) + \frac{1}{2} \|y - x\|^2$$

Proximity operator: definition



Moreau envelope

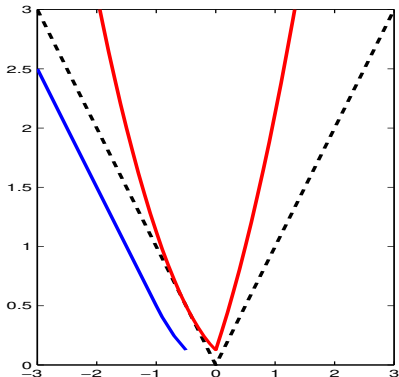
$$\gamma f(x) = \inf_{y \in \mathcal{H}} f(y) + \frac{1}{2\gamma} \|y - x\|^2$$



Proximity operator

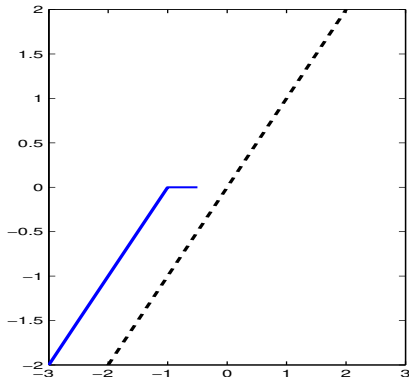
$$\text{prox}_f(x) = \underset{y \in \mathcal{H}}{\operatorname{argmin}} f(y) + \frac{1}{2} \|y - x\|^2$$

Proximity operator: definition



Moreau envelope

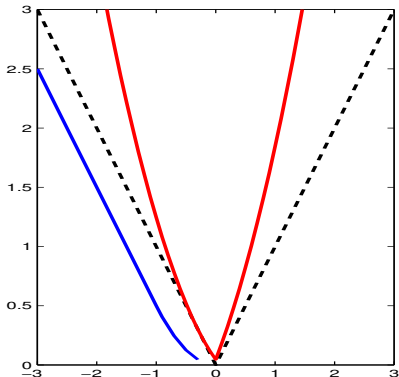
$$\gamma f(x) = \inf_{y \in \mathcal{H}} f(y) + \frac{1}{2\gamma} \|y - x\|^2$$



Proximity operator

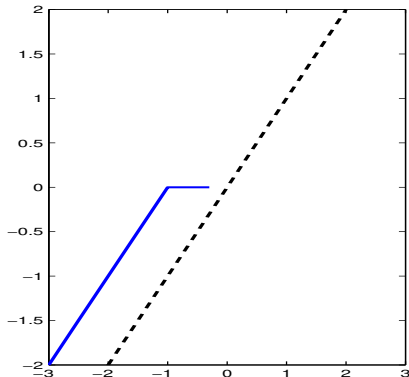
$$\text{prox}_f(x) = \underset{y \in \mathcal{H}}{\operatorname{argmin}} f(y) + \frac{1}{2} \|y - x\|^2$$

Proximity operator: definition



Moreau envelope

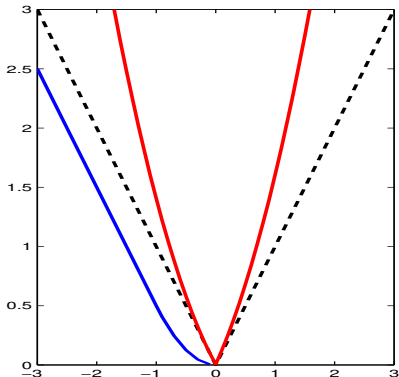
$$\gamma f(x) = \inf_{y \in \mathcal{H}} f(y) + \frac{1}{2\gamma} \|y - x\|^2$$



Proximity operator

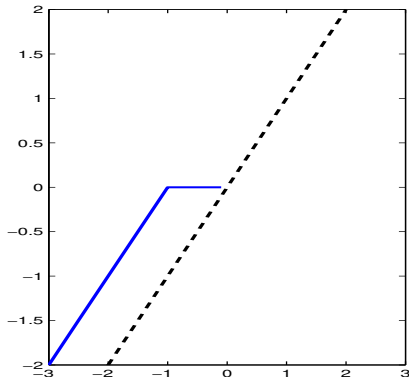
$$\text{prox}_f(x) = \underset{y \in \mathcal{H}}{\operatorname{argmin}} f(y) + \frac{1}{2} \|y - x\|^2$$

Proximity operator: definition



Moreau envelope

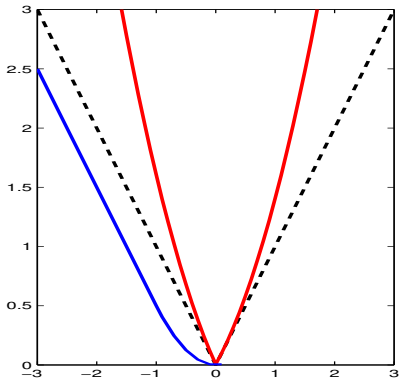
$$\gamma f(x) = \inf_{y \in \mathcal{H}} f(y) + \frac{1}{2\gamma} \|y - x\|^2$$



Proximity operator

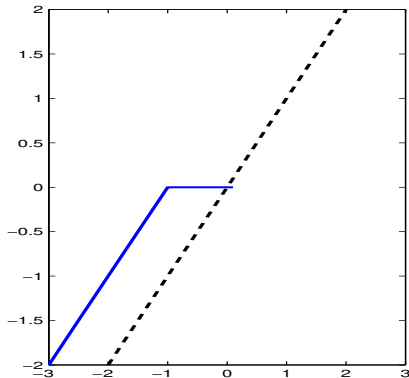
$$\text{prox}_f(x) = \argmin_{y \in \mathcal{H}} f(y) + \frac{1}{2} \|y - x\|^2$$

Proximity operator: definition



Moreau envelope

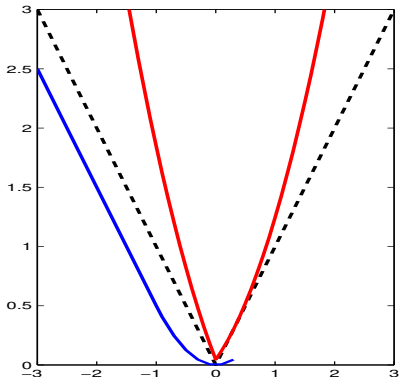
$$\gamma f(x) = \inf_{y \in \mathcal{H}} f(y) + \frac{1}{2\gamma} \|y - x\|^2$$



Proximity operator

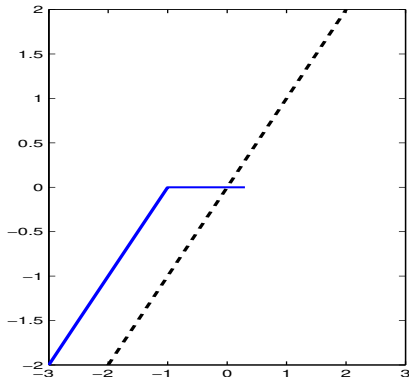
$$\text{prox}_f(x) = \underset{y \in \mathcal{H}}{\operatorname{argmin}} f(y) + \frac{1}{2} \|y - x\|^2$$

Proximity operator: definition



Moreau envelope

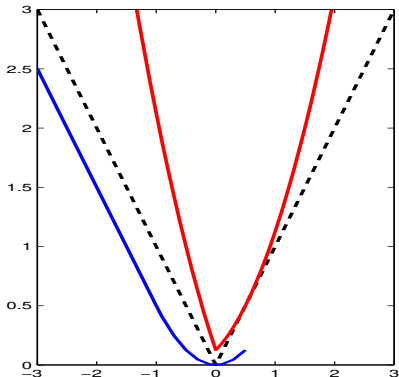
$$\gamma f(x) = \inf_{y \in \mathcal{H}} f(y) + \frac{1}{2\gamma} \|y - x\|^2$$



Proximity operator

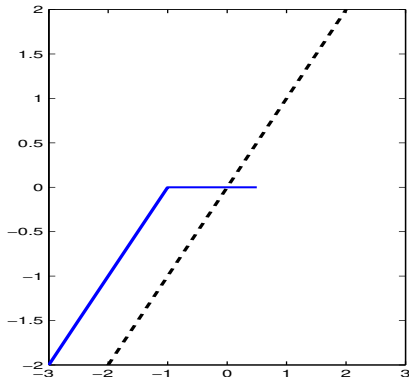
$$\text{prox}_f(x) = \underset{y \in \mathcal{H}}{\operatorname{argmin}} f(y) + \frac{1}{2} \|y - x\|^2$$

Proximity operator: definition



Moreau envelope

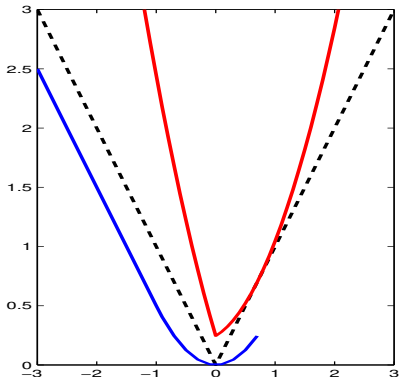
$$\gamma f(x) = \inf_{y \in \mathcal{H}} f(y) + \frac{1}{2\gamma} \|y - x\|^2$$



Proximity operator

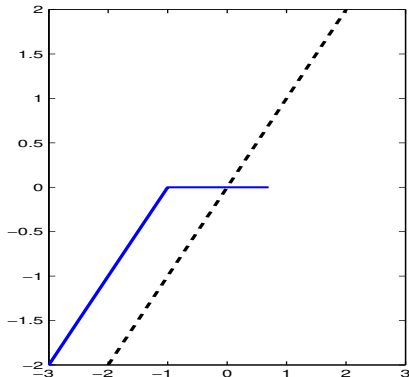
$$\text{prox}_f(x) = \argmin_{y \in \mathcal{H}} f(y) + \frac{1}{2} \|y - x\|^2$$

Proximity operator: definition



Moreau envelope

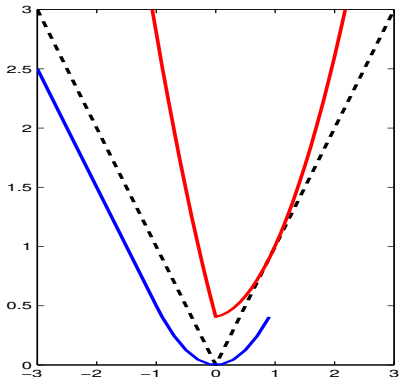
$$\gamma f(x) = \inf_{y \in \mathcal{H}} f(y) + \frac{1}{2\gamma} \|y - x\|^2$$



Proximity operator

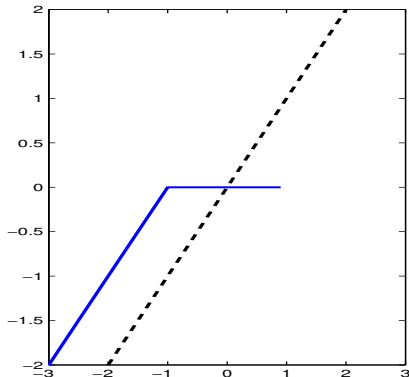
$$\text{prox}_f(x) = \underset{y \in \mathcal{H}}{\operatorname{argmin}} f(y) + \frac{1}{2} \|y - x\|^2$$

Proximity operator: definition



Moreau envelope

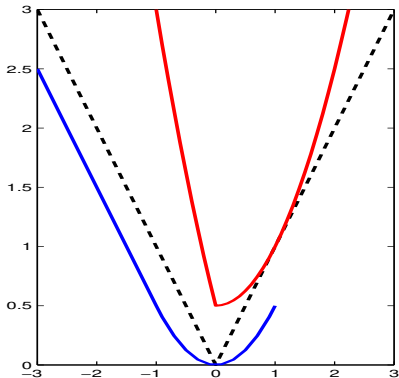
$$\gamma f(x) = \inf_{y \in \mathcal{H}} f(y) + \frac{1}{2\gamma} \|y - x\|^2$$



Proximity operator

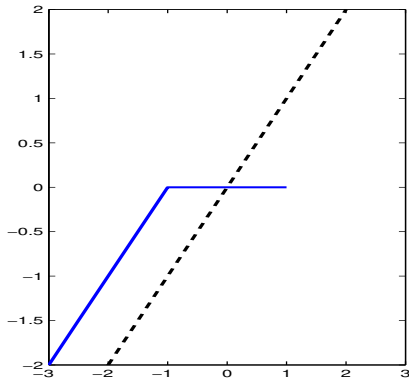
$$\text{prox}_f(x) = \underset{y \in \mathcal{H}}{\operatorname{argmin}} f(y) + \frac{1}{2} \|y - x\|^2$$

Proximity operator: definition



Moreau envelope

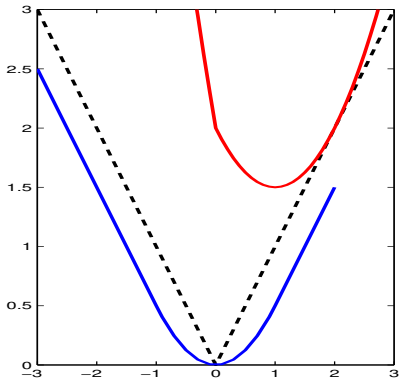
$$\gamma f(x) = \inf_{y \in \mathcal{H}} f(y) + \frac{1}{2\gamma} \|y - x\|^2$$



Proximity operator

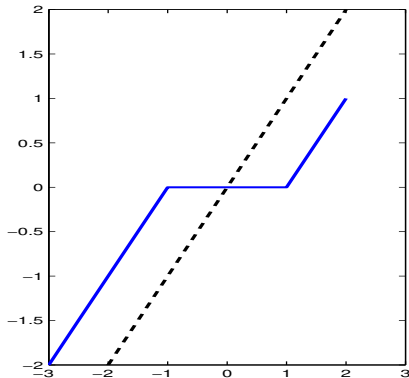
$$\text{prox}_f(x) = \underset{y \in \mathcal{H}}{\operatorname{argmin}} f(y) + \frac{1}{2} \|y - x\|^2$$

Proximity operator: definition



Moreau envelope

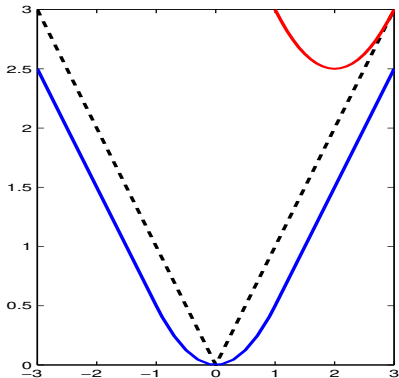
$$\gamma f(x) = \inf_{y \in \mathcal{H}} f(y) + \frac{1}{2\gamma} \|y - x\|^2$$



Proximity operator

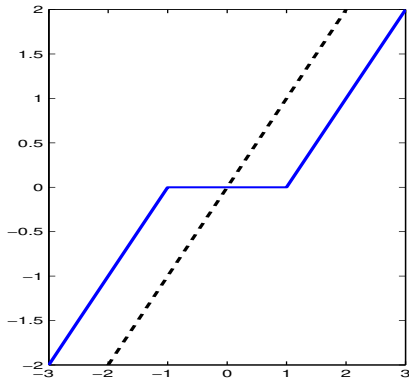
$$\text{prox}_f(x) = \underset{y \in \mathcal{H}}{\operatorname{argmin}} f(y) + \frac{1}{2} \|y - x\|^2$$

Proximity operator: definition



Moreau envelope

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



Proximity operator

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

Proximity operator: characterization

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

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

Proximity operator: characterization

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

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

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

$$\begin{aligned} p &= \arg \min_{y \in \mathcal{H}} f(y) + \frac{1}{2} \|y - x\|^2 \\ \Leftrightarrow \quad 0 &\in \partial \left(f + \frac{1}{2} \|\cdot - x\|^2 \right) (p) \\ \Leftrightarrow \quad 0 &\in \partial f(p) + p - x \\ \Leftrightarrow \quad x &\in (\operatorname{Id} + \partial f)(p). \end{aligned}$$

Proximity operator: examples

Let \mathcal{H} be a Hilbert space and $C \subset \mathcal{H}$.

The indicator function of C is

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

Projection :

If C be a nonempty closed convex subset of \mathcal{H} , then

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

Proximity operator: examples

Power q function with $q \geq 1$:

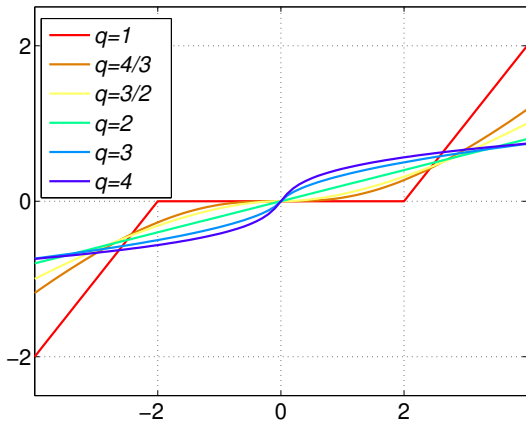
Let $\chi > 0$, $q \in [1, +\infty[$ and $\varphi: \mathbb{R} \rightarrow]-\infty, +\infty] : \xi \mapsto \chi|\xi|^q$.

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

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

Proximity operator: examples

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



Proximity operator: examples

Quadratic function :

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

Let $L \in \mathcal{B}(\mathcal{G}, \mathcal{H})$, $\gamma \in]0, +\infty[$ and $z \in \mathcal{G}$.

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

Proximity operator: properties

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

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

Proximity operator: properties

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

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

then

$$(\forall x = (x_1, \dots, x_n) \in \mathcal{H}_1 \times \dots \times \mathcal{H}_n) \quad \text{prox}_f(x) = (\text{prox}_{f_i}(x_i))_{1 \leq i \leq n}.$$

Proximity operator: properties

Let \mathcal{H} be a separable Hilbert space.

Let $(b_i)_{i \in I}$ be an orthonormal basis of \mathcal{H} .

For every $i \in I$, let $\varphi_i \in \Gamma_0(\mathbb{R})$ such that $\varphi_i \geq 0$. For every $x \in \mathcal{H}$, if

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

then

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

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

Proximity operator: properties

Let \mathcal{H} be a separable Hilbert space.

Let $(b_i)_{i \in I}$ be an orthonormal basis of \mathcal{H} .

For every $i \in I$, let $\varphi_i \in \Gamma_0(\mathbb{R})$ such that $\varphi_i \geq 0$. For every $x \in \mathcal{H}$, if

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

then

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

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

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

Proximity operator: properties

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

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

Proximity operator: properties

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

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

Proof: $LL^* = \mu \text{Id} \Rightarrow \text{ran } L = \mathcal{H}$ is closed, hence

$V = \text{ran } (L^*) = (\ker L)^\perp$ is closed. The orthogonal projection onto V is $P_V = L^*(LL^*)^{-1}L = \mu^{-1}L^*L$.

For every $x \in \mathcal{H}$, $p = \text{prox}_{f \circ L} x \Leftrightarrow x - p \in \partial(f \circ L)(p) = L^* \partial f(Lp)$ (since $\text{ran } L = \mathcal{H}$). Thus, $x - p \in V$.

It can be deduced that $P_{V^\perp} p = P_{V^\perp} x = x - P_V x = x - \mu^{-1} L^* L x$.

Furthermore,

$$x - p \in L^* \partial f(Lp) \Rightarrow Lx - Lp \in \mu \partial f(Lp) \Leftrightarrow Lp = \text{prox}_{\mu f}(Lx).$$

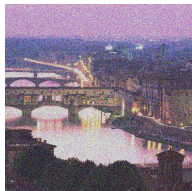
We have thus $P_V p = \mu^{-1} L^* L p = \mu^{-1} L^* \text{prox}_{\mu f}(Lx)$ and

$$p = P_V p + P_{V^\perp} p = x - \mu^{-1} L^* (\text{Id} - \text{prox}_{\mu f})(Lx).$$

Proximity operator: properties

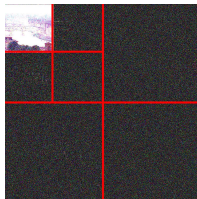
Particular case : $L \in \mathcal{B}(\mathcal{H}, \mathcal{H})$ unitary, $\text{prox}_{f \circ L} = L^* \text{prox}_f L$.

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



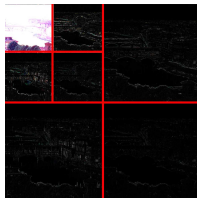
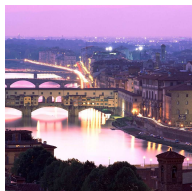
L

→



L^*

←



$\text{prox}_{\lambda \|\cdot\|_1}$

←

Useful websites

- ▶ Exhaustive list of proximity operators, Matlab and Python codes:
`http://proximity-operator.net/`
authors: Chierchia, Chouzenoux, Combettes, Pesquet
- ▶ On Github: `https://github.com/cvxgrp/proximal`
authors: Parikh, Chu, Boyd
- ▶ SPAMS: `http://spams-devel.gforge.inria.fr/`
authors: Mairal, Bach, Ponce, Sapiro, Jenatton, Obozinski
- ▶ Fast implementation:
`https://www.gipsa-lab.grenoble-inp.fr/~laurent.condat/software.html`
author: Condat